Задачи анализа временных рядов, теория метода «Анализ Сингулярного Спектра» SSA

Яковлев Денис Михайлович, гр.21.Б04-мм

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

3 курс (бак.) «Производственная практика
(научно-исследовательская работа)»
(Семестр 6)

Санкт-Петербург, 2024

Введение

- Цель работы — решение теоретических и прикладных задач анализа временных рядов с применением метода SSA.
- Используемый метод — SSA (Singular Spectrum Analysis), или «Анализ Сингулярного Спектра» (сокращенно, АСС).
- Задача — исследование поведения ошибок восстановления ряда в асимптотическом случае.

Обозначения

Определение 1 (Сигнал)

Вещественный сигнал $\mathrm{H} = (h_0, h_1, \ldots, h_n, \ldots)$, *управляется линейной рекуррентной формулой (ЛРФ) порядка* d

$$h_n = \sum_{k=1}^{d} a_k h_{n-k},\ n \geqslant d,\ a_d > 0.$$

$\mathrm{H}_N = (h_0, h_1, \ldots, h_N)$.

Определение 2 (Помеха)

Помеха — $\mathrm{E}_N = (e_0, \ldots, e_N)$.

Обозначения

Определение 3 (Траекторная матрица)

Пусть $1 \leqslant L < N$ *— длина окна,* $K = N - L + 1$. *Тогда траекторной матрицей сигнала* $\mathrm{H}_N = (h_0, h_1, \ldots, h_N)$ *называется матрица* $\mathbf{H}$ *размера* $L \times K$ *вида:*

$$\begin{pmatrix} h_0 & h_1 & h_2 & \ldots & h_{K-1} \\ h_1 & h_2 & h_3 & \ldots & h_K \\ h_2 & h_3 & h_4 & \ldots & h_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{L-1} & h_L & h_{L+1} & \ldots & h_{N-1} \end{pmatrix}$$

Замечание 1 (Размеры матриц)

Считаем, что все рассматриваемые матрицы размера $L \times K$, *если не оговорено иначе.*

Обозначения

Замечание 2 (Ранг матрицы)

$d = \operatorname{rank} \mathbf{H} < \min(L, K)$ — *ранг траекторной матрицы* $\mathbf{H}$, *образованной от сигнала* H, *управляемого ЛРФ порядка* d.

Определение 4 (Матрица возмущённого сигнала $\mathbf{H}(\delta)$)

Пусть $\delta \in \mathbb{R}$ — *формальный параметр возмущения,* $\mathbf{E}$ — *траекторная матрица* E_N, $\mathbf{H}$ — H_N *соответственно. Тогда* $\mathbf{H}(\delta) = \mathbf{H} + \delta\mathbf{E}$ — *матрица возмущённого сигнала.*

Замечание 3

$\mathbf{H}(0) = \mathbf{H}$.

Обозначения

Определение 5 (Возмущённая матрица для сингулярного разложения $\mathbf{A}(\delta)$)

$$\mathbf{A}(\delta) = \mathbf{H}(\delta)\mathbf{H}(\delta)^{\mathrm{T}} = \mathbf{H}\mathbf{H}^{\mathrm{T}} + \delta(\mathbf{H}\mathbf{E}^{\mathrm{T}} + \mathbf{E}\mathbf{H}^{\mathrm{T}}) + \delta^2\mathbf{E}\mathbf{E}^{\mathrm{T}},$$

$\mathbf{A}(\delta)$ — *возмущённая матрица размера* $L \times L$.

Замечание 4

$\mathbf{A}(0) = \mathbf{A} = \mathbf{H}\mathbf{H}^{\mathrm{T}}$.

Определение 6 (Наименьшее и наибольшее ненулевые собственные числа $\mu_{\min}, \mu_{\max}$)

Пусть $\{\mu_n\}_{n=1}^{L}$ *— собственные числа* $\mathbf{A}$. *Тогда* $\mu_{\min}, \mu_{\max}$ *— наименьшее и наибольшее ненулевые собственные числа.*

Обозначения

Определение 7 (Собственные подпространства матрицы $\mathbf{A}$)

Пусть $\{U_n\}_{n=1}^{L}$ *— собственные вектора размера* $L \times 1$ *матрицы* $\mathbf{A}$. *Тогда* $\mathbb{U}_0$ *— собственное подпространство, соответствующее нулевым собственным числам, а* $\mathbb{U}_0^{\perp}$ *— соответствует ненулевым собственным числам.*

Определение 8 (Ортогональные проекторы собственных подпространств $\mathbf{P}_0^{\perp}, \mathbf{P}_0$)

Пусть $\mathbf{I}$ *— единичная матрица размера* $L \times L$. *Тогда* $\mathbf{P}_0$ *— ортогональный проектор на* $\mathbb{U}_0$, *а* $\mathbf{P}_0^{\perp} = \mathbf{I} - \mathbf{P}_0$ *— ортогональный проектор на* $\mathbb{U}_0^{\perp}$. $\mathbf{P}_0$, $\mathbf{P}_0^{\perp}$ *— матрицы размера* $L \times L$.

Замечание 5

$\mathbf{P}_0^{\perp}\mathbf{H} = \mathbf{H}$.

Обозначения

Замечание 6 (Про матричные нормы)

Полагаем $\|\cdot\|$ *— спектральная норма. В иных случаях явно указываем норму.*
Например, максимальная норма $\|\cdot\|_{\max}$.

Определение 9 (Псевдообратная матрица $\mathbf{S}_0$)

Пусть $\mathbf{S}_0$ *— матрица размера* $L \times L$, *псевдообратная к* $\mathbf{H}\mathbf{H}^{\mathrm{T}}$.
Положим $\mathbf{S}_0^{(0)} = -\mathbf{P}_0$ *и* $\mathbf{S}_0^{(k)} = \mathbf{S}_0^k$ *для* $k \geqslant 1$, $\left\|\mathbf{S}_0^{(k)}\right\| = 1/\mu_{min}^k$.

Определение 10 (Возмущение матрицы $\mathbf{A}(\delta)$)

Зависящую от δ *часть матрицы* $\mathbf{A}(\delta)$ *будем обозначать* $\mathbf{B}(\delta)$.
Тогда $\mathbf{B}(\delta) = \delta(\mathbf{H}\mathbf{E}^{\mathrm{T}} + \mathbf{E}\mathbf{H}^{\mathrm{T}}) + \delta^2\mathbf{E}\mathbf{E}^{\mathrm{T}}$.

Обозначения

Определение 11 (Ошибка восстановления $r_i(N)$)

Пусть $\widetilde{\mathrm{H}}_N(\delta) = (\widetilde{h}_0(\delta), \widetilde{h}_1(\delta), \ldots, \widetilde{h}_N(\delta))$ *— восстановленный сигнал из* $\mathrm{H}_N(\delta) = (h_0 + \delta e_0, h_1 + \delta e_1, \ldots, h_N + \delta e_N)$*, полученный методом SSA.*
Тогда $r_i(N) = \widetilde{h}_i(\delta) - h_i,\ 0 \leqslant i < N$ *— ошибка восстановления.*

Определение 12 (Оператор ганкелевизации $\mathcal{S}$)

$\mathcal{S}$ *— оператор диагонального усреднения (ганкелевизации).*

Замечание 7 (Оператор ганкелевизации для траекторной матрицы сигнала)

$\mathcal{S}\mathbf{H} = \mathbf{H}$*.*

Постановка общей задачи

Постановка 1 (Общая задача)

Пусть $\mathbf{N}$ *— некоторая матрица размера* $L \times L$*. Так как* $\|\mathcal{S}\mathbf{A}\|_{\max} \leqslant \|\mathbf{A}\|_{\max}$ *и* $\|\mathbf{A}\|_{\max} \leqslant \|\mathbf{A}\|$ *для любой конечномерной матрицы* $\mathbf{A}$*:*

$$\max_{0 \leqslant n < N} |r_i(N)| = \|\mathcal{S}\Delta_\delta(\mathbf{H})\|_{\max} \leqslant \|\Delta_\delta(\mathbf{H})\|_{\max}$$
$$\leqslant \left\|(\mathbf{P}_0^\perp(\delta) - \mathbf{P}_0^\perp - \mathbf{N})\mathbf{H}(\delta)\right\| + \left\|\mathbf{N}\,\mathbf{H}(\delta) + \delta\mathbf{P}_0^\perp\mathbf{E}\right\|_{\max}, \quad (1)$$

где $\Delta_\delta(\mathbf{H}) = \mathbf{P}_0^\perp(\delta)\mathbf{H}(\delta) - \mathbf{P}_0^\perp\mathbf{H}$*.*
Общая задача: подобрать такую $\mathbf{N}$*, чтобы правая часть* (1) *стремилась к нулю.*

Формулировка

Формулировка 1 (Оценка нормы)

Оценить выражение сверху
$\forall n \in \mathbb{N}: \left\| \mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \sum\limits_{p=1}^{n} \mathbf{W}_p(\delta) \right\|$, $\mathbf{W}_p(\delta)$ *— матрицы размера* $L \times L$ *такие, что*

$$\mathbf{W}_p(\delta) = (-1)^p \sum_{l_1+\dots+l_{p+1}=p,\, l_j \geqslant 0} \mathbf{W}_p(l_1, \dots, l_{p+1}; \delta),$$

а

$$\mathbf{W}_p(l_1, \dots, l_{p+1}; \delta) = \mathbf{S}_0^{(l_1)} \mathbf{B}(\delta) \mathbf{S}_0^{(l_2)} \dots \mathbf{S}_0^{(l_p)} \mathbf{B}(\delta) \mathbf{S}_0^{(l_{p+1})}.$$

Известные результаты

Воспользуемся вспомогательными теоремами и леммами из [2].

Теорема 1 (Теорема 2.1)

Пусть $\delta_0 > 0$ *и*

$$\|\mathbf{B}(\delta)\| < \mu_{min}/2$$

для всех $\delta \in (-\delta_0, \delta_0)$. *Тогда для возмущённого проектора* $\mathbf{P}_0^{\perp}(\delta)$ *верно представление:*

$$\mathbf{P}_0^{\perp}(\delta) = \mathbf{P}_0^{\perp} + \sum_{p=1}^{\infty} \mathbf{W}_p(\delta).$$

Известные результаты

Теорема 2 (Теорема 2.3)

Если $\delta_0 > 0$ *и* $\dfrac{\|\mathbf{B}(\delta)\|}{\mu_{min}} < \dfrac{1}{4}$ *для всех* $\delta \in (-\delta_0, \delta_0)$, *то проектор* $\mathbf{P}_0^{\perp}(\delta)$ *существует и*

$$\left\|\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp}\right\| \leqslant 4C\frac{\|\mathbf{S}_0\mathbf{B}(\delta)\mathbf{P}_0\|}{1 - 4\,\|\mathbf{B}(\delta)\|\,/\mu_{min}},$$

где $C = e^{1/6}/\sqrt{\pi} \approx 0.667$.

Лемма 1 (Лемма 6.1)

Если $0 < \beta < 1/4$, $k \geqslant 0$, *то* $\sum_{p=k}^{\infty} \binom{2p}{p} \beta^p \leqslant C\dfrac{(4\beta)^k}{1 - 4\beta}$, *где* $C = e^{1/6}/\sqrt{\pi}$.

Полученные результаты

Следствие 1

В условиях Теоремы 2, $\forall n \in \mathbb{N}$*:*

$$\left\| \mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \sum_{p=1}^{n} \mathbf{W}_p(\delta) \right\| \leqslant 4^{n+1} C \left(\frac{\|\mathbf{B}(\delta)\|}{\mu_{min}} \right)^{n+1} \frac{1}{1 - 4\left\|\mathbf{B}(\delta)\right\| / \mu_{min}},$$

где $C = e^{1/6}/\sqrt{\pi}$*.*

Постановка

Постановка 2 (Линейный сигнал с гармониками)

Рассмотрим при $n = 0, 1, \ldots, N-1$ *линейный сигнал* $h_n = \theta_1 n + \theta_0$, *где* $\theta_1 \neq 0$, *и помеху, которая является линейной комбинацией гармоник*

$$e_n = \sum_{l=1}^{r} \tau_l \cos(2\pi n \omega_l + \varphi_l),$$

где $\tau_l \neq 0, \tau_l \neq \tau_p$ *и* $\omega_l \neq \omega_p$ *при* $l \neq p$ *и* $0 < \omega_l < 1/2$.
Пусть $\widetilde{h}_0(\delta), \ldots, \widetilde{h}_{N-1}(\delta)$ *— результаты восстановления ряда* $\{h_n + \delta e_n\}_{n=0}^{N-1}$ *с помощью метода SSA, а* $r_n(N) = \widetilde{h}_n(\delta) - h_n$.
Тогда теорема из [3] утверждает, что для любого $\delta \in \mathbb{R}$ *при* $N \to \infty$.

$$\max_{0 \leqslant n < N} |r_n(N)| = O(N^{-1}).$$

Формулировка и известные результаты

Формулировка 2

Обобщить результат [3] с $L = K$ *при* $\mathbf{N} = \mathbf{W}_1$ *до* $L/N \to \alpha \in (0, 1)$ *с помощью выбора* $\mathbf{N} = \mathbf{W}_1 + \mathbf{W}_2$*.*

Воспользуемся вспомогательными теоремами и леммами из [3]

Лемма 2 (Лемма 1)

При $N \to \infty$ *имеет место соотношение* $\|\mathbf{H}\mathbf{E}^{\mathrm{T}}\|_{\max} = O(N)$*.*

Лемма 3 (Лемма 2)

При $N \to \infty$ *имеет место соотношение* $\|\mathbf{P}_0^{\perp}\mathbf{E}\|_{\max} = O(N^{-1})$*.*

Лемма 4 (Лемма 3)

При $N \to \infty$ *имеет место соотношение* $\|\mathbf{S}_0\mathbf{E}\| = O(N^{-4})$*.*

Полученные результаты

Замечание 8

При $\mathbf{N} = \mathbf{W}_1 + \mathbf{W}_2$ *рассматривается неравенство*

$$\max_{0\leqslant i<N} |r_i(N)| \leqslant \left\| \Big(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - (\mathbf{W}_1 + \mathbf{W}_2)\Big)\,\mathbf{H}(\delta) \right\| + \left\| (\mathbf{W}_1 + \mathbf{W}_2)\,\mathbf{H}(\delta) + \delta \mathbf{P}_0^{\perp}\mathbf{E} \right\|_{\max}.$$

Предложение 1

Пусть $L/N \to \alpha \in (0,1)$. *Тогда для любого* δ

$$\left\| \Big(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \mathbf{W}_1(\delta) - \mathbf{W}_2(\delta)\Big)\,\mathbf{H}(\delta) \right\| = O(N^{-1}).$$

Предложение 2

В условиях Предложения 1
$\|(\mathbf{W}_1(\delta) + \mathbf{W}_2(\delta))\mathbf{H}(\delta)\|_{\max} = O(N^{-1})$.

Приложение. Результаты вычислительных экспериментов

Формулировка 3

Проиллюстрировать теоретический результат задачи №2 для ряда

$$\widetilde{h}_n = n + 3\cos(2\pi n\omega + \varphi),$$

где $\omega = 1/4$, $\varphi = \pi/8$, $n = 0, 1, \ldots, N-1$, $N = 9 \ldots 200$, *длина окна* $L = \lfloor N/3 \rfloor$.

Приложение. Результаты вычислительных экспериментов

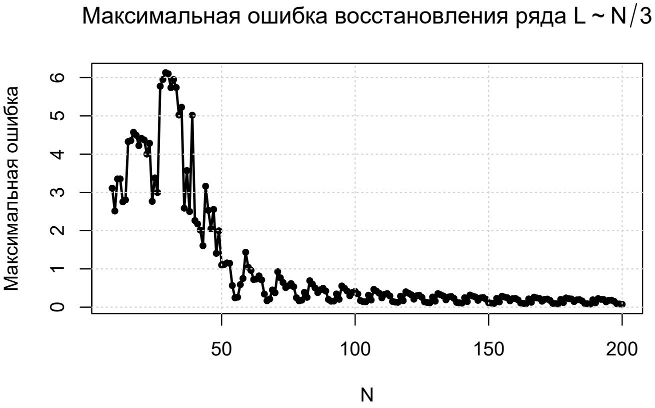

Figure 1: Максимальные ошибки восстановления ряда в зависимости от длины ряда при $\widetilde{h}_n = n + 3\cos(\pi n/2 + \pi/8)$.

Приложение. Результаты вычислительных экспериментов

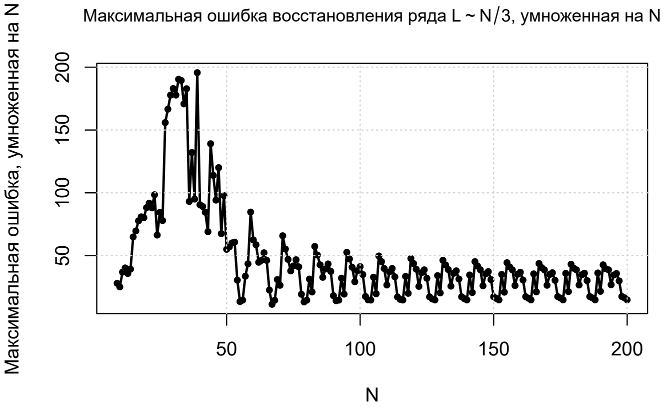

Figure 2: Максимальные ошибки восстановления ряда, умноженные на N, в зависимости от длины ряда N для

Заключение

1. Поставлена общая теоретическая задача;
2. Дана оценка $\|\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \sum_{p=1}^{n} \mathbf{W}_p(\delta)\| \; \forall n \in \mathbb{N}$;
3. Обобщён результат [3] асимптотической разделимости линейного сигнала с линейной комбинацией гармоник с $L = K$ до $L/N \to \alpha \in (0, 1)$;
4. Проведён вычислительный эксперимент, подтверждающий полученный результат.

Список литературы

- Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure. SSA and Related Techniques.* Champan & Hall/CRC, Boca Raton-London-New York-Washington D.C.
- Nekrutkin, V. (2010). *Perturbation expansions of signal subspaces for long signals.* Statistics and Its Interface. **3**, 297–319.
- *Н.Зенкова, В.Некруткин.* Об асимптотической разделимости линейных сигналов с гармониками методом анализа сингулярного спектра, Вестник СПбГУ. Математика. Механика. Астрономия. 2022. Т. 9 (67). Вып. 2.