TENSORS FOR SIGNAL AND FREQUENCY ESTIMATION IN SUBSPACE-BASED METHODS: WHEN THEY ARE USEFUL?

N.A. Khromov[1], N.E. Golyandina[2]

[1 2] *St. Petersburg State University*
St. Petersburg, Russia
e-mail: [1]hromovn@mail.ru, [2]n.golyandina@spbu.ru

В работе рассматриваются тензорные модификации singular spectrum analysis для решения двух задач, выделения сигнала и оценки частот в зашумленной сумме экспоненциально-модулированных синусоид. Рассматриваются модификации с использованием High-Order SVD. Проводится численное сравнение. Численно показано, что для задачи выделения сигнала тензорные модификации проигрывают тензорным в большинстве случаев в одномерном случае, но могут выигрывать у многомерного SSA для системы рядов. Для оценки частот тензорные модификации, как правило, выигрывают.

Keywords: time series, signal, frequency estimation, tensor, singular spectrum analysis

1 Introduction

Одним из методов анализа временных рядов является singular spectrum analysis (SSA) [?], в котором исходный временной ряда трансформируется в матрицу, называемую траекторной, по заданной длине окна L и далее анализируется сингулярное разложение (SVD) этой матрицы. Если идет речь об оценке сигнала и его свойствах по наблюдаемому зашумленному ряду, то рассматриваются первые r компонент SVD, где r — ранг траекторной матрицы сигнала. На основе выбранных компонент строится оценка сигнала, при этом отличительной чертой метода является то, что он не требует задания модели сигнала. Однако одновременно SSA позволяет работать с параметрической моделью сигнала в виде суммы произведений полиномов экспонент и синусоид. Особую роль играет оценка частот. На основе оценки подпространства сигнала с помощью первых r левых сингулярных векторов методом ESPRIT (HSVD для LS версии и HTLS для TLS) [?] строится оценка частот, присутствующих в сигнале. Пусть сигнал задан в виде суммы синусоид (или комплексных экспонент в комплексном случае). Свойства оценок частот и оценок сигнала сильно различаются. В частности, дисперсия оценки сигнала имеет порядок $1/N$, в то время как дисперсия оценок частот имеет порядок $1/N^3$, где N — длина временного ряда и шум белый, гауссовский.

В ряде работ предлагаются тензорные модификации методов SSA и ESPRIT, где исходно ряд трансформируется не в матрицу, а в тензор, как правило, размерности три. Одним из распространенных вариантов тензорных разложений является HO-SVD, являющийся обобщением разложения SVD.

Целью данной работы является численное сравнение тензорных и матричных модификаций SSA для решения двух задач, оценки сигнала и оценки частот.

Будем рассматривать тензорные модификации, предлагаемые в работах ..., расширенные для выделения сигнала.

2 Методы

2.1 Схема Tensor SSA для выделения сигнала

Общая структура тензорных SSA алгоритмов на основе HO-SVD следующая (обычный SSA является его частным случаем). Пусть X - наблюдаемый объект. Вместо длины окна рассматриваются порядки тензора по каждой из трех размерностей I, J and K, где часть из них выражается через другие или фиксируется. Параметрами метода являются три значения R_1, R_2 и R_3, например, равные r, но не всегда.

1. Вложение $\mathbf{X} = \mathcal{T}(\mathsf{X})$ - траекторный тензор.
2. Разложение $\mathbf{X} = \sum_{l=1}^{I} \sum_{k=1}^{J} \sum_{p=1}^{K} \mathcal{Z}_{lkp} U_l^{(1)} \circ U_k^{(2)} \circ U_p^{(3)}$.
3. Группировка $\hat{\mathbf{X}} = \sum_{l=1}^{R_1} \sum_{k=1}^{R_2} \sum_{p=1}^{R_3} \mathcal{Z}_{lkp} U_l^{(1)} \circ U_k^{(2)} \circ U_p^{(3)}$.
4. Получение из $\hat{\mathbf{X}}$ оценки сигнала $\hat{\mathsf{X}}$ на основе структуры траекторного тензора и операции, обратной к вложению.

Далее будем рассматривать два варианта исходных объектов, одноканальные и многоканальные временные ряды.

2.2 Тензоры вложения

Пусть $\mathsf{X} = (x_1, x_2, \ldots, x_N)$ — (одноканальный) временной ряд длины N, $x_n \in \mathbb{C}$.

Definition 1 (Оператор вложения временного ряда в тензор)**.** Оператором вложения временного ряда в тензор с длинами окна I и L: $1 < I, L < N$, $I + L < N + 1$ будем называть отображение $\mathcal{T}_{I,L}$, переводящее ряд X в тензор $\mathcal{X} \in \mathbb{C}^{I \times L \times J}$ $(J = N - I - L + 2)$ по правилу $\mathcal{X}_{ilj} = x_{i+l+j-2}$, где $i \in \overline{1:I}$, $l \in \overline{1:L}$, $j \in \overline{1:J}$.

Пусть $\mathsf{X} = (\mathsf{X}^{(1)}, \mathsf{X}^{(2)}, \ldots, \mathsf{X}^{(P)})$ — P-канальный временной ряд, состоящий из P одноканальных временных рядов, также называемых каналами.

Definition 2 (Оператор вложения многоканального ряда в тензор)**.** Оператором вложения многоканального ряда в тензор с длиной окна L: $1 < L < N$ будем называть отображение $\mathcal{T}_L$, переводящее P-канальный ряд X в тензор $\mathcal{X} \in \mathbb{C}^{L \times K \times P}$ $(K = N - L + 1)$ по правилу $x^{(p)}_{l+k-1}$, где $l \in \overline{1:L}$, $k \in \overline{1:K}$, $p \in \overline{1:P}$.

2.3 Методы для оценки параметров сигнала.

Рассмотрим в общем случае P-канальный временной ряд (возможно $P = 1$) с элементами

$$x_n^{(p)} = \sum_{r=1}^{R} a_r^{(p)} e^{\alpha_r n} e^{\mathrm{i}\left(2\pi\omega_r n + \varphi_r^{(p)}\right)},$$

где параметрами модели являются амплитуды $a_j^{(p)} \in \mathbb{C} \setminus \{0\}$, фазы $\varphi_j^{(p)} \in [0, 2\pi)$, частоты $\omega_j \in [0, 1/2]$ и степени затухания $\alpha_j \in \mathbb{R}$. Алгоритм HO-ESPRIT, оценивающий частоты и степени затухания ряда, определяется следующим образом. После шага разложения строится матрица $\mathbf{U} = \mathbf{U}_d = \left[U_1^{(d)} : U_2^{(d)} : \ldots : U_{R_d}^{(d)}\right]$ для некоторого $d \in \{1, 2, 3\}$, и решается уравнение

$$\mathbf{U}^{\uparrow} = \mathbf{U}_{\downarrow}\mathbf{Z}$$

относительно матрицы $\mathbf{Z}$, где запись $\mathbf{U}^{\uparrow}$ означает матрицу $\mathbf{U}$ без первой строки, а $\mathbf{U}_{\downarrow}$ — без последней. R наибольших собственных чисел матрицы $\mathbf{Z}$ считаются оценками $\lambda_r = e^{\alpha_r + 2\pi\mathrm{i}\omega_r}$, из которых можно получить параметры α_r и ω_r.

2.4 Dstack модификация

В работе [?] для ускорения работы метода предлагается преобразование одномерного ряда в многомерный перед применением тензорной модификации: $x_m^{(d)} = x_{(m-1)D+d}$, где $m \in \overline{1 : (N/D)}$. В [?] это применяется только для модификации ESPRIT, называемой HTLSDstack, но мы будем применять данное преобразование временного ряда и для оценки сигнала, метод назовем SSADstack. Тензорные модификации строятся как для многоканального ряда.

3 Сравнение тензорных методов с матричными

Все численные сравнения были проведены для временных рядов в виде суммы синусоид.

Для одномерных временных рядов и задачи выделения сигнала было проведено сравнение следующих методов: SSA, HO-SSA, SSADstack, HO-SSADstack с $R_3 = \max$ и HO-SSADstack с $R_3 = 1$. Было получено, что в большинстве случаев метод SSA существенно выигрывает по точности, а если проигрывает, то незначительно и только в очень узком диапазоне параметров, что делает это небольшое преимущество нереализуемым на практике.

Для одномерных временных рядов и задачи оценки частот рассматривался сигнал в виду двух синусоид с близкими частотами. Сравнивались методы ESPRIT, HO-ESPRIT, HTLSDstack, HO-HTLSDstack с $R_3 = \max$ и HO-HTLSDstack с $R_3 = 1$. Было получено, что при низком уровне шума ESPRIT работает точнее, однако при среднем и большом уровне шума HO-ESPRIT становится точнее при оптимальном выборе параметров, а HO-HTLSDstack с $R_3 = 1$ обыгрывает все методы.

Для многоканальных временных рядов было получено, что в случае, когда ряды являются суммой синусоид с одинаковыми частотами, тензорные модификации дают более точный результат, как в задаче выделения сигнала, так и в задаче оценивания частот.

4 Conclusion

Проведенное численное сравнение показало разный эффект от тензорной HO-SVD модификации для временных рядов. Для выделения сигнала для одномерных временных рядов матричный вариант однозначно лучше. Для многоканальных временных рядов с одинаковыми частотами в каналах и для задачи оценки частот тензорный вариант может давать выигрыш в точности.

References

[1] Papy J.M., De Lathauwer L., Van Huffel S. (2005). Exponential data fitting using multilinear algebra: the single-channel and multi-channel case. *Linear Algebra with Applications*. Vol. **12**, Num. **8**, pp. 809-826.

[2] Papy J.M., De Lathauwer L., Van Huffel S. (2009). Exponential data fitting using multilinear algebra: the decimative case. *Journal of Chemometrics*. Vol. **23**, Num. **7-8**, pp. 341-351s.

[3] Jacobs P.A., Lewis P.A.W. (1983). Stationary Discrete Autoregressive-Moving Average Time Series Generated by Mixtures. *Journal of Time Series Analysis*. Vol. **4**, Num. **1**, pp. 19-36.

[4] Johnson N.L., Kotz S., Balakrishnan N. (1997). *Discrete Multivariate Distributions*. Wiley: New York.

[5] Worldometers.info [Electronic resource] Mode of access: `https://www.worldometers.info/coronavirus`. Date of access: 27.02.2022.