Аппроксимация временных рядов рядами конечного ранга

Звонарев Никита

Санкт-Петербургский государственный университет
Кафедра статистического моделирования

Санкт-Петербург
2015г.

Постановка задачи

Дан временной ряд $\mathbb{X} = \mathbb{S} + \mathbb{N}$, $\mathbb{X} = (x_1, \ldots, x_N)$,
$\mathbb{N}$ — белый шум.
$\mathbb{S}$ — сигнал, имеющий *определённую структуру*, например:

1. $\mathbb{S}$ удовлетворяет линейной рекуррентной формуле ($s_n = a_1 s_{n-1} + a_2 s_{n-2}$)
2. $\mathbb{S}$ удовлетворяет параметрической модели ($s_n = Ce^{\alpha n}\cos(2\pi\omega n + \psi)$ или $s_n = C_1 e^{\alpha_1 n} + C_2 e^{\alpha_2 n}$)

Задача: построить оценку $\hat{\mathbb{S}}$ сигнала $\mathbb{S}$.

Выбираем первый подход

Постановка задачи

Дан временной ряд $\mathbb{X} = \mathbb{S} + \mathbb{N}$, $\mathbb{X} = (x_1, \ldots, x_N)$,
$\mathbb{N}$ — белый шум.
$\mathbb{S}$ — сигнал, имеющий *определённую структуру*, например:

1. $\mathbb{S}$ удовлетворяет линейной рекуррентной формуле ($s_n = a_1 s_{n-1} + a_2 s_{n-2}$)
2. $\mathbb{S}$ удовлетворяет параметрической модели ($s_n = Ce^{\alpha n}\cos(2\pi\omega n + \psi)$ или $s_n = C_1 e^{\alpha_1 n} + C_2 e^{\alpha_2 n}$)

Задача: построить оценку $\hat{\mathbb{S}}$ сигнала $\mathbb{S}$.
Выбираем первый подход

Модель сигнала

Общий вид:

Определение

Ряд, управляемый минимальной линейной рекуррентной формулой порядка r:
$\mathbb{S} = (s_1, \ldots, s_N), \quad s_n = \sum_{i=1}^{r} a_i s_{n-i},\ n = r+1, \ldots, N,\ a_r \neq 0.$

Параметрический вид рядов, управляемых ЛРФ:

$$s_n = \sum_i P_i(n) \exp(\alpha_i n) \cos(2\pi\omega_i n + \psi_i).$$

Параметрический подход не работает из-за неустойчивости

Идея: траекторная матрица

Возьмём $\mathbb{S} = (s_1, \dots, s_N)$ — ряд, управляемый минимальной ЛРФ порядка r: $s_n = \sum_{i=1}^{r} a_i s_{n-i}$.
$1 < L < N$ — *длина окна*, $K = N - L + 1$.
Траекторная матрица ряда $\mathbb{S}$:

$$\mathbf{S} = \mathcal{T}(\mathbb{S}) = \begin{pmatrix} s_1 & s_2 & \dots & s_K \\ s_2 & s_3 & \dots & s_{K+1} \\ \vdots & \vdots & \vdots & \vdots \\ s_L & s_{L+1} & \dots & s_N \end{pmatrix}.$$

Заметим, что $\mathbf{S}$ — ганкелева (*structured*, $\mathbf{S} \in \mathcal{H}$) и $\operatorname{rank} \mathbf{S} = r$ (*low-rank*, $\mathbf{S} \in \mathcal{M}_r$).
$\mathbb{X} = \mathbb{S} + \mathbb{N}$, $\mathbf{X} = \mathcal{T}(\mathbb{X}) = \mathbf{S} + \mathbf{N}$
Оценка $\hat{\mathbb{S}} \Leftrightarrow$ **SLRA** (Structured (Hankel) Low Rank Approximation) матрицы $\mathbf{X}$ в классе $\mathcal{H} \cap \mathcal{M}_r$.

Эквивалентные задачи

$\mathbb{X} = (x_1, \ldots, x_N)$, $\mathbf{X} = \mathcal{T}(\mathbb{X})$.
Задача аппроксимации матрицы (SLRA):

$$\|\mathbf{X} - \mathbf{Y}\|_F^2 \to \min_{\mathbf{Y} \in \mathcal{M}_r \cap \mathcal{H}},$$

Эквивалентная задача аппроксимации рядов:

$$\sum_{i=1}^{N} w_i (x_i - y_i)^2 \to \min_{\mathbb{Y} \in \mathcal{X}_N^r},$$

$\mathcal{X}_N^r$ — множество рядов, управляемых ЛРФ,

$$w_i = \begin{cases} i & \text{для } i = 1, \ldots, L-1, \\ L & \text{для } i = L, \ldots, K, \\ N - i + 1 & \text{для } i = K+1, \ldots, N, \end{cases} \quad —$$

трапециевидные веса.

Методы Cadzow и SSA

SLRA:

$$\|\mathbf{X} - \mathbf{Y}\|_F^2 \to \min_{\mathbf{Y} \in \mathcal{M}_r \cap \mathcal{H}}$$

Решение — метод Cadzow (1988):

$$\mathbf{Y}_k = (\Pi_{\mathcal{H}} \circ \Pi_{\mathcal{M}_r})^k(\mathbf{X}),$$

$\Pi_{\mathcal{H}}$ — проектор на ганкелевы матрицы, $\Pi_{\mathcal{M}_r}$ — проектор на матрицы ранга $\leq r$, k — число итераций.
$k = 1$ — метод SSA (Singular Spectrum Analysis, ссылки). Не дает оценку из $\mathcal{M}_r \cap \mathcal{H}$.

Что известно про Cadzow?

Методы Cadzow и SSA

SLRA:

$$\|\mathbf{X} - \mathbf{Y}\|_F^2 \to \min_{\mathbf{Y} \in \mathcal{M}_r \cap \mathcal{H}}$$

Решение — метод Cadzow (1988):

$$\mathbf{Y}_k = (\Pi_{\mathcal{H}} \circ \Pi_{\mathcal{M}_r})^k(\mathbf{X}),$$

$\Pi_{\mathcal{H}}$ — проектор на ганкелевы матрицы, $\Pi_{\mathcal{M}_r}$ — проектор на матрицы ранга $\leq r$, k — число итераций.
$k = 1$ — метод SSA (Singular Spectrum Analysis, ссылки). Не дает оценку из $\mathcal{M}_r \cap \mathcal{H}$.

Что известно про Cadzow?

Метод Cadzow

$$\mathbf{Y}_k = (\Pi_{\mathcal{H}} \circ \Pi_{\mathcal{M}_r})^k(\mathbf{X}).$$

Ничего не известно ни про сходимость к глобальному, ни к локальному минимуму (понятие “локального экстремума” отсутствует, т.к. его не ввести естественным образом)

Но:

Теорема (Cadzow, 1988)

1. $\|\mathbf{Y}_k - \Pi_{\mathcal{M}_r}\mathbf{Y}_k\| \to 0$ *при* $k \to +\infty$, $\|\Pi_{\mathcal{M}_r}\mathbf{Y}_k - \mathbf{Y}_{k+1}\| \to 0$ *при* $k \to +\infty$.
2. *Существует сходящаяся подпоследовательность точек* $\mathbf{Y}_{i_1}, \mathbf{Y}_{i_2}, \ldots$ *такая, что ее предел* $\mathbf{Y}^*$ *лежит в* $\mathcal{M}_r \cap \mathcal{H}$.

Напоминание: эквивалентные задачи

Аппроксимация рядов:

$$\sum_{i=1}^{N} w_i (x_i - y_i)^2 \to \min_{\mathbb{Y} \in \mathcal{X}_N^r},$$

$$w_i = \begin{cases} i & \text{для } i = 1, \dots, L-1, \\ L & \text{для } i = L, \dots, K, \\ N - i + 1 & \text{для } i = K+1, \dots, N \end{cases}$$

Наша задача — с помощью аппроксимации ряда оценить сигнал.

Можно ли обобщить Cadzow и использовать другие веса вместо w_i?

Варианты (полу)норм

Ряды: $\|\mathbb{X}\|_q^2 = \sum_{i=1}^N q_i x_i^2$, все $q_i \geq 0$. *Но как аппроксимировать по* $\|\mathbb{X}\|_q$?

Решение: подобрать эквивалентную матричную норму

1. $\|\mathbf{X}\|_{\mathbf{D}_1,\mathbf{D}_2}^2 = \mathrm{tr}(\mathbf{D}_1\mathbf{X}\mathbf{D}_2\mathbf{X}^{\mathrm{T}})$, $\mathbf{D}_1 \in \mathsf{R}^{L\times L}$, $\mathbf{D}_2 \in \mathsf{R}^{K\times K}$ — симметричные, неотрицательно определённые матрицы (+ диагональные в нашем случае) (Zhigljavsky, 2014)
2. Частный случай предыдущего пункта: $\mathbf{D}_1$, $\mathbf{D}_2$ — единичные матрицы (Фробениусовская норма $\|\mathbf{X}\|_F$)

Проблемы:

1. Как вычислить $\Pi_{\mathcal{H}}$, $\Pi_{\mathcal{M}_r}$ по нужной норме?
2. Как подобрать $\mathbf{D}_1$ и $\mathbf{D}_2$, соответствующие q_i: $\|\mathbb{X}\|_q = \|\mathbf{X}\|_{\mathbf{D}_1,\mathbf{D}_2}$?

Проекторы $\Pi_{\mathcal{M}_r}$ и $\Pi_{\mathcal{H}}$

$\Pi_{\mathcal{M}_r}$: через косоугольное SVD-разложение $\mathbf{X}$. Сводится к стандартному SVD-разложению матрицы $\mathbf{O}_1\mathbf{X}\mathbf{O}_2^\mathrm{T}$, $\mathbf{O}_i^\mathrm{T}\mathbf{O}_i = \mathbf{D}_i$, $i = 1, 2$.

$\Pi_{\mathcal{H}}$: через взвешенное диагональное усреднение

$$\mathbf{X} = \Pi_{\mathcal{H}}(\mathbf{Y}), \quad x_{l,k} = \frac{\sum_{i,j:\, i+j=l+k} d_{1,i} d_{2,j} y_{i,j}}{\sum_{i,j:\, i+j=l+k} d_{1,i} d_{2,j}}.$$

Эквивалентность и равенство норм

$\|\mathbb{X}\|_q^2 = \sum_{i=1}^N q_i x_i^2$, все $q_i > 0$
$\|\mathbf{X}\|^2_{\mathbf{D}_1,\mathbf{D}_2} = \mathrm{tr}(\mathbf{D}_1\mathbf{X}\mathbf{D}_2\mathbf{X}^{\mathrm{T}})$
$\mathbf{D}_1 = \mathrm{diag}(d_{1,1},\ldots,d_{1,L})$, $\mathbf{D}_2 = \mathrm{diag}(d_{2,1},\ldots,d_{2,K})$.

Теорема

Полунорма $\|\cdot\|_{\mathbf{D}_1,\mathbf{D}_2}$ является нормой $\Leftrightarrow$ все $d_{1,i} > 0$, $i = 1,\ldots,L$, $d_{2,j} > 0$, $j = 1,\ldots,K$.

Векторы D_1, D_2, $Q \in \mathsf{R}^N$, $D_1 = (d_{1,1},\ldots,d_{1,L},0,\ldots,0)^{\mathrm{T}}$, $D_2 = (d_{2,1},\ldots,d_{2,K},0,\ldots,0)^{\mathrm{T}}$, $Q = (q_1,\ldots,q_N)^{\mathrm{T}}$.
Свертка: $C = A \star B$, $c_i = \sum_{j=0}^{N-1} a_j b_{i-j}$, $A = (a_0,\ldots,a_{N-1})^{\mathrm{T}}$ ...

Теорема (Zhigljavsky, 2015)

$\|\mathcal{T}(\mathbb{X})\|_{\mathbf{D}_1,\mathbf{D}_2} = \|\mathbb{X}\|_q$ *для любого* $\mathbb{X} \Leftrightarrow Q = D_1 \star D_2$.

Единичные веса ряда

$$\sum_{i=1}^{N} q_i (x_i - y_i)^2 \to \min_{\mathbb{Y} \in \mathcal{X}_N^r}$$

Естественные веса: $q_i = 1$.

Вопрос: как найти невырожденные $\mathbf{D}_1$, $\mathbf{D}_2$ такие, что $D_1 \star D_2 = (1, \ldots, 1)^{\mathrm{T}}$?
Anatoly Zhigljavsky, 2015: D_1 и D_2 могут состоять только из нулей и единиц $\Rightarrow$ таких невырожденных D_1 и D_2 не существует, значит, невозможно найти эквивалентную матричную задачу.

Пример: $N = 12$, $L = 4$, $K = 9$. $\mathbf{D}_1 = \mathrm{diag}(1, 1, 1, 1)$,
$\mathbf{D}_2 = \mathrm{diag}(1, 0, 0, 0, 1, 0, 0, 0, 1)$.

Поиск весов

$$Q = D_1 \star D_2.$$

Идея: потерять точность равенства, но избавиться от вырожденности.

Зафиксируем $\mathbf{D}_1$ — единичная матрица. Аппроксимация требуемых единичных весов: зафиксируем параметр $0 < \alpha < 1$, минимизация по диагональной положительно определённой $\mathbf{D}_2$:

$$\|D_1 \star D_2 - (1, \ldots, 1)^{\mathrm{T}}\|_. \to \min_{\mathrm{cond}\,\mathbf{D}_2 \le \frac{1}{\alpha}}$$

Condition number (число обусловленности)
$\mathrm{cond}\,\mathbf{D} = \|D\| \|D\|^{-1} = \frac{\max_i d_i}{\min_i d_i}$, если $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N)$.

План

$$\|D_1 \star D_2 - (1,\dots,1)^{\mathrm{T}}\|_{\cdot} \to \min_{\operatorname{cond} \mathbf{D}_2 \le \frac{1}{\alpha}} \qquad (1)$$

1. Как решить задачу (1)? Если $\|\cdot\|_{\cdot} = \|\cdot\|_1$ или $\|\cdot\|_\infty$ — сводится к задаче линейного программирования $\|\cdot\|_{\cdot} = \|\cdot\|_2$ — квадратичное программирование, есть очень быстрое решение при $L \approx N/2$ ("Active set" метод, оптимальный выбор начальной точки)
2. Как решить задачу (1), если не фиксировать $\mathbf{D}_1$? Открытый вопрос: невыпуклая целевая функция (приближенное решение — эвристика), неясные ограничения.
3. Какой выигрыш дают примерно единичные веса?

План

$$\|D_1 \star D_2 - (1, \ldots, 1)^{\mathrm{T}}\|_{\cdot} \to \min_{\operatorname{cond} \mathbf{D}_2 \le \frac{1}{\alpha}} \qquad (1)$$

1. Как решить задачу (1)? Если $\|\cdot\|_{\cdot} = \|\cdot\|_1$ или $\|\cdot\|_\infty$ — сводится к задаче линейного программирования $\|\cdot\|_{\cdot} = \|\cdot\|_2$ — квадратичное программирование, есть очень быстрое решение при $L \approx N/2$ (“Active set” метод, оптимальный выбор начальной точки)
2. Как решить задачу (1), если не фиксировать $\mathbf{D}_1$? Открытый вопрос: невыпуклая целевая функция (приближенное решение — эвристика), неясные ограничения.
3. Какой выигрыш дают примерно единичные веса?

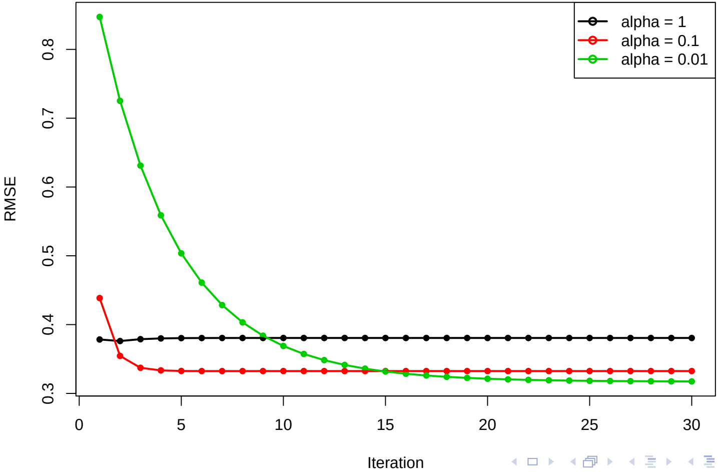
План

$$\|D_1 \star D_2 - (1,\ldots,1)^{\mathrm{T}}\|_{\cdot} \to \min_{\mathrm{cond}\,\mathbf{D}_2 \le \frac{1}{\alpha}} \qquad (1)$$

1. Как решить задачу (1)? Если $\|\cdot\|_{\cdot} = \|\cdot\|_1$ или $\|\cdot\|_\infty$ — сводится к задаче линейного программирования $\|\cdot\|_{\cdot} = \|\cdot\|_2$ — квадратичное программирование, есть очень быстрое решение при $L \approx N/2$ (“Active set” метод, оптимальный выбор начальной точки)
2. Как решить задачу (1), если не фиксировать $\mathbf{D}_1$? Открытый вопрос: невыпуклая целевая функция (приближенное решение — эвристика), неясные ограничения.
3. Какой выигрыш дают примерно единичные веса?

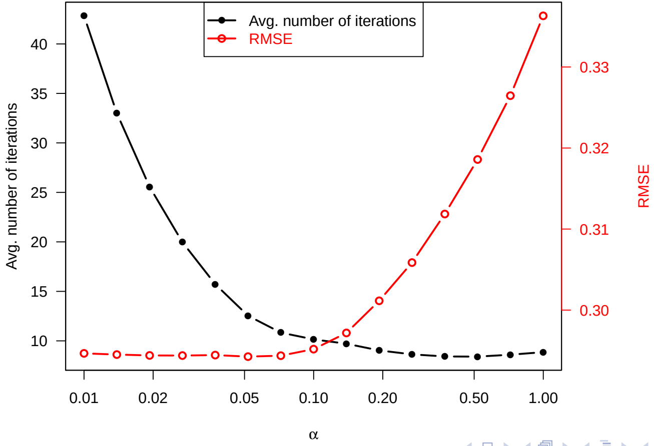
Сравнение (зависимость RMSE от числа итераций)

Задача оценки сигнала: $N = 40$, $L = 20$, $r = 2$, $\mathbb{X} = (x_1, \ldots, x_N)$, $x_k = 5\sin\frac{2k\pi}{6} + \varepsilon_k$; ε_k, $k = 1, \ldots N$, — гауссовский белый шум, $\mathbb{E}\varepsilon_k = 0$, $\mathbb{D}\varepsilon_k = 1$, 1000 реализаций.



Сравнение по трудоемкости

Задача оценки сигнала: $N = 40,\ L = 20,\ r = 2,$
$\mathbb{X} = (x_1, \ldots, x_N),\ x_k = 5\sin\frac{2k\pi}{6} + \varepsilon_k.$
Критерий остановки: $\frac{\|\mathcal{T}^{-1}(\mathbf{Y}_k) - \mathcal{T}^{-1}(\mathbf{Y}_{k+1})\|^2}{N} < 10^{-8}$

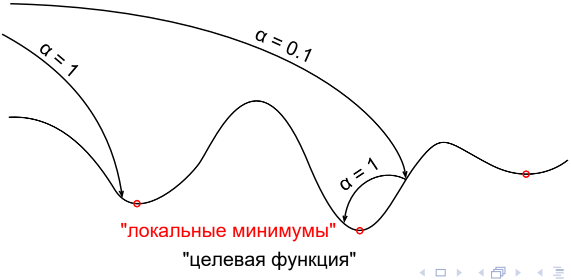


Чередование проекций при различном параметре α

Идея: совместить малые и большие α, чтобы получить и быструю, и точную оценку.
Целевая функция имеет большое число “локальных минимумов”.

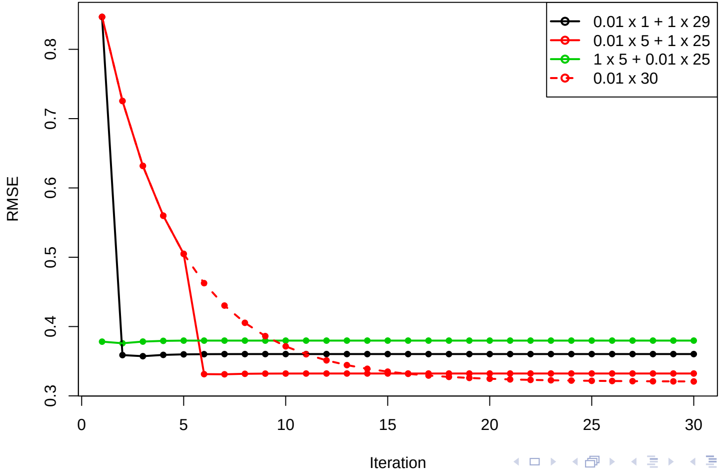
При большом α ($\alpha = 1$): быстрая сходимость, но далеко от оптимальной окрестности

При малом α: медленная сходимость, но близко к оптимальной окрестности



Сравнение (RMSE при чередовании α)

Задача оценки сигнала: $N = 40$, $L = 20$, $r = 2$,
$\mathbb{X} = (x_1, \ldots, x_N)$, $x_k = 5\sin\frac{2k\pi}{6} + \varepsilon_k$; ε_k, $k = 1, \ldots N$, — гауссовский белый шум, $\mathbb{E}\varepsilon_k = 0$, $\mathbb{D}\varepsilon_k = 1$, 1000 реализаций.



Планы на будущее

- Теоретически установить соотношение между скоростью сходимости и разделимостью
- Оценить разделимость теоретически через $\operatorname{cond} \mathbf{D}_2$ в общем случае
  (разделимость $\leftrightarrow$ α: есть результат для весов определённого вида при $N \vdots L$ (Zvonarev, 2015))
- Использование двусторонних весов (без фиксирования $\mathbf{D}_1$).
  Включает в себя два предыдущих вопроса
- Использование недиагональных $\mathbf{D}_1$ и $\mathbf{D}_2$ (например, для красного шума)