

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. Распирение 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 временные 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].

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

Алгоритм CSSA

Алгоритм 1.

Вход: Временной ряд $\mathbb{X} = (x_1, \dots, x_N)$, длина окна L , ранг сигнала r .

Выход: Оценка сигнала $\tilde{\mathbb{S}}$.

Алгоритм:

1. **Вложение.** Построим $\mathbf{X} \in \mathbb{R}^{L \times K}$, L -траекторную матрицу ряда \mathbb{X} : $\mathbf{X} = \mathcal{T}_L \mathbb{X} = [X_1 : \dots : X_K]$, где $K = N - L + 1$, а X_i — векторы L -вложения: $X_i = (x_i, \dots, x_{i+L-1})^T \in \mathbb{R}^L$.
2. **Разложение.** Построим SVD-разложение матрицы \mathbf{X} : $\mathbf{X} = \sum_{k=1}^{\text{rank } \mathbf{X}} \sqrt{\lambda_k} U_k V_k^H = \sum_{k=1}^{\text{rank } \mathbf{X}} \hat{\mathbf{X}}_k$, 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 $\hat{\mathbf{S}}$: $\hat{\mathbf{S}} = \sum_{k=1}^r \hat{\mathbf{X}}_k$.
4. **Диагональное усреднение.** Применим процедуру диагонального усреднения (проекция в норме Фробениуса на линейное пространство ганкелевых матриц): $\tilde{\mathbf{S}} = \mathcal{H} \hat{\mathbf{S}}$, затем сопоставим полученным Ганкелевым матрицам ряды длины N : $\tilde{\mathbb{S}} = \mathcal{T}_L^{-1} \tilde{\mathbf{S}}$.

Для дальнейших рассуждений потребуется знание рангов конкретных рядов.

Из [3] известно, что ранг комплексного сигнала, состоящего из двух синусоид равен 2, если сдвиг между синусоидами не равен $\pi/2$ и равен 1 в случае мнимой экспоненты. Ранг вещественного сигнала, состоящего из синусоиды, равен 2. Ранги же комплексной и вещественной констант равны 1.

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 [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 [8] для теоремы 2.1 из [6] была получена следующая формула для $\mathbf{H}^{(1)}$ в случае достаточно маленького возмущения.

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

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:

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

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

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

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

Тогда

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

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.

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

□

Следствие 1. Пусть выполнены условия теоремы 1.

Тогда

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

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, проведённые в [1], показывают, что для сигнала в виде комплексной экспоненты MSE CSSA-оценка сигнала равна полусумме MSE SSA-оценок сигнала его вещественной и мнимой частей.

Замечание 2. Численные эксперименты, проведённые для данной работы показывают, что для сигнала в виде комплексной экспоненты MSE CSSA-оценки сигнала поточечно равны полусумме MSE SSA-оценок сигнала его вещественной и мнимой частей.

Случай константных сигналов с выбросом

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

В работе [7] была получен частный случай формулы (1) для вещественных сигналов единичного ранга.

$$\mathbf{H}^{(1)} = -U^T \mathbf{E} V U V^T + U U^T \mathbf{E} + \mathbf{E} V V^T,$$

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 3. 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.

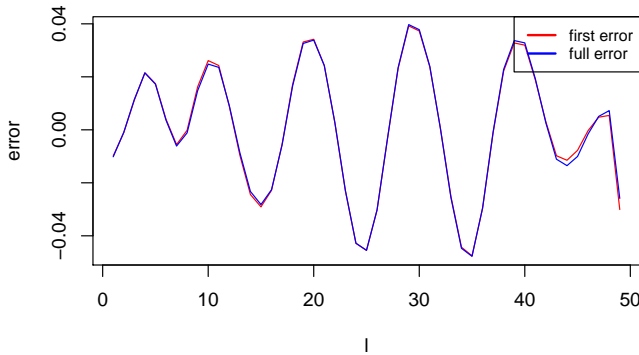


Рис. 1: Вещественные части первого порядка и полной ошибок.

Для случая возмущения в виде выброса был рассмотрен пример с сигналом $s_l = 1 + 1i$, с возмущением в виде выброса $a_1 + ia_2 = 10 + i10$ на позиции $k = L - 1$. 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] T. Kato. Perturbation theory for linear operators. *Springer-Verlag*, 1966.
- [6] V. Nekrutkin. Perturbation expansions of signal subspaces for long signals. *Statistics and Its Interface.*, Vol.3, P. 297-319, 2010.
- [7] V. Nekrutkin. Perturbations in SSA. *Manuscript*, 2008.
- [8] А. Константинов. Некоторые задачи анализа временных рядов (теория методов "Singal Subspace"). *Курсовая работа, науч. рук. В. Некруткин*, 2018.