Использование метода поиска наилучшей проекции для анализа сингулярного спектра

Ломтев Максим Александрович, гр. 522

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

Научный руководитель: к.ф.-м.н., доцент Голяндина Н. Э.
Рецензент: к.ф.-м.н., доцент Коробейников А. И.

Санкт-Петербург
2013г.

Постановка задачи

$F_N = (f_0, \ldots, f_{N-1})$ — временной ряд. Пусть $F_N = F_N^{(1)} + F_N^{(2)}$.

Задача: по наблюдаемому ряду F_N найти, может быть, приближенно, $F_N^{(1)}$ и $F_N^{(2)}$.

Говорим, что метод разделяет ряды, если он может решить задачу нахождения слагаемых.

Метод: SSA (анализ сингулярного спектра).

Требуется: рассмотреть модификацию SSA, которая в ряде случаев улучшает разделимость рядов.

Общая схема методов

Вход — ряд $F_N = (f_0, \ldots, f_{N-1})$ длины N.

Схема методов, связанных с SSA:

- Выбрать *длину окна* L и сопоставить ряду *траекторную матрицу* $\mathbf{X}$ размера $L \times K$, $L \leq K = N - L + 1$, из отрезков ряда длины L в качестве столбцов. Траекторной матрице соответствует пространство $\mathcal{L}$ строк размерности $d \leq L$.
- Разложить траекторную матрицу в сумму матриц $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$, или, эквивалентно, найти базис $\{S_1, ..., S_d\}$ пространства $\mathcal{L}$.
- Сгруппировать матрицы в две группы или, эквивалентно, разбить базис на две части.
- Получить разложение $F = F^{(1)} + F^{(2)}$ на основе группировки.

Методы SSA и ICA

SSA: Базис находится с помощью сингулярного разложения траекторной матрицы.

ICA: Базис находится с помощью поиска наилучшей проекции (Projection Pursuit, PP) для решения задачи анализа независимых компонент (Independent Component Analysis, ICA).

SSA-ICA: ICA применяется к подпространству траекторного пространства, найденному с помощью SSA.

Пусть $F = F^{(1)} + F^{(2)} + F^{(3)}$.

Результат SSA: $F = \widetilde{F}^{(1,2)} + \widetilde{F}^{(3)}$,
Результат ICA: $\widetilde{F}^{(1,2)} = \widetilde{F}^{(1)} + \widetilde{F}^{(2)}$,
Результат SSA-ICA: $F = \widetilde{F}^{(1)} + \widetilde{F}^{(2)} + \widetilde{F}^{(3)}$.

Далее для простоты будем считать, что $F^{(3)} = 0$, $F^{(1,2)} = F$.

Разделимость как разделение подпространств

Ряды $F_N = F_N^{(1)} + F_N^{(2)}$ — ряды длины N, L — длина окна.
$\mathbf{X}$, $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ — траекторные матрицы $L \times K$, $\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$.
$\mathcal{L}$, $\mathcal{L}^{(1)}$, $\mathcal{L}^{(2)}$ — пространства строк траект. матриц, $\mathcal{L} = \mathcal{L}^{(1)} \bigoplus \mathcal{L}^{(2)}$.
r, r_1, r_2 — ранги траекторных матриц, они же размерности пространств строк, $r = r_1 + r_2$.

Пусть метод SSA находит базис пространства $\mathcal{L}$.
Будем говорить, что базис $\mathcal{L}$ разделяет подпространства, если его можно разбить на базисы подпространств $\mathcal{L}^{(1)}$ и $\mathcal{L}^{(2)}$.

Задача метода SSA-ICA — найти такой разделяющий базис $\mathcal{L}$.
Если разделяющий базис найден, то мы получаем $F_N^{(1)}$ и $F_N^{(2)}$ точно.
Назовем вектора “правильными”, если они составляют разделяющий базис.

Слабая и сильная разделимость

Проблема: результат метода не единственен (например, при совпадающих сингулярных числах сингулярное разложение не единственно).
Слабая разделимость: среди всех базисов, которые являются результатом метода, ***существует разделяющий***.
Сильная разделимость: ***любой*** базис, являющийся результатом метода, ***разделяющий***.
Для применения нужна сильная разделимость.

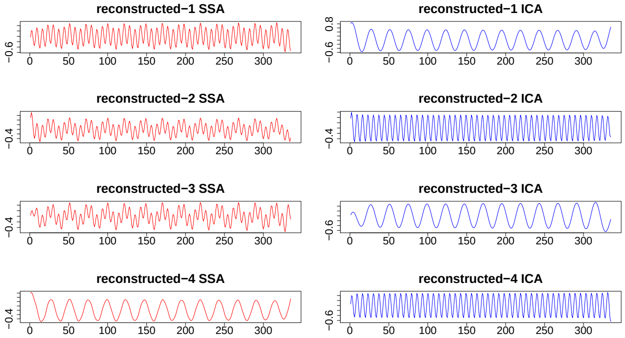
SSA приближенно слабо разделяет тренд и периодическую компоненты, гармоники с разными периодам, сигнал и шум.
Но: компоненты с одинаковым вкладом могут смешаться (нет сильной разделимости).

ICA: не умеет отделять сигнал от шума, несколько хуже разделяет детерминированные компоненты.
Но: может разделять компоненты с одинаковым вкладом.

Пример отсутствия сильной разделимости

$$f_n = A \sin\left(2\pi\omega_1 n + \gamma_1\right) + A \sin\left(2\pi\omega_2 n + \gamma_2\right), \tag{1}$$

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_4. \tag{2}$$



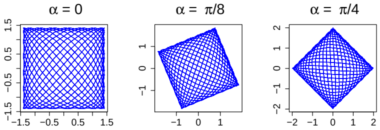
Структура работы

- Сначала задача разделения решается для разделения подпространств, не обязательно являющихся траекторными. Если размерность подпространств равна 1, то можно говорить о разделении векторов. Результаты продемонстрированы на примере разделения двух векторов.
- Формулируются алгоритмы SSA-ICA для рядов.
- Полученные для разделения подпространств результаты используются для получения условий точной и асимптотической разделимости рядов.
- Примеры.

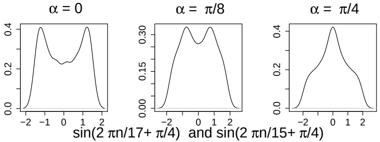
Будем формально вводить определения и формулировать результаты только для точной разделимости.

Применение PP для разделения векторов, пример

S_1 и S_2 — синусы с периодами T_1 и T_2.
$S_1(\alpha) = S_1 \cos\alpha + S_2 \sin\alpha, \;\; S_2(\alpha) = -S_1 \sin\alpha + S_2 \cos\alpha.$
Разделяющий базис $\{S_1(\alpha), \; S_1(\alpha)\}$ соответствует $\alpha = 0$.
Применим SVD к $[S_1(\alpha) : S_2(\alpha)]$ — получим случайное направление.
Формально применим ICA — получим $\alpha = 0$.



Скаттерплоты $(S_1(\alpha), \; S_2(\alpha))$



Формы плотностей проекций на ось X

PP для векторов: Критерий из ICA

$\|X\|_P = \sqrt{\frac{1}{K}\sum_{n=1}^{K} x_n^2}$.
$Z = (z_1, \ldots, z_K)^{\mathrm{T}}, \bar{Z} = \frac{1}{K}\sum_{n=1}^{K} z_n = 0,\ \|Z\|_P = 1.$

Критерий "правильности" вектора:

$$J(Z) = \left[-\frac{1}{K}\sum_{n=1}^{K} G(z_n) + C_\nu\right]^2, C_\nu = \mathbb{E}G(\nu), \nu \in \mathrm{N}(0,1),$$

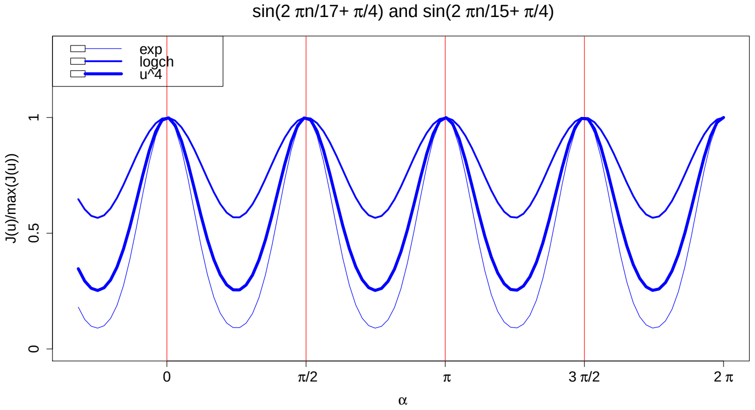
$G(u)$ — чётная, трижды дифференцируемая.

Пусть базис $\{S_1, S_2\}$ — разделяющий, центрированный, ортонормированный. Тогда $\{\pm S_{1\,\text{или}\,2}, \pm S_{2\,\text{или}\,1}\}$ тоже разделяющий.
$S_1(\alpha) = S_1\cos\alpha + S_2\sin\alpha,\ \ S_2(\alpha) = -S_1\sin\alpha + S_2\cos\alpha.$
Тогда базис $\{S_1(\alpha), S_2(\alpha)\}$ является разделяющим при $\alpha = 0, \pi/2, \pi, 3\pi/2$.

Вопрос: максимумы $J(S_1(\alpha))$ достигаются на α, соответствующих векторам из разделяющего базиса?

ICA для векторов: Пример

S_1 и S_2 — синусы с периодами T_1 и T_2.
Поведение $J(S_1(\alpha))$ для разных G . Соответствие максимумов векторам из разделяющего базиса ($\alpha = 0, \pi/2, \pi, 3\pi/2$).



Будем рассматривать $G(u) = u^4$.

ICA для векторов: условия разделения

Поиск разделяющего базиса по ICA — последовательная задача максимизации функционала $J(Z)=\left[-\frac{1}{K}\sum_{n=1}^{K}G(z_n)+C_\nu\right]^2$, где $Z\in\mathcal{L}$.

Пусть ортонормированный и центрированный базис $\{S_1, S_2\}$ является разделяющим, $S_1(\alpha) = S_1\cos\alpha + S_2\sin\alpha$.

Утверждение

Для того чтобы максимум $J(S_1(\alpha))$ достигался при $\alpha = 0, \pi/2$, необходимо и достаточно:

$$\sum_{n=1}^{K} G'(s_{n,1})\, s_{n,2} = 0,\ \sum_{n=1}^{K} G'(s_{n,2})\, s_{n,1} = 0,$$

$$\left[-\frac{1}{K}\sum_{n=1}^{K}G(s_{n,1})+\mathbb{E}G(\nu)\right]\sum_{n=1}^{K}\left[G'(s_{n,1})\, s_{n,1}-G''(s_{n,1})\, s_{n,2}^2\right]<0,$$

$$\left[-\frac{1}{K}\sum_{n=1}^{K}G(s_{n,2})+\mathbb{E}G(\nu)\right]\sum_{n=1}^{K}\left[G'(s_{n,2})\, s_{n,2}-G''(s_{n,2})\, s_{n,1}^2\right]<0.$$

ICA для векторов: итоги и проблемы

Для подпространств, заданных центрированными и ортонормированными базисами:

- Были получены условия слабой и асимптотической разделимости подпространств в общем виде.
- Условия были проверены для разделения центрированных подпространств, порожденных гармониками (сильная), линейной функцией и экспонентой (асимптотическая).

Следующая задача: использование результатов для траекторных пространств рядов.

Если траекторные матрицы имеют нулевые средние по строкам, то ICA применяется прямо к траекторной матрице (к базису строк).

Если траекторная матрица $\mathbf{X}$ не центрирована по строкам, то есть два способа применения ICA:

- SSA-ICA с центрированием: применение ICA к центрированной по строкам матрице,
- SSA-ICA: без внешнего центрирования, центрирование внутри максимизируемого функционала.

Базовый метод SSA-ICA

Функционал: $J_{\mathrm{nc}}(Q) = \left(-\frac{1}{K}\sum_{l=1}^{K} G\left(\frac{q_l - \bar{Q}}{\mathrm{s.d}(Q)}\right) + \mathbb{E}G(\nu)\right)^2.$

Пусть ранг $\mathbf{X} - \bar{\mathbf{X}}$ равен рангу $\mathbf{X}$. Обозначим ранг r.

A. Оптимизационная задача с внутренним центрированием для $\mathbf{X}$:

$$J_{\mathrm{nc}}\left(\mathbf{X}^{\mathrm{T}}\widehat{H}\right) \to \max_{\widehat{H}:\mathrm{s.d.}\left(\mathbf{X}^{\mathrm{T}}\widehat{H}\right)=1}. \tag{3}$$

Ищем последовательные решения $\widehat{\mathbf{W}} = \left[\widehat{W}_1 : \ldots : \widehat{W}_r\right]$ такие, что $\operatorname{cov}\left(\mathbf{X}^{\mathrm{T}}\widehat{W}_j, \mathbf{X}^{\mathrm{T}}\widehat{W}_i\right) = 0, i \neq j.$

Результат SSA-ICA:

$$S_j = \mathbf{X}^{\mathrm{T}}\widehat{W}_j, j = 1, \ldots, r, \tag{4}$$

$$\mathbf{X} = \sum_{i=1}^{r} \widehat{U}_i S_i^{\mathrm{T}}, \text{ где } \widehat{\mathbf{U}} = \left(\widehat{\mathbf{W}}^{\mathrm{T}}\right)^{-1}. \tag{5}$$

SSA-ICA с центрированием

Функционал: $J\left(Z\right) = \left(-\frac{1}{K}\sum_{l=1}^{K} G\left(z_l\right) + \mathbb{E}G\left(\nu\right)\right)^2.$

Обозначим r_c ранг $\mathbf{X} - \bar{\mathbf{X}}$.

B. Оптимизационная задача без центрирования для $\mathbf{X} - \bar{\mathbf{X}}$:

$$J\left(\left(\mathbf{X} - \bar{\mathbf{X}}\right)^{\mathrm{T}} \widehat{H}\right) \to \max_{\widehat{H}: \mathrm{s.d.}\left(\left(\mathbf{X}-\bar{\mathbf{X}}\right)^{\mathrm{T}} \widehat{H}\right)=1} \tag{6}$$

Ищем последовательные решения $\widehat{\mathbf{W}} = \left[\widehat{W}_1 : \ldots : \widehat{W}_{r_c}\right]$ такие, что $\mathrm{cov}\left(\left(\mathbf{X} - \bar{\mathbf{X}}\right)^{\mathrm{T}} \widehat{W}_j, \left(\mathbf{X} - \bar{\mathbf{X}}\right)^{\mathrm{T}} \widehat{W}_i\right) = 0, i \neq j$.

Результат SSA-ICA с центрированием:

$$S_j^{(c)} = \left(\mathbf{X} - \bar{\mathbf{X}}\right)^{\mathrm{T}} \widehat{W}_j, j = 1, \ldots, r_c, \tag{7}$$

$$\mathbf{X} = \bar{\mathbf{X}} + \sum_{i=1}^{r_c} \widehat{U}_i \left(S_i^{(c)}\right)^{\mathrm{T}}, \text{ где } \widehat{\mathbf{U}} = \left(\widehat{\mathbf{W}}^{\mathrm{T}}\right)^{-1}. \tag{8}$$

SSA-ICA: Эквивалентность оптимизационных задач

Пусть ранги $\mathbf{X}$ и $\mathbf{X} - \bar{\mathbf{X}}$ равны

Утверждение

Решения $\widehat{\mathbf{W}}$ оптимизационных задач A и B, с центрированием и без, совпадают.

Утверждение

Пусть $\mathbf{X} - \bar{\mathbf{X}} = \mathbf{U}^{(c)} \left(\mathbf{\Lambda}^{(c)}\right)^{1/2} \left(\mathbf{V}^{(c)}\right)^{\mathrm{T}}$ — SVD разложение. Матрица $\mathbf{W}$ — решение оптимизационной задачи без центрирования, применённой к $\mathbf{V}^{(c)}$. Тогда

$$\widehat{\mathbf{W}} = \mathbf{U}^{(c)} \left(\mathbf{\Lambda}^{(c)}\right)^{-1/2} \mathbf{W}$$

решение обеих оптимизационных задач A и B.

Следствие: результаты о разделении центрированных подпространств можно перенести на случай нецентрированных.

SSA-ICA: Примеры разделимости рядов

Утверждение

$f_n^{(1)} = A_1 \sin(2\pi\omega_1 n + \gamma_1)$, $f_n^{(2)} = A_2 \sin(2\pi\omega_2 n + \gamma_2)$, $0 \le n \le N-1$. *Длина окна* L *такая, что* $K\omega_j, j = 1, 2$ *целые,* $K = N - L + 1$, $0 < \omega_j < \frac{1}{2}, j = 1, 2$, $\omega_1 \ne \omega_2$, *и* $\omega_1 \ne 3\omega_2, \omega_2 \ne 3\omega_1$, *и* $\omega_1 + \omega_2 \ne \frac{1}{2}$. *Тогда ряды сильно разделимы с помощью SSA-ICA.*

Утверждение

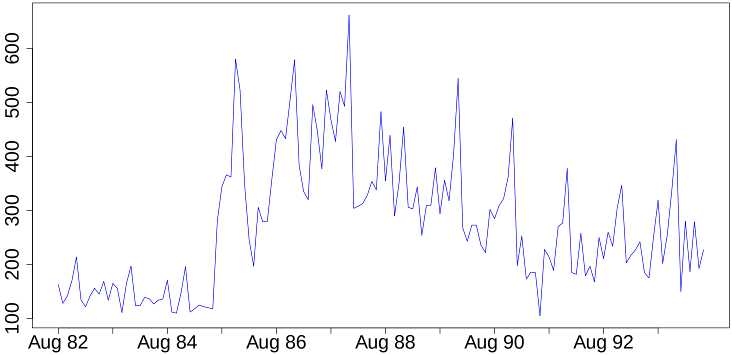
Пусть $f_n^{(N,1)} = \exp\frac{\alpha n}{N}$ *и* $f_n^{(2)} = A\sin(2\pi\omega n + \gamma)$, $0 \le n \le N-1$. *Тогда ряды асимптотически разделимы с помощью SSA-ICA при* $K \to \infty$.

Утверждение

Пусть $f_n^{(1)} = a(n+1) + b$ *и* $f_n^{(2)} = A\sin(2\pi\omega n + \gamma)$, $0 \le n \le N-1$. *Тогда ряды асимптотически разделимы с помощью SSA-ICA с центрированием при* $K \to \infty$.

Реальный ряд

Ряд ежемесячных продаж белого сладкого вина в Австралии в тысячах литров в период с августа 1982 года по июнь 1994 года.

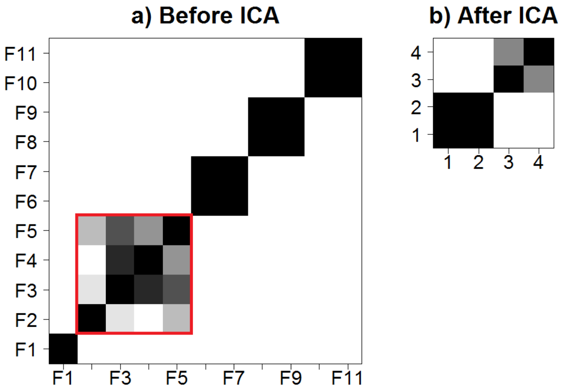


Реальный ряд: Взвешенные корреляции

$F = \sum F^{(k)}$

$\{\rho_w(F^{(i)}, F^{(j)})\}$ показывают степень разделения $F^{(i)}, F^{(j)}$.

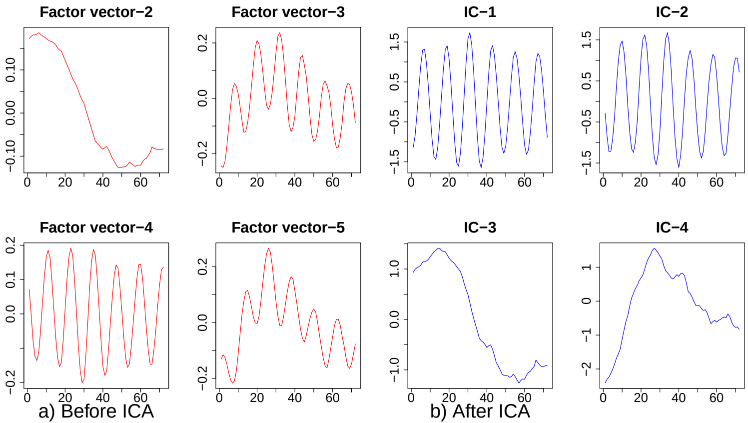
Около нуля (белый цвет) — хорошая отделимость, большие значения (черный или серый) — плохая.



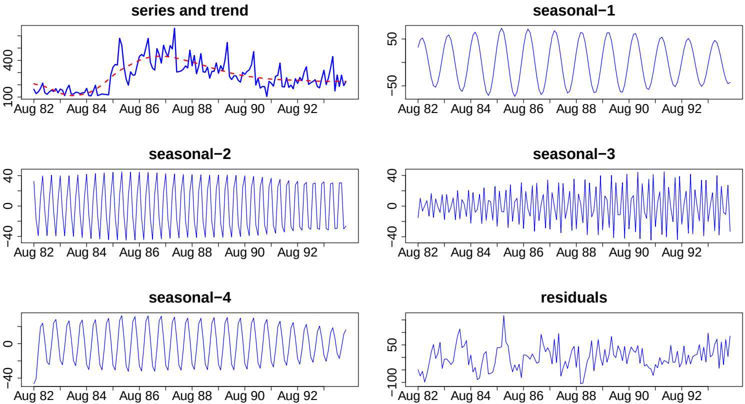
Реальный ряд: базисы траекторных пространств строк

SSA
базис не разделяет

SSA-ICA
базис приближенно разделяет



Реальный ряд: Результат SSA-ICA



Результаты

1. Выведены условия слабой и асимптотической разделимости центрированных подпространств.
2. Условия были проверены для подпространств, представляющих: синусы с разными частотами (сильная разделимость), линейную функцию и синус, экспоненту и синус (асимпт. разделимость).
3. Были предложены две модификации метода SSA для разделения смешавшихся компонент: базовый вариант SSA-ICA и вариант SSA-ICA с центрированием.
4. Выведены условия разделимости рядов для SSA-ICA.
5. Проверены условия разделимости для следующих рядов: синусов с разными частотами (сильная разделимость), линейной функции и синуса, экспоненты и синуса (асимптотическая).
6. Продемонстрировано применение метода SSA-ICA на примере реальных данных.
7. Алгоритмы SSA-ICA и SSA-ICA с центрированием были реализованы на языке программирования R с использованием пакетов Rssa, fastICA.