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

$\mathsf{X}_q = \left(x_1^{(q)}, x_2^{(q)}, \ldots, x_N^{(q)}\right)^{\mathrm{T}}$, $x_i \in \mathbb{R}$ — одномерный временной ряд

$\mathsf{X} = \left(\mathsf{X}^{(1)} : \mathsf{X}^{(2)} : \ldots : \mathsf{X}^{(Q)}\right)$ — многомерный временной ряд

$\mathsf{X} = \mathsf{T} + \mathsf{P} + \mathsf{E}$
T — медленно меняющаяся компонента (тренд)
P — периодическая компонента (сезонность)
E — случайная компонента (шум)

Возможные задачи:

1. Выделение сигнала из ряда: нахождение $\mathsf{S} = \mathsf{T} + \mathsf{P}$
2. Разделение сигнала: нахождение компонент 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

$\mathsf{X} = \sum_{m=1}^{M} \mathsf{S}_m + \mathsf{E}$ — одномерный временной ряд,

$\mathsf{S} \stackrel{\mathrm{def}}{=} \sum_{m=1}^{M} \mathsf{S}_m$ — сигнал.

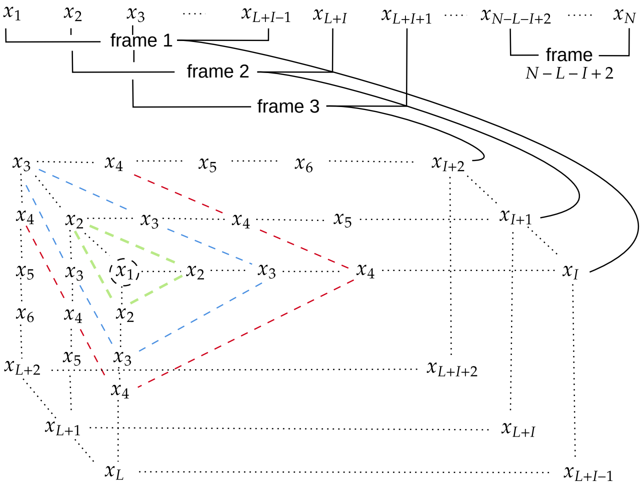
Параметры: $I, L < N$ — длины окна, $I + L \leqslant N + 1$, $J = N - I - L + 2$, R — число элементов разложения, относимых к сигналу, $\mathfrak{I}_1, \ldots, \mathfrak{I}_M \subseteq \{1, 2, \ldots, R\}$, $\mathfrak{I}_i \cap \mathfrak{I}_j = \varnothing$ — наборы индексов, относимых к компонентам сигнала.

Схема алгоритма HO-SSA для разделения компонент сигнала

1. Вложение $\mathsf{X} \xmapsto{I,L} \mathcal{X} \in \mathbb{R}^{I \times L \times J}$ — траекторный тензор,
2. Разложение
$$\mathcal{X} = \sum_{i=1}^{d_1} \sum_{l=1}^{d_2} \sum_{j=1}^{d_3} \mathcal{Z}_{ilj} U_i^{(1)} \circ U_l^{(2)} \circ U_j^{(3)}, \quad d_1 \leqslant I,\, d_2 \leqslant L,\, d_3 \leqslant J$$
3. Группировка
$$\widetilde{\mathcal{S}}_m = \sum_{i \in \mathfrak{I}_m} \sum_{l \in \mathfrak{I}_m} \sum_{j \in \mathfrak{I}_m} \mathcal{Z}_{ilj} U_i^{(1)} \circ U_l^{(2)} \circ U_j^{(3)}, \quad R \leqslant \min(d_1, d_2, d_3)$$
4. Восстановление усреднение $\widetilde{\mathcal{S}}_m$ вдоль обобщённых антидиагоналей $i + l + j = \mathrm{const}$.

Результат алгоритма $\widetilde{\mathsf{S}}_m$ — оценки компонент S_m

HO-SSA: траекторный тензор



I, L — параметры длины окна

HO-SSA: разложение и группировка

Определение

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

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

Способы приближения меньшими рангами:

1. Усечение HOSVD:

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

2. Higher-Order Orthogonal Iteration (HOOI):

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

HO-SSA: разделимость и ранги рядов в терминах SSA

- Разделимость: пусть $\mathsf{S} = \mathsf{S}_1 + \mathsf{S}_2$, $\mathbf{S}$, $\mathbf{S}_1$, $\mathbf{S}_2$ — траекторные матрицы этих сигналов с длиной окна L (траекторные тензоры с I=1)

Определение (Разделимость в терминах SSA)

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

$$\mathbf{S} = \sum_{j=1}^{R} \sqrt{\lambda_j} U_j V_j^{\mathrm{T}} = \overbrace{\sum_{j\in\mathfrak{I}_1} \sqrt{\lambda_j} U_j V_j^{\mathrm{T}}}^{\mathbf{S}_1} + \overbrace{\sum_{j\in\mathfrak{I}_2} \sqrt{\lambda_j} U_j V_j^{\mathrm{T}}}^{\mathbf{S}_2}$$

- Ранг ряда: нахождение ранга аппроксимации R?
  $\Lambda^{(L)}(\mathsf{S})$ — пространство столбцов $\mathbf{S}$

Определение (Ранг сигнала в терминах SSA)

S *имеет ранг R, если $\forall L$: $R \leqslant \min(L, N-L+1)$ $\quad \dim \Lambda^{(L)}(\mathsf{S}) = R$*

HO-SSA: разделимость

Пусть $\mathsf{S} = \mathsf{S}_1 + \mathsf{S}_2$,
$\mathcal{S}$, $\mathcal{S}_1$, $\mathcal{S}_2$ — траекторные тензоры этих сигналов с длинами окна I и L

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

HOSVD *тензора* $\mathcal{S}$ *можно представить в виде суммы* HOSVD *тензоров* $\mathcal{S}_1$ *и* $\mathcal{S}_2$ *тогда и только тогда, когда сигналы* S_1 *и* S_2 *слабо* 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.

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*

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

$\mathrm{rank}_1(\mathcal{S}) = \mathrm{rank}_2(\mathcal{S}) = \mathrm{rank}_3(\mathcal{S}) = R$, *где* $\mathcal{S}$ *— 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.

HOSVD-MSSA: алгоритм

Пусть $\mathsf{X} = \sum_{m=1}^{M} \mathsf{S}_m + \mathsf{E}$ — Q-мерный временной ряд

Параметры: L — длина окна, $K = N - L + 1$,
R, $\mathfrak{I}_1, \ldots, \mathfrak{I}_M$ — как в HO-SSA,
R_3 — число элементов разложения по третьему направлению,
$\mathfrak{Q}_1, \ldots, \mathfrak{Q}_M \subseteq \{1, 2, \ldots R_3\}$, $\mathfrak{Q}_i \cap \mathfrak{Q}_j = \varnothing$ — индексы группировки по третьему направлению

- Вложение $\mathsf{X} \overset{L}{\longmapsto} \mathcal{X} \in \mathbb{R}^{L \times K \times Q}$ — 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

$$\widetilde{\mathcal{S}}_m = \sum_{l \in \mathfrak{I}_m} \sum_{k \in \mathfrak{I}_m} \sum_{q \in \mathfrak{Q}_m} \mathcal{Z}_{lkp} \mathbf{U}_l^{(1)} \circ \mathbf{U}_k^{(2)} \circ \mathbf{U}_q^{(3)}$$

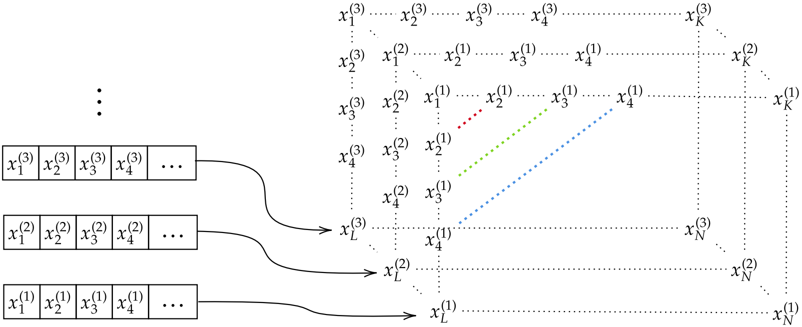
- Восстановление усреднение сечений третьего направления $\widetilde{\mathcal{S}}_m$ по побочным диагоналям

Результат алгоритма: $\widetilde{\mathsf{S}}_m$ — оценка S_m

HOSVD-MSSA: траекторный тензор многомерного ряда

X — многомерный временной ряд длины N

L — длина окна, $K = N - L + 1$



HOSVD-MSSA: разделимость

Пусть $\widehat{\mathsf{S}} = \mathsf{S}_1 + \mathsf{S}_2$,
$\mathcal{S}$, $\mathcal{S}_1$, $\mathcal{S}_2$ — траекторные тензоры этих сигналов с длиной окна L

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

Теорема

HOSVD *тензора* $\mathcal{S}$ *можно представить в виде суммы* HOSVD *тензоров* $\mathcal{S}_1$ *и* $\mathcal{S}_2$ *тогда и только тогда, когда* $\Lambda^{(L)}(\mathsf{S}_1) \perp \Lambda^{(L)}(\mathsf{S}_2)$ *и* $\Lambda^{(K)}(\mathsf{S}_1) \perp \Lambda^{(K)}(\mathsf{S}_2)$

Замечание

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

HOSVD-MSSA: n-ранги траекторного тензора

Теорема

Пусть $\mathsf{S} = \left(\mathsf{S}^{(1)} : \ldots : \mathsf{S}^{(Q)}\right)$, *тогда справедливы следующие утверждения.*

1. S *имеет ранг* R *в терминах теории* MSSA *тогда и только тогда, когда для траекторного тензора* $\mathcal{S}$, *построенного по любой длине окна* $L < N$ *такой, что* $R \leqslant \min(L, K)$ *выполняется*
$$\operatorname{rank}_1(\mathcal{S}) = \operatorname{rank}_2(\mathcal{S}) = R.$$
2. $\operatorname{rank}_3(\mathcal{S})$ *равен рангу матрицы, в строках которой содержатся одномерные сигналы* $\mathsf{S}^{(q)}$.

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.

Численные результаты: сравнение HO-SSA с SSA

Пусть временной ряд имеет вид

$$\mathsf{X} = (s_1 + \varepsilon_1, s_2 + \varepsilon_2, \ldots, s_N + \varepsilon_N),$$

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

$$\mathsf{X} = (\mathsf{X}_1 : \ldots : \mathsf{X}_Q), \quad \mathsf{X}_q = \left(x_1^{(q)}, \ldots, x_N^{(q)}\right)^{\mathrm{T}}, \quad x_n^{(q)} = \hat{s}_n^{(q)} + \tilde{s}_n^{(q)} + \varepsilon_n^{(q)},$$

где $\varepsilon_n^{(q)} \sim \mathrm{N}(0, 0.01)$ 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

$\mathsf{X} = \widehat{\mathsf{S}} + \widetilde{\mathsf{S}} + \mathsf{E}$, 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 $\varepsilon_n^{(p)} \sim \mathrm{N}(0,1)$ и независимы.

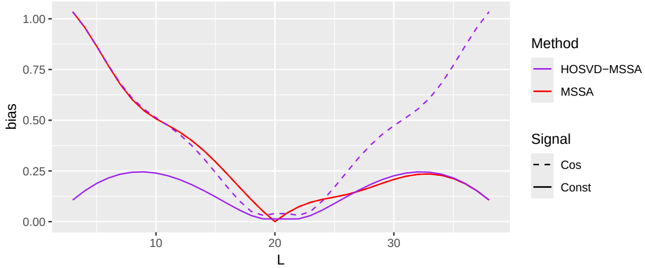


Figure: Смещение оценок каждым методом (RMSE без шума)

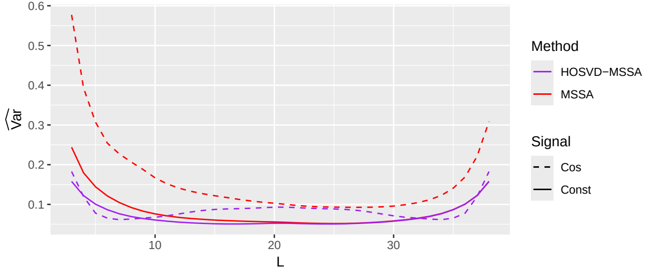
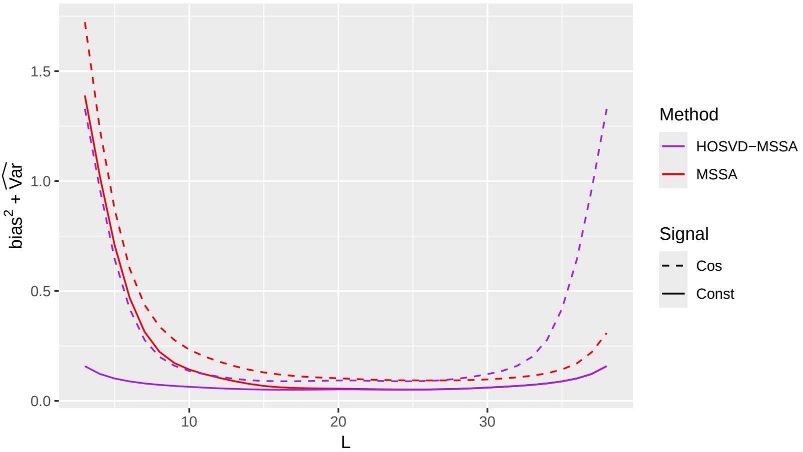


Figure: Дисперсии оценок каждым методом

Численные результаты: сумма смещения и дисперсии



Оценки методом HOSVD-MSSA имеют меньшую дисперсию, чем методом MSSA, однако могут проигрывать в точности за счёт большого 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, и над 2D-SSA для гармонических сигналов с равными амплитудами
   - Возможное преимущество HOSVD-MSSA над MSSA в условиях приблизительной разделимости для HOSVD-MSSA и точной для MSSA за счёт меньшей дисперсии оценки тензорным методом
4. Реализация методов HO-SSA и HOSVD-MSSA на языке R в стиле пакета Rssa (исходный код опубликован в репозитории Zenodo)