

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

Постановка задачи

$X_p = (x_1^{(p)}, x_2^{(p)}, \dots, x_N^{(p)})$, $x_i \in \mathbb{R}$ — одномерный временной ряд

$X = (X^{(1)} : X^{(2)} : \dots : X^{(P)})$ — многомерный временной ряд

$$X = T + P + E$$

T — медленно меняющаяся компонента (тренд)

P — периодическая компонента (сезонность)

E — случайная компонента (шум)

Возможные задачи:

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

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

[Golyandina, Nekrutkin et al. (2001)],

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

[Голяндина, Степанов (2005)]

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.

Схема методов SSA и MSSA

$\mathbf{X} = \sum_{k=1}^m \mathbf{X}_k + \mathbf{E}$ — (многомерный) временной ряд,

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

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, относимых к компонентам сигнала.

Схема алгоритма MSSA для разделения компонент сигнала

❶ **Вложение** $\mathbf{X} \xrightarrow{L} \mathbf{X} = [\mathbf{X}^{(1)} : \mathbf{X}^{(2)} : \dots : \mathbf{X}^{(P)}] \in \mathbb{R}^{L \times KP}$,

$$\mathbf{X}^{(p)} = [X_1^{(p)} : X_2^{(p)} : \dots : X_K^{(p)}] \in \mathbb{R}^{L \times K},$$

$$X_i^{(p)} = (x_i^{(p)}, x_{i+1}^{(p)}, \dots, x_{i+L-1}^{(p)})^T$$

❷ **Разложение** $\mathbf{X} = \sum_{j=1}^d \sqrt{\lambda_j} U_j V_j^T$, $d \leq \min(L, KP)$

❸ **Группировка** $\tilde{\mathbf{X}}_k = \sum_{j \in \mathcal{I}_k} \sqrt{\lambda_j} U_j V_j^T$, $R \leq d$

$$\tilde{\mathbf{X}}_k = [\tilde{\mathbf{X}}_k^{(1)} : \tilde{\mathbf{X}}_k^{(2)} : \dots : \tilde{\mathbf{X}}_k^{(P)}]$$

❹ **Восстановление** усреднение $\tilde{\mathbf{X}}_k^{(p)}$ вдоль побочных диагоналей

Результат алгоритма $\tilde{\mathbf{X}}_k$ — оценки компонент \mathbf{X}_k

Разделимость и ранги рядов

- **Разделимость:** пусть $X = X_1 + X_2$

$$X = \sum_{j=1}^d \sqrt{\lambda_j} U_j V_j^T \stackrel{?}{=} \overbrace{\sum_{j \in \mathcal{I}_1} \sqrt{\lambda_j} U_j V_j^T}^{X_1} + \overbrace{\sum_{j \in \mathcal{I}_2} \sqrt{\lambda_j} U_j V_j^T}^{X_2}$$

Определение

Траекторное пространство $\Lambda^{(1,L)}(X)$ — пространство столбцов матрицы X , построенной по длине окна L , $\Lambda^{(2,K)}(X)$ — пространство её строк.

X_1 и X_2 L -разделимы $\iff \Lambda^{(1,L)}(X_1) \perp \Lambda^{(1,L)}(X_2)$ и $\Lambda^{(2,K)}(X_1) \perp \Lambda^{(2,K)}(X_2)$

- **Ранг ряда:** выбор параметра R ?

Определение

X имеет ранг d ($\text{rank } X = d$), если $\forall L: d \leq \min(L, N - L + 1)$
 $\dim \Lambda^{(L)}(X) = d$

Рекомендуемый выбор $R = \text{rank } X$

Переход к тензорам

Базовый алгоритм: ряд $X \Rightarrow$ матрица $\mathbf{X} \Rightarrow$ SVD \mathbf{X}

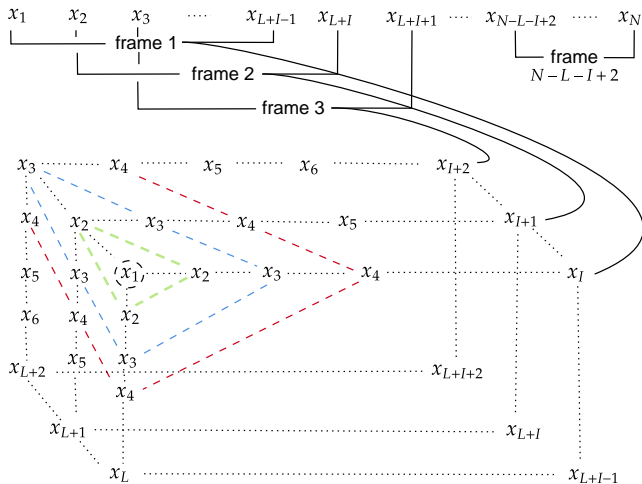
Тензорный алгоритм: ряд $X \Rightarrow$ тензор $\mathcal{X} \Rightarrow$ тензорное разложение \mathcal{X}

Тензорные разложения, расширяющие SVD [Sidiropoulos et al. (2016)]:

- Higher-order singular value decomposition (HOSVD)
- Canonical polyadic decomposition (CPD)
- $(L_r, L_r, 1)$ -decomposition

Тензорный переход с HOSVD для оценки параметров сигнала особого вида: [Pary et al. (2005)]

Траекторный тензор



I, L — параметры длины окна

Higher-Order SSA

$X = \sum_{k=1}^m X_k + E$ — одномерный временной ряд

Параметры алгоритма: $I, L < N$ — длины окна, $I + L < N + 1$,
 $J = N - I - L + 2$; $R, \mathfrak{I}_1, \dots, \mathfrak{I}_m$ — как в SSA

Схема алгоритма HO-SSA для разделения компонент сигнала

- ① Вложение $X \xrightarrow{I, L} \mathcal{X}$ — траекторный тензор,
- ② Разложение

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

- ③ Группировка

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

- ④ Восстановление усреднение $\tilde{\mathcal{X}}_k$ вдоль побочных сечений
 $i + l + j = \text{const}$

Результат алгоритма \tilde{X}_k — оценки компонент X_k

Разделимость и ранг в HO-SSA

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

Временные ряды \tilde{X} и \hat{X} длины N слабо I - и L -разделимы в терминах SSA тогда и только тогда, когда существует такое HOSVD траекторного тензора \mathcal{X} ряда $X = \tilde{X} + \hat{X}$, что его можно представить в виде суммы HOSVD траекторных тензоров рядов \tilde{X} и \hat{X} .

Определение

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

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

Пусть ряд X имеет ранг d в терминах SSA.

Тогда для любых значений параметров I и L таких, что

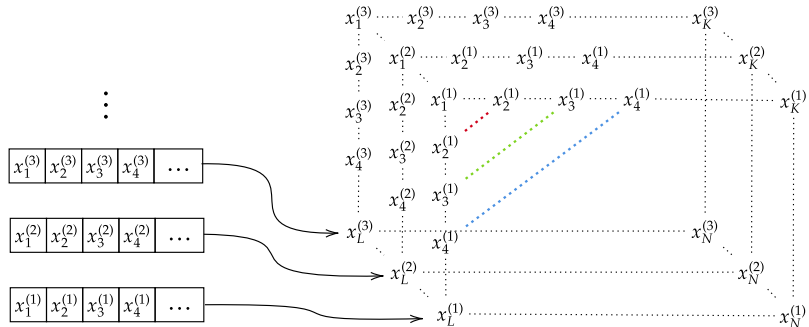
$$d \leq \min(I, L, N - I - L + 2),$$

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

Траекторный тензор многомерного ряда

X — многомерный временной ряд длины N

L — длина окна, $K = N - L + 1$



HOSVD-MSSA для разделения компонент сигнала

Пусть $X = \sum_{m=1}^M X_m + E$ — P -мерный временной ряд

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

$R, \mathfrak{I}_1, \dots, \mathfrak{I}_M$ — как в MSSA

R_3 — число элементов разложения по третьему направлению

$\mathfrak{P}_1, \dots, \mathfrak{P}_M \subseteq \{1, 2, \dots, R_3\}$, $\mathfrak{P}_i \cap \mathfrak{P}_j = \emptyset$ — индексы группировки по третьему направлению

- **Вложение** $X \xrightarrow{L} \mathcal{X}$ — траекторный тензор
- **Разложение**

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

- **Группировка**

$$\tilde{\mathcal{X}}_m = \sum_{l \in \mathfrak{I}_m} \sum_{k \in \mathfrak{I}_m} \sum_{p \in \mathfrak{P}_m} \mathcal{Z}_{lkp} U_l^{(1)} \circ U_k^{(2)} \circ U_p^{(3)}$$

- **Восстановление** усреднение сечений третьего направления $\tilde{\mathcal{X}}_m$ по побочным диагоналям

Результат алгоритма: \tilde{X}_m — оценка X_m

Ранг сигнала в HOSVD-MSSA

Теорема

Пусть X — P -мерный временной ряд длины N , тогда справедливы следующие утверждения.

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

$$\text{rank}_1(\mathcal{X}) = \text{rank}_2(\mathcal{X}) = d.$$

- 2 $\text{rank}_3(\mathcal{X})$ равен рангу матрицы, в строках которой содержатся одномерные временные ряды, составляющие заданный многомерный ряд.

Разделимость в HOSVD-MSSA

\hat{X}, \tilde{X}, X — многомерные временные ряды, $X = \hat{X} + \tilde{X}$

$\hat{\mathcal{X}}, \tilde{\mathcal{X}}, \mathcal{X}$ — траекторные тензоры рядов \hat{X}, \tilde{X}, X с длиной окна L

$$\Lambda_I(X) = \text{span} \left\{ \left(x_i^{(p)}, x_{i+1}^{(p)}, \dots, x_{i+I-1}^{(p)} \right) \right\}$$

Теорема

Такие HOSVD тензоров $\hat{\mathcal{X}}$ и $\tilde{\mathcal{X}}$, что их сумма является HOSVD тензора \mathcal{X} существуют тогда и только тогда, когда $\Lambda_L(\hat{X}) \perp \Lambda_L(\tilde{X})$ и $\Lambda_K(\hat{X}) \perp \Lambda_K(\tilde{X})$

- Строже, чем для MSSA!

Численные сравнения в задаче разделения компонент

$$\begin{aligned} \mathbf{X} &= (\mathbf{X}_1 : \mathbf{X}_2 : \dots : \mathbf{X}_P), \\ \mathbf{X}_p &= \left(x_1^{(p)}, x_2^{(p)}, \dots, x_N^{(p)} \right)^T, \\ x_n^{(p)} &= \hat{s}_n^{(p)} + \tilde{s}_n^{(p)} + \varepsilon_n^{(p)}, \end{aligned}$$

$\varepsilon_n^{(p)}$ — независимые случайные величины с распределением $N(0, 0.01)$

Оценка точности — среднеквадратичное отклонение (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m \text{MSE}(\mathbf{S}, \tilde{\mathbf{S}}_i)}, \quad \text{MSE}(\mathbf{S}, \tilde{\mathbf{S}}) = \frac{1}{N} \sum_{j=1}^N (s_j - \tilde{s}_j)^2,$$

Варианты сигналов

- ❶ Равные по каналам 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 Различные по каналам амплитуды: $N = 44, P = 12,$

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

- ❸ Линейные по каналам 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 сигналов: $N = 44, P = 12,$

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

- ❺ Разделимость константы от гармоник: $N = 44, P = 12,$

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

- ❻ Различные по каналам частоты: $N = 59, P = 12,$

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

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

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

Среднеквадратичные отклонения оценок компонент

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