Исследование ошибок восстановления в методе «Гусеница» с помощью теории возмущений

Власьева Екатерина Михайловна, гр. 522

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н. Голяндина Н.Э.
Рецензент: к.ф.-м.н. Некруткин В.В.

Санкт-Петербург
2008г.

Наблюдается ряд «Сигнал + шум» $F = S(\delta) = S + \delta E$ длины N, где

$$S = (s_0, \ldots, s_{N-1}), \quad E = (\varepsilon_0, \ldots, \varepsilon_{N-1}), \quad \mathbb{D}\varepsilon_i = \sigma^2, \ \mathbb{E}\varepsilon_i = 0,$$

Метод «Гусеница»-SSA:
L — длина окна (параметр метода), $1 < L < N$.

$$S + \delta E \xrightarrow{L} \mathbf{S} + \delta\mathbf{E} \xrightarrow{\mathbf{P}} \widetilde{\mathbf{S}} = \mathbf{S} + \delta\mathbf{S}^{(1)} + \delta^2\mathbf{S}^{(2)} + \ldots \xrightarrow{\mathcal{H}}$$
$$\xrightarrow{\mathcal{H}} \widetilde{S} = S + \underbrace{\delta S^{(1)} + \delta^2 S^{(2)} + \ldots}_{\text{ошибка восстановления}}$$

$S^{(1)}$ — первый член ошибки восстановления.
$\delta S^{(1)}$ адекватно описывает $\widetilde{S} - S$ — полную ошибку восстановления.
Задача:

1. Найти дисперсию первого порядка ошибки восстановления $\mathbb{D}S^{(1)}$;
2. Исследовать зависимость дисперсии ошибки от σ^2, L, N.

Постановка задачи

Рассматривался метод «Гусеница» для анализа одного ряда и системы двух рядов.

SSA	MSSA
$S = (s_0, \ldots, s_{N-1})$	$\begin{cases} S_{(1)} = \left(s_{(1)0}, \ldots, s_{(1)N-1}\right) \\ S_{(2)} = \left(s_{(2)0}, \ldots, s_{(2)N-1}\right) \end{cases}$

Исследования проводились на

1. $s_i \equiv c$	$\begin{cases} s_{(1)i} \equiv c_{(1)} \\ s_{(2)i} \equiv c_{(2)} \end{cases}$	аналитически.
2. $s_i = \cos(2\pi i/5)$	$\begin{cases} s_{(1)i} = \cos(2\pi i/5) \\ s_{(2)i} = A\cos(2\pi i/5) \end{cases}$	моделированием.

Алгоритм SSA: $F = S + \delta E$

1. Шаг 1: Вложение: $F \xrightarrow{L} \mathbf{F}$.
2. Шаг 2: Сингулярное разложение: $\mathbf{F}$, $\lambda_1 \geqslant \ldots \geqslant \lambda_d > 0$ — собственные числа $\mathbf{X} = \mathbf{F}\mathbf{F}^{\mathrm{T}}$, $U_1, \ldots, U_d$ — соответствующие собственные векторы.
3. Шаг 3: Выбор набора собственных векторов и проектирование на собственное подпространство: $\mathbf{F} \xrightarrow{\mathbf{P}} \widetilde{\mathbf{S}}$, $\mathbf{P}$ — проектор на $\mathbb{U} = \operatorname{span}_{i \in J} \{U_i\}$, $J \subseteq \{1, \ldots, d\}$.
4. Шаг 4: Восстановление (диагональное усреднение): $\widetilde{S} = \mathcal{H}\left(\widetilde{\mathbf{S}}\right)$.

Основная операция — это проектирование.

Вопрос

Как влияет возмущение матрицы на результат проектирования?

$\mathbf{X}$ — симметричная неотрицательно определенная матрица,
$\mu_1 > \mu_2 > \ldots > \mu_r > \mu_{r+1} = 0$ — различные собственные числа $\mathbf{X}$,
$\mathbb{U}_1, \ldots, \mathbb{U}_r, \mathbb{U}_{r+1}$ — соответствующие ортогональные собственные подпространства, $\mathbf{P}_k \colon \mathbb{R}^L \to \mathbb{U}_k$
Пусть $\mathbf{X}(\delta) = \mathbf{X} + \delta \mathbf{X}^{(1)} + \delta^2 \mathbf{X}^{(2)} + \ldots$.
Тогда (формальный ряд) $\mathbf{P}_k(\delta) = \mathbf{P}_k + \delta \mathbf{P}_k^{(1)} + \delta^2 \mathbf{P}_k^{(2)} + \ldots$.
Нас будет интересовать $\mathbf{P}_k^{(1)}$ — первый формальный порядок возмущения, который имеет вид:

$$\mathbf{P}_k^{(1)} = \sum_{i \neq k,\ 1 \leqslant i \leqslant r+1} \frac{1}{\mu_k - \mu_i} \left(\mathbf{P}_k \mathbf{X}^{(1)} \mathbf{P}_i + \mathbf{P}_i \mathbf{X}^{(1)} \mathbf{P}_k \right).$$

Пусть $\mathbf{S}$ — траекторная матрица сигнала, $\mathbf{X} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$,
$\mathbf{S}(\delta) = \mathbf{S} + \delta\mathbf{E}$, где $\mathbf{E}$ — траекторная матрица шума, и $\mathbf{X}(\delta) = \mathbf{S}(\delta)\mathbf{S}(\delta)^{\mathrm{T}}$,
Тогда

$$\mathbf{X}(\delta) = \mathbf{X} + \delta\mathbf{X}^{(1)} + \delta^2\mathbf{X}^{(2)},$$

где $\mathbf{X}^{(1)} = \mathbf{S}\mathbf{E}^{\mathrm{T}} + \mathbf{E}\mathbf{S}^{\mathrm{T}}$ и $\mathbf{X}^{(2)} = \mathbf{E}\mathbf{E}^{\mathrm{T}}$.

Нас интересует результат проекции $\mathbf{P}_k$ на собственное подпространство $\mathbb{U}_k$, т.е. $\mathbf{S}_k(\delta) = \mathbf{P}_k(\delta)\mathbf{S}(\delta)$:

$$\mathbf{S}_k(\delta) = \mathbf{S}_k + \delta\mathbf{S}_k^{(1)} + \ldots.$$

Пример $(d = r = 1,\ k = 1)$:

$$\mathbf{S}_1^{(1)} = \frac{1}{\mu_1}\left(\mathbf{X}^{(1)}\mathbf{X}_1 - \mathbf{P}_1\mathbf{X}^{(1)}\mathbf{X}_1\right) + \mathbf{P}_1\mathbf{E},$$

где $\mathbf{X}_1 = \mathbf{P}_1\mathbf{S}$, $\mathbf{P}_1$ — проектор на $\mathbb{U}_1$.

Одномерный метод «Гусеница»

$$S \xrightarrow{L} \mathbf{S}$$

$$E \xrightarrow{L} \mathbf{E}$$

Двумерный метод «Гусеница»

$$S_{(1)} \xrightarrow{L} \mathbf{S}_{(1)} \; S_{(2)} \xrightarrow{L} \mathbf{S}_{(2)}$$
$$\mathbf{S} = \left(\mathbf{S}_{(1)} : \mathbf{S}_{(2)}\right)$$
$$E_{(1)} \xrightarrow{L} \mathbf{E}_{(1)} \; E_{(2)} \xrightarrow{L} \mathbf{E}_{(2)}$$
$$\mathbf{E} = \left(\mathbf{E}_{(1)} : \mathbf{E}_{(2)}\right)$$

$$\mathbf{S}(\delta) = \mathbf{S} + \delta \mathbf{E}$$

$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_d > 0$ — собственные числа $\mathbf{X} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$, $d = \operatorname{rank} \mathbf{S}$

U_i, $i = 1, \ldots, d$ — собственные векторы матрицы $\mathbf{X}$, соответствующие λ_i

V_i, $i = 1, \ldots, d$ — собственные векторы матрицы $\mathbf{X}^{\mathrm{T}}$, соответствующие λ_i

$d = 1$

$$\mathbf{S}_1^{(1)} = -\left(U_1^{\mathrm{T}} \mathbf{E} V_1\right) U_1 V_1^{\mathrm{T}} + U_1 U_1^{\mathrm{T}} \mathbf{E} + \mathbf{E} V_1 V_1^{\mathrm{T}} = = -\frac{U_1^{\mathrm{T}} \mathbf{E} V_1}{\sqrt{\lambda_1}} \mathbf{F} + U_1 U_1^{\mathrm{T}} \mathbf{E} + \mathbf{E} V_1 V_1^{\mathrm{T}}.$$

$d = 2$

$$\mathbf{S}_1^{(1)} = U_1 U_1^{\mathrm{T}} \mathbf{E} + U_2 U_2^{\mathrm{T}} \mathbf{E} + \mathbf{E} V_1 V_1^{\mathrm{T}} + \mathbf{E} V_2 V_2^{\mathrm{T}} - - \left(\alpha_{11} U_1 V_1^{\mathrm{T}} + \alpha_{12} U_1 V_2^{\mathrm{T}} + \alpha_{21} U_2 V_1^{\mathrm{T}} + \alpha_{22} U_2 V_2^{\mathrm{T}}\right),$$

где $\alpha_{ij} = U_i \mathbf{E} V_j^{\mathrm{T}}$.

$F = S(\delta) = S + \delta E$,
N — длина сигнала,
L — длина окна,
$\widetilde{S} = S + \delta S^{(1)} + \delta^2 S^{(2)} + \ldots$,
$S^{(1)}$ — первый порядок ошибки восстановления сигнала,
l — номер точки ряда,
$\lambda = \lim\limits_{N\to+\infty} 2l/N$ — положение точки относительно края (относительная координата точки).

Одномерный метод «Гусеница», $L = \alpha N$

$S \equiv c$, $\mathbb{D}\varepsilon_i = \sigma^2$.

Дисперсия первого порядка ошибки восстановления (для $\delta = 1$):

$$\mathbb{D}s_l^{(1)} \underset{N\to+\infty}{\sim} F(\alpha,\lambda,N) = \frac{\sigma^2}{N} \begin{cases} P_2(\alpha,\lambda), \ 0 \leqslant \lambda < 2(1-2\alpha) \\ P_4(\alpha,\lambda), \ 2(1-2\alpha) \leqslant \lambda < 2\alpha, \\ 2/(3\alpha), 2\alpha \leqslant \lambda \leqslant 1. \end{cases} .$$

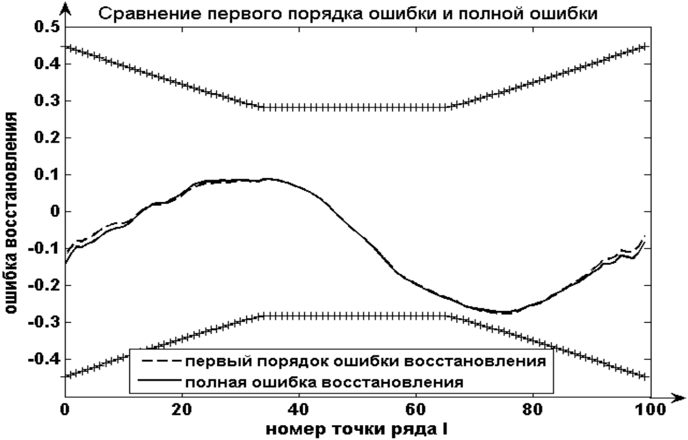
Двумерный метод «Гусеница», $L = 0.5N$

$S_{(1)} \equiv c_{(1)}, \ \ S_{(2)} \equiv c_{(2)}, \ \ \mathbb{D}\varepsilon_{(1)i} = \sigma^2_{(1)}, \ \ \mathbb{D}\varepsilon_{(2)i} = \sigma^2_{(2)}$.

Дисперсия первого порядка ошибки восстановления для первого ряда:

$$\mathbb{D}s_l^{(1)} \underset{N\to+\infty}{\sim} G(\lambda,N) = \frac{2}{3N(c^2_{(1)} + c^2_{(2)})^2} P_2(c_{(1)}, c_{(2)}, \sigma_{(1)}, \sigma_{(2)}, \lambda).$$

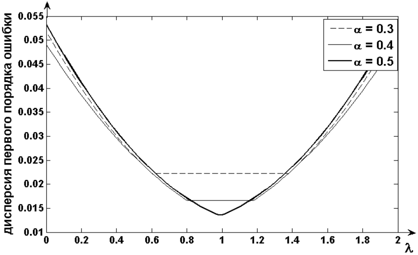
Первый порядок и полная ошибки восстановления

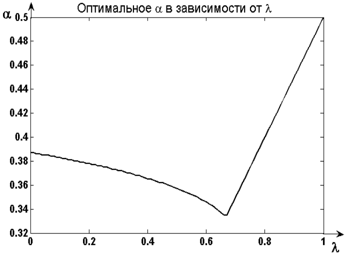


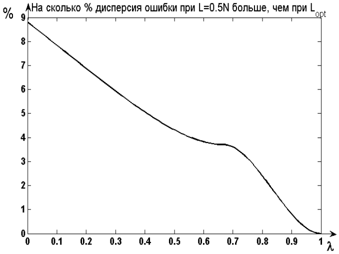
$$c = 1,\ \sigma = 1,\ L = \lceil N/3 \rceil$$

$$\delta S^{(1)} \approx \widetilde{S} - S$$

Константный сигнал. Аналитические результаты

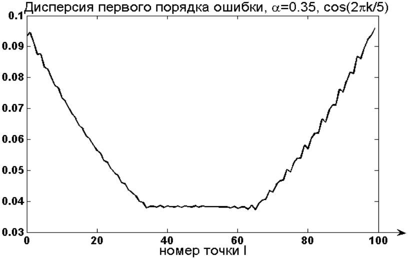






Сигнал синусоида. Моделирование

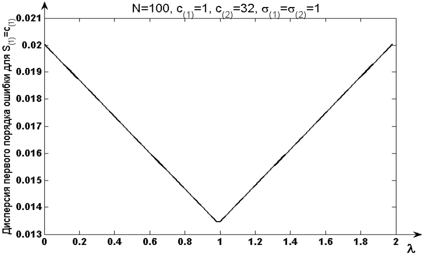
- Зависимость дисперсии первого порядка ошибки от l — номера точки:



- l — номер точки, $L = \alpha N$, $\alpha \leq 1/2$. Можно выбрать $\alpha < 1/2$, т.ч. дисперсия первого порядка ошибки в фиксированной точке l будет меньше, чем при $\alpha = 1/2$.

Константные сигналы. $L = 0.5N$. Аналитические результаты

1. Дисперсия первого порядка ошибки в зависимости от λ:



2. Для любых $c_{(1)}$, $c_{(2)}$, $\sigma_{(1)}$, $\sigma_{(2)}$ дисперсия первого порядка ошибки в центральной точке равна $4\sigma^2_{(1)}/3N$.
3. Сравнение с одномерным методом «Гусеница».
   Пусть $\sigma_{(1)} = \sigma$. Если $\sigma_{(2)} = \sigma_{(1)}$, то для любого $c_{(2)} \neq 0$ применение MSSA к системе из двух рядов лучше (дисперсия ошибки меньше), чем SSA для каждого ряда по-отдельности.
   Пример: если $\sigma_{(i)} = 1$ и $c_{(i)} = 1$, то применение MSSA в крайней точке дает выигрыш 31%.

Константный сигнал. $L = 0.5N$. Исскуственное добавление второго ряда

$$c,\ \sigma \longrightarrow \begin{cases} c_{(1)} = c,\ \ \sigma_{(1)} = \sigma \\ c_{(2)} \to +\infty,\ \ \sigma_{(2)} = 0 \end{cases}$$

При $c_{(2)} \to +\infty$ $G(\lambda, N) \to 2\sigma^2_{(1)}(3-\lambda)/3N$
На краях дисперсия первого порядка ошибки уменьшается в $8/3$ раза.

Сигналы синусоиды. Моделирование.

1. Если рассматривать две гармоники с одинаковыми периодом и шумом, то при применении MSSA к системе из двух рядов дисперсия первого порядка ошибки меньше, чем применение SSA к каждому ряду по-отдельности.
2. Искусственное добавление к одномерному ряду гармоники с тем же периодом и большой амплитудой с $\sigma_{(2)} = 0$:

На краях дисперсия первого порядка ошибки уменьшается примерно в 2.5 раза.

Заключение

1. В работе были получены аналитические формулы для дисперсии первого порядка ошибки восстановления зашумленного сигнала на примере константы. Формулы позволили получить интересные результаты, которые было бы трудно получить на практике ввиду трудоемкости моделирования.

2. Моделирование по формулам для первого порядка возмущения подтвердило, что результаты, полученные для константных рядов, имеют место и в случае синусоид.