Алгоритмы идентификации комплексных экспоненциально-модулированных гармоник в рамках метода анализа сингулярного спектра[1]

Жорникова П. Г., студент кафедры статистического моделирования СПбГУ, polina.zhornikova@mail.ru

Голяндина Н.Э., к.ф.-м.н., доцент кафедры статистического моделирования СПбГУ, n.golyandina@spbu.ru

Аннотация

Рассматривается задача идентификации периодических компонент в комплексно-значных временных рядах с помощью метода SSA (Singular Spectrum Analysis). Результатом работы первого этапа метода SSA является сингулярное разложение ряда на элементарные компоненты, среди которых требуются найти те, что относятся к периодическим компонентам. Для вещественного случая методы идентификации таких компонент хорошо разработаны. Однако, они оказываются неприменимыми в комплексном случае. В данной работе предлагаются и обосновываются алгоритмы идентификации периодических компонент в комплексном случае.

Введение

Рассмотрим задачу идентификации периодических компонент в комплекснозначных временных рядах. Одним из методов разложения временных рядов на тренд, периодические компоненты и шум, не требующим априорного задания модели ряда, является метод анализа сингулярного спектра (SSA, Singular Spectrum Analysis) [4]. Первый шаг метода SSA состоит из построения на основе временного ряда так называемой траекторной матрицы, столбцами которой являются отрезки временного ряда длины L, называемой длиной окна. На втором шаге траекторная матрица раскладывается в элементарные матрицы ранга 1 с помощью сингулярного разложения. Каждая элементарная матрица задаётся сингулярным числом и левым и правым сингулярными векторами. Третий шаг состоит из идентификации элементарных компонент, соответствующих интересующим нас компонентам, и их группировки. И, наконец, на последнем шаге сгруппированные компоненты преобразуются обратно во временные ряды. Соответственно, если мы правиль-

[1]Работа поддержана грантом РФФИ 16-04-000821

но идентифицировали компоненты, относящиеся к тренду, то получим трендовую составляющую ряда; аналогично, правильная группировка позволяет выделить периодические компоненты.

Метод SSA хорошо согласуется с выделение экспоненциально-модулированных (э.-м.) гармоник вида $x_k = ae^{\alpha k}\cos(2\pi\omega k + \phi)$, поэтому задачей данной статьи будем считать выделение обобщённых периодик, являющихся суммой экспоненциально-модулированных гармоник. Будем рассматривать случай $0 < \omega < 0.5$.

Из теории метода SSA известно, что в вещественном случае э.-м. гармоника порождает в разложении (в условиях её приближённой отделимости) две элементарные компоненты, причём соответствующие сингулярные вектора также имеют вид э.-м. гармоник, возможно, приближённо, и фазы отличаются примерно на $\pi/2$. Поэтому двумерная диаграмма, где по оси X откладывается один сингулярный вектор (для определённости, будем рассматривать левые сингулярные вектора), а по оси Y — второй, будет иметь регулярный вид. В частности, если $\alpha = 0$, то точки будут лежать на окружности, например, образуя правильный многоугольник, а в общем случае точки лежат на спирали. Такие пары векторов можно как находить визуально, так и строить алгоритмы их количественной идентификации [1].

Заметим, что в вещественном случае сингулярные вектора определены с точность до знака, что не портит их характерные черты и не влияет на алгоритмы идентификации. Однако в комплексном случае [5, 3, 2], когда рассматривается комплексное сингулярное разложение, сингулярные вектора определены с точностью до умножения на $e^{\mathrm{i}2\pi t}$, где t — произвольное число от 0 до 1. Эта свобода в определении сингулярных векторов портит структуру двумерных диаграмм.

Задачей данной работы является построение и обоснование алгоритмов нахождения t, такого что сингулярные вектора снова становятся идентифицируемыми.

Теоретические сведения и результаты

Метод Complex SSA

Опишем коротко, согласно [3], базовый алгоритм метода Complex SSA. На вход алгоритма поступает комплекснозначный ряд $F_N = (f_1, \ldots, f_N)$ длины $N > 2$. Метод можно разделить на два этапа: разложение и восстановление.

1. На этапе разложения выбирается параметр $1 < L < N$, называемый

«длиной окна». Затем ряду F_N сопоставляется его траекторная (ганкелева) матрица $\mathbf{X}$ размера $L \times K$, $K = N - L + 1$, $\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = [X_1 : \ldots : X_K]$, $X_i = (f_{i-1}, \ldots, f_{i+L-2})^{\mathrm{T}}$, $1 \le i \le K$. Далее происходит сингулярное разложение (SVD) матрицы $\mathbf{X}$

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d, \tag{1}$$

где $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^*$, $d = \max\{i : \lambda_i > 0\}$, $\lambda_1, \ldots, \lambda_d$ — положительные собственные числа матрицы $\mathbf{X}\mathbf{X}^*$, взятые в неубывающем порядке, и $U_1, \ldots, U_d$ — соответствующие им ортонормированные собственные вектора, $V_i = \mathbf{X}^* U_i / \sqrt{\lambda_i}$, $*$ обозначает эрмитово сопряжение. Числа $\sqrt{\lambda_i}$ называются *сингулярными числами*, U_i и V_i — *левыми* и *правыми сингулярными векторами* матрицы $\mathbf{X}$ соответственно.

2. Этап восстановления начинается с процедуры группировки, которая делит все множество индексов $1, \ldots, d$, полученных в результате разложения (1), на m не пересекающихся подмножеств $I_1, \ldots, I_m$. Результирующая матрица $\mathbf{X}_I$, соответствующая группе $I = \{i_1, \ldots, i_p\}$, определяется как $\mathbf{X}_I = \mathbf{X}_{i_1} + \ldots + \mathbf{X}_{i_p}$. Вычислив такие матрицы для всех групп $I = I_1, \ldots, I_m$, получаем сгруппированный вид представления (1):

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}. \tag{2}$$

Далее происходит процедура диагонального усреднения. Для произвольной матрицы $\mathbf{Y}$ размера $L \times K$ она состоит в следующем: матрице $\mathbf{Y}$ сопоставляется ряд $Y_N = (y_1, \ldots, y_N)$ такой, что $N = L + K - 1$, а y_i — среднее элементов y_{jk} матрицы $\mathbf{Y}$, удовлетворяющих $j + k = i + 1$. Применяя диагональное усреднение к результирующим матрицам $\mathbf{X}_{I_k}$, получаем восстановленные ряды $\check{F}_N^{(k)} = (\check{f}_1^{(k)}, \ldots, \check{f}_N^{(k)})$. Таким образом, исходный комплекснозначный ряд $F_N = (f_1, \ldots, f_N)$ раскладывается в сумму m комплекснозначных рядов

$$F_n = \sum_{k=1}^{m} \check{F}_n^k = \sum_{k=1}^{m} (\check{F}_n^{(1,k)} + \mathrm{i}\check{F}_n^{(2,k)}).$$

Э.-м. гармонический ряд

Элементы комплекснозначного экспоненциально-модулированного гармонического ряда S_N имеют вид

$$s_k = e^{\alpha k}(a \cos(2\pi\omega k + \phi_1) + \mathrm{i}\, b \cos(2\pi\omega k + \phi_2)), \tag{3}$$

где $0 < \omega < 0.5$, $0 \le \phi_1, \phi_2 < 2\pi$, $a, b \ne 0$.

Особым случаем является

$$a = b, \qquad |\phi_1 - \phi_2| = \pi/2 \mod \pi. \tag{4}$$

L-Ранг ряда d — это ранг его траекторной матрицы и, следовательно, количество левых или правых сингулярных векторов, соответствующих ненулевым сингулярным числам траекторной матрицы $\mathbf{X}$. Для определённости будем рассматривать левые сингулярные вектора $U_1, \ldots, U_d$, т.е. собственные вектора матрицы $\mathbf{X}\mathbf{X}^*$.

Лемма 1 1. *Если условие* (4) *не выполнено, то ранг ряда S_N равен 2 и элементы собственных векторов U_1 и U_2 имеют вид*

$$u_k^{(1)} = e^{\alpha k}(c_{11}\cos(2\pi\omega k + \phi_1) + \mathrm{i}\, c_{12}\cos(2\pi\omega k + \psi_1)), \tag{5}$$

$$u_k^{(2)} = e^{\alpha k}(c_{21}\cos(2\pi\omega k + \phi_2) + \mathrm{i}\, c_{22}\cos(2\pi\omega k + \psi_2)), \tag{6}$$

где $1 \le k \le L$, $0 \le \phi_1, \phi_2, \psi_1, \psi_2 < 2\pi$, $c_{11}, c_{12}, c_{21}, c_{22} \neq 0$.

2. *Если условие* (4) *выполнено, то ранг ряда S_N равен 1 и элементы собственного вектора U_1 имеют вид*

$$u_k^{(1)} = c\, e^{\mathrm{i}2\pi\omega k + \alpha k}, \quad c \neq 0.$$

Замечание 1 *По свойствам сингулярного разложения, сингулярные вектора определены неоднозначно. В частности, если U — собственный вектор, то и $e^{\mathrm{i}2\pi t}U$ — тоже собственный вектор.*

Введем теперь для пары векторов $A = (a_1, \ldots, a_L)^{\mathrm{T}}$, $B = (b_1, \ldots, b_L)^{\mathrm{T}}$ и некоторого числа $0 \leqslant t \leqslant 1$ два функционала. Пусть $W(t) = e^{\mathrm{i}2\pi t}B$. Определим

$$\tau_1(A, B; t) := \frac{1}{L}\sum_{k=1}^{L}(x_k - \bar{x})^2 + \frac{1}{L}\sum_{k=1}^{L}(y_k - \bar{y})^2,$$

где $x_k = (\operatorname{Re} a_k)^2 + (\operatorname{Re} w_k(t))^2$, $y_k = (\operatorname{Im} a_k)^2 + (\operatorname{Im} w_k(t))^2$, и

$$\tau_2(A, B; t) := \frac{1}{L-1}\sum_{k=1}^{L-1}\left(\Phi_k - \bar{\Phi}\right)^2 + \frac{1}{L-1}\sum_{k=1}^{L-1}(\Psi_k - \bar{\Psi})^2,$$

где Φ_k — угол между векторами $(\operatorname{Re} a_k, \operatorname{Re} w_k(t))^{\mathrm{T}}$ и $(\operatorname{Re} a_{k+1}, \operatorname{Re} w_{k+1}(t))^{\mathrm{T}}$; а Ψ_k — угол между $(\operatorname{Im} a_k, \operatorname{Im} w_k(t))^{\mathrm{T}}$ и $(\operatorname{Im} a_{k+1}, \operatorname{Im} w_{k+1}(t))^{\mathrm{T}}$.

Замечание 2 1. *Равенство* $\tau_1(A, B; t) = 0$ *означает, что точки, составленные из вещественных частей* $(\mathrm{Re}\, a_k, \mathrm{Re}\, w_k(t))^{\mathrm{T}}$*, лежат на одной окружности, и то же самое верно для точек* $(\mathrm{Im}\, a_k, \mathrm{Im}\, w_k(t))^{\mathrm{T}}$.

2. *Равенство* $\tau_2(A, B; t) = 0$ *означает, что все углы между последовательными точками двумерной диаграммы мнимых и двумерной диаграммы вещественных частей векторов* A *и* $W(t)$ *одинаковые. Если при этом нормы этих векторов медленно меняются по* k*, то на двумерных диаграммах получаются «спирали».*

Следующие утверждения следуют из леммы 1 и ортонормированности собственных векторов.

Утверждение 1 *Пусть для ряда* S_N *не выполнено условие* (4). *Тогда*

1. *если* $\alpha = \alpha_N = C/N$*, где* C *— некоторая константа, и* $L = [\beta N]$*, где* $0 < \beta < 1$*, то существует такое* $t = t(U_1, U_2, L)$*,* $0 \leqslant t \leqslant 1$*, что* $\lim_{L\to\infty} \tau_2(U_1, U_2; t) = 0$*;*

2. *если* $\alpha = 0$*, то существует такое* $t = t(U_1, U_2, L)$*,* $0 \leqslant t \leqslant 1$*, что* $\lim_{L\to\infty} \tau_1(U_1, U_2; t) = 0$ *и* $\lim_{L\to\infty} \tau_2(U_1, U_2; t) = 0$*;*

3. *если* $\alpha = 0$ *и* $L\omega$ *— целое, то существует такое* $0 \leqslant t \leqslant 1$*, что* $\tau_1(U_1, U_2; t) = 0$ *и* $\tau_2(U_1, U_2; t) = 0$.

Утверждение 2 *Пусть для ряда* S_N *выполнено условие* (4). *Тогда*

1. $\tau_2(\mathrm{Re}\, U_1, \mathrm{Im}\, U_1; 0) = 0$*;*

2. *если* $\alpha = 0$*, то* $\tau_1(\mathrm{Re}\, U_1, \mathrm{Im}\, U_1; 0) = 0$.

Алгоритм идентификации

Результатом этапа разложения алгоритма Complex SSA является набор собственных векторов $U_1, \ldots, U_d$. Опишем алгоритм идентификации пар собственных векторов, соответствующих э.-м. гармоникам ранга 2. Пусть τ — рассматриваемый функционал (τ_1 или τ_2).

1. Для каждой пары U_i, U_j, $i < j$, решается оптимизационная задача
$$\tau(U_i, U_j; t) \longrightarrow \min_t .$$
Пусть минимум достигается на $t_{i,j}$ и минимальное значение функционала равно $\tau_{i,j}$.

2. Если идентификация визуальная, то среди двумерных диаграмм отдельно вещественных и мнимых частей U_i и $e^{\mathrm{i}2\pi t_{i,j}}$ ищутся те, где изображение обладает свойствами, близкими к свойствам, соответствующим нулевому значению функционала τ (см. замечание 2).

3. Если идентификация автоматическая, то отбираются те пары U_i, U_j, у которых значение функционала $\tau_{i,j}$ меньше заданного порога, если неизвестно число э.-м. гармоник, и пары с минимальными значениями функционала, если их количество известно.

На этапе восстановления, идентифицированные пары собственных векторов образуют группы, на основе которых делается восстановление.

Для идентификации э.-м. гармоники ранга 1 оптимизация не нужна. В остальном принцип идентификации тот же.

Пример

Рассмотрим ряд $F_N = S_N^{(1)} + S_N^{(2)} + S_N^{(3)} + R_N$ с

$$\begin{aligned} s_k^{(1)} &= 10\cos(2\pi k/25) + 20\mathrm{i}\cos(2\pi k/25), \\ s_k^{(2)} &= e^{0.004k}(\cos(2\pi k/5) + 70\mathrm{i}\cos(2\pi k/5 + \pi/4)), \\ s_k^{(3)} &= e^{0.005k}(\cos(2\pi k/10) + \mathrm{i}\sin(2\pi k/10)), \\ r_k &= \varepsilon_k^{(1)} + \mathrm{i}\,\varepsilon_k^{(2)}, \end{aligned}$$

где $\varepsilon_k^{(1)}$ и $\varepsilon_k^{(2)}$ — независимые реализации стандартного нормального распределения $N(0,1)$, $1 \le k \le N$, $N = 199$. Здесь $S_N^{(1)}$ — гармоника с рангом 2, $S_N^{(2)}$ — э.-м. гармоника с рангом 2, $S_N^{(3)}$ — э.-м. гармоника с рангом 1, а R_N — шум. Возьмем длину окна $L = 100$. Тогда $K = 100$ тоже.

Посмотрим на результат работы алгоритма с функционалом τ_2 для этого ряда. Для реализации метода Complex SSA использовался R-пакет Rssa [6].

На рис. 1 изображены двумерные диаграммы пар собственных векторов (их мнимых и вещественных частей, на одном и том же рисунке), полученных в результате применения Complex SSA. После работы алгоритма идентификации с минимизацией функционала τ_2 диаграммы имеют вид, изображенный на рис. 2. Видно, что компоненты 1–2 и 3–4 относятся к э.-м. гармоникам с $d = 2$. Минимальные значения функционала τ_2 для первых восьми последовательных пар равны 1.99e-05, 0.65, 2.86e-05, 0.32, 1.06, 1.11, 0.59, 1.47 и подтверждают визуальный выбор.

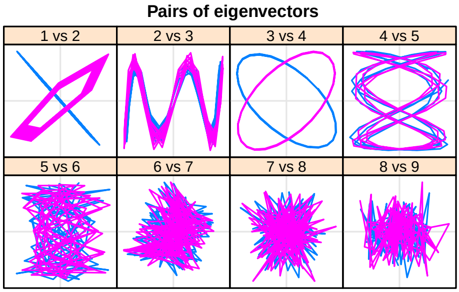

Рис. 1: Двумерные диаграммы собственных векторов для ряда F_N.

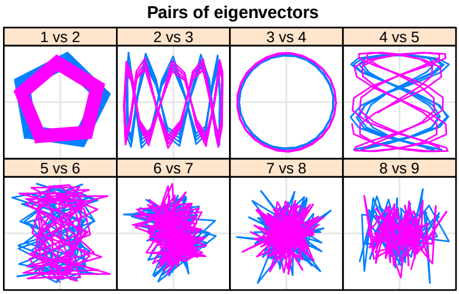

Рис. 2: Двумерные диаграммы собственных векторов для ряда F_N после работы алгоритма.

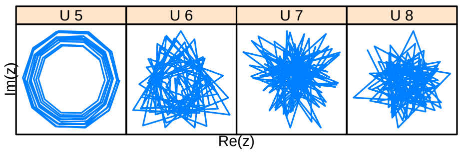

Рис. 3: Двумерные диаграммы вещ. и мним. частей собственных векторов для ряда F_N.

Далее ищем регулярные изображения в виде спирали на двумерной диаграмме вещественной и мнимой частей каждого собственного вектора отдельно (Рис. 1), исключая из рассмотрения компоненты 1–4. Видно, что вектор 5 относится к компоненте с $d = 1$.

Заключение

В работе был представлен новый алгоритм идентификации экспоненциально-модулированных гармоник с помощью комплексной модификации метода SSA. На примерах была продемонстрирована работоспособность метода.

Литература

[1] Th. Alexandrov, N. Golyandina. Automatic extraction and forecast of time series cyclic components within the framework of SSA. In *Proceedings of the 5th St.Petersburg Workshop on Simulation*, pages 45–50. St. Petersburg State University, 2005.

[2] K. Eftaxias, S. Enshaeifar, O. Geman, S. Kouchaki, S. Sanei. Detection of Parkinson's tremor from EMG signals; a singular spectrum analysis approach. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*. pages 398–402, 2015.

[3] Н. Голяндина, В. Некруткин, Д.Степанов. Варианты метода 'Гусеница'-SSA для анализа многомерных временных рядов. В: *Труды II Международной конференции 'Идентификация систем и проблемы управления'*, с. 2139–2168. Москва, 2003.

[4] N. Golyandina, V. Nekrutkin, A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman&Hall/CRC, 2001.

[5] C. Keppenne, U. Lall. Complex singular spectrum analysis and multivariate adaptive regression splines applied to forecasting the southern oscillation. *Exp. Long-Lead Forcst. Bull.*, 1996.

[6] A. Korobeynikov, A. Shlemov, K. Usevich, N. Golyandina. *Rssa: A collection of methods for singular spectrum analysis*, 2015. R package version 0.13-1.