Некоторые задачи анализа сингулярного спектра

Киселев Михаил Михайлович, гр. 20.Б04-мм

Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — к. ф.-м. н., доцент **В. В. Некруткин**
Рецензент — Разработчик, Onfido Ltd, Великобритания **Е. В. Иванова**

Санкт-Петербург
2024 г.

Общая постановка задачи выделения сигнала из суммы сигнала и помехи методом Singular Spectrum Analysis

Дано:

$$\begin{cases} \mathsf{F}_N = (f_0, \ldots, f_{N-1}) \text{ — «сигнал»}, \\ \mathsf{E}_N = (e_0, \ldots, e_{N-1}) \text{ — «помеха»}, \\ \mathsf{X}_N = (f_0 + \delta e_0, \ldots, f_{N-1} + \delta e_{N-1}) \text{ — наблюдаемый временной ряд}, \end{cases}$$

δ — параметр возмущения.

$$\left(\text{ЛРФ: } f_n = \sum_{k=1}^{d} a_k f_{n-k},\ n \geq d \right)$$

Задача SSA: построить приближение $\widetilde{\mathsf{F}}_N$ для F_N.

Этап 1: разложение. Вложение и сигнулярное разложение

1. L — длина окна $(1 < L < N)$,
2. $X_i = (x_i, \ldots, x_{i+L-1})^T$, $0 \le i < K := N - L + 1$.

Траекторная матрица ряда X_N:

$$\mathbf{H}(\delta) = [X_1 : \ldots : X_K] = \begin{pmatrix} x_0 & x_1 & \ldots & x_{K-1} \\ x_1 & x_2 & \ldots & x_K \\ \vdots & \vdots & \ddots & \vdots \\ x_{L-1} & x_L & \ldots & x_{N-1} \end{pmatrix}.$$

$\mathrm{H} = \mathrm{H}_N$, $\mathrm{E} = \mathrm{E}_N$ — траекторные матрицы рядов F_N и $E_N \Longrightarrow \mathbf{H}(\delta) := \mathbf{H} + \delta\mathbf{E}$.
После SVD:

$$\widetilde{\mathbf{H}}(\delta) = \sum_{i=1}^{d} \lambda_i U_i V_i^T, \text{ где}$$

- $d := \max\{i : \lambda_i > 0\}$.
- $\lambda_1, \ldots, \lambda_L$ — собственные числа $\mathbf{H}(\delta)\mathbf{H}(\delta)^\mathsf{T}$, $\lambda_i \ge \lambda_{i+1}$,
  $U_1, \ldots, U_L$ — соответсвующие собственные вектора.
- $V_i := \mathbf{H}(\delta)^T U_i / \sqrt{\lambda_i}, \quad i = 1, \ldots, d.$

Этап 2: восстановление.

1. **Диагональное усреднение**

$$\hat{\mathbf{H}}(\delta) = \mathcal{S}\widetilde{\mathbf{H}}(\delta), \text{ то есть } \forall i + j = const \ \left(\hat{\mathbf{H}}(\delta)\right)_{ij} := \overline{\left(\widetilde{\mathbf{H}}(\delta)\right)_{ij}}$$

2. **Восстановление сигнала**

$$\tilde{\mathsf{F}}_N(\delta) = (f_0(\delta), \ldots, f_{N-1}(\delta)) \text{ — приближение } \mathsf{F}_N.$$

Обозначим:
$R_N(\delta) := (r_0(\delta), \ldots, r_{N-1}(\delta))$, $r_i(\delta) = f_i(\delta) - f_i$ — ряд ошибок восстановления,

$$\Delta_\delta(\mathbf{H}) = \hat{\mathbf{H}}(\delta) - \mathbf{H} \text{ — матрица ошибок восстановления.}$$

В ряде работ[1] анализ ошибок восстановления проводится в рамках их линейного приближения по параметру возмущения.

Задача ВКР состоит в численной проверке этого подхода для ряда сигналов и помех при больших длинах ряда N.

[1] Например Senov M. *Robust versions of the SSA method.* — 2022

$$\mathbf{H}(\delta) = \mathbf{H} + \delta\mathbf{E}, \quad \widetilde{\mathbf{H}}(\delta) - \mathbf{H}?$$

$\mathbb{U}_0^{\perp}$ — линейное пространство столбцов $\mathbf{H}$. $\mathbf{P}_0$ — ортогональный проектор на $\mathbb{U}_0$ и $\mathbf{P}_0^{\perp} = \mathbf{I} - \mathbf{P}_0$ — ортогональный проектор на $\mathbb{U}_0^{\perp}$.

$$\mathbf{P}_0^{\perp} = \sum_{i=1}^{d} U_i U_i^{\mathsf{T}}.$$

$\mathbf{Q}_0^{\perp}$ ортогональный проектор на пространство столбцов матрицы $\mathbf{H}^{\mathsf{T}}$,

$$\mathbf{Q}_0^{\perp} = \sum_{i=1}^{d} V_i V_i^{\mathsf{T}}.$$

$$\mathbf{A}(\delta) = \mathbf{H}(\delta)\mathbf{H}(\delta)^{\mathsf{T}} = \mathbf{A} + \delta\mathbf{A}^{(1)} + \delta^2\mathbf{A}^{(2)} = \mathbf{A} + \mathbf{B}(\delta), \text{ где}$$

- $\mathbf{A} \coloneqq \mathbf{H}\mathbf{H}^{\mathsf{T}}$,
- $\mathbf{A}^{(1)} = \mathbf{H}\mathbf{E}^{\mathsf{T}} + \mathbf{E}\mathbf{H}^{\mathsf{T}}$ и $\mathbf{A}^{(2)} = \mathbf{E}\mathbf{E}^{\mathsf{T}}$,
- $\mathbf{B}(\delta) \coloneqq \delta\mathbf{A}^{(1)} + \delta^2\mathbf{A}^{(2)}$.

$\mathsf{B}(\delta) = \delta\mathbf{A}^{(1)} + \delta^2\mathbf{A}^{(2)}$, $\mu_{\min} \coloneqq \min\{\mu \in \Sigma : \mu > 0\}$, $\mathbf{S}_0^{(0)} = -\mathbf{P}_0$ и $\mathbf{S}_0^{(k)} = \mathbf{S}_0^k$.

Теорема (Nekrutkin V., 2010)

Пусть $\delta_0 > 0$ *и* $\|\mathbf{B}(\delta)\| < \mu_{\min}/2$ *для всех* $\delta \in (-\delta_0, \delta_0)$. *Тогда*

$$\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} = \sum_{n=1}^{\infty} \delta^n \mathbf{V}_0^{(n)}, \qquad (1)$$

где

$$\mathbf{V}_0^{(n)} = \sum_{p=\lceil n/2 \rceil}^{n} (-1)^p \sum_{\substack{s_1+\ldots+s_p=n,\ s_i=1,2 \\ l_1+\ldots+l_{p+1}=p,\ l_j \geq 0}} \mathbf{V}_0^{(n)}(s,\ell),$$

$\mathbf{s} = (s_1, \ldots, s_p)$, $\ell = (l_1, \ldots, l_{p+1})$, *и*

$$\mathbf{V}_0^{(n)}(s,\ell) = \mathbf{S}_0^{(l_1)}\mathbf{A}^{(s_1)}\mathbf{S}_0^{(l_2)} \ldots \mathbf{A}^{(s_p)}\mathbf{S}_0^{(l_{p+1})},$$

ряд в правой части (1) сходится по спектральной норме.

$$\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} = \sum_{n=1}^{\infty} \delta^n \mathbf{V}_0^{(n)} \text{ из теоремы (Nekrutkin V., 2010).}$$

По следствию теоремы,

$$\mathbf{V}_0^{(1)} = \mathbf{P}_0 \mathbf{A}^{(1)} \mathbf{S}_0 + \mathbf{S}_0 \mathbf{A}^{(1)} \mathbf{P}_0 = \mathbf{P}_0 \mathbf{E} \mathbf{H}^{\mathsf{T}} \mathbf{S}_0 + \mathbf{S}_0 \mathbf{H} \mathbf{E}^{\mathsf{T}} \mathbf{P}_0$$

1.
$$\widetilde{\mathbf{H}}(\delta) - \mathbf{H} = (\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp})\mathbf{H}(\delta) + \delta \mathbf{P}_0^{\perp} \mathbf{E}$$

2.
$$\mathbf{H}^{\mathsf{T}} \mathbf{S}_0 \mathbf{H} = \mathbf{Q}_0^{\perp}$$

Отсюда, подставляя в (1) получим, что линейный по параметру возмущения член разности $\widetilde{\mathbf{H}}(\delta) - \mathbf{H}$ имеет вид (см., например, Zenkova N., Nekrutkin V., 2022).

$$\delta\left(\mathbf{P}_0^{\perp}\mathbf{E} + \mathbf{E}\mathbf{Q}_0^{\perp} - \mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp}\right). \tag{2}$$

N — длина ряда, $L = \lfloor N/2 \rfloor$ — длина окна.

$$\max_{n=1,\ldots,N} |r_n(\delta)| \text{ — ошибки восстановления SSA,} \tag{3}$$

$$\max_{i=1,\ldots,N} \left| \left[\mathcal{S}(\mathbf{P}_0\mathbf{E} + \mathbf{E}\mathbf{Q}_0^{\perp} - \mathbf{P}_0^{\perp}\mathbf{E}\mathbf{Q}_0^{\perp} \right]_i \right| \text{ — полное линейное приближение,} \tag{4}$$

$$\max_{i=1,\ldots,N} \left| \left[\mathcal{S}(\mathbf{P}_0\mathbf{E} + \mathbf{E}\mathbf{Q}_0^{\perp} \right]_i \right| \text{ — неполное линейное приближение.} \tag{5}$$

Сигналы:

1. Константный — $f_n \equiv 1$, $d = 1$,
2. Линейный — $f_n = n$, $d = 2$,
3. Синусоида — $f_n = \cos(2\pi\omega n + \gamma)$, $d = 2$,
4. Растущий экспоненциальный сигнал[2] — $f_n = a^n$, $a > 1$, $d = 1$.

Помехи:

1. Единичный выброс — $e_n = 1$ при $n = n_0$, и $e_n = 0\ \forall n \neq n_0$.
2. Синусоида — $e_n = \cos(2\pi/\sqrt{99}n + \pi)$,
3. Абсолютно сходящийся ряд — $e_n = 0.99^n$, $0 < a < 1$,
4. Гауссовский белый шум — $e_n \in \mathsf{N}(0,\ 1)$,
5. Равномерный белый шум — $e_n \in \mathsf{U}(-\sqrt{6},\ \sqrt{6})$.

Для случайных помех: $\max\limits_{n=1,\ldots,N} \mathsf{E}|r_n(\delta)|$ и $\max\limits_{n=1,\ldots,N} \mathsf{D}|r_n(\delta)|$, $\Longrightarrow$

$$\max_{n=1,\ldots,N} \widehat{\mathsf{E}}|r_n(\delta)| = \max_{n=1,\ldots,N} \overline{|r_n(\delta)|}, \quad \text{где} \quad \overline{|r_n(\delta)|} = \frac{1}{100}\sum_{j=1}^{100} |r_n^{(j)}(\delta)|,$$

$$\max_{n=1,\ldots,N} \widehat{\mathsf{D}}|r_n(\delta)| = \max_{n=1,\ldots,N} \frac{1}{100}\sum_{j=1}^{100} \left(|r_n^{(j)}(\delta)| - \overline{|r_n(\delta)|}\right)^2$$

[2] с или без дискретизации

Линейный сигнал. Единичный выброс

$f_n = n.$ $(d = 2)$.

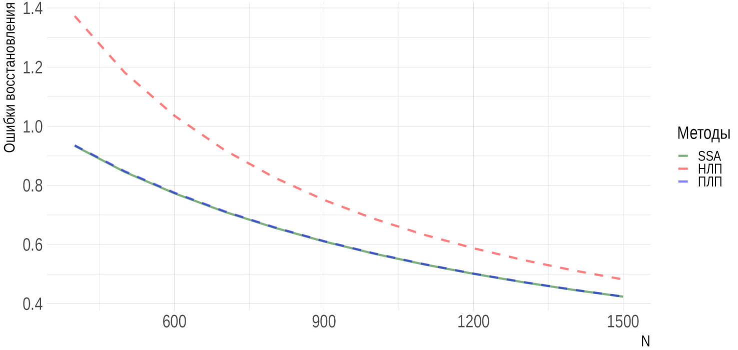

Рис.: Помеха — $e_n = 1$ при $n = 50,$ и $e_n = 0$ $\forall n \neq 50$, $\delta = 5$.

Линейный сигнал. Синусоидная помеха

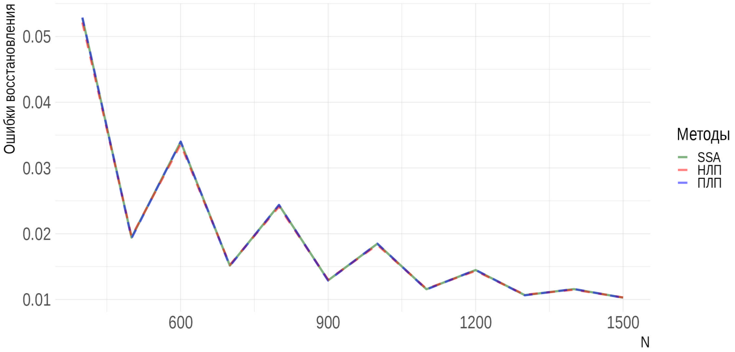

Рис.: Помеха — $e_n = \cos(2\pi/\sqrt{99}n + \pi)$, $\delta = 0.3$.

Линейный сигнал. Сходящийся ряд

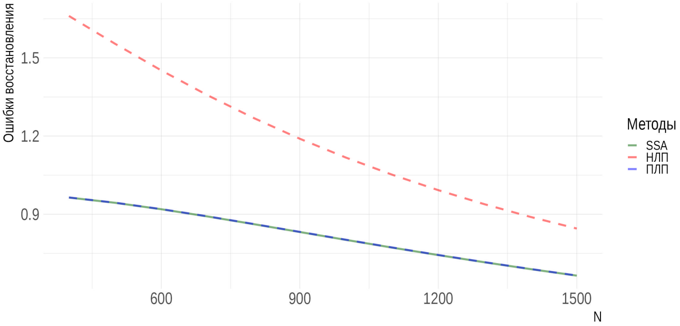

Рис.: Помеха — $e_n = 0.99^n$, $\delta = 1$.

Линейный сигнал. Гауссовский белый шум

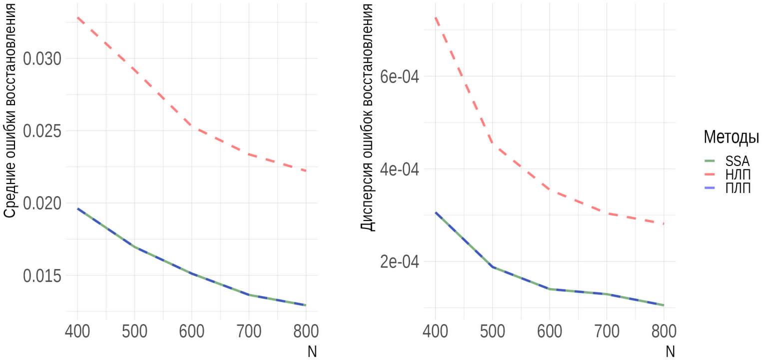

Рис.: $\max \widehat{\mathsf{E}}|r_n(\delta)|$ и $\max \widehat{\mathsf{D}}|r_n(\delta)|$. Помеха — $e_n \in \mathsf{N}(0,\ 1)$, $\delta = 0.2$.

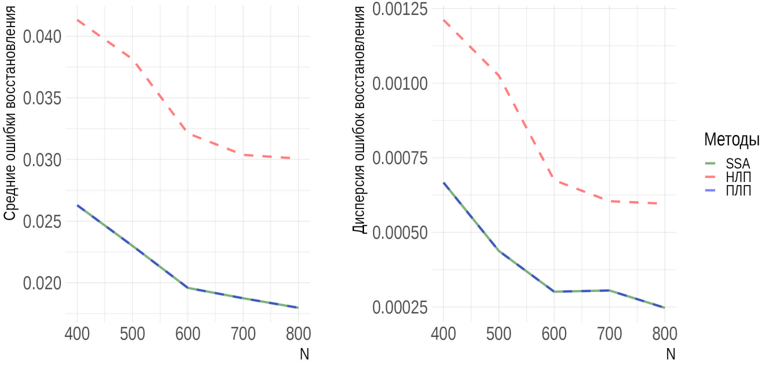

Рис.: $\max \widehat{\mathsf{E}}|r_n(\delta)|$ и $\max \widehat{\mathsf{D}}|r_n(\delta)|$. Помеха — $e_n \in \mathsf{U}(-\sqrt{6},\ \sqrt{6})$, $\delta = 0.2$.

$f_n = \cos(2\pi/\sqrt{6})$. $(d = 2)$.

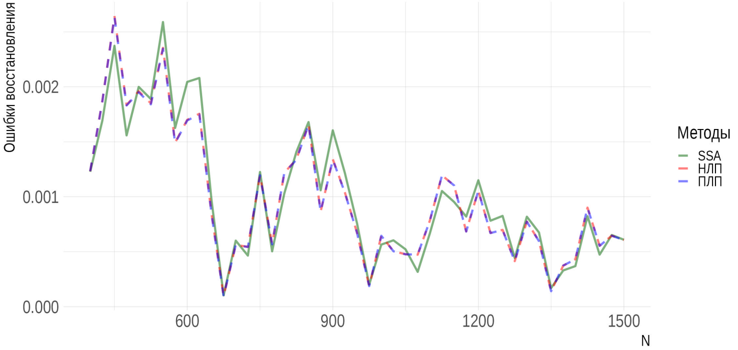

Рис.: Помеха: $e_n = \cos(2\pi/\sqrt{99}n + \pi)$, $\delta = 0.15$.

$f_n = 1.011^n$. $(d = 1)$

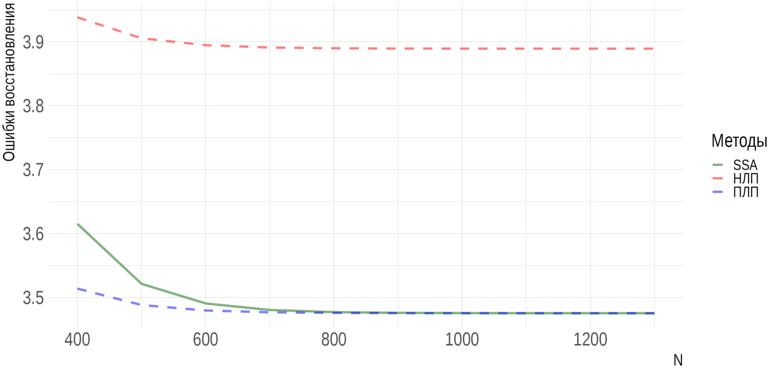

Рис.: Помеха — $e_n = 1$ при $n = N - 10$, и $e_n = 0\ \forall n \neq N - 10$, $\delta = 100$.

Растущий экспоненциальный сигнал. Дискретизация. Единичный выброс в конце ряда

$f_n = a^{(n-1)T/N}$, где $T = const$, N — длина ряда. ($a = 1.05$, $T = 10$, $d = 1$)

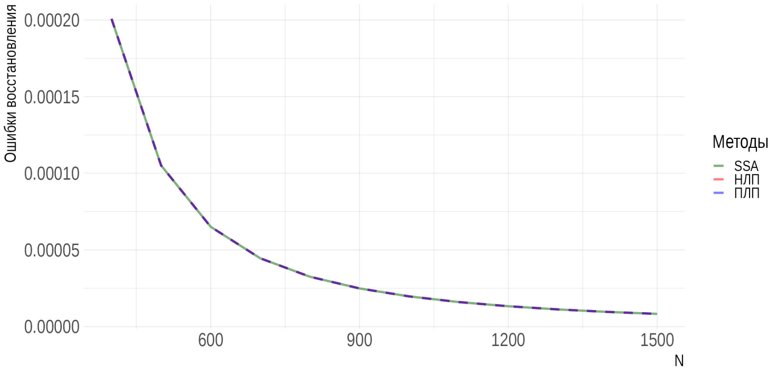

Рис.: Помеха — $e_n = 1$ при $n = N - 10,$ и $e_n = 0$ $\forall n \neq N - 10$, $\delta = 50$.

Растущий экспоненциальный сигнал. Дискретизация. Синусоидная помеха

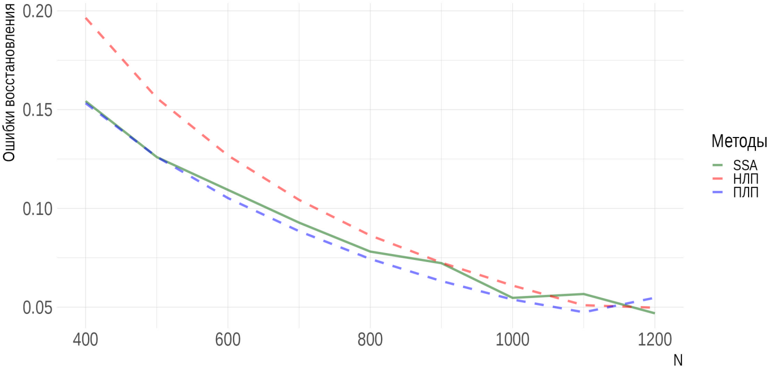

Рис.: Помеха — $e_n = \cos(2\pi/\sqrt{99}n + \pi)$, $\delta = 0.15$.

Растущий экспоненциальный сигнал. Дискретизация. Гауссовский белый шум

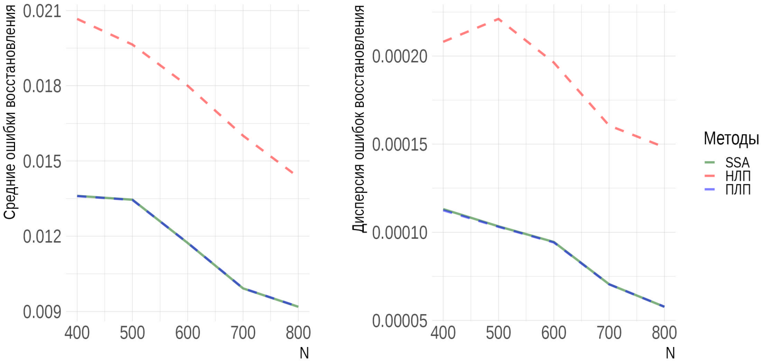

Рис.: $\max \widehat{\mathsf{E}}|r_n(\delta)|$ и $\max \widehat{\mathsf{D}}|r_n(\delta)|$. Помеха — $e_n \in \mathsf{N}(0,\ 1)$, $\delta = 0.1$.

Заключение

- Проведены численные эксперименты для сравнения ошибок восстановления метода SSA с его линейной аппроксимацией при больших длинах ряда.
- Во всех экспериментах при больших N полное линейное приближение очень хорошо аппроксимирует ошибки восстановления, при этом даже при отсутствии сходимости ряда $R_N(\delta)$ к нулю.
- Полная линейная аппроксимация даёт лучшие приближения ряда ошибок восстановления относительно неполного варианта.
- При равномерном или гауссовском белом шуме полное линейное приближение позволяет получить хорошие оценки ошибок восстановления в терминах математического ожидания и дисперсии.
- Рассмотрен неполный член линейного приближения, который дает близкие аппроксимации с ПЛП в случае, если есть априорная информация, что помеха имеет синусоидную форму.
- Эксперименты выполнены на языке R с пакетом RSSA, исходные файлы для воспроизведения результатов доступны по ссылке https://doi.org/10.5281/zenodo.11373477.