Локальные модификации метода анализа сингулярного спектра

Сандалов Сергей, группа 15.Б04-мм

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

Научный руководитель: к.ф.-м.н., доцент Голяндина Н.Э.

Рецензент: к.ф.-м.н., системный программист Звонарев Н.К.

Санкт-Петербург
2019г.

Введение. Постановка задачи

Вещественный временной ряд $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$, где $\mathbb{S}_N = (s_1, s_2, \ldots, s_N)$ — сигнал с $s_n = A(n)\cos(2\pi\omega(n)\cdot n + \phi)$, $A(n) > 0$, а $\mathbb{R}_N = (r_1, r_2, \ldots, r_N)$ — стационарный процесс.

Цель: по наблюдаемому ряду $\mathbb{F}_N$ оценить $\mathbb{S}_N$.

Метод: SSA (Singular spectrum analysis)[Analysis of Time Series Structure: SSA and Related Techniques, Golyandina et. al. 2001].

Известно: если $A(n) = Ae^{\gamma n}$, $\omega(n) = \mathrm{const}$, а $\gamma, A \in$ R, то SSA хорошо умеет выделять сигнал.

Задача: в рамках SSA предложить подход лучше, чем базовый метод SSA, справляющийся с оценкой сигнала при более сложной модуляции как амплитуды, так и частоты.

Введение. Алгоритм SSA

$\mathbb{F}_N = (f_1, f_2, \dots, f_N)$ — временной ряд.
Параметры алгоритма: L — длина окна, r — количество компонент для получения оценки сигнала.

- **Вложение**: $1 < L < N$, $K = N - L + 1$, ряд переводится в траекторную матрицу $\mathbf{X} = [X_1 : \ldots : X_K]$, где $X_i = (f_i, \dots, f_{i+L-1})^{\mathrm{T}}$, где $1 \le i \le K$.
- **Сингулярное разложение**: $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$, $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$, где $1 \le i \le d = \mathrm{rk}(\mathbf{X})$ и $\sqrt{\lambda_i}$ — сингулярные числа, упорядоченные по невозрастанию.
- **Группировка**: $\hat{\mathbf{X}} = \sum\limits_{i=1}^{r} \mathbf{X}_i$.
- **Диагонализация**: $\hat{\mathbf{X}}$ переводится во временной ряд $\widetilde{\mathbb{S}}_N$.

Введение. Основные определения

Рассмотрим модель сигнала $\mathbb{S}_N$. $\mathbf{S}$ – траекторная матрица $\mathbb{S}_N$.

Ранг временного ряда

Если равенство $\mathrm{rk}(\mathbf{S}) = d < N/2$ имеет место для любого $L : d \leq \min(L, K)$, то будем говорить, что ряд $\mathbb{S}_N$ имеет ранг d (обозн. $\mathrm{rk}(\mathbb{S}_N) = d$; говорят $\mathbb{S}_N$ — ряд конечного ранга).

- Если $\mathrm{rk}(\mathbb{F}_N) = d$ и для SSA взять $L \geq d$ и $r = d$, то $\widetilde{\mathbb{S}}_N = \mathbb{F}_N$.
- Если $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$, где $\mathbb{R}_N$ — шум, а $\mathrm{rk}(\mathbb{S}_N) = d$, то рекомендуется брать $L \approx \frac{N}{2}$ и $r = d$.

Примеры:

- $s_n = Ae^{\alpha n}\cos(2\pi\omega n + \phi)$ — ранг 2.
- $s_n = \sum_{i=0}^{m} c_i n^i$ — ранг $m + 1$.

Введение. Основные определения

Число обусловленности

Числом обусловленности матрицы сигнала $\mathbb{S}_N$ ранга d называется $\mathrm{cond}(\mathbf{S}) = \frac{\lambda_1^{sig}}{\lambda_d^{sig}}$, где λ_i^{sig} – собственные числа матрицы $\mathbf{S}$, взятые в неубывающем порядке.

Отношение сигнал/шум

Пусть $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$, где $\mathbb{S}_N = (s_1, s_2, \ldots, s_N)$ — сигнал, а $\mathbb{R}_N = (r_1, r_2, \ldots, r_N)$ — стац. процесс. SNR $= \frac{\frac{1}{N}\sum_{i=1}^N s_i^2}{\mathrm{E}r_1^2}$.

Замечание: чем больше SNR (Signal-to-Noise Ratio), тем сильнее сигнал по отношению к шуму.

Амплитудная модуляция

Рассмотрим два амплитудно-модулированных сигнала:

- Экспоненциально-модулированный косинус $\mathbb{S}_N^{(1)} = e^{-\frac{n}{80}} \cos(2\pi \frac{1}{12} n)$, ранг 2.
- Квадратично-модулированный косинус $\mathbb{S}_N^{(2)} = (15 \cdot 10^{-5} n^2 - 0.02n + 1) \cos(2\pi \frac{1}{12} n)$, ранг 6.

Рис.: $\mathbb{S}_N^{(1)}$, $\mathrm{cond}(\mathbf{S}^{(1)}) = 1.008$

Рис.: $\mathbb{S}_N^{(2)}$, $\mathrm{cond}(\mathbf{S}^{(2)}) = 126.187$

Добавим шум с SNR $= 5$ к сигналам и применим SSA с $L = 60$.

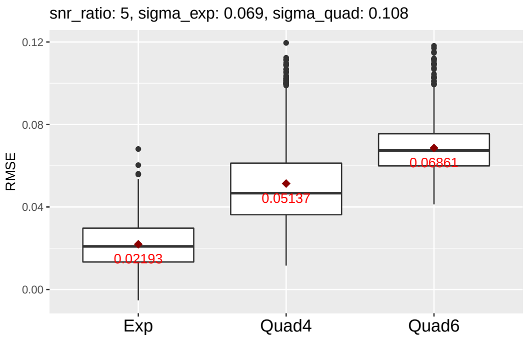
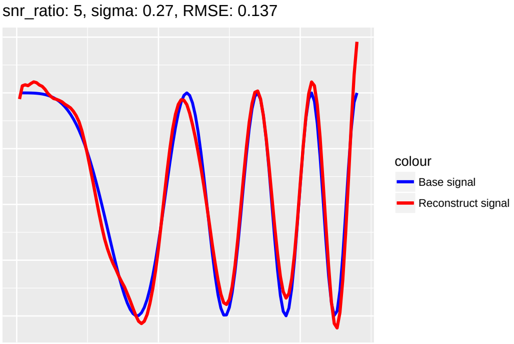


Рис.: Средн. ошибка оценки $\mathbb{S}_N^{(1)}$ по 2-ум комп., $\mathbb{S}_N^{(2)}$— по 4-м и 6-ти комп.

Вывод: для сигнала с полиномиальной модуляцией выбор $4 < 6{=}\mathrm{rk}(\mathbb{S}_N^{(2)})$ оказался лучше.

Пусть имеется временной ряд $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$, где $\mathbb{S}_N = \cos(2\pi(\frac{n}{60})^2)$ — сигнал , а $\mathbb{R}_N$ — белый гауссовский шум с $\mathrm{SNR} = 5$.

При *оптимальных параметрах L=34*, $r = 5$ оценка получается плохая.



Общие черты локальных алгоритмов

Рассмотрим следующие локальные алгоритмы:

- Overlap SSA [Leles et. al. 2017]
- Sliding SSA [Harmouche et. al. 2018]
- Averaging SSA — обобщение Overlap SSA

Общая идея: применять SSA на небольших отрезках ряда, а потом объединять результаты.

На вход алгоритмам подаются:

- $\mathbb{F}_N = (f_1, f_2, \ldots, f_N)$ – временной ряд
- Z – длина локального сегмента ряда
- q – количество элементов внутри локального сегмента, используемое для восстановления
- L – длина окна
- r – количество компонент для оценки сигнала

Общие соотношения на параметры: $L, q < Z$, $r \leq L$.

Описание локальных алгоритмов

Схема локальных алгоритмов:

- С помощью SSA происходит анализ не всего ряда, а только его отрезков, длины Z.
- Рассматриваются скользящие отрезки ряда с заданным сдвигом.
- Из оценки сигнала берутся не все значения, а только некоторые, расположенные специальным образом.
- Результаты по отрезкам объединяются некоторым образом для получения оценки сигнала всего ряда.

	Выбор	Сдвиг	Объед-е
Overlap SSA	q точек из середины	q	стык
Sliding SSA	q точек из левой половины	q	стык
Averaging SSA	q точек из середины	1	усредн-е

Соотношения между алгоритмами

- Если положить $q = 1$, то все предложенные локальные алгоритмы совпадают между собой.
- Если положить $Z = N$, где N – длина ряда (нечетно), то все предложенные локальные алгоритмы совпадают с глобальным методом SSA, примененным ко всему ряду целиком.

Постановка численного эксперимента

Рассмотрим вещественный временной ряд $\mathbb{F}_N = \mathbb{S}_N + \mathbb{R}_N$, где $\mathbb{S}_N = (s_1, s_2, \ldots, s_N)$ с $s_n = A(n)\cos(2\pi\omega(n)\cdot n + \phi)$ — сигнал, а $\mathbb{R}_N = (r_1, r_2, \ldots, r_N)$ — белый гауссовский шум.

Задача: сравнить локальные алгоритмы и стандартный SSA.

Примеры сигналов:

- Экспоненциально-модулированный косинус $s_n = e^{-\frac{n}{80}}\cos(2\pi\frac{1}{12}n)$, ранг 2.
- Квадратично-модулированный косинус $s_n = (15\cdot 10^{-5}n^2 - 0.02n + 1)\cos(2\pi\frac{1}{12}n)$, ранг 6.
- Частотно-модулированный косинус $s_n = \cos(2\pi\cdot(\frac{n}{60})^2)$, ряд не конечного ранга.

Постановка численного эксперимента

1. Выбрать вид выделяемого сигнала.
2. Зафиксировать $\mathrm{SNR} = 4$ и $N = 121$ — длина ряда.
3. Промоделировать $k = 5000$ реализаций белого гауссовского шума при выбранном SNR и добавить их к сигналу. Получится выборка из временных рядов.
4. Для каждого из указанных методов найти оптимальные параметры, при которых среднее MSE оценки сигнала минимально на полученной выборке. Оптимальные параметры находятся полным перебор по решетке значений.
5. Сравнить методы при найденных оптимальных параметрах на выборке из временных рядов, полученной в пункте 3.

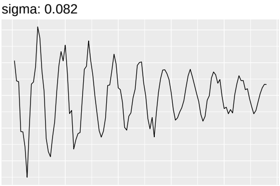


Рис.: Эксп.-модулир. сигнал с SNR=4

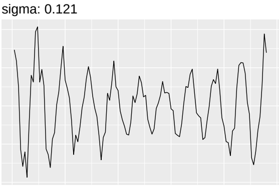


Рис.: Квадр.-модулир. сигнал с SNR=4

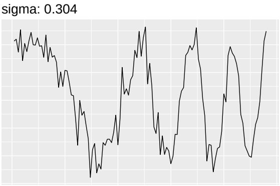


Рис.: Частотно-модулир. сигнал с SNR=4

Экспоненциальная модуляция

Таблица: Экспоненциально-модулированный косинус

	Z	q	L	r	RMSE
SSA			60	2	0.03906
Overlap SSA	121	$\forall$	60	2	0.03906
Sliding SSA	121	$\forall$	60	2	0.03906
Averaging SSA	121	$\forall$	60	2	0.03906

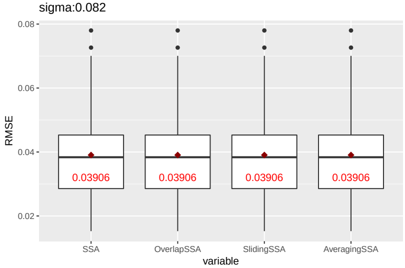


Таблица: Квадратично-модулированный косинус

	Z	q	L	r	RMSE
SSA			27	2	0.06784
Overlap SSA	41	1	20	2	0.05587
Sliding SSA	41	1	20	2	0.05887
Averaging SSA	41	1	20	2	0.05887

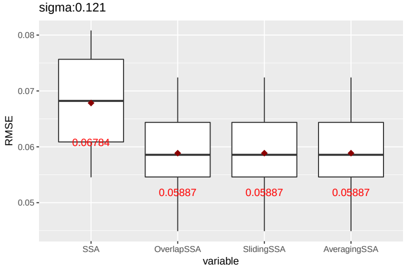
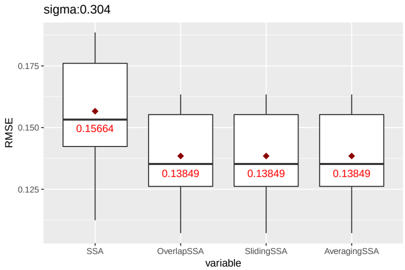


Таблица: Частотно-модулированный косинус

	Z	q	L	r	RMSE
SSA			34	5	0.15664
Overlap SSA	29	1	13	2	0.13849
Sliding SSA	29	1	13	2	0.13849
Averaging SSA	29	1	13	2	0.13849



Выводы

- Все три локальных метода совпали между собой при выборе оптимальных параметров.
- Для экспоненциальной модуляции лучше всех справился стандартный SSA.
- В случае квадратичной модуляции локальные методы показали значимое преимущество перед SSA.
- Для частотной модуляции преимущество локальных методов еще сильнее.
- Оптимальным значением q^* для всех алгоритмов оказалось значение $q = 1$, т.е. используется только центральная точка восстановленного ряда для каждого отрезка.
- В результате моделирования выяснено, что чем сложнее вид сигнала для SSA, тем меньше должно быть Z.

Заключение

Что сделано:

- Изучена теория метода SSA.
- Разобраны и реализованы локальные алгоритмы Overlap SSA, Sliding SSA и Averaging SSA на языке R.
- Проведено численное сравнение локальных алгоритмов.
- Сделаны рекомендации по выбору параметров локальных модификаций на качественном уровне.