Экспоненциальный сигнал и гармоническая помеха в теории анализа сингулярного спектра временных рядов

Иванова Е.В., студентка кафедры
статистического моделирования СПбГУ,
imst563@gmail.com
Некруткин В.В., кандидат физико-математических наук,
доцент кафедры статистического моделирования СПбГУ,
vnekr@statmod.ru

Аннотация

Метод анализа сингулярного спектра (SSA) используется для разложения временного ряда на интепретируемые аддитивные компоненты. Среди задач, которые решает SSA, можно выделить отделение сигнала от аддитивного шума (помехи). Настоящий доклад посвящен случаю растущего экспоненциального сигнала и гармонической помехи при большой длине ряда. Точнее, с помощью результатов теории возмущений линейных операторов Т. Като изучены ошибки восстановления сигнала в различных точках ряда.

Введение

SSA (Singular Spectrum Analysis) — метод анализа временных рядов, использующийся для разложения исходного ряда на несколько различных интерпретируемых аддитивных компонент. Подробное его описание можно найти в [1].

В статье [2] рассматривается вариант SSA, посвященный отделению сигнала $F_N = (f_0, \ldots, f_{N-1})$ от аддитивной помехи $E_N = (e_0, \ldots, e_{N-1})$ при больших N в случае, когда сигнал F_N управляется минимальной линейной рекуррентной формулой порядка d. Тем самым предполагается, что наблюдаемый временной ряд записывается в виде $F_N(\delta) = F_N + \delta E_N$, где δ — формальный параметр возмущения.

В этом случае выбором “длины окна” $1 < L < N$ ряд $F_N(\delta)$ каноническим образом преобразуется в ганкелеву матрицу $\mathbf{H}(\delta)$ с L строками и $K = N + 1 - L$ столбцами, причем $\mathbf{H}(\delta) = \mathbf{H} + \delta\mathbf{E}$, где ганкелевы матрицы $\mathbf{H}$ и $\mathbf{E}$ строятся аналогичным образом исходя из рядов F_N и E_N. При всех достаточно больших L, K ранг матрицы $\mathbf{H}$ равен d.

В рассматриваемой ситуации процедуру SSA можно описать следующим образом.

1. производится сингулярное разложение матрицы $\mathbf{H}(\delta)$;

2. элементарные матрицы этого разложения, соответствующие d наибольшим сингулярным числам, суммируются. В результате получается матрица $\widetilde{\mathbf{H}}$;

3. производится ганкелизация матрицы $\widetilde{\mathbf{H}}$, то есть строится ганкелева матрица, наиболее близкая к $\widetilde{\mathbf{H}}$ по норме Фробениуса;

4. наконец, эта матрица взаимно-однозначным образом преобразуется во временной ряд $\widetilde{F}_N(\delta)$, который объявляется приближением к F_N. В дальнейшем оператор перехода от произвольной матрицы $\mathbf{G}$ к порождаемому ею временному ряду G обозначается $\mathcal{S}$, так что $G = \mathcal{S}(\mathbf{G})$.

В [2] показано, что ряд $\widetilde{F}_N(\delta) - F_N$ (то есть ряд из погрешностей полученной аппроксимации сигнала F_N) имеет следующий вид. Обозначим $\mathbf{P}_0^{\perp}$ и $\mathbf{P}_0^{\perp}(\delta)$ ортогональные проекторы на линейные пространства столбцов матриц $\mathbf{H}$ и $\mathbf{H}(\delta)$. Тогда

$$\Delta(F_N) \stackrel{\text{def}}{=} \widetilde{F}_N(\delta) - F_N = \mathcal{S}\big(\big(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp}\big)\,\mathbf{H}(\delta) + \delta\,\mathbf{P}_0^{\perp}\mathbf{E}\big)\,. \qquad (1)$$

Представление (1), в частности, было применено в [2] для растущего экспоненциального сигнала $f_n = a^n$ с $a > 1$ и постоянной помехи $e_n = 1$. При этом оказалось, что максимальная по модулю погрешность аппроксимации элементов сигнала F_N не будет стремиться к нулю, если $L \sim \alpha N$ с $\alpha \in (0,1)$ при $N \to \infty$. Аналогичный результат, доказательство которого можно найти в [4], получается и для "пилообразной" помехи $e_n = (-1)^n$, что, возможно, несколько противоречит полуинтуитивным представлениям о свойствах метода SSA.

Настоящий доклад посвящен распространению этих результатов на гармоническую помеху вида $e_n = \sin(2\pi\omega n + \varphi)$ с $\omega \in (0, 1/2)$ и $\varphi \in [0, 2\pi)$.

Как и в [2], используется классическая техника рядов возмущений Т. Като [3], примененная к разности проекторов $\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp}$. Отметим, что скорость сходимости спектральной нормы разности этих проекторов имеет самостоятельный интерес, так как на этой сходимости основаны некоторые (subspace-based) методы обработки временных рядов.

Основные результаты

Основной результат, представленный в докладе, может быть сформулирован следующим образом.

Пусть $f_n = a^n$ с $a > 1$ и $e_n = \sin(2\pi\omega n + \varphi)$ с $\omega \in (0, 1/2)$, $\varphi \in [0, 2\pi)$. Обозначим j-й член ряда погрешностей (1) как $r_j = r_j(N)$.

Теорема 1 *Если $N \to \infty$ и $L/N \to \alpha \in (0,1)$, то для любого δ*

1.

$$\frac{a^N}{\sqrt{N}} \left\| \mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} \right\| \to |\delta| \, \frac{(a^2-1)}{a} \sqrt{\frac{\alpha(a^2-1)}{2(a^2+1-2a\cos(2\pi\omega))}} \,. \qquad (2)$$

2. Если $\varepsilon \in (0,1)$, то $r_j \to 0$ равномерно по $j \in [0, N(1-\varepsilon))$.

3. Если $j = N - \ell$, где $\ell = O(1)$, то существует такая подпоследовательность $N = N_k \to \infty$, что r_j не стремится к нулю.

Замечание 1 *1. Более тщательный анализ показывает, что скорость сходимости r_j к нулю зависит от положения j на промежутке $[0, N(1-\varepsilon))$ — с ростом j эта скорость, вообще говоря, уменьшается от экспоненциальной до $O(N^{-1})$.*
2. Если $j = N - \ell$, где $\ell = \mathrm{const}$, то множество предельных точек последовательности $r_j = r_j(N)$ зависит от того, является ли частота ω рациональной или нет.

При $\omega = p/q$ предельное множество состоит из конечного числа t точек, причем $2 \le t \le q$, а обе границы являются достижимыми.

В случае иррациональной частоты это множество заполняет некоторый интервал, а соответствующее распределение предельных точек является вариантом распределения арксинуса.

Общий ход доказательства

Приведем общий ход доказательства Теоремы 1, опуская довольно громоздкие технические детали.

Прежде всего, используется разложение Като, позволяющее при легко проверяемых условиях представить разность $\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp}$ в виде бесконечной суммы некоторых операторов, зависящих, в частности от матриц $\mathbf{H}$ и $\mathbf{E}$, см. [2].

Для рассматриваемых сигнала и помехи линейный по δ член этого разложения имеет вид

$$\delta \mathbf{V}_0^{(1)} = \delta \left(\mathbf{P}_0 \mathbf{E} \mathbf{H}^{\mathrm{T}} \mathbf{P}_0^{\perp} + \mathbf{P}_0^{\perp} \mathbf{H} \mathbf{E}^{\mathrm{T}} \mathbf{P}_0 \right) / \mu,$$

где μ — единственное ненулевое собственное число матрицы $\mathbf{H}$, а $\mathbf{P}_0$ — ортогональный проектор на ядро этой матрицы.

Оператор $\delta\mathbf{V}_0^{(1)}$ оказывается главным членом разности $\mathbf{P}_0^{\perp}(\delta)-\mathbf{P}_0^{\perp}$ в том смысле, что для любого δ

$$\|\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \delta\mathbf{V}_0^{(1)}\| = o\big(\|\mathbf{V}_0^{(1)}\|\big), \quad N \to \infty.$$

Относительно простой вид $\delta\mathbf{V}_0^{(1)}$ позволяет посчитать норму этого оператора и проанализировать ее асимптотику, что приводит к (2).

Затем доказывается, что

$$\Big\|\big(\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp} - \delta\mathbf{V}_0^{(1)}\big)\mathbf{H}(\delta)\Big\| \to 0.$$

Отсюда следует, что для изучения предельного поведения погрешностей $r_j(N)$ вместо (1) достаточно рассматривать

$$\delta\mathcal{S}\big(\mathbf{V}_0^{(1)}\mathbf{H}(\delta) + \mathbf{P}_0^{\perp}\mathbf{E}\big).$$

Поскольку вид входящих в это выражение матриц выписывается явно, это дает возможность доказать два последних утверждения Теоремы 1, а также провести более точный анализ погрешностей r_j, обозначенный в Замечании 1.

Замечание 2 *В [4] доказано, что при замене экспоненциального ряда $f_n = a^n$ на его “дискретизованный” вариант $f_n = a^{nT/N}$ с $T = \mathrm{const}$ и рассмотрении “пилообразной” помехи $e_n = (-1)^n$ погрешности $r_j = r_j(N)$ равномерно по j стремятся к нулю.*

Следует ожидать, что подобный результат имеет место и для гармонической помехи $e_n = \cos(2\pi\omega n + \varphi)$ с $\omega \in (0, 1/2)$. Доказательство этого факта, однако, усложняется тем, что здесь главный член разности проекторов $\mathbf{P}_0^{\perp}(\delta) - \mathbf{P}_0^{\perp}$ имеет гораздо более сложный вид, чем $\delta\mathbf{V}_0^{(1)}$.

Литература

[1] Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and related techniques. — Chapman and Hall/CRC, 2001. — 320 p.

[2] Nekrutkin V. Perturbation expansions of signal subspaces for long signals // Stat. Interface. — 2010. — Vol. 3. — P. 297–319.

[3] Като Т. Теория возмущения линейных операторов. — Мир, М., 1972. — 740 с.

[4] Иванова Е.В. Об одном примере, относящемся к анализу сингулярного спектра временных рядов. Бакалаврская работа. — СПбГУ, СПб, 2015. — 80 с.