

Тензорный анализ сингулярного спектра

Хромов Никита Андреевич, гр.20.Б04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Научный руководитель: д.ф.-м.н. Голяндина Н.Э.

Рецензент: к.ф.-м.н. Усевич К.Д.

Санкт-Петербург,
2024

Введение

$X^{(q)} = (x_1^{(q)}, x_2^{(q)}, \dots, x_N^{(q)})^T$, $x_i \in \mathbb{R}$ — временной ряд

$X = (X^{(1)} : X^{(2)} : \dots : X^{(Q)})$ — Q -канальный временной ряд

$$X = T + P + E$$

T — медленно меняющаяся составляющая (тренд)

P — периодическая составляющая (сезонность)

E — случайная составляющая (шум)

Возможные задачи:

1. Выделение сигнала из ряда: нахождение $S = T + P$
2. Разделение сигнала: нахождение составляющих T и P

Возможный метод решения: Singular Spectrum Analysis ([SSA](#))

[Broomhead, King (1986a)], [Golyandina, Nekrutkin, Zhigljavsky (2001)],

и его многомерное расширение Multivariate SSA ([MSSA](#))

[Broomhead, King (1986b)]

Цель: реализация тензорных расширений методов SSA и MSSA, исследование их свойств с точки зрения точности выделения сигнала и разделения компонент, сравнение расширений с базовыми методами.

HO-SSA: алгоритм

$X = \sum_{m=1}^M S_m + E$ — временной ряд,

$S \stackrel{\text{def}}{=} \sum_{m=1}^M S_m$ — сигнал.

Параметры: $I, L < N$ — длины окна, $I + L \leq N + 1$, $J = N - I - L + 2$,
 R — число элементов разложения, относимых к сигналу,
 $\mathcal{I}_1, \dots, \mathcal{I}_M \subseteq \{1, 2, \dots, R\}$, $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$ — наборы индексов элементов
 разложения, относимых к составляющим сигнала.

Схема алгоритма HO-SSA для разделения компонент сигнала

- ① Вложение $X \xrightarrow{I, L} \mathcal{X} \in \mathbb{R}^{I \times L \times J}$ — траекторный тензор,
- ② Разложение

$$\text{HOSVD}(\mathcal{X}) = \sum_{i=1}^{d_1} \sum_{l=1}^{d_2} \sum_{j=1}^{d_3} \mathcal{Z}_{ilj} U_i^{(1)} \circ U_l^{(2)} \circ U_j^{(3)}, \quad d_1 \leq I, d_2 \leq L, d_3 \leq J$$

- ③ Группировка

$$\tilde{S}_m = \sum_{i \in \mathcal{I}_m} \sum_{l \in \mathcal{I}_m} \sum_{j \in \mathcal{I}_m} \mathcal{Z}_{ilj} U_i^{(1)} \circ U_l^{(2)} \circ U_j^{(3)}, \quad R \leq \min(d_1, d_2, d_3)$$

- ④ Восстановление усреднение \tilde{S}_m вдоль обобщённых антидиагоналей
 $i + l + j = \text{const.}$

Результат алгоритма \tilde{S}_m — оценки составляющих S_m

HO-SSA: Траекторная матрица

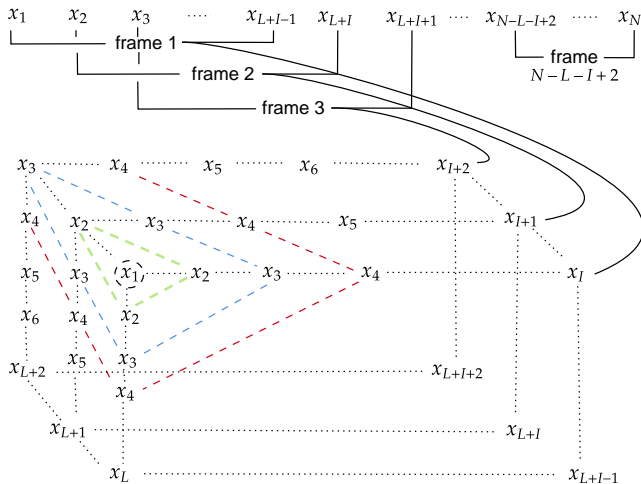
$$\mathbf{X} = (x_1, x_2, \dots, x_N)$$

$$1 < L < N, K = N - L + 1.$$

Определение (Траекторная матрица ряда \mathbf{X} с длиной окна L)

$$\mathbf{X} \xrightarrow{L} \mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}.$$

HO-SSA: траекторный тензор



I, L — параметры длины окна

HO-SSA: разложение и группировка

Определение

n -ранг тензора \mathcal{X} ($\text{rank}_n(\mathcal{X})$) — размерность пространства n -столбцов \mathcal{X} .

Идея выделения сигнала: приближение \mathcal{X} тензором $\tilde{\mathcal{X}}$ меньших n -рангов.

Способы приближения меньшими рангами:

① Усечение HOSVD:

$$\mathcal{X} = \sum_{i=1}^{d_1} \sum_{l=1}^{d_2} \sum_{j=1}^{d_3} \mathcal{Z}_{ilj} U_i^{(1)} \circ U_l^{(2)} \circ U_j^{(3)} \mapsto \sum_{i=1}^{R_1} \sum_{l=1}^{R_2} \sum_{j=1}^{R_3} \mathcal{Z}_{ilj} U_i^{(1)} \circ U_l^{(2)} \circ U_j^{(3)} = \tilde{\mathcal{X}}$$

② Higher-Order Orthogonal Iteration (HOOI):

$$\tilde{\mathcal{X}} = \text{HOSVD} \left(\arg \min_{\mathcal{Y}} \|\mathcal{X} - \mathcal{Y}\|_F \right), \quad \text{rank}_n(\mathcal{Y}) = R_n$$

В HO-SSA $R_1 = R_2 = R_3 = R$

HO-SSA: разделимость и ранги рядов в терминах SSA

- **Выделение сигнала:** нахождение ранга ряда

Пусть S — сигнал, S_L — его траекторная матрица с длиной окна L .

Определение (Ранг сигнала в терминах SSA)

S имеет ранг R , если $R \leq N/2$ и $\forall L: R \leq \min(L, N - L + 1)$

$$\text{rank}(S_L) = R$$

- **Разделение составляющих сигнала** пусть $S = S_1 + S_2$,
 S, S_1, S_2 — траекторные матрицы этих сигналов с длиной окна L

Определение (Разделимость в терминах SSA)

Сигналы S_1 и S_2 L -разделимы, если существуют такие $\mathcal{J}_1, \mathcal{J}_2$, что

$$S = \sum_{j=1}^R \sqrt{\lambda_j} U_j V_j^T = \overbrace{\sum_{j \in \mathcal{J}_1} \sqrt{\lambda_j} U_j V_j^T}^{S_1} + \overbrace{\sum_{j \in \mathcal{J}_2} \sqrt{\lambda_j} U_j V_j^T}^{S_2}$$

HO-SSA: n -ранги траекторного тензора

Теорема (О связи рангов рядов в SSA и HO-SSA)

Пусть сигнал S имеет ранг R в терминах SSA.

Тогда для любых значений параметров I и L таких, что

$$R \leq \min(I, L, N - I - L + 2),$$

$\text{rank}_1(S) = \text{rank}_2(S) = \text{rank}_3(S) = R$, где S — траекторный тензор S , построенный по длинам окна I, L .

Замечание

Теорема позволяет использовать известные результаты о рангах сигналов из теории SSA для аппроксимации траекторного тензора на шагах разложения и группировки в методе HO-SSA.

HO-SSA: разделимость

Пусть $S = S_1 + S_2$,

S, S_1, S_2 — траекторные тензоры этих сигналов с длинами окна I и L

Теорема (О связи разделимости в SSA и HO-SSA)

HOSVD тензора S можно представить в виде суммы HOSVD тензоров S_1 и S_2 тогда и только тогда, когда сигналы S_1 и S_2 I - и L -разделимы в терминах SSA.

Замечание

Теорема позволяет выделить класс сигналов, которые можно разделить методом HO-SSA, а также даёт рекомендации к выбору параметров I и L .

HOSVD-MSSA: алгоритм

Пусть $X = \sum_{m=1}^M S_m + E$ — Q -канальный временной ряд

Параметры: L — длина окна, $K = N - L + 1$,

$R, \mathfrak{I}_1, \dots, \mathfrak{I}_M$ — как в HO-SSA,

R_3 — число элементов разложения по третьему направлению,

$\Omega_1, \dots, \Omega_M \subseteq \{1, 2, \dots, R_3\}$, $\Omega_i \cap \Omega_j = \emptyset$ — индексы группировки по третьему направлению

- **Вложение** $X \xrightarrow{L} \mathcal{X} \in \mathbb{R}^{L \times K \times Q}$ — траекторный тензор
- **Разложение**

$$\mathcal{X} = \sum_{l=1}^{d_1} \sum_{k=1}^{d_2} \sum_{q=1}^{d_3} \mathcal{Z}_{lkp} U_l^{(1)} \circ U_k^{(2)} \circ U_q^{(3)}$$

- **Группировка**

$$\tilde{S}_m = \sum_{l \in \mathfrak{I}_m} \sum_{k \in \mathfrak{I}_m} \sum_{q \in \Omega_m} \mathcal{Z}_{lkp} U_l^{(1)} \circ U_k^{(2)} \circ U_q^{(3)}$$

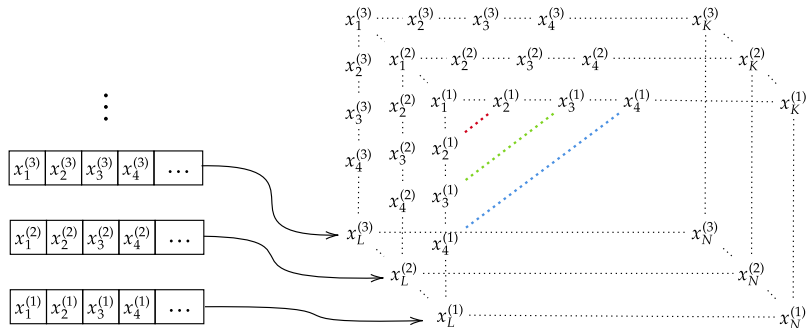
- **Восстановление** усреднение сечений третьего направления \tilde{S}_m по побочным диагоналям

Результат алгоритма: \tilde{S}_m — оценка S_m

HOSVD-MSSA: траекторный тензор многоканального ряда

X — многоканальный временной ряд длины N

L — длина окна, $K = N - L + 1$



HOSVD-MSSA: n -ранги траекторного тензора

Теорема (О n -рангах траекторного тензора многоканального ряда)

Пусть $S = (S^{(1)} : \dots : S^{(Q)})$, тогда справедливы следующие утверждения.

- 1 S имеет ранг R в терминах теории MSSA тогда и только тогда, когда для траекторного тензора S , построенного по любой длине окна $L < N$ такой, что $R \leq \min(L, K)$ выполняется

$$\text{rank}_1(S) = \text{rank}_2(S) = R.$$

- 2 $\text{rank}_3(S)$ равен рангу матрицы, в строках которой содержатся каналы сигнала $S^{(q)}$.

Замечание

Теорема позволяет использовать известные результаты о рангах сигналов из теории MSSA для аппроксимации траекторного тензора по первым двум направлениям, а также даёт рекомендации к выбору ранга аппроксимации по третьему направлению на шагах разложения и группировки в методе HOSVD-MSSA.

HOSVD-MSSA: разделимость

Пусть $S = S_1 + S_2$,

S, S_1, S_2 — траекторные тензоры этих сигналов с длиной окна L

$$\Lambda^{(I)}(S) = \text{span} \left\{ \left(s_i^{(q)}, s_{i+1}^{(q)}, \dots, s_{i+I-1}^{(q)} \right) \right\}$$

Теорема (О разделимости методом HOSVD-MSSA)

HOSVD тензора S можно представить в виде суммы HOSVD тензоров S_1 и S_2 тогда и только тогда, когда $\Lambda^{(L)}(S_1) \perp \Lambda^{(L)}(S_2)$ и $\Lambda^{(K)}(S_1) \perp \Lambda^{(K)}(S_2)$

Замечание

Теорема позволяет выделить класс сигналов, которые можно разделить методом HOSVD-MSSA, а также даёт рекомендации к выбору параметра L .

Численные результаты: сравнение HO-SSA с SSA

Пусть временной ряд имеет вид

$$X = (s_1 + \varepsilon_1, s_2 + \varepsilon_2, \dots, s_N + \varepsilon_N),$$

где $N = 71$, $s_n = 30 \cos(2\pi n/12)$, ε_n — шум.

Table: RMSE оценки сигнала: SSA.

Вид шума	$L = 12$	$L = 24$	$L = 30$	$L = 36$
Белый, $\sigma^2 = 25$	1.82	1.42	1.40	1.42
Красный, $\delta^2 = 5$, $\varphi = 0.5$	1.31	1.03	1.01	1.03
Красный, $\delta^2 = 5$, $\varphi = 0.9$	1.88	1.37	1.34	1.36

Table: RMSE оценки сигнала: HO-SSA.

Вид шума \ $I \times L$	19×30	12×31	7×36	12×37	12×49
Белый, $\sigma^2 = 25$	1.62	1.56	1.49	1.53	1.63
Красный, $\delta^2 = 5$, $\varphi = 0.5$	1.19	1.14	1.08	1.12	1.17
Красный, $\delta^2 = 5$, $\varphi = 0.9$	1.51	1.44	1.39	1.42	1.56

Численные результаты: сравнение HOSVD-MSSA с MSSA и 2D-SSA

$$X = (X_1 : \dots : X_Q), \quad X_q = \left(x_1^{(q)}, \dots, x_N^{(q)} \right)^T, \quad x_n^{(q)} = \hat{s}_n^{(q)} + \tilde{s}_n^{(q)} + \varepsilon_n^{(q)},$$

где $\varepsilon_n^{(q)} \sim N(0, 0.01)$ и независимы, модель: $s_n^{(q)} = C_q \cos(2\pi n \omega_q + \varphi_q)$

Вид сигнала	MSSA	HOSVD-MSSA	2D-SSA
Равные сигналы	0.026	0.019	0.014
	0.025	0.016	0.014
Различие амплитуд	0.029	0.019	0.086
	0.029	0.019	0.083
Линейные фазы	0.026	0.025	0.117
	0.025	0.025	0.114
Произвольные фазы	0.026	0.025	0.034
	0.025	0.025	0.033
Разделимость с const	0.017	0.017	0.023
	0.025	0.019	0.033
Различие частот	0.024	0.018	0.012
	0.024	0.018	0.031
	0.024	0.014	0.026
Ортогональность по каналам	0.031	0.023	0.025
	0.030	0.022	0.024

Численные результаты: смещение и дисперсия

$X = \hat{S} + \tilde{S} + E$, где $\hat{s}_n^{(1)} = 3$, $\hat{s}_n^{(2)} = -1.5$, $\tilde{s}_n^{(1)} = \cos(2\pi n/20)$ и $\tilde{s}_n^{(2)} = 2 \cos(2\pi n/20)$, а $\varepsilon_n^{(p)} \sim N(0, 1)$ и независимы.

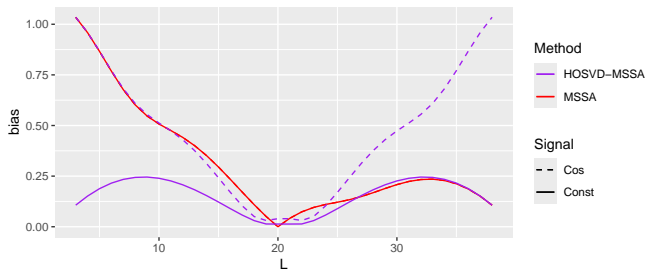


Figure:
Смещение оценок каждым методом (RMSE без шума)

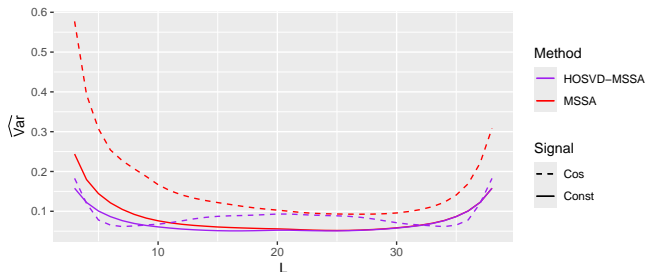
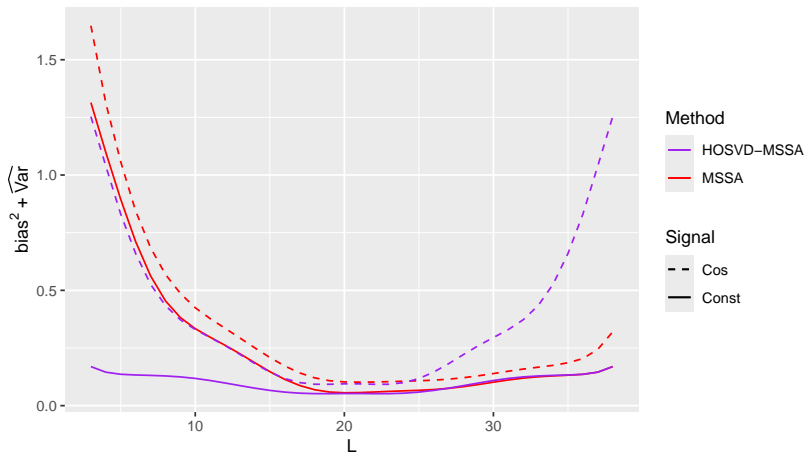


Figure:
Дисперсии оценок каждым методом

Численные результаты: сумма смещения и дисперсии



Оценки методом HOSVD-MSSA могут иметь меньшую дисперсию, чем методом MSSA, однако могут проигрывать в точности за счёт бóльшего смещения при некоторых L .

Результаты

1 Свойства HO-SSA

- Критерий разделимости HO-SSA, связь с разделимостью SSA
- Связь рангов траекторного тензора с SSA-рангом ряда
- Трудоёмкость алгоритма

2 Свойства HOSVD-MSSA

- Критерий разделимости HOSVD-MSSA
- Теорема о рангах траекторного тензора, связь с MSSA-рангом и 3-ранг
- Утверждение о симметричности траекторного тензора относительно замены длины окна L на K

3 Численные выводы

- Отсутствие преимущества HO-SSA над SSA
- Преимущество HOSVD-MSSA над MSSA в условиях разделимости
- Возможное преимущество HOSVD-MSSA над MSSA в условиях приближённой разделимости для HOSVD-MSSA в зависимости от соотношения смещения и дисперсии

4 Реализация методов HO-SSA и HOSVD-MSSA на языке R в стиле пакета Rssa (исходный код опубликован в репозитории Zenodo)



Broomhead D. S., King G. P. Extracting qualitative dynamics from experimental data // Physica D: Nonlinear Phenomena. — 1986. — Vol. 20, no. 2–3. — P. 217–236.



Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of time series structure: SSA and related techniques. — Chapman & Hall/CRC, 2001.



Broomhead D. S., King G. P. On the Qualitative Analysis of Experimental Dynamical Systems // Nonlinear Phenomena and Chaos. — 1986. — P. 113–144.

Трудоёмкости алгоритмов

- **HOSVD-SSA:** вычисление HOSVD тензора размерности $I \times L \times J$ имеет трудоёмкость порядка

$$O(ILJ(\min(I, LJ) + \min(L, IJ) + \min(J, IL))).$$

Если требуется вычислить только усечение HOSVD с n -рангами (r_1, r_2, r_3) , то трудоёмкость можно уменьшить до порядка

$$O(ILJ(r_1 + r_2 + r_3)).$$

- **HOOI-SSA:** HOOI — итеративный алгоритм. Начальное приближение: усечение HOSVD. Трудоёмкость каждой итерации имеет порядок

$$O(r_1 r_2 r_3 (I + L + J)),$$

а скорость сходимости алгоритма линейная. Итого для достижения точности ε :

$$O\left(ILJ(r_1 + r_2 + r_3) + \frac{1}{\varepsilon} r_1 r_2 r_3 (I + L + J)\right)$$

Варианты сигналов в сравнении HOSVD-MSSA с MSSA и 2D-SSA

- ❶ Равные сигналы: $N = 44, P = 12$,

$$\hat{s}_n^{(p)} = 2 \cos(2\pi n/5), \quad \tilde{s}_n^{(p)} = \cos(2\pi n/3).$$

- ❷ Различие амплитуд: $N = 44, P = 12$,

$$\hat{s}_n^{(p)} = 2c_1^{(p)} \cos(2\pi n/5), \quad \tilde{s}_n^{(p)} = c_2^{(p)} \cos(2\pi n/3).$$

- ❸ Линейные фазы: $N = 44, P = 12$,

$$\hat{s}_n^{(p)} = 2c_1^{(p)} \cos(2\pi n/5 + p\pi/6), \quad \tilde{s}_n^{(p)} = c_2^{(p)} \cos(2\pi n/3 + p\pi/9).$$

- ❹ Произвольные фазы: $N = 44, P = 12$,

$$\hat{s}_n^{(p)} = 2c_1^{(p)} \cos(2\pi n/5 + \varphi_1^{(p)}), \quad \tilde{s}_n^{(p)} = c_2^{(p)} \cos(2\pi n/3 + \varphi_2^{(p)}).$$

- ❺ Разделимость с константой: $N = 44, P = 12$,

$$\hat{s}_n^{(p)} = 3c_1^{(p)}, \quad \tilde{s}_n^{(p)} = c_2^{(p)} \cos(2\pi n/3).$$

- ❻ Различие частот: $N = 59, P = 12$,

$$\hat{s}_n^{(p)} = 2 \cos(2\pi n/5), \quad \tilde{s}_n^{(p)} = \begin{cases} \cos(2\pi n/3), & 1 \leq p \leq 10, \\ 0.4 \cos(2\pi n/6), & 11 \leq p \leq 12. \end{cases}$$

- ❼ Ортогональность по каналам: $N = 29, P = 12$,

$$\hat{s}_n^{(p)} = 2 \cos(2\pi n/5) \cos(2\pi p/3), \quad \tilde{s}_n^{(p)} = 0.5 \cos(2\pi n/3) \cos(2\pi p/6).$$