Робастные варианты метода SSA

Сенов Михаил Андреевич

Санкт-Петербургский государственный университет
Уровень образования: бакалавриат
Направление 01.03.02 «Прикладная математика и информатика»
Основная образовательная программа СВ.5004.2018
«Прикладная математика и информатика»
Профессиональная траектория «Вычислительная стохастика и статистические модели»

Научный руководитель: к.ф.-м.н., доц. Голяндина Н.Э.
Рецензент: к.ф.-м.н. Пепелышев А.Н.

Санкт-Петербург, 2022

Введение

$\mathsf{X} = (x_1, \ldots, x_N)$ временной ряд длины N.

Модель: $\mathsf{X} = \mathsf{S} + \mathsf{R}$, S сигнал, R возмущение (шум или выброс).

Задача: Оценить сигнал $\tilde{\mathsf{S}} = F(\mathsf{X})$, F — используемый метод.

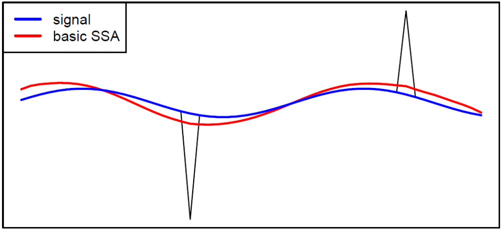
Метод: SSA (Singular Spectrum Analysis) для вещественных рядов, CSSA (Complex Singular Spectrum Analysis) — обобщение для комплексных рядов.
Golyandina, Nekrutkin, Zhigljavsky (2001). Analysis of time series structure: SSA and related techniques.

Вопросы:

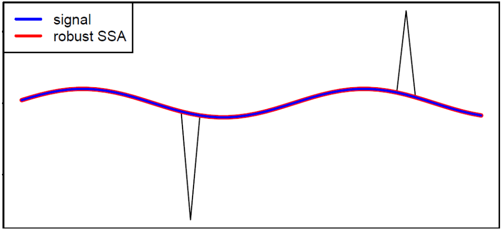
1. Устойчивые к выбросам (робастные) модификации CSSA? Робастные модификации SSA (вещ. случай) были предложены ранее (Третьякова, 2020).
2. Что лучше, с точки зрения величины ошибки $\tilde{\mathsf{S}} - \mathsf{S}$, $\mathsf{SSA}(\mathsf{X}_{\mathrm{Re}}) + \mathrm{i}\mathsf{SSA}(\mathsf{X}_{\mathrm{Im}})$ или $\mathsf{CSSA}(\mathsf{X}_{\mathrm{Re}} + \mathrm{i}\mathsf{X}_{\mathrm{Im}})$?

Часть I. Robust CSSA: Введение

Базовый SSA: реагирует на выбросы.



Робастный SSA: не реагирует на выбросы.



Часть I: Обозначения

Рассмотрим временной ряд $\mathsf{X} = (x_1, \ldots, x_N) = \mathsf{S} + \mathsf{R}$, L длина окна, $K = N - L + 1$.

Оператор вложения $\mathcal{T}_L : \mathbb{R}^N \to \mathcal{M}_{\mathcal{H}} : \mathcal{T}_L(\mathsf{X}) = \mathbf{X}$,

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \ldots & x_K \\ x_2 & x_3 & \ldots & x_{K+1} \\ \vdots & \vdots & & \vdots \\ x_L & x_{L+1} & \ldots & x_N \end{pmatrix}, K = N - L + 1,$$

$\mathbf{X}$ — L-траекторная матрица X.

Ранг сигнала $\operatorname{rk} \mathcal{T}_L(\mathsf{S}) = r$.

$\mathcal{M}_r$ — пространство матриц размера $L \times K$ ранга не больше r.
$\mathcal{M}_{\mathcal{H}}$ — пространство ганкелевых матриц $L \times K$.
$\mathcal{T}_L(\mathsf{S}) \in \mathcal{M}_{\mathcal{H}} \cap \mathcal{M}_r$.

$\Pi_r : \mathcal{M} \to \mathcal{M}_r$, $\Pi_{\mathcal{H}} : \mathcal{M} \to \mathcal{M}_{\mathcal{H}}$ — проекторы на $\mathcal{M}_r$ и $\mathcal{M}_{\mathcal{H}}$ по некоторой норме.

Часть I: Выделение сигнала

Временной ряд $\mathsf{X} = (x_1, \ldots, x_N) = \mathsf{S} + \mathsf{R}$, $\operatorname{rk} \mathcal{T}_L(\mathsf{S}) = r$.

Алгоритм для выделения сигнала

$$\tilde{\mathsf{S}} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}} \Pi_r \mathcal{T}_L(\mathsf{X}).$$

Нормы для Π_r и $\Pi_{\mathcal{H}}$:

- $\mathbb{L}_2$(Фробениус): выделение сигнала базовым SSA

$$\|\mathbf{X}\|_{\mathrm{F}} = \sqrt{\sum_{i=1} \sum_{j=1} |x_{ij}|^2}.$$

- $\mathbb{L}_1$: робастная версия SSA

$$\|\mathbf{X}\|_1 = \sum_{i=1} \sum_{j=1} |x_{ij}|.$$

- weighted $\mathbb{L}_2$: робастная версия SSA

$$\|\mathbf{W}^{1/2} \odot \mathbf{X}\|_{\mathrm{F}} = \sqrt{\sum_{i=1} \sum_{j=1} w_{ij} |x_{ij}|^2},$$

где выбросам соответствуют маленькие веса.

Часть I: Результаты

Алгоритмы робастных версий CSSA были построены и реализованы на языке R.

Алгоритмы основаны на решении следующих задач.
Траекторная матрица $\mathbf{Y} \in \mathbb{C}^{L\times K}$ на вход. Разложение $\mathbf{M}\mathbf{V}^{\mathrm{H}}$ на выходе, где $\mathbf{M} \in \mathbb{C}^{L\times r}, \mathbf{V} \in \mathbb{C}^{K\times r}, \mathbf{W} \in \mathbb{R}^{L\times K}$.

- $\mathbb{L}_2$ CSSA:
$$\|\mathbf{Y} - \mathbf{M}\mathbf{V}^{\mathrm{H}}\|_{\mathrm{F}} \longrightarrow \min_{\mathbf{M},\mathbf{V}},$$
имеет решение в замкнутой форме.
- $\mathbb{L}_1$ CSSA:
$$\|\mathbf{Y} - \mathbf{M}\mathbf{V}^{\mathrm{H}}\|_{1} \longrightarrow \min_{\mathbf{M},\mathbf{V}}.$$
решение вычисляется итеративно.
- weighted $\mathbb{L}_2$ CSSA:
$$\|\mathbf{W}^{1/2} \odot (\mathbf{Y} - \mathbf{M}\mathbf{V}^{\mathrm{H}})\|_{\mathrm{F}} \longrightarrow \min_{\mathbf{M},\mathbf{V}},$$
решение вычисляется итеративно, $\mathbf{W}$ обновляется на каждой итерации согласно величине остатков разложения.

Часть II. Ошибки в SSA и CSSA: Введение

$\mathsf{X} = (x_1, \ldots, x_N)$ временной ряд длины N.

Модель: $\mathsf{X} = \mathsf{S} + \mathsf{R}$, S сигнал, R возмущение (шум или выброс).

Задача: Оценить сигнал $\tilde{\mathsf{S}} = F(\mathsf{X})$, F — используемый метод.

Метод: SSA (Singular Spectrum Analysis) для вещественных рядов, CSSA (Complex Singular Spectrum Analysis) — обобщение для комплексных рядов.
Golyandina, Nekrutkin, Zhigljavsky (2001). Analysis of time series structure: SSA and related techniques.

Вопрос: Что лучше, с точки зрения величины ошибки $\tilde{\mathsf{S}} - \mathsf{S}$, $\mathsf{SSA}(\mathsf{X}_{\mathrm{Re}}) + \mathrm{i}\mathsf{SSA}(\mathsf{X}_{\mathrm{Im}})$ или $\mathsf{CSSA}(\mathsf{X}_{\mathrm{Re}} + \mathrm{i}\mathsf{X}_{\mathrm{Im}})$?

Рассматриваем теорию возмущений (Kato, 1966). Будем вычислять первый порядок ошибки (Nekrutkin, 2008), в предположении, что первый порядок достаточно точно описывает полную ошибку.

Часть II. Ошибки в SSA и CSSA: Структура

Структура Части 2:

1. Первый порядок ошибки для общего случая, теоретическое сравнение SSA и CSSA.
2. Пример: сигнал — гармоники, возмущение — шум. Особый случай комплексной экспоненты.
3. Пример: сигнал — константный сигнал, возмущение — выброс. Явный вид ошибки.
4. Численное сравнение первого порядка ошибки и полной ошибки.

Часть II: Обозначения

Временной ряд $\mathsf{X} = (x_1, \ldots, x_N)$, L — длина окна, r — ранг оцениваемого сигнала (ранга траекторной матрицы сигнала).

Алгоритм SSA (CSSA) выделения сигнала

$$\tilde{\mathsf{S}} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}} \Pi_r \mathcal{T}_L(\mathsf{X}).$$

Модель: $\mathsf{X} = \mathsf{S}(\delta)$, где $\mathsf{S}(\delta) = \mathsf{S} + \delta \mathsf{R}$ длины N
$\mathbf{H} = \mathcal{T}_L(\mathsf{S})$.

$\tilde{\mathsf{S}} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}}(\mathbf{H} + \delta \mathbf{H}^{(1)} + \delta^2 \mathbf{H}^{(2)})$ из (Nekrutkin, 2008).

$\mathsf{F} = \tilde{\mathsf{S}} - \mathsf{S} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}}(\delta \mathbf{H}^{(1)} + \delta^2 \mathbf{H}^{(2)})$ ошибка восстановления.

Рассматриваем $\delta = 1$,

$\boxed{\mathsf{F}^{(1)} = \mathcal{T}_L^{-1} \Pi_{\mathcal{H}}(\mathbf{H}^{(1)})}$ первый порядок ошибки восстановления.

Часть II: Формула для $\mathbf{H}^{(1)}$

Итак, рассматриваем $\delta = 1$, $\mathsf{X} = \mathsf{S}(1) = \mathsf{S} + \mathsf{R}$,
оценку сигнала $\tilde{\mathsf{S}} = \mathcal{T}_L^{-1}\Pi_{\mathcal{H}}(\mathbf{H} + \mathbf{H}^{(1)} + \mathbf{H}^{(2)}))$.
Хотим найти $\mathbf{H}^{(1)}$, поскольку ошибка $\mathsf{F} \approx \mathsf{F}^{(1)} = \mathcal{T}_L^{-1}\Pi_{\mathcal{H}}(\mathbf{H}^{(1)})$.

Утверждение

Пусть R достаточно мало. Тогда

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

где $\mathbf{P}_0$ — проектор на пространство столбцов $\mathbf{H}$,
$\mathbf{Q}_0$ — проектор на пространство строк $\mathbf{H}$,
$\mathbf{P}_0^{\perp} = \mathbf{I} - \mathbf{P}_0$, $\mathbf{I}$ — единичная матрица,
$\mathbf{E} = \mathcal{T}_L(\mathsf{R})$.

Получено на основе результатов из (Константинов, 2018) и (Некруткин, 2010).

Часть II: Теорема

Первые порядки ошибки восстановления:

- CSSA-F$^{(1)}$: сигнал S, возмущение R, метод CSSA,
- SSA-F$_{\mathrm{Re}}^{(1)}$: сигнал $\mathrm{Re}(\mathsf{S})$, возмущение $\mathrm{Re}(\mathsf{R})$, метод SSA,
- SSA-F$_{\mathrm{Im}}^{(1)}$: сигнал $\mathrm{Im}(\mathsf{S})$, возмущение $\mathrm{Im}(\mathsf{R})$, метод SSA.

Теорема

Пусть пространства столбцов траекторных матриц рядов S, $\mathrm{Re}(\mathsf{S})$ и $\mathrm{Im}(\mathsf{S})$ совпадают и то же самое верно для пространств строк. Тогда при любом достаточно малом возмущении R

$$\text{CSSA-F}^{(1)} = \text{SSA-F}_{\mathrm{Re}}^{(1)} + \mathrm{i}\,\text{SSA-F}_{\mathrm{Im}}^{(1)}.$$

Получается из линейности вхождения $\mathbf{E}$ в формулу

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

Часть II: Пример, две зашумлённые синусоиды

Сигнал:

$$s_l = A\cos(2\pi\omega l + \phi_1) + \mathrm{i}B\cos(2\pi\omega l + \phi_2),$$

где $0 < \omega \leq 0.5$ и $0 \leq \phi_i < 2\pi$.

Особый случай: При $|\phi_2 - \phi_1| = \pi/2$ и $A = B$, комплексная экспонента,

$$s_l = Ae^{\pm\mathrm{i}(2\pi\omega l + \phi_1)}.$$

Возмущение: Случайный стационарный процесс с нулевым матожиданием и достаточно малой дисперсией.

Обозначения:

CSSA-$\mathrm{F}^{(1)} = ($CSSA-$f_1^{(1)}, \ldots,$ CSSA-$f_N^{(1)})$,

SSA-$\mathrm{F}_{\mathrm{Re}}^{(1)} = ($SSA-$f_{\mathrm{Re},1}^{(1)}, \ldots,$ SSA-$f_{\mathrm{Re},N}^{(1)})$,

SSA-$\mathrm{F}_{\mathrm{Im}}^{(1)} = ($SSA-$f_{\mathrm{Im},1}^{(1)}, \ldots,$ SSA-$f_{\mathrm{Im},N}^{(1)})$.

Часть II: Пример, две зашумлённые синусоиды. MSE

Следствие

Для сигнала $s_l = A\cos(2\pi\omega l + \phi_1) + \mathrm{i}B\cos(2\pi\omega l + \phi_2)$, не являющегося комплексной экспонентой, с достаточно малым возмущением R выполняется

$$\mathbb{D}(\text{CSSA-}f_l^{(1)}) = \mathbb{D}(\text{SSA-}f_{\mathrm{Re},l}^{(1)}) + \mathbb{D}(\text{SSA-}f_{\mathrm{Im},l}^{(1)}).$$

Показано, используя (Степанов, Голяндина, 2005).

Предположение

Для сигнала $s_l = Ae^{\pm\mathrm{i}(2\pi\omega l + \phi_1)}$, с достаточно малым возмущением R выполняется

$$\mathbb{D}(\text{CSSA-}f_l^{(1)}) \stackrel{?}{=} \frac{1}{2}[\mathbb{D}(\text{SSA-}f_{\mathrm{Re},l}^{(1)}) + \mathbb{D}(\text{SSA-}f_{\mathrm{Im},l}^{(1)})].$$

Показано эмпирически.

Сигнал:

$$s_l = c_1 + \mathrm{i}c_2.$$

Возмущение: $\mathsf{R} = (0, \ldots, a_1 + \mathrm{i}a_2, \ldots, 0)$ — достаточно малый выброс на позиции k.

Тракторные пространства сигнала совпадают, возмущение достаточно малое, справедлива теорема

$$\text{CSSA-F}^{(1)} = \text{SSA-F}^{(1)}_{\mathrm{Re}} + \mathrm{i}\text{SSA-F}^{(1)}_{\mathrm{Im}}.$$

Достаточно уметь вычислять $\text{SSA-F}^{(1)}_{\mathrm{Re}} = \mathcal{T}_L^{-1}\Pi_{\mathcal{H}}(\mathbf{H}^{(1)}_{\mathrm{Re}})$.

Для $\mathbf{H}^{(1)}_{\mathrm{Re}}$ известна формула (Nekrutkin, 2008)

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

где $U = \{1/\sqrt{L}\}_{i=1}^{L}$, $V = \{1/\sqrt{K}\}_{i=1}^{K}$, $K = N - L + 1$.

Часть II: Пример, константный сигнал с выбросом, явный вид SSA-$f_{\mathrm{Re},l}^{(1)}$

SSA-$f_{\mathrm{Re},l}^{(1)} = \left(\mathcal{T}_L^{-1}\Pi_{\mathcal{H}}(\mathbf{H}_{\mathrm{Re}}^{(1)})\right)_l$

Приведем результат для случая $k \le \min(L/2, K-L)$ и $L < K$, где $K = N - L + 1$:

$$\text{SSA-}f_{\mathrm{Re},l}^{(1)} = \frac{a}{LK}\begin{cases} (L+K-k), & 1 \le l \le k \\ \frac{1}{l}(L+K-l)k, & k < l \le L \\ \frac{1}{L}K(L+k-l), & L < l < L+k \\ 0, & L+k \le l \le K \\ \frac{1}{N-l+1}(K-l)(L-k), & K < l < K+k \\ -k, & K+k \le l \le N \end{cases}.$$

Замечание: При фиксированном L первый порядок ошибки не стремится к 0 с ростом N.

Часть II: Сравнение первого порядка и полной ошибок

Зашумлённые синусоиды: Численно было показано, первый порядок адекватно описывает полную ошибку.

Константный сигнал с выбросом: Численно было показано, первый порядок адекватно оценивает полную ошибку при $L = \alpha N$ для больших N. При фиксированном L это не так.

Пример с выбросом:

$$s_l = 1 + \mathrm{i},$$

R — выброс $10 + \mathrm{i}10$ на позиции $k = L - 1$.

Таблица: 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

Основные результаты

- Реализации на R: Робастные модификации были обобщены на комплексный случай.
- Теория: При совпадении траекторных пространств комплексного ряда первые порядки ошибок для CSSA и SSA совпадают.
- Пример:
  - $s_l = A\cos(2\pi\omega l + \phi_1) + \mathrm{i}B\cos(2\pi\omega l + \phi_2)$, сдвиг не $\pi/2$
  $$\mathbb{D}(\text{CSSA-}f_l^{(1)}) = \mathbb{D}(\text{SSA-}f_{\mathrm{Re},l}^{(1)}) + \mathbb{D}(\text{SSA-}f_{\mathrm{Im},l}^{(1)}).$$
  - $s_l = Ae^{\pm\mathrm{i}(2\pi\omega l + \phi_1)}$
  $$\mathbb{D}(\text{CSSA-}f_l^{(1)}) \stackrel{?}{=} \frac{1}{2}[\mathbb{D}(\text{SSA-}f_{\mathrm{Re},l}^{(1)}) + \mathbb{D}(\text{SSA-}f_{\mathrm{Im},l}^{(1)})].$$
- Пример: Для $s_l = c_1 + \mathrm{i}c_2$ с выбросом была получена аналитическая формула первого порядка. При $L = \alpha N$ ошибка стремится к 0.
- Численные эксперименты: Рассмотрен вопрос приближения полной ошибки первым порядком.