Ошибки оценивания сигнала с помощью комплексного анализа сингулярного спектра [1]

Голяндина Н. Э., доцент кафедры Статистического Моделирования СПбГУ, n.golyandina@spbu.ru

Сенов М. А., студент, СПбГУ, senov.mikhail@gmail.com

Аннотация

Анализ Сингулярного Спектра (Singular Spectrum Analysis, SSA) — непараметрический метод для разложения временного ряда на сумму интерпретируемых компонент, таких как тренд, периодики и шум. Расширение SSA на комплексный случай называется CSSA. В комплексных временных рядах одним из часто встречающихся случаев является сигнал, состоящий из суммы мнимых экспонент. В работе проведено теоретическое сравнение CSSA и SSA, примененного отдельно к вещественной и мнимой части, для случая синусоидальных сигналов, частным случаем которых является мнимая экспонента. Для константного сигнала получен явный вид первого порядка ошибки его оценки в случае наличия в ряде выброса.

Введение

Анализ Сингулярного Спектра (Singular Spectrum Analysis, SSA) [2] — мощный метод анализа временных рядов, не требующий предварительного задания параметрической модели ряда. Однако есть класс сигналов, а именно временный ряды, управляемые линейными рекуррентными соотношениями, который позволяет получать теоретические результаты.

В данной работе рассматриваются случаи константного и двух синусоидальных сигналов. Для анализа ошибки оценивания сигнала используется теория возмущений [5], которая была применена для выделения сигнала методом SSA в ряде работ, см., например, [6].

[1] Работа выполнена при поддержке гранта РФФИ 20-01-00067

Алгоритм CSSA

Алгоритм 1.
Вход: *Временной ряд* $\mathbb{X} = (x_1, \ldots, x_N)$*,* длина окна L*, ранг сигнала* r*.*
Выход: *Оценка сигнала* $\widetilde{\mathbb{S}}$*.*
Алгоритм:

1. ***Вложение.*** *Построим* $\mathbf{X} \in \mathsf{R}^{L\times K}$*,* L-траекторную матрицу *ряда* $\mathbb{X}$*:* $\mathbf{X} = \mathcal{T}_L\mathbb{X} = [X_1 : \ldots X_K]$*, где* $K = N - L + 1$*, а* X_i — векторы L-вложения: $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}} \in \mathsf{R}^L$*.*

2. ***Разложение.*** *Построим SVD-разложение матрицы* $\mathbf{X}$*:* $\mathbf{X} = \sum_{k=1}^{\operatorname{rank}\mathbf{X}} \sqrt{\lambda_k} U_k V_k^{\mathrm{H}} = \sum_{k=1}^{\operatorname{rank}\mathbf{X}} \widehat{\mathbf{X}}_k$*, где* U_k*,* V_k *— правые и левые сингулярные векторы матрицы* $\mathbf{X}$ *соответственно,* $\sqrt{\lambda_k}$ *— сингулярные числа.*

3. ***Группировка.*** *Сгруппируем матрицы компонент сигнала* $\widehat{\mathbf{S}}$*:* $\widehat{\mathbf{S}} = \sum_{k=1}^{r} \widehat{\mathbf{X}}_k$*.*

4. ***Диагональное усреднение.*** *Применим процедуру диагонального усреднения (проекции в норме Фробениуса на линейное пространство ганкелевых матриц):* $\widetilde{\mathbf{S}} = \mathcal{H}\widehat{\mathbf{S}}$*, затем сопоставим полученным Ганкелевым матрицам ряды длины* N*:* $\widetilde{\mathbb{S}} = \mathcal{T}_L^{-1}\widetilde{\mathbf{S}}$*.*

Применение теории возмущений к SSA и CSSA

Наблюдаем комплексный временной ряд $\mathbb{X}$ длины N и длиной окна L, данный ряд представляется как $\mathbb{X} = \mathbb{S} + \mathbb{R}$, где $\mathbb{S}$ — сигнал, $\mathbb{R}$ — возмущение.

В работе [6] вводится разложение восстановления сигнала в модели $\mathbb{S}(\delta) = \mathbb{S} + \delta\mathbb{R}$, что соответствует $\mathbf{H}(\delta) = \mathbf{H} + \delta\mathbf{E}$, где $\mathbf{H}(\delta)$ — траекторная матрица $\mathbb{S}(\delta)$, $\mathbf{H}$ — траекторная матрица $\mathbb{S}$, $\delta\mathbf{E}$ — траекторная матрица возмущения $\delta\mathbb{R}$ и рассматривается линейный по δ член ошибки восстановления, называемый первым порядком ошибки восстановления.

Рассмотрим возмущение ряда $\mathbb{R}$ с $\delta = 1$, его траекторная матрица $\mathbf{E}$. Первый порядок ошибки восстановления обозначим как $\mathbb{F}^{(1)} = \mathcal{H}(\mathbf{H}^{(1)})$.

В работе [7] была получена формула для $\mathbf{H}^{(1)}$

$$\mathbf{H}^{(1)} = \mathbf{P}_0\mathbf{E}\mathbf{Q}_0^{\perp} + \mathbf{P}_0^{\perp}\mathbf{E}, \tag{1}$$

где $\mathbf{P}_0^{\perp}$ — проектор на пространство столбцов $\mathbf{H}$, $\mathbf{Q}_0^{\perp}$ — проектор на пространство строк $\mathbf{H}$, $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_0^{\perp}$, $\mathbf{I}$ — единичная матрица.

Сравнение CSSA и SSA в случае совпадающих пространств сигналов

Обозначим за:

$\mathbb{F}^{(1)} = (f_1^{(1)}, \ldots, f_N^{(1)})$ первый порядок ошибки восстановления $\mathbb{S}$ с возмущением $\mathbb{R}$ метода CSSA,

$\mathbb{F}_{\mathrm{Re}}^{(1)} = (f_{\mathrm{Re},1}^{(1)}, \ldots, f_{\mathrm{Re},N}^{(1)})$ первый порядок ошибки восстановления $\mathrm{Re}(\mathbb{S})$ с возмущением $\mathrm{Re}(\mathbb{R})$ метода SSA,

$\mathbb{F}_{\mathrm{Im}}^{(1)} = (f_{\mathrm{Im},1}^{(1)}, \ldots, f_{\mathrm{Im},N}^{(1)})$ первый порядок ошибки восстановления $\mathrm{Im}(\mathbb{S})$ с возмущением $\mathrm{Im}(\mathbb{R})$ метода SSA.

Теорема 1. *Пусть траекторные пространства $\mathbb{S}$, $\mathrm{Re}(\mathbb{S})$ и $\mathrm{Im}(\mathbb{S})$ совпадают.*

Тогда

$$\mathbb{F}^{(1)} = \mathbb{F}_{\mathrm{Re}}^{(1)} + i\mathbb{F}_{\mathrm{Im}}^{(1)}.$$

Линейность вхождения $\mathbf{E}$ в формулу (1) позволяет доказать данное утверждение.

Случайное возмущение

Рассмотрим случайное возмущение $\mathbb{R}$.

Для дальнейших рассуждений приведём известный результат.

Лемма 1. *Пусть $\zeta = \xi + i\eta$. Тогда $\mathbb{D}\zeta = \mathbb{D}\xi + \mathbb{D}\eta$.*

Доказательство.

$$\mathbb{D}\zeta = \mathbb{E}(|\zeta - \mathbb{E}\zeta|^2) = \mathbb{E}(|(\xi - \mathbb{E}\xi) + i((\eta - \mathbb{E}\eta))|^2) =$$
$$= \mathbb{E}((\xi - \mathbb{E}\xi)^2 + (\eta - \mathbb{E}\eta)^2) = \mathbb{E}(\xi - \mathbb{E}\xi)^2 + \mathbb{E}(\eta - \mathbb{E}\eta)^2 = \mathbb{D}\xi + \mathbb{D}\eta.$$

□

Следствие 1. *Пусть выполнены условия теоремы 1.*

Тогда

$$\mathbb{D}f_l^{(1)} = \mathbb{D}f_{\mathrm{Re},l}^{(1)} + \mathbb{D}f_{\mathrm{Im},l}^{(1)}. \tag{2}$$

Получается автоматически из теоремы 1 и леммы 1.

Случай двух зашумленных синусоид

Следствие 2. *Для комплексного ряда, состоящего из двух зашумленных синусоид со сдвигом, не равным $\pi/2$, выполняется формула* (2).

Совпадение траекторных пространств сигналов для такого ряда было показано в [3].

Замечание 1. *Численные эксперименты показывают, что для сигнала в виде комлексной экспоненты MSE CSSA-оценки сигнала поточечно равна полусумме MSE SSA-оценок сигнала его вещественной и мнимой частей.*

Случай константных сигналов с выбросом

Рассматриваем сигнал $\mathbb{S} = (c_1 + ic_2, \ldots, c_1 + ic_2)$, возмущённый выбросом $a_1 + ia_2$ на позиции k. Исходя из теоремы 1, достаточно уметь вычислять первый порядок ошибки восстановления сигнала $\mathbb{S} = (c, \ldots, c)$, возмущённого выбросом a на позиции k.

В работе [6] была получен частный случай формулы (1) для сигналов единичного ранга.

$$\mathbf{H}^{(1)} = -U^{\mathrm{H}}\mathbf{E}VUV^{\mathrm{H}} + UU^{\mathrm{H}}\mathbf{E} + \mathbf{E}VV^{\mathrm{H}},$$

где U, V — сингулярные вектора матрицы $\mathbf{H}$.

Подставляя $U = \{1/\sqrt{L}\}_{i=1}^{L}$, $V = \{1/\sqrt{K}\}_{i=1}^{K}$, $K = N - L + 1$ и в предположении, что $L \leq K$, можно получить явный вид первого порядка ошибки восстановления.

Выпишем его для случая $k \leq \min(L/2, K - L)$

$$f_l^{(1)} = \frac{a}{LK}\begin{cases} (L+K-k), & 1 \leq l \leq k \\ \frac{1}{l}(L+K-l)k, & k < l \leq L \\ \frac{1}{L}K(L+k-l), & L < l < L+k \\ 0, & L+k \leq l \leq K \\ \frac{1}{N-l+1}(K-l)(L-k), & K < l < K+k \\ -k, & K+k \leq l \leq N \end{cases}.$$

Замечание 2. *Из данной формулы видно, что при фиксированном L, не зависящим от N, первый порядок ошибки не стремится к* 0 *с ростом N, тогда как численные эксперименты показывают, что полная ошибка восстановления стремится к* 0 *с ростом N.*

Численное сравнение первого порядка ошибки и полной ошибки оценивания сигнала

Для случая зашумленных гармоник рассмотрен пример с сигналом $s_l = \cos(2\pi l/10) + i\cos(2\pi l/10 + \pi/2)$, $\sigma^2 = 0.1$, $N = 49$, $L = 20$. Результат представлен на Рис. 1.

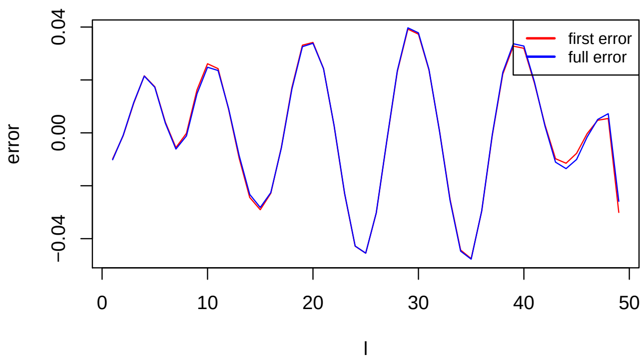

Рис. 1: Вещественные части первого порядка и полной ошибок.

Для случая возмущения в виде выброса был рассмотрен пример с сигналом $s_l = 1 + 1i$, возмущённым выбросом $a_1 + ia_2 = 10 + i10$. Результаты представлены в таблице 1.

Таблица 1: Максимальное различие первого порядка и полной ошибок.

N	50	100	400	1600
$L = N/2$	0.1313	0.0419	0.0033	0.0002
$L = 20$	0.3074	0.1965	0.5655	0.6720

Численные результаты показывают, что для случая зашумленных гармоник первый порядок адекватно оценивает полную ошибку восстановления сигнала в каждой точке.

Однако для случая возмущения в виде выброса это верно только когда L и K стремятся к бесконечности с ростом N.

Все численные результаты были получены при помощи пакета [4].

Заключение

В работе удалось подвести теоретическую базу под имеющиеся ранее численные результаты ([1]) по сравнению CSSA и SSA для двух зашумленных гармоник с одинаковой частотой и сдвигом, не равным $\pi/2$. Для мнимой экспоненты был получен более общий, нежели имеющиеся ранее, численный результат. Для константного ряда с выбросами был получен явный вид ошибок первого порядка оценки сигнала в каждой точке, а также численно исследовано соотношение между ошибкой первого порядка и полной ошибкой.

Литература

[1] N. Golyandina, A. Korobeynikov, A. Shlemov, and K. Usevich. Multivariate and 2D extensions of singular spectrum analysis with the Rssa package. *Journal of Statistical Software*, 67(2):1–78, 2015.

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

[3] Д. Степанов, Н. Голяндина. Варианты метода "Гусеница"-SSA для прогноза многомерных временных рядов. Труды IV Международной конференции "Идентификация систем и задачи управления" SICPRO'05. Москва, 2005, с. 1831-1848.

[4] A. Korobeynikov, A. Shlemov, K. Usevich, and N. Golyandina. *Rssa: A collection of methods for singular spectrum analysis http://CRAN.R-project.org/package=Rssa*, 2021. R package version 1.04.

[5] Kato

[6] V.V. Nekrutkin. *Perturbations in SSA*. Manuscript, 2008.

[7] Константинов