Исследование условий для поддерживающих временных рядов в MSSA

Ткаченко Егор Андреевич, гр.19.Б04-мм

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

Отчет по производственной практике (семестр 6)

Санкт-Петербург, 2022

Научный руководитель к.ф.-м.н., доцент Голяндина Нина Эдуардовна, кафедра статистического моделирования

Введение

В данной работе решается задача прогноза временного ряда, с помощью алгоритмов SSA и MSSA.
Цель работы — выяснить при каких параметрах алгоритм MSSA дает результат лучше SSA.

Базовые определения

Временной ряд

Вещественный временной ряд длины N:

$$\mathsf{F} = (f_1, \ldots, f_N),\ \ f_j \in \mathbb{R}.$$

Многомерный временной ряд

Многомерный временной ряд $\vec{\mathsf{F}}$ — набор s временных рядов $\mathsf{F}^{(p)}$ длин N_p:

$$\vec{\mathsf{F}} = \{\mathsf{F}^{(p)} = (f_1^{(p)}, \ldots, f_{N_p}^{(p)}),\ p = 1, \ldots, s\}.$$

Определения

Траекторная матрица

L-Траекторная матрица ряда F:

$$\mathcal{T}_{\mathsf{SSA}}(\mathsf{F}) = \begin{pmatrix} f_1 & f_2 & \dots & f_K \\ f_2 & f_3 & \dots & f_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & \dots & f_N \end{pmatrix}.$$

для многомерного ряда $\vec{\mathsf{F}}$:

$$\mathcal{T}_{\mathsf{MSSA}}(\vec{\mathsf{F}}) = [\mathcal{T}_{\mathsf{SSA}}(\mathsf{F}^{(1)}) : \ldots : \mathcal{T}_{\mathsf{SSA}}(\mathsf{F}^{(s)})].$$

Из траекторной матрицы можно восстановить ряд.

L-Ранг ряда

L-Ранг ряда — это ранг его траекторной матрицы:

$$r_p = \mathrm{rank}_L\ \mathsf{F} = \mathrm{rank}\ \mathcal{T}_{\mathsf{SSA}}(\mathsf{F}), \qquad \mathrm{rank}_L\ \vec{\mathsf{F}} = \mathrm{rank}\ \mathcal{T}_{\mathsf{MSSA}}(\vec{\mathsf{F}}).$$

Ранг ряда

Ряд называется рядом конечного ранга r, если его L-ранг равен r для любой длины окна L и любой достаточно большой длины N.

Алгоритмы SSA и MSSA

Вход: Ряд F_1 для SSA или многомерный ряд F для MSSA; длина окна $L \le N_1$ для SSA или $L \le N_p$ для MSSA.

Алгоритм

1. Вложение. Временной ряд переводится в L-траекторную матрицу $\mathbf{X}$
2. Сингулярное разложение. Методом SVD матрица $\mathbf{X}$ раскладывается на сумму d матриц $\mathbf{X}_i$ ранга 1.
3. Группировка. Множество индексов $\{1, \ldots, d\}$ делится на m непересекающихся множеств $I_1, \ldots, I_m$. Матрицы $\mathbf{X}_i$ суммируются в m матриц $\mathbf{X}_{I_j}$
4. Восстановление. Сгруппированные матрицы $\mathbf{X}_{I_j}$ диагональным усреднением восстанавливаются в ряды.

Выход: Разложение исходного ряда на сумму m рядов.

Линейная рекуррентная формула; управляемый ЛРФ ряд

Ряд $\mathsf{F}_\mathsf{p} = (f_i)_{i=1}^{N_p}$ — управляемый ЛРФ, если существуют такие $a_1, \ldots, a_d$, что:

$$f_{i+d} = \sum_{k=1}^{d} a_k f_{i+d-k},\ 1 \leq i \leq N_p - d,\ a_d \neq 0,\ d < N_p - 1.$$

Прогноз ряда

Прогноз вещественного временного ряда F_p:

$$\tilde{\mathsf{f}}_{N_p} = \sum_{k=1}^{L-1} a_k f_{N_p - k}.$$

Задача

Пусть имеется временной ряд $F_1 = S_1 + R_1$, где

- Сигнал S_1 — ряд управляемый ЛРФ.
- Шум R_1 — ряд без структуры.

Задача: спрогнозировать сигнал S_1.
Пусть помимо ряда F_1 имеется временной ряд F_2.
Идея: использование ряда F_2 может улучшить прогноз сигнала S_1.

- Второй ряд дает алгоритму больше данных, которые могут улучшить ЛРФ.
- Второй ряд может сделать прогноз хуже, если его структура отличается от первого.

Ошибка прогноза $\widetilde{\mathsf{S}}$ сигнала S_1

$\mathsf{MSE}(\widetilde{\mathsf{S}}, \mathsf{S}_1) = \frac{1}{N_f}\sum_{i=N+1}^{N+N_f}(\widetilde{s}_i - s_i)^2$

Поддерживающий ряд (для прогноза)

Ряд F_2 — поддерживающий, если

$\mathsf{MSE}(\widetilde{\mathsf{S}}_{\mathsf{MSSA}}, \mathsf{S}_1) < \mathsf{MSE}(\widetilde{\mathsf{S}}_{\mathsf{SSA}}, \mathsf{S}_1)$

Вопрос: Как понять, что ряд поддерживающий?

Согласованность

- Сигналы $\mathsf{S}_1, \mathsf{S}_2$ полностью согласованы, если $r_{MSSA} = r_1 = r_2$
- Сигналы $\mathsf{S}_1, \mathsf{S}_2$ полностью не согласованы, если $r_{MSSA} = r_1 + r_2$

Относительная ошибка

Относительная ошибка прогноза (восстановления)

$$error_{rel} = \frac{error_{\mathsf{SSA}} - error_{\mathsf{MSSA}}}{error_{\mathsf{SSA}} + error_{\mathsf{MSSA}}},$$

где $error_{\mathsf{SSA}}, error_{\mathsf{MSSA}}$ — ошибки прогноза (восстановления) методами SSA и MSSA соответственно.

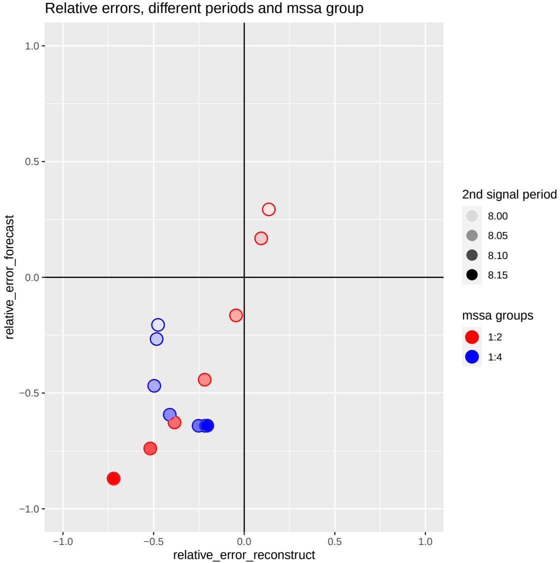
Как интерпретировать значения относительной ошибки?

- значения больше 0 значат, что что MSSA лучше SSA;
- значения меньше 0 значат, что что MSSA хуже SSA;
- значения около 0 значат что ошибки примерно равны;
- значения далеко от 0 значат, что ошибки сильно отличаются.

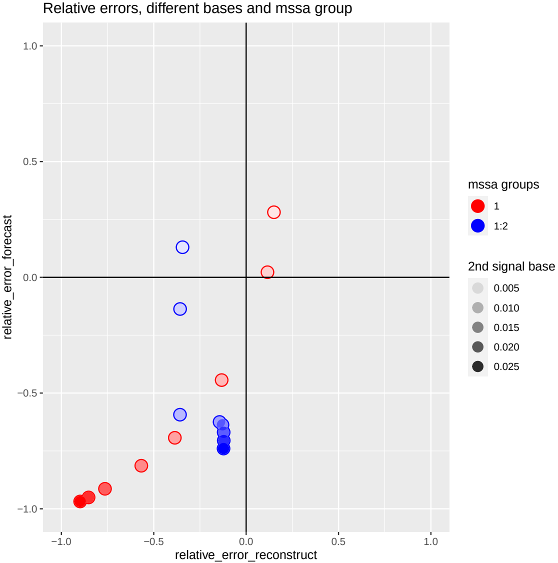
Выбор количества компонент для MSSA

Когда сигналы похожи, их можно считать согласованными и лучше использовать (при прогнозе или восстановлении сигнала) ранг равный рангу одного сигнала. Когда сигналы отличаются, их следует считать не согласованными и использовать ранг равный сумме рангов сигналов.
Будет ли ошибка MSSA меньше при таком выборе ранга для алгоритма MSSA в восстановлении и прогнозе первого ряда. И будут ли при этом вторые ряды поддерживающими.

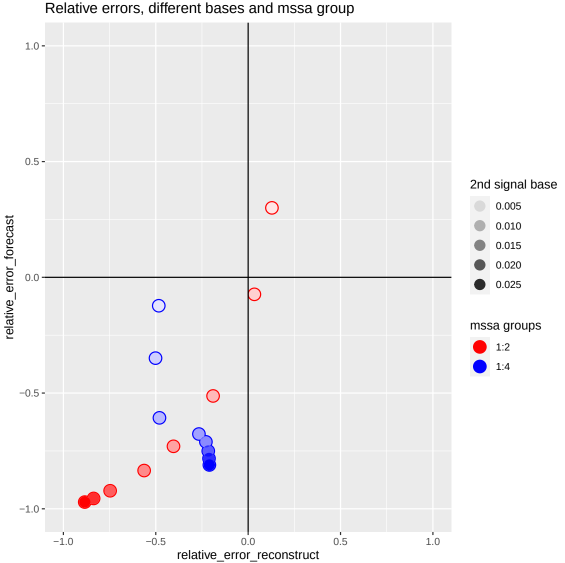
Относительные ошибки для косинуса



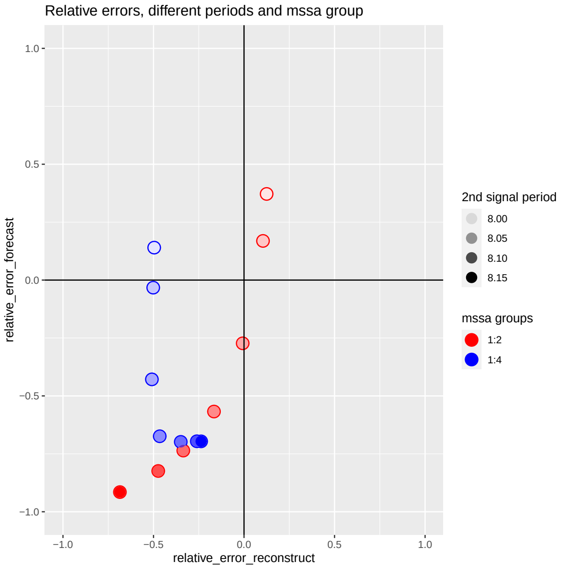
Относительные ошибки для экспоненты



Относительные ошибки для экспоненты



Относительные ошибки для экспоненты



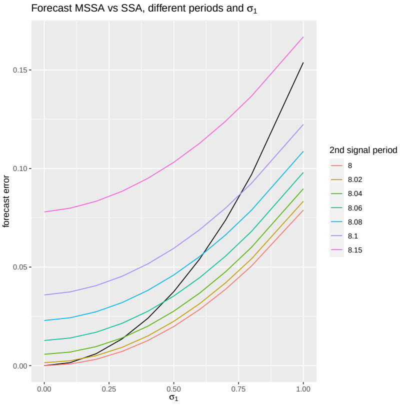
Результат первого эксперимента

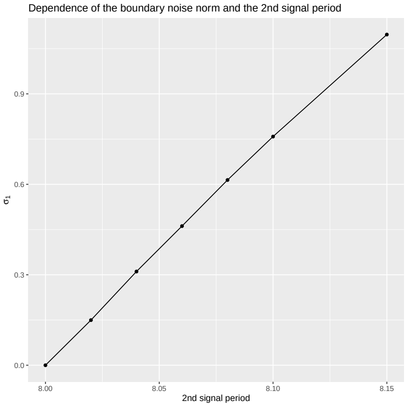
Для всех видов сигналов при отклонении второго сигнала от первого всегда наступал момент, когда использование удвоенного ранга дает меньшие ошибки прогноза и восстановления.
Но при этом, второй ряд редко оказывался поддерживающим, потому что большая часть наблюдений находилась в нижней левой четверти.

Ошибки прогноза для разных шумов первого ряда и параметров второго ряда

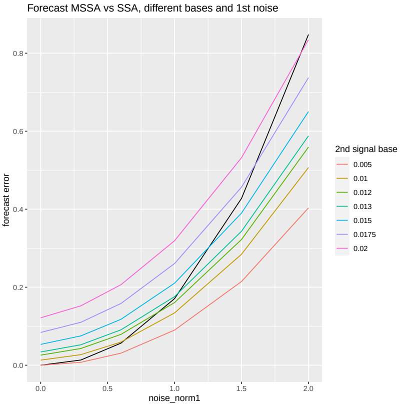
Гипотеза: при увеличении шума первого ряда, MSSA станет лучше для любого отклонения второго ряда. Если это так, то можно найти зависимость граничного значения σ_1 (при котором SSA становится хуже MSSA) от изменения параметра второго сигнала.

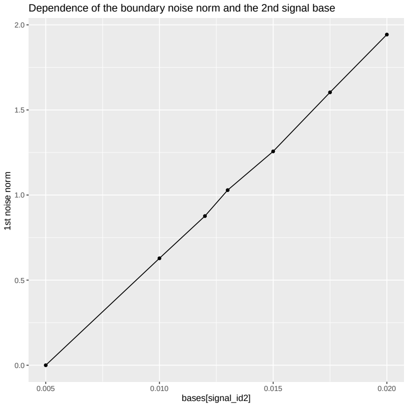
Сигнал косинус





Сигнал экспонента

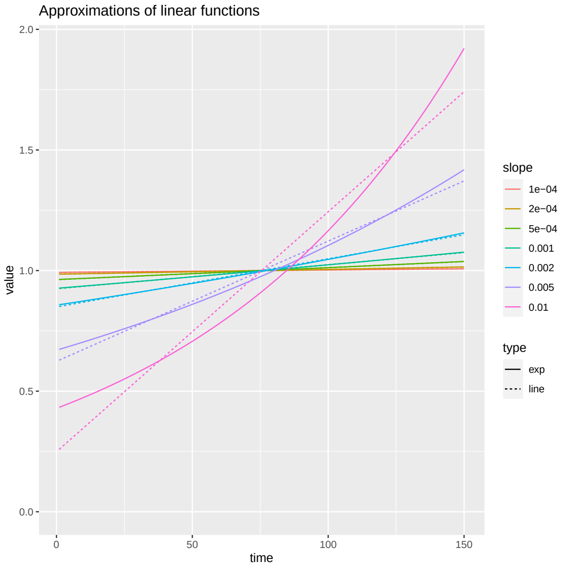




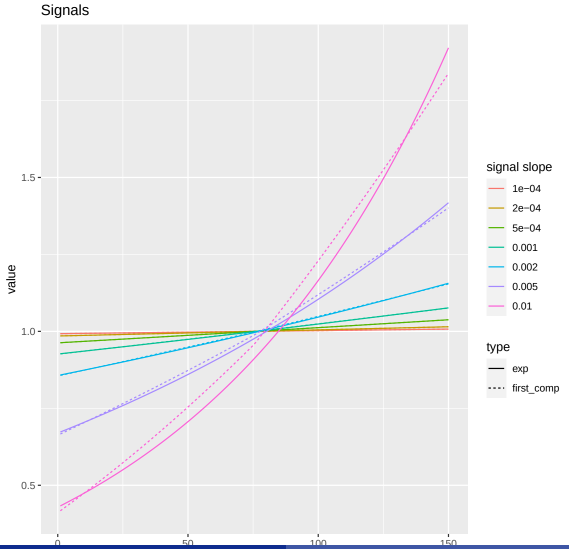
Результат второго эксперимента

Гипотеза подтверждена, зависимость граничных значений σ_1 от отклонения второго сигнала линейная.

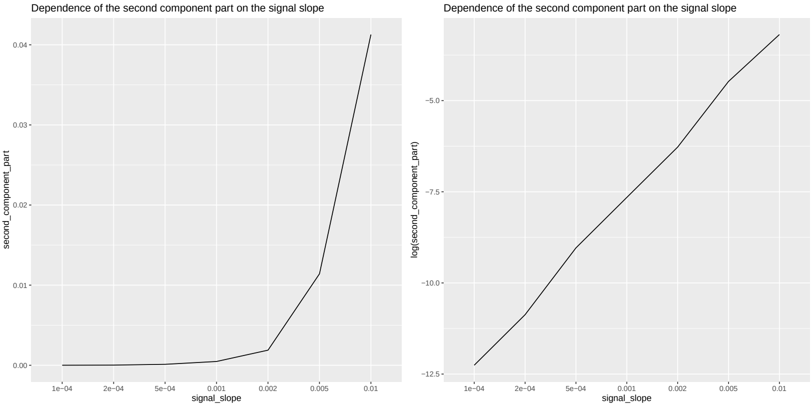
Третий эксперимент: линейные сигналы



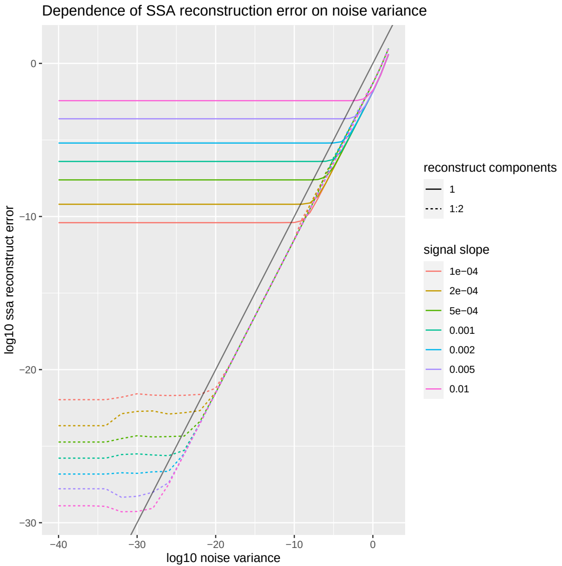
Является ли первая компонента разложения линейного ряда показательной функцией?



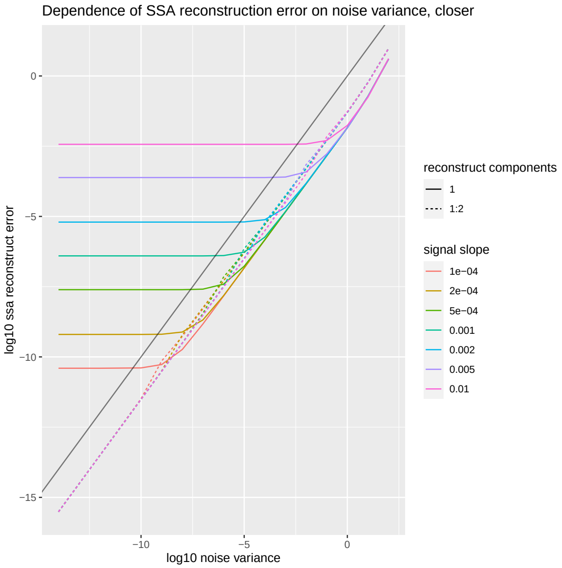
Зависимость доли второй компоненты от угла наклона



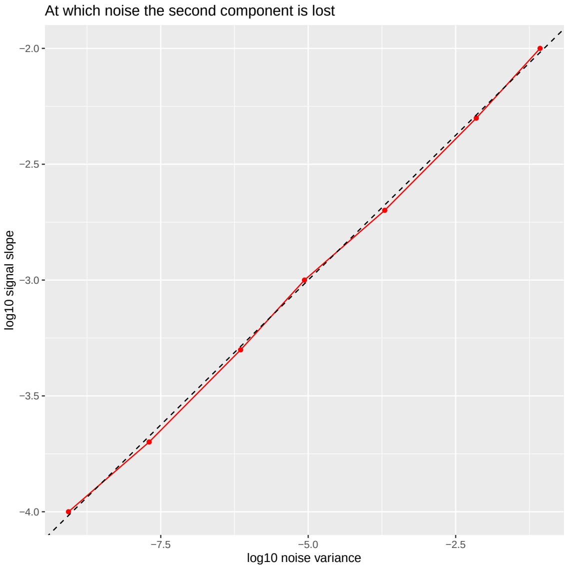
При каком шуме вторая компонента теряется



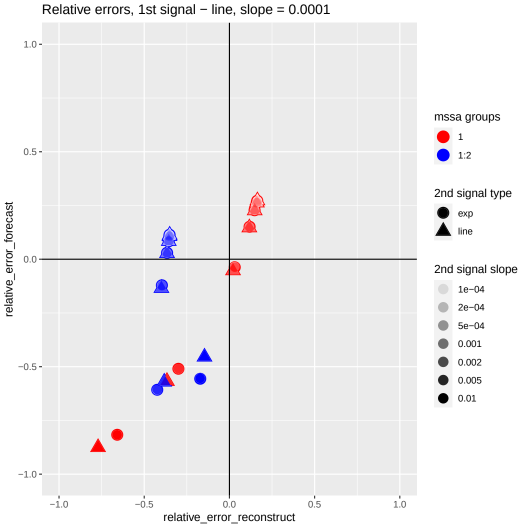
При каком шуме вторая компонента теряется



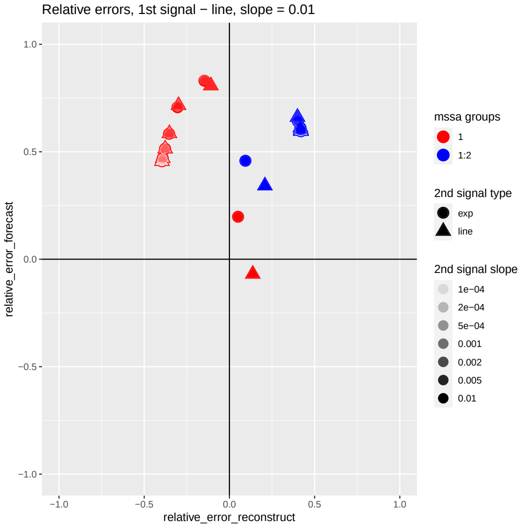
При каком шуме вторая компонента теряется



Сравнение линейного ряда и его аппроксимации как поддерживающих рядов



Сравнение линейного ряда и его аппроксимации как поддерживающих рядов



Результат третьего эксперимента

Так как первая компонента разложения линейного ряда оказалась не экспонентой, это значит, что сигналы не полностью согласованы.
Для линейных функций с большим наклоном алгоритм $/MSSA$ дает результат лучше чем $/SSA$, а маленьким наклоном наоборот.

Заключение

Найдено много интересных зависимостей.

Экспоненциальную аппроксимацию линейного ряда можно использовать в качестве поддерживающего ряда для линейных рядов с большим наклоном.

Список литературы

- Golyandina N, Korobeynikov A, Zhigljavsky A. Singular Spectrum Analysis with R. — Springer, 2018. — P. 272.
- Федоров Н. Поддерживающие временные ряды в анализе сингулярного спектра. — 2020. — выпускная квалификационная работа магистра, СПбГУ, СПб.