Применение комплексно-значного метода анализа сингулярного спектра

Жорникова Полина Георгиевна, гр. 422

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

Научный руководитель: к.ф.-м.н., доцент Голяндина Н.Э.
Рецензент: ассистент Шлемов А.Ю.

Санкт-Петербург
2016г.

Введение: Общая постановка задачи

Комплексный временной ряд $F_N = F_N^{(1)} + \mathrm{i}F_N^{(2)}$, где $F_N^{(i)} = (f_1^{(i)}, \ldots, f_N^{(i)})$, $i = 1, 2$, — вещественные.

Проблема: найти компоненты разложения $F_N = T_N + S_N + R_N$, где $T_N = T_N^{(1)} + \mathrm{i}T_N^{(2)}$, $S_N = S_N^{(1)} + \mathrm{i}S_N^{(2)}$, $R_N = R_N^{(1)} + \mathrm{i}R_N^{(2)}$,
$T_N^{(i)}$ — тренд, $S_N^{(i)}$ — периодика и $R_N^{(i)}$ — шум для ряда $F_N^{(i)}$.

Метод: **SSA (Singular Spectrum Analysis)** [Golyandina N., Nekrutkin V., Zhigljavsky A., 2001] — для вещественных рядов,
Complex SSA [Keppenne C., Lall U., 1996] — для комплексных рядов.

Две задачи:

- Выделение периодики для вещественного и комплексного случаев.
- Применение метода Complex SSA: исследование и обоснование алгоритма выделения прямых линий с зашумленного изображения.

Введение: Алгоритм метода Complex SSA

$F_N = F_N^{(1)} + \mathrm{i}F_N^{(2)}$ — комплексный временной ряд.

1. **Вложение:** $0 < L < N$ — *длина окна*, $K = N - L + 1$, ряд переводится в *траекторную матрицу* $\mathbf{X} = [X_1 : \ldots : X_K]$, где $X_i = (f_i, \ldots, f_{i+L-1})^{\mathrm{T}}$.
2. **Сингулярное разложение:** $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$, $\mathbf{X}_i = \mu_i U_i V_i^*$,
   μ_i — сингулярные числа,
   U_i — левые сингулярные вектора,
   V_i — правые сингулярные вектора.
3. **Группировка:** матрицы $\mathbf{X}_i$ группируются в три группы, соотв. тренду, периодике и шуму; внутри каждой группы матрицы суммируются.
4. **Диагональное усреднение:** сгруппированные матрицы переводятся в ряды.

В результате получаем:

$$F_N = T_N + {\color{red}S_N} + R_N,$$

где T_N — тренд, ${\color{red}S_N}$ — периодика и R_N — шум.

Задача 1: Идентификация периодики для вещественного и комплексного случаев.

Разложение, которое ищем: $F_N = T_N + S_N + R_N$, $\quad F_N \leftrightarrow \mathbf{X}$.

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \underbrace{(\mathbf{X}_k + \ldots + \mathbf{X}_m)}_{\leftrightarrow S_N} + \ldots + \mathbf{X}_d, \qquad \mathbf{X}_i = \mu_i U_i V_i^*.$$

Вопрос: как найти группу матриц $\mathbf{X}_k, \ldots, \mathbf{X}_m$, относящуюся к периодике?

Идентификация периодики происходит с помощью сингулярных чисел μ_i, левых сингулярных векторов U_i, правых сингулярных векторов V_i.

Периодика (обобщенная) описывается суммой экспоненциально-модулированных гармоник, задаваемых формулой:

- $A\, e^{\beta k} \cos(2\pi\omega k + \phi)$ — в вещественном случае,
- $A\, e^{\beta k} (\cos(2\pi\omega k + \phi) + \mathrm{i} B \cos(2\pi\omega k + \psi))$ — в комплексном случае.

Задача: научиться идентифицировать вещественную и комплексную э.-м. гармоники.

P, Q – два вещественных вектора длины L. Два функционала:

1. $\tau_1(P,Q) := \frac{1}{L-1}\sum_{k=1}^{L-1}\left(\theta_k - \bar{\theta}\right)^2$,

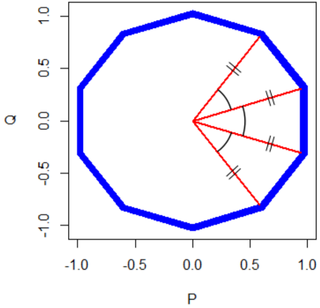
   где θ_k — угол между $(p_k, q_k)^{\mathrm{T}}$ и $(p_{k+1}, q_{k+1})^{\mathrm{T}}$, $\bar{\theta} = \frac{1}{L-1}\sum_{k=1}^{L-1}\theta_k$;

2. $\tau_2(P,Q) := \frac{1}{L}\sum_{k=1}^{L}(x_k - \bar{x})^2$,

   где $x_k = (p_k)^2 + (q_k)^2$, $\bar{x} = \sum_{k=1}^{L} x_k / L$.

Свойства регулярности:

1. $\tau_1(P,Q) = 0 \Leftrightarrow$ углы между $(p_k, q_k)^{\mathrm{T}}$ и $(p_{k+1}, q_{k+1})^{\mathrm{T}}$ равны.
2. $\tau_2(P,Q) = 0 \Leftrightarrow$ точки $(p_k, q_k)^{\mathrm{T}}$ лежат на окружности.



Вещественная экспоненциально-модулированная гармоника S_N:

$s_n = e^{\alpha n} A \cos(2\pi\omega n + \phi)$, $n = 1, \ldots, N$, $0 < \omega < 0.5$.

Сингулярное разложение траекторной матрицы: $\mathbf{S} = \mu_1 U_1 V_1^\mathrm{T} + \mu_2 U_2 V_2^\mathrm{T}$.

L — длина окна, U_1 и U_2 — вектора длины L.

$\tau_1(P, Q) = 0 \Leftrightarrow$ углы между $(p_k, q_k)^\mathrm{T}$ и $(p_{k+1}, q_{k+1})^\mathrm{T}$ равны.

$\tau_2(P, Q) = 0 \Leftrightarrow$ точки $(p_k, q_k)^\mathrm{T}$ лежат на окружности.

Теорема

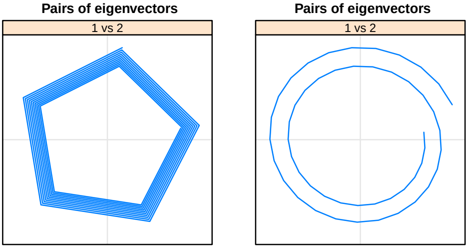
Для сингулярных векторов U_1 и U_2, порождаемых рядом S_N, верны следующие утверждения.

1. *Если $\alpha = C/N$ и $L = [\beta N]$, где C — константа, $0 < \beta < 1$, то $\lim_{L\to\infty} \tau_1(U_1, U_2) = 0$.*
2. *Если $\alpha = 0$ и $L = [\beta N]$, где $0 < \beta < 1$, то $\lim_{L\to\infty} \tau_1(U_1, U_2) = 0$ и $\lim_{L\to\infty} \tau_2(U_1, U_2) = 0$.*
3. *Если $\alpha = 0$ и $L\omega$ — целое, то $\tau_1(U_1, U_2) = 0$ и $\tau_2(U_1, U_2) = 0$.*

Вещественная э.-м. гармоника: $s_n = e^{\alpha n} A \cos(2\pi\omega n + \phi)$.

$$\mathbf{S} = \mu_1 U_1 V_1^{\mathrm{T}} + \mu_2 U_2 V_2^{\mathrm{T}}.$$

Если $\alpha = C/N$ и $L = [\beta N]$, где C — константа, $0 < \beta < 1$, то $\lim_{L\to\infty} \tau_1(U_1, U_2) = 0$.



(a) $\omega = 0.2$, $\alpha = 0.005$, $C = 0.495$, $\tau_1 =$1.3e-05.

(b) $\omega = 0.04$, $\alpha = 0.009$, $C = 0.891$, $\tau_1 =$4.2e-05.

Рис.: Двумерные диаграммы с.в. вещ. э.-м. гармоник, $N = 99$, $L = 50$, $\beta = 0.5$.

Сингулярное разложение: $\mathbf{X} = \mu_1 U_1 V_1^{\mathrm{T}} + \ldots + \mu_d U_d V_d^{\mathrm{T}}$,
$U_1, \ldots, U_d$ — левые сингулярные вектора.

τ — рассматриваемый функционал (τ_1 или τ_2).

1. Визуальная идентификация: среди двумерных диаграмм векторов U_j и U_{j+1} ищутся те, где изображение обладает регулярным видом («спирали»).
2. Автоматическая идентификация:
    1. **число э.-м. гармоник известно**: отбираем пары U_j, U_{j+1} с минимальным значением функционала $\tau(U_j, U_{j+1})$;
    2. **число э.-м. гармоник неизвестно**: отбираем пары U_j, U_{j+1}, у которых значение функционала $\tau(U_j, U_{j+1})$ меньше заданного порога.

Комплексная экспоненциально-модулированная гармоника:

$s_n = e^{\alpha n}(A\cos(2\pi\omega n+\phi)+\mathrm{i}B\cos(2\pi\omega n+\psi))$, $n=1,\ldots,N$, $0<\omega<0.5$.

Ранг ряда $d=1$, если $A=B$ и $|\phi-\psi|\equiv\pi/2 \pmod{\pi}$, иначе $d=2$.

В отличие от вещественного случая:

- Синг. разложение имеет одно $(d=1)$ или два $(d=2)$ слагаемых.
- С.в. определены неоднозначно, с точностью до *комплексного поворота*, т.е. умножения на $e^{\mathrm{i}2\pi t}$, $0\leqslant t<1$.

Комплексная экспоненциально-модулированная гармоника:

$$s_n = e^{\alpha n}(A\cos(2\pi\omega n + \phi) + \mathrm{i}B\cos(2\pi\omega n + \psi)),\ n = 1, \ldots, N,\ 0 < \omega < 0.5.$$

Ранг ряда $d = 1$, если $A = B$ и $|\phi - \psi| \equiv \pi/2 \pmod{\pi}$, иначе $d = 2$.

В отличие от вещественного случая:

- Синг. разложение имеет одно $(d = 1)$ или два $(d = 2)$ слагаемых.
- С.в. определены неоднозначно, с точностью до *комплексного поворота*, т.е. умножения на $e^{\mathrm{i}2\pi t}$, $0 \leqslant t < 1$.

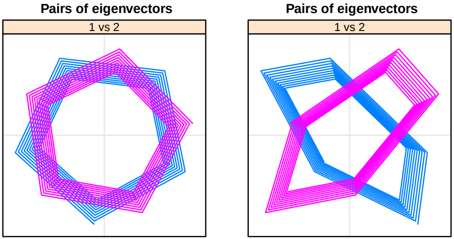


Рис.: Примеры двумерных диаграмм с.в. э.-м. комплексной гармоники, $N = 99$, $L = 50$, $\omega = 0.2$, $\alpha = 0.007$.

Идентификация периодики: Комплексный случай: Теория

Комплексная экспоненциально-модулированная гармоника S_N:
$s_n = e^{\alpha n}(A\cos(2\pi\omega n + \phi) + \mathrm{i}B\cos(2\pi\omega n + \psi))$, $n = 1, \ldots, N$, $0 < \omega < 0.5$.

В случае ранга $d = 2$: $\mathbf{S} = \mu_1 U_1 V_1^* + \mu_2 U_2 V_2^*$.

Теорема

Для с.в. U_1 и U_2, порожденных рядом S_N с $d = 2$, верны след. утв-я.

1. *Если $\alpha = C/N$ и $L = [\beta N]$, где C — константа, $0 < \beta < 1$, то существует пос-ть $t = t_L$, принимающая значения в $[0, 0.5)$, такая что*
$$\lim_{L\to\infty}(\tau_1(\operatorname{Re} U_1, \operatorname{Re} e^{\mathrm{i}2\pi t} U_2) + \tau_1(\operatorname{Im} U_1, \operatorname{Im} e^{\mathrm{i}2\pi t} U_2)) = 0.$$
2. *Если $\alpha = 0$ и $L = [\beta N]$, где $0 < \beta < 1$, то существует пос-ть $t = t_L$, принимающая значения в $[0, 0.5)$, такая что*
$$\lim_{L\to\infty}(\tau_k(\operatorname{Re} U_1, \operatorname{Re} e^{\mathrm{i}2\pi t} U_2) + \tau_k(\operatorname{Im} U_1, \operatorname{Im} e^{\mathrm{i}2\pi t} U_2)) = 0, \quad k = 1, 2.$$
3. *Если $\alpha = 0$ и $L\omega$ — целое, то существует $0 \leqslant t < 0.5$, такое что*
$$\tau_k(\operatorname{Re} U_1, \operatorname{Re} e^{\mathrm{i}2\pi t} U_2) + \tau_k(\operatorname{Im} U_1, \operatorname{Im} e^{\mathrm{i}2\pi t} U_2) = 0, \quad k = 1, 2.$$

Комплексная э.-м. гармоника: $e^{\alpha n}(A\cos(2\pi\omega n+\phi)+\mathrm{i}B\cos(2\pi\omega n+\psi))$.

$\mathbf{S}=\mu_1 U_1 V_1^* + \mu_2 U_2 V_2^*$.

Идентификация: фиксируем U_1, для U_2 ищем поворот $2\pi t$, $0\leqslant t<0.5$, решая оптимизационную задачу:

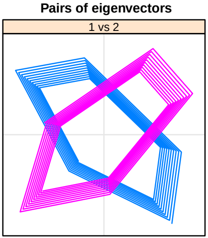
$$\tau_k(\operatorname{Re}U_1,\operatorname{Re}e^{\mathrm{i}2\pi t}U_2)+\tau_k(\operatorname{Im}U_1,\operatorname{Im}e^{\mathrm{i}2\pi t}U_2)\longrightarrow\min_t,\quad k=1 \text{ или } 2.$$

Комплексная э.-м. гармоника: $e^{\alpha n}(A\cos(2\pi\omega n+\phi)+\mathrm{i}B\cos(2\pi\omega n+\psi))$.

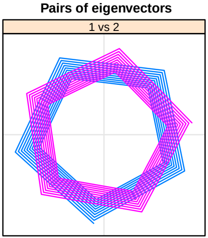
$\mathbf{S}=\mu_1 U_1 V_1^* + \mu_2 U_2 V_2^*$.

Идентификация: фиксируем U_1, для U_2 ищем поворот $2\pi t$, $0\leqslant t<0.5$, решая оптимизационную задачу:

$$\tau_k(\operatorname{Re}U_1,\operatorname{Re}e^{\mathrm{i}2\pi t}U_2)+\tau_k(\operatorname{Im}U_1,\operatorname{Im}e^{\mathrm{i}2\pi t}U_2)\longrightarrow\min_t,\quad k=1 \text{ или } 2.$$



(a) До поворота, $\tau_1=0.62$.



(b) После поворота, $\tau_1=5.1\text{e-}5$.

Рис.: Двумерные диаграммы с.в. э.-м. комплексной гармоники, $N=99$, $L=50$, $\omega=0.2$, $\alpha=0.007$.

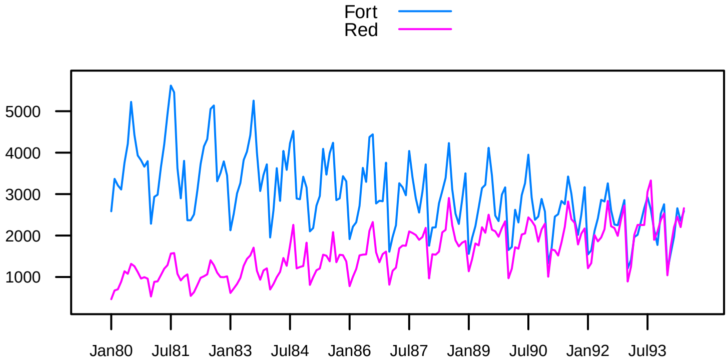
Сингулярное разложение: $\mathbf{X} = \mu_1 U_1 V_1^* + \ldots + \mu_d U_d V_d^*$,
$U_1, \ldots, U_d$ — левые сингулярные вектора.

Идентификация э.-м. гармоник с рангом $d = 2$.

1. Для пары $(U_i,\ U_j)$, $i < j$, решается оптимизационная задача
$$\tau_k(\mathrm{Re}\, U_i, \mathrm{Re}\, e^{\mathrm{i}2\pi t} U_j) + \tau_k(\mathrm{Im}\, U_i, \mathrm{Im}\, e^{\mathrm{i}2\pi t} U_j) \longrightarrow \min_t, \quad k = 1 \text{ или } 2.$$
На $t_{i,j}$ достигается минимум, $\tau_{i,j}$ — минимальное значение.
2. Визуальная идентификация: среди двумерных диаграмм вещественных и мнимых частей U_i и $e^{\mathrm{i}2\pi t_{i,j}} U_j$ ищутся те, где изображение обладает регулярным видом («спирали»).
3. Автоматическая идентификация:
   1. **число э.-м. гармоник известно**: отбираем пары $\left(U_i,\ e^{\mathrm{i}2\pi t_{i,j}} U_j\right)$ с минимальными значениями функционала $\tau_{i,j}$.
   2. **число э.-м. гармоник неизвестно**: отбираем пары $\left(U_i,\ e^{\mathrm{i}2\pi t_{i,j}} U_j\right)$, у которых значение функционала $\tau_{i,j}$ меньше заданного порога;

Для случая $d = 1$ рассматривается $\tau_k(\mathrm{Re}\, U_1, \mathrm{Im}\, U_1)$.

Два ряда ежемесячных продаж крепленого (Fort) и красного (Red) вина в Австралии в тысячах литров в период с января 1980 года по июнь 1994 года.



Рассматриваем ряд Fort+iRed.

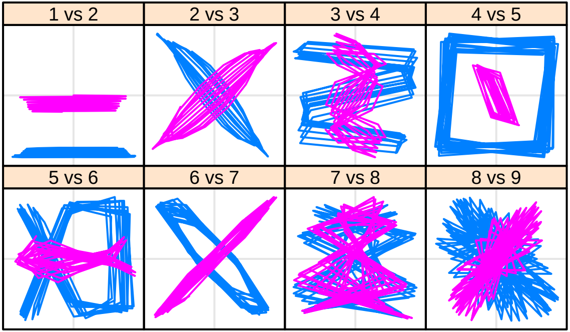


Рис.: Двумерная диаграмма исходных сингулярных векторов для ряда Fort+iRed.

Идентификации периодики: Пример

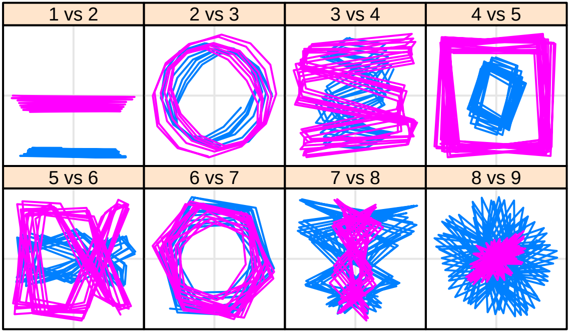


Рис.: Двумерная диаграмма сингулярных векторов для ряда Fort+iRed после работы алгоритма с τ_1.

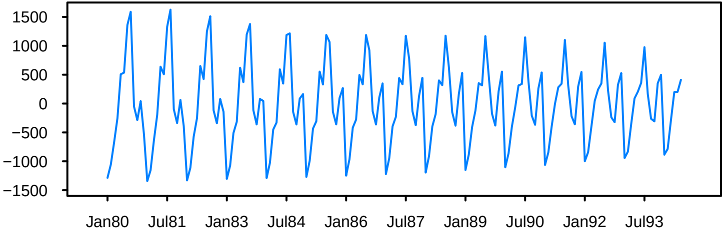


Рис.: Обобщенная периодическая компонента ряда Fort.

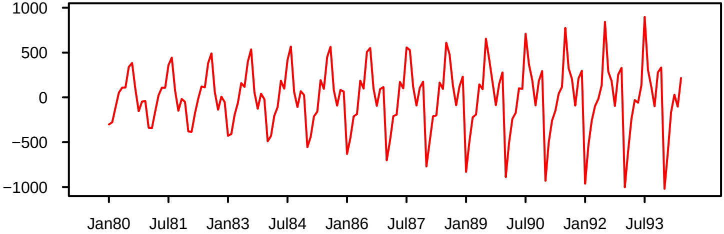


Рис.: Обобщенная периодическая компонента ряда Red.

Задача 2: Применение метода Complex SSA.

Пусть наблюдаем матрицу $\mathbf{X}$ размера $N_t \times N_c$:

$$\mathbf{X} = \mathbf{S} + \mathbf{R},$$

где $\mathbf{S}$ — сигнал: матрица из «0» и «1», все «1» образуют прямые линии; $\mathbf{R} = \{\varepsilon_{ij}\}_{i,j}^{N_t,N_c}$ — шум, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Простейший пример матрицы $\mathbf{S}$:

$$\begin{matrix} 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \end{matrix}$$

Проблема: по матрице $\mathbf{X}$ оценить сигнал $\mathbf{S}$ (выделить прямые).

Метод: Алгоритм [Trickett, 2003].

Задача: доказать корректность алгоритма, рассмотреть несколько его модификаций и выделить лучший вариант алгоритма.

Вход: матрица $\mathbf{X}$, система из N_c временных рядов длины N_t:
$\mathbf{X} = \{x_{ct} = x_c(t) \mid c = 1, \ldots, N_c,\ t = 1, \ldots, N_t\}$.

1. Дискретное преобразование Фурье (DFT) для каждого ряда (для каждой строчки матрицы $\mathbf{X}$). Результат — матрица Ξ.

   DFT для ряда $(0, \ldots, \underbrace{1}_{m}, \ldots, 0)$: $\quad (y_0, \ldots, y_{N-1}),\ y_k = e^{\frac{-2\pi \mathrm{i}}{N} km}$.

   $e^{\frac{-2\pi \mathrm{i}}{N} km} = \sin(\frac{-2\pi}{N} km) + \mathrm{i} \cos(\frac{-2\pi}{N} km)$ — **компл. гармоника!**
2. Метод Complex SSA одним из двух способов:
   1. для каждого столбца матрицы Ξ (COL$_1$),
   2. для каждой строчки матрицы Ξ (ROW$_1$).

   Выбираем длину окна L, идентифицируем M компонент, относящихся к сигналу, и получаем матрицу $\Xi^{(1)}(k)$.
3. Обратное преобразование Фурье одним из двух способов:
   1. для каждого столбца матрицы $\Xi^{(1)}$ (COL$_2$),
   2. для каждой строчки матрицы $\Xi^{(1)}$ (ROW$_2$).

   Результат: матрица $\mathbf{X}^{(1)}$.

Итоговый результат: $\mathbf{X}^{(1)} = \{x_c^{(1)}(t) \mid c = 1, \ldots, N_c,\ t = 1, \ldots, N_t\}$.

Вывод из сравнения 4 вариантов: алгоритм COL_1–ROW_2 — лучший.

Пусть сигнал $\mathbf{S}$ состоит из m прямых с уравнениями $c = a_j\, t + b_j$, $j = 1, \ldots, m$, и с некоторыми условиями на a_j и b_j.

$\mathbf{\Xi} = \{\xi_t(c)\}_{t=1,c=1}^{N_t,N_c}$ —результат первого этапа DFT, примененного к $\mathbf{S}$.

Тогда $\xi_t(c) = \sum_{j=1}^{m} e^{-2\pi \mathrm{i}\frac{(t-1)(c/a_j-1-b_j/a_j)}{N_t}}$, $c = 1, \ldots, N_c$, является суммой комплексных гармоник!

$\mathbf{X} = \mathbf{S} + \mathbf{R}$, $\mathbf{R} = \{\varepsilon_{ij}\}_{i,j}^{N_t,N_c}$, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

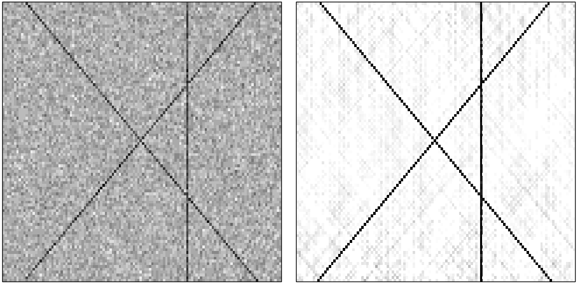
Выделение m прямых из матрицы $\mathbf{X}$: на втором этапе Complex SSA восстанавливаем сигнал на основе первых m сингулярных векторов.

$\mathbf{X} = \mathbf{S} + \mathbf{R}$;

$\mathbf{S}$ — сигнал из трех прямых с уравнениями $c = t + 10$, $c = 110 - t$ и $c = 70$;

$\mathbf{R} = \{\varepsilon_{ij}\}_{i,j}^{N_t, N_c}$, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

$N_c = 100$, $N_t = 120$, $\sigma = 0.2$.



(a) Исходное изображение. (b) Выделенный сигнал.

Рис.: Пример с тремя линиями и шумом до и после работы алгоритма.

Результаты

- Построены и обоснованы алгоритмы идентификации периодики для вещественных и комплексных рядов. Предложено два варианта идентификации: визуальная и автоматическая.
- Проведено численное сравнение вариантов алгоритма с разными функционалами.

- Исследован алгоритм Триккета выделения прямых линий из зашумленного изображения и три аналогичные модификации алгоритма.
- Показано, что вариант, предложенный Триккетом, при некоторых не очень ограничивающих условиях является лучшим, эффективно и корректно работающим для любого количества линий.

- Проиллюстрированы примеры работы алгоритмов идентификации и алгоритмов выделения прямых линий.
- Алгоритмы идентификации и алгоритмы выделения прямых линий реализованы на языке программирования R с использованием пакетов Rssa и lattice.