

Прогноз и нахождение формулы сигнала конечного ранга на основе SSA и итераций Кэдзоу

Курышев Иван Владимирович, 522-я группа

Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра статистического моделирования

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



Санкт-Петербург
2007г.

Ряды конечного ранга

- $F_N = (f_0, \dots, f_{N-1}), 1 < L < N, K = N - L + 1$

- $F_N \xleftrightarrow{L,K} \mathbf{X} = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & \dots & \dots & \dots \\ f_2 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ f_{L-1} & f_L & \dots & \dots & f_{N-1} \end{pmatrix}$

траекторная матрица, ганкелева, $f_{i+j} = \text{const}$

- $\mathbf{X} = \mathcal{T} F_N, F_N = \mathcal{T}^{-1} \mathbf{X}$

- О рядах конечного ранга:

- F_N — ряд конечного ранга $d < N/2$, если $\forall L : d < \min(L, K), \text{rank } \mathbf{X} = d$

- Ряд, управляемый линейной рекуррентной формулой (ЛРФ) $f_{d+i} = \sum_{l=1}^d a_l f_{i+d-l}, a_d \neq 0$

- Сумма произведений полиномов, экспонент и гармоник

Постановка задачи

- $F_N = S_N + R_N$, R_N — шум, S_N — ряд конечного ранга p (p — известно)
- Задачи:
 - 1. Выделить S_N
Метод «Гусеница»-SSA, итерации Cadzow (аппроксимация траекторной матрицы ряда ганкелевой матрицей ранга p)
 - 2. Сделать прогноз S_N
Различные методы на основе ЛРФ, на основе продолжения в заданном пространстве размерности p .
Задача: сравнить по точности на модельных примерах.
 - 3. Найти аналитическую формулу сигнала.
Реализация алгоритма.

Итерации Cadzow

- $\mathcal{M}_{L,K}$ — пр-во вещественных матриц размера $L \times K$,
 $\mathcal{M}_{L,K}^{(\mathcal{H})}$ — пр-во вещественных ганкелевых матриц $L \times K$,
 $\mathcal{M}_{L,K}^{(p)}$ — пр-во вещественных матриц ранга p $L \times K$.
- Ортогональные проекторы (по норме Фробениуса)
 - 1. $\mathcal{R}^{(p)} : \mathcal{M}_{L,K} \longrightarrow \mathcal{M}_{L,K}^{(p)}$, $\mathbf{Y} = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T$ — сингулярное разложение матрицы \mathbf{Y} .
 $\mathcal{R}^{(p)} \mathbf{Y} = \sum_{i=1}^p \sqrt{\lambda_i} U_i V_i^T$
 - 2. $\mathcal{H} : \mathcal{M}_{L,K} \longrightarrow \mathcal{M}_{L,K}^{(\mathcal{H})}$ — ганкелизация
 (усреднение по диагоналям $i + j = \text{const}$)
- Из теоремы (Cadzow, 1988):
 Для последовательности матриц $(\mathcal{H}\mathcal{R}^{(p)})^m \mathbf{X}$, \exists -ет подпоследовательность, которая сходится к
 $\mathbf{X}_{\mathcal{H}}^{(p)} \in \mathcal{M}_{L,K}^{(\mathcal{H})} \cap \mathcal{M}_{L,K}^{(p)} : \|\mathbf{X} - \mathbf{X}_{\mathcal{H}}^{(p)}\|_F \longrightarrow \min$.
- Т. о., алгоритм аппроксимации временных рядов рядом ранга p :
 $F_N^{(p)} = T^{-1}(\mathcal{H}\mathcal{R}^{(p)})^m T F_N$
 $m = 1$ — метод «Гусеница»-SSA

Прогноз

- $\{U_i\}_{i=1}^p$ — ортонорм. система, $U_i \in \mathbb{R}^L, p < d$.
- $\mathcal{L}_p = \text{span}(U_1, \dots, U_p)$
- Формула для выражения одних координат вектора через другие:
 Для $X = \begin{pmatrix} Y \\ Z \end{pmatrix} \in \mathcal{L}_p$, $\mathcal{I} = \{1, \dots, L\}, \mathcal{P} = \{L - P + 1, \dots, L\}$,
 $\mathbf{U} = [U_1, \dots, U_p]$

$$Z = (\mathbf{E}_P - \mathbf{U}|_{\mathcal{P}}(\mathbf{U}|_{\mathcal{P}})^T)^{-1} \mathbf{U}|_{\mathcal{P}}(\mathbf{U}|_{\mathcal{I} \setminus \mathcal{P}})^T Y.$$

Прогноз на P точек вперед

$$\bullet S = (s_0, \dots, s_{N-1}, \underbrace{\ominus, \ominus}_{P=2}) \longrightarrow \left(\begin{array}{ccc|cc} \star & \dots & \star & \star & \star \\ \star & \dots & \star & \star & \star \\ \dots & \dots & \dots & \dots & \dots \\ \star & \dots & \star & \star & \ominus \\ \star & \dots & \star & \ominus & \ominus \end{array} \right)$$

$$\bullet \begin{array}{ccc} \text{тр. matr.} & & \text{восст. matr.} \\ F_N & \xrightarrow{\mathcal{R}^{(p)}} & \text{ранга } p \\ (\mathbf{X}|\ominus) & & (\mathbf{X}^{(p)}|\ominus) \end{array} \xrightarrow{\mathcal{H}} \begin{array}{c} \text{матрица} \\ \text{восст. ряда} \\ (\mathbf{X}_H^{(p)}|\ominus) \end{array}$$

Методы прогноза:

- последовательный: по 1-ой точке на основе $\mathbf{X}_H^{(p)}$ (после ганкелизации)
- одновременный: сразу на P точек вперед на основе $\mathbf{X}_H^{(p)}$ (после ганкелизации)
- векторный: по 1-ой точке, на основе $\mathbf{X}^{(p)}$ (до ганкелизации)

Численное сравнение методов прогноза

- Пространство \mathcal{L}_p строится на основе M итераций Cadzow
- U_1, \dots, U_p — первые p левых сингулярных векторов разложения матрицы $\mathbf{X} = \mathcal{T}F_N$
- Задача:
сравнить
 - последовательный (R), $M = 1$
 - одновременный (S), $M = 1$
 - векторный (V), $M = 1$
 - Cadzow (C), $M = 10$

Результаты моделирования

- Сравнение проводилось на следующих рядах:
 - $s_n = A \sin(2\pi n/T)$, $A = 1$, $T = 25, \dots, 200$
 - $s_n = \exp(An) \sin(2\pi n/T)$, $A = \pm 0.01$, $T = 25, \dots, 200$
 - $s_n = \exp(An)$, $A = \pm 0.005, \dots, \pm 0.02$
 - $s_n = An + 1$, $A = \pm 0.005, \dots, \pm 0.1$
- $N = 100$, $L = 50$
- $\varepsilon_n \sim N(0, \sigma^2)$, $\sigma = 0.05, \dots, 1$
- $P = 10, \dots, 49$
- $Q = 10000$ повторов.
-

$$\text{MSD}(S_{N+P}, \tilde{S}_{N+P}) = \sqrt{\frac{1}{Q} \sum_{q=1}^Q \left(\frac{1}{P} \sum_{i=N}^{N+P-1} (s_i - \tilde{s}_i)^2 \right)}$$

- Основной результат: V лучше (R, S, C)

Нахождение аналитического вида сигнала

- Аппроксимация сигнала рядом заданного ранга с помощью итераций Cadzow
 - Ряд ранга $d \implies$ линейная рекуррентная формула (ЛРФ)

$$f_{d+i} = \sum_{l=1}^d a_l f_{d+i-l},$$

$$(a_d, \dots, a_1) = \frac{1}{1-\nu^2} \sum_{i=1}^d \pi_i U_i^\nabla,$$
 где U_i — i -тый левый сингулярный вектор в SVD $(d+1)$ -траекторной матрицы восстановленного сигнала,
 $\nu^2 = \pi_1^2 + \dots + \pi_d^2$
 - ЛРФ \implies характеристический полином

$$P_d(\lambda) = \lambda^d - \sum_{i=1}^d a_i \lambda^{d-i}$$
 - Корни характеристического полинома $\lambda_1, \dots, \lambda_p$ кратности k_1, \dots, k_p определяют вид ряда:

$$f_n = \sum_{m=1}^p \left(\sum_{j=0}^{k_m-1} c_{mj} n^j \right) \lambda_m^n$$
 с точностью до c_{mj} .
 - Коэффициенты c_{mj} находим с помощью МНК.

Кластеризация корней

- Особенность алгоритма:
численный метод \Rightarrow нет одинаковых корней
- Алгоритм кластеризации:
 - На входе: $G = \{\lambda_m\}_{m=1}^d, \varepsilon_r, \varepsilon_c$.
 - Отделение друг от друга вещественных и комплексных корней:
 $G_{real} = \{\lambda_m : |\operatorname{Im} \lambda_m| < \varepsilon_r\},$
 $G_{comp} = \{\lambda_m : |\operatorname{Im} \lambda_m| \geq \varepsilon_r\}.$
 - Кластеризация:
Вещественные корни:
 Сортировка G_{real} по возрастанию;
 $G_{\lambda_1} = \{\lambda_m : |\lambda_m - \lambda_1| < \varepsilon_c, \lambda_m \in G_{real}\}$
 $\mu_1 = \sqrt[n]{\prod_{i \in G_{\lambda_1}} \lambda_i}$
 $k_1 = |G_{\lambda_1}|$
 $G_{real} := G_{real} \setminus G_{\lambda_1}$
 - На выходе $\{\mu_m\}_{m=1}^{p_1}$ с кратностями $\{k_m\}_{m=1}^{p_1}$
Комплексные корни:
 $G_c = \{\lambda_m : \operatorname{Im} \lambda_m > 0, \lambda_m \in G_{comp}\};$ аналогично случаю с вещественными корнями.

Пример: сумма гармоник

• Формула находится

$$\sigma = 0.1$$

Исходный сигнал	Восстановленный сигнал
$\sin(2\pi n/5) +$ $\sin(2\pi n/10) +$ $\sin(2\pi n/25) +$ $\sin(2\pi n/50)$	$1.020 \cos(2\pi n/4.998 - 1.570) +$ $0.983 \cos(2\pi n/9.996 - 1.570) \exp(0.001n) +$ $0.960 \cos(2\pi n/25.001 - 1.571) \exp(0.001n) +$ $0.972 \cos(2\pi n/50.032 - 1.570) \exp(0.001n)$

- $\text{MSD}(I, F) = 1.e-3$, $\text{MSD}(I, R) = 1.e-3$, $\text{MSD}(R, F) = 1.e-3$
- Собственные числа: 677.2, ..., 6.5e-03, 5.6e-06, 5.6e-13

• Формула не находится

Исходный сигнал:

$$s_n = \sin(2\pi n/35) + \sin(2\pi n/40) + \sin(2\pi n/45), \sigma = 0.001$$

- Восстановить формулу не удастся
- Собственные числа 1403.3, ..., 5.8e-10, 1.3e-10, 3.5e-11

Пример: сумма экспонент

- Формула находится**

$$\sigma = 0.001$$

Исходный сигнал	Восстановленный сигнал
$\exp(-0.005n) +$ $\exp(-0.03n) +$ $\exp(-0.07n)$	$0.979 \exp(-0.005n) +$ $1.030 \exp(-0.030n) +$ $0.991 \exp(-0.071n)$

- $\text{MSD}(I, F) = 1.e-8, \text{MSD}(I, R) = 1.e-8, \text{MSD}(R, F) = 1.e-9$

- Собственные числа:

$$719.1, 0.063, 1.6e-07, 1.4e-13$$

- Формула не находится**

Исходный сигнал:

$$s_n = \exp(0.005n) + \exp(0.01n) + \exp(0.015n), \sigma = 0$$

- Восстановить формулу не удастся

- Собственные числа:

$$11855.9, 0.002, 3.2e-11, 7.5e-12$$

Пример: полином

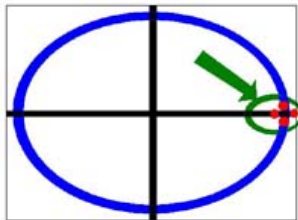
- Формула находится

$$\sigma = 0.001, \varepsilon_r = 0.1, \varepsilon_c = 0.1$$

Исходный сигнал	Восстановленный сигнал
$-0.5 +$ $0.03n +$ $0.0006n^2 +$ $0.000004n^3$	$-0.50202 \exp(-0.00053n) +$ $0.03027n \exp(-0.00053n) +$ $-0.00061n^2 \exp(-0.00053n) +$ $0.000004n^3 \exp(-0.00053n)$

- $\text{MSD}(I, F) = 1.e-8$, $\text{MSD}(I, R) = 1.e-8$, $\text{MSD}(R, F) = 1.e-7$
- Собственные числа: 13.5, 0.1, $2.3e-05$, $9.2e-09$, $1.0e-14$

Корни характеристич.полинома на комплексной плоскости:



Пример: полином

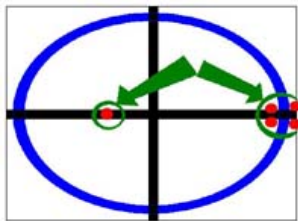
- **Формула не находится**

Исходный сигнал:

$$s_n = 8(0.01n - 0.5)^4, \sigma = 0, \varepsilon_r = 0.1, \varepsilon_c = 0.1$$

- Восстановить формулу не удается
- Собственные числа: 10.7, 0.2, $6.6e-05$, $7.5e-08$, $9.1e-12$, $4.9e-12$

Корни характеристич.полинома на комплексной плоскости:



Заключение

Итак, в дипломной работе

- С помощью моделирования было получено, что для прогноза зашумленного сигнала конечного ранга векторный метод прогноза наилучший.
- Был реализован метод нахождения аналитической формулы зашумленного сигнала и исследованы границы его применимости.