

Поиск весов для задачи оценивания сигнала конечного ранга в присутствии случайного шума*

Н. К. Звонарев

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: Звонарев Н. К. Поиск весов для задачи оценивания сигнала конечного ранга в присутствии случайного шума // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2021. Т. 8 (66). Вып. 3. С. 417–429.
<https://doi.org/10.21638/spbu01.2021.304>

Рассматривается задача взвешенной аппроксимации временного ряда рядом конечного ранга с целью оценивания сигнала в модели «сигнал плюс шум», где обратная ковариационная матрица шума является $(2p + 1)$ -диагональной. Решается задача поиска весов, которые приводят к улучшению точности оценок сигнала. Построен и теоретически обоснован эффективный численный метод поиска весов. Проведено численное моделирование для различных примеров шума, показывающее улучшение точности метода оценивания сигнала.

Ключевые слова: временные ряды, метод Кэдзоу, ряды конечного ранга, взвешенный метод наименьших квадратов, дивергенция Кульбака–Лейблера, методы оптимизации.

1. Введение. 1.1. Сигналы конечного ранга. Рассмотрим задачу извлечения сигнала $\mathbf{S} = (s_1, \dots, s_N)^T$ из зашумленного временного ряда $\mathbf{X} = \mathbf{S} + \mathbf{R}$, где \mathbf{S} управляется *линейной рекуррентной формулой* (ЛРФ) порядка r : $s_n = \sum_{i=1}^r a_i s_{n-i}$, $n = r+1, \dots, N$, $a_r \neq 0$, \mathbf{R} – шум с нулевым матожиданием и ковариационной матрицей $\mathbf{\Sigma} = \mathbf{\Sigma}_N \in \mathbb{R}^{N \times N}$. Хорошо известен результат, который задает явный вид управляемых ЛРФ рядов в параметрическом виде: $s_n = \sum_i P_i(n) \exp(\alpha_i n) \cos(2\pi\omega_i n + \psi_i)$, где $P_i(n)$ – многочлены от n (см. [1, теорема 5.3] и [2]).

Известно, что для выделения такого рода сигналов хорошо работают методы, основанные на оценке сигнального подпространства (subspace-based methods), и, в частности, анализ сингулярного спектра (см. [2–4]). Идея методов состоит в следующем: зафиксируем *длину окна* L , $1 < L < N$, положим $K = N - L + 1$ и определим *траекторную матрицу* для ряда $\mathbf{S} \in \mathbb{X}_N$ (\mathbb{X}_N – множество вещественных временных рядов длины N):

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

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-01-00067).

© Санкт-Петербургский государственный университет, 2021

где \mathcal{T}_L обозначает биекцию между \mathbb{X}_N и \mathcal{H} , и \mathcal{H} — множество ганкелевых матриц порядка $L \times K$ с одинаковыми значениями на побочных диагоналях $i + j = \text{const}$.

Скажем, что ряд \mathbf{S} , у которого траекторная матрица \mathbf{S} имеет неполный ранг r для любого $r < L < N - r + 1$, называется *рядом ранга r* . Известно, что определение рядов ранга r эквивалентно равенству ранга r у траекторной матрицы с $L = r + 1$, если $N > 2r$ [5, следствие 5.1]. Множество временных рядов ранга r длины N обозначим как $\mathcal{D}_r^N = \mathcal{D}_r$.

1.2. Сведение к задаче аппроксимации. Рассмотрим задачу аппроксимации временного ряда рядом ранга r по косоугольной евклидовой норме:

$$\mathbf{Y}^* = \arg \min_{\mathbf{Y} \in \overline{\mathcal{D}}_r} \|\mathbf{X} - \mathbf{Y}\|_{\mathbf{W}}^2, \quad (1)$$

где $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^N$ — временные ряды длины N , $\|\mathbf{Z}\|_{\mathbf{W}}$ — косоугольная евклидова норма, порожденная скалярным произведением

$$\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbf{W}} = \mathbf{X}^T \mathbf{W} \mathbf{Y}, \quad (2)$$

$\mathbf{W} \in \mathbb{R}^{N \times N}$ — симметричная положительно определенная матрица весов (назовем ее «весовой матрицей»), $\overline{\mathcal{D}}_r$ — замыкание множества \mathcal{D}_r .

Полученное решение \mathbf{Y}^* задачи (1) можно использовать как оценку сигнала \mathbf{S} рядом конечного ранга. Более того, полученная оценка позволяет решить множество связанных задач: оценка коэффициентов ЛРФ и параметров ряда, прогнозирование ряда, заполнение пропусков и так далее (см. [1]). Заметим, что при $\mathbf{W} = \mathbf{\Sigma}^{-1}$ полученная оценка сигнала \mathbf{Y}^* в предположении о гауссовости шума \mathbf{R} является оценкой максимального правдоподобия: логарифм функции правдоподобия в многомерной нормальной модели выглядит как $-\frac{N}{2} \log(2\pi) + \frac{1}{2} \det(\mathbf{W}) - \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_{\mathbf{W}}^2$ (например, см. [6, формула (8.19)]), где ряд \mathbf{Y} удовлетворяет параметрической модели $\mathbf{Y} \in \overline{\mathcal{D}}_r$; выражение с точностью до константных слагаемых, множителя $\frac{1}{2}$ и знака минус совпадает с (1).

1.3. Проблема подбора весов. Для решения задачи (1) рассмотрим связанную с ней задачу ганкелевой структурной аппроксимации матрицей неполного ранга (Structured Low-Rank Approximation, SLRA):

$$\mathbf{Y}^* = \arg \min_{\mathbf{Y} \in \mathcal{M}_r \cap \mathcal{H}} f_{\mathbf{L}, \mathbf{R}}(\mathbf{Y}), \quad f_{\mathbf{L}, \mathbf{R}}(\mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_{\mathbf{L}, \mathbf{R}}^2, \quad (3)$$

где $\|\mathbf{X}\|_{\mathbf{L}, \mathbf{R}}$ — порожденная скалярным произведением

$$\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbf{L}, \mathbf{R}} = \text{tr}(\mathbf{L} \mathbf{X} \mathbf{R} \mathbf{Y}^T) \quad (4)$$

матричная норма [7], $\mathcal{M}_r \subset \mathbb{R}^{L \times K}$ — множество матриц ранга, не превосходящего r , $\mathbf{L} \in \mathbb{R}^{L \times L}$, $\mathbf{R} \in \mathbb{R}^{K \times K}$ — положительно определенные матрицы весов, а \mathbf{X} и \mathbf{Y} — матрицы, связанные с задачей (1) соотношениями $\mathbf{X} = \mathcal{T}_L(\mathbf{X})$ и $\mathbf{Y} = \mathcal{T}_L(\mathbf{Y})$.

Между весовой матрицей \mathbf{W} в задаче (1) и матрицами \mathbf{L} , \mathbf{R} в (3) существует соотношение, при котором нормы $\|\cdot\|_{\mathbf{W}}$ и $\|\cdot\|_{\mathbf{L}, \mathbf{R}}$ являются эквивалентными [8].

Основная проблема состоит в том, чтобы найти матрицы \mathbf{L} , \mathbf{R} (назовем их «матричными весами»), которые бы соответствовали матрице $\mathbf{W}_0 = \mathbf{\Sigma}^{-1}$, что естественно для модели с такой ковариационной матрицей шума и приводит к задаче оценивания сигнала по методу максимального правдоподобия. Для случая авторегрессионного шума (и для белого шума в частности) в [8] показано, что если это

и возможно, то только при вырожденности либо матрицы \mathbf{L} , либо \mathbf{R} , что недопустимо при решении задачи методами типа Oblique Cadzow (см. [9, замечание 4]). Следовательно, приходится жертвовать точным равенством \mathbf{W} и Σ^{-1} и использовать невырожденные матрицы \mathbf{L} , \mathbf{R} , которые дают матрицу \mathbf{W} , приближенную к Σ^{-1} . Частный случай этой задачи для модели с белым шумом был разобран в [10].

Цель статьи — рассмотреть подходы к поиску матриц \mathbf{L} , \mathbf{R} в случае, когда матрица Σ^{-1} является ленточной и содержит $(2p + 1)$ ненулевые диагонали, $p \geq 0$. Такие матрицы возникают при нестационарном обобщении авторегрессионной дискретной модели [11]. Важные частные случаи такой модели — классический стационарный процесс авторегрессии порядка p (AR(p)) [12] и модель нестационарного белого шума.

1.4. Подход к подбору весов. Известно, что если взять $(2p + 1)$ -диагональную матрицу \mathbf{L} и диагональную матрицу \mathbf{R} , то соответствующая им матрица \mathbf{W} будет $(2p + 1)$ -диагональной [8]. Этим фактом можно воспользоваться, чтобы упростить подход. В предположении, что матрица \mathbf{L} зафиксирована, положительно определена и имеет сходную с матрицей Σ^{-1} структуру, задача поиска весов сводится только к подбору матрицы \mathbf{R} . Например, если $\Sigma = \Sigma_N(\mathbf{b})$, где $\Sigma_M(\mathbf{b}) \in \mathbb{R}^{M \times M}$ — матрица ковариации стационарного процесса AR(p) длины M с коэффициентами $\mathbf{b} \in \mathbb{R}_p$, то зафиксируем $\mathbf{L} = \Sigma_L^{-1}(\mathbf{b})$. Для матрицы \mathbf{R} ограничим число обусловленности сверху, чтобы гарантировать ее невырожденность. Как и в работах [9, 10], мы рассмотрим диагональную матрицу $\mathbf{R} = \text{diag}(r_1, \dots, r_K)$, введем параметр $0 < \alpha \leq 1$ и потребуем, чтобы выполнялось неравенство $r_i \geq 0$ и

$$\frac{\min_i r_i}{\max_i r_i} \geq \alpha, \quad (5)$$

что эквивалентно неравенству $\text{cond } \mathbf{R} \leq 1/\alpha$. После этого сформулируем задачу поиска нужной матрицы \mathbf{R} при помощи минимизации дивергенции Кульбака — Лейблера (являющейся естественной функцией различия между распределениями, задающими требуемую и приближенную модели шума) между многомерными нормальными распределениями с нулевыми средними и ковариационными матрицами Σ и \mathbf{W}^{-1} .

Полученные в результате матричные веса \mathbf{L} , \mathbf{R} применяются в методе итераций Кэдзоу (Cadzow) [13], который дает приближенное численное решение задачи SLRA (3). Определим следующую последовательность матриц, начинающуюся с заданной матрицы \mathbf{X} :

$$\mathbf{X}_0 = \mathbf{X}, \quad \mathbf{X}_{k+1} = \Pi_{\mathcal{H}} \Pi_{\mathcal{M}_r} \mathbf{X}_k, \quad k \geq 0, \quad (6)$$

$\Pi_{\mathcal{H}}$ и $\Pi_{\mathcal{M}_r}$ — проекторы на соответствующие множества по матричной норме $\|\cdot\|_{\mathbf{L}, \mathbf{R}}$, порожденной скалярным произведением (4). В [9] показано, что параметр α влияет на скорость сходимости метода Кэдзоу (6): чем больше $\text{cond}(\mathbf{R})$ (что равнозначно меньшему значению α), тем медленнее сходится метод.

Ключевое отличие от предыдущего подхода, описанного в [10], состоит в обобщении на случай неединичной матрицы Σ и замене целевой функции на выражение, обоснованное с точки зрения математической статистики и определенное для произвольных положительно определенных матриц. Решение задачи сводится к оптимизации гладкой функции в многомерном кубе. Для подобных задач созданы эффективные численные алгоритмы, например алгоритм L-BFGS-B [14].

1.5. Структура работы. Раздел 2 содержит известную информацию о связи между матричными весами и матрицей \mathbf{W} , необходимую для изучения задачи поиска весов. В разделе 3 формулируется задача поиска весов как оптимизационная задача и рассматривается выбор матрицы \mathbf{L} на основе обобщения модели авторегрессии. Раздел 4 посвящен вычислительному аспекту решения задачи поиска весов: задача сводится к оптимизации гладкой функции в многомерном кубе, выведены выражения для вычисления значения и градиента функции в точке. Приведена временная трудоемкость одной итерации для метода оптимизации при различных матрицах Σ . Раздел 5 содержит численное сравнение полученных весов с результатом работы [10] по ошибке оценки сигнала с помощью метода Кэдзоу на модельных примерах. В [10] для случая единичных весов показано, что чем ближе веса к требуемым, тем точнее решение задачи. В разделе 5 показано, что эти соотношения выполняются и для неединичных весов, полученных предложенным методом.

2. Соотношение между весами рядов и матричными весами. Рассмотрим соотношение, связывающее (полу)норму $\|\mathbf{Z}\|_{\mathbf{W}}$, порожденную скалярным произведением (2) и (полу)норму $\|\mathbf{Z}\|_{\mathbf{L},\mathbf{R}}$, которая порождена скалярным произведением (4) (полу)нормы получаются, если не требовать невырожденность матриц \mathbf{W} , \mathbf{L} , \mathbf{R}). Для эквивалентности задач (1) и (3) нужно, чтобы отображение \mathcal{T}_L сохраняло норму, то есть чтобы выполнялись равенства $\|\mathbf{Z}\|_{\mathbf{W}} = \|\mathbf{Z}\|_{\mathbf{L},\mathbf{R}}$ для любого \mathbf{Z} , $\mathbf{Z} = \mathcal{T}_L(\mathbf{Z})$, где $\mathbf{L} = (l_{i,j})$, $\mathbf{R} = (r_{i,j})$ — матричные веса, $\mathbf{W} = (w_{i,j})$ — весовая матрица.

Определим свертку матриц $\mathbf{C} = \mathbf{A} * \mathbf{B}$ для $\mathbf{A} \in \mathbb{R}^{L \times L}$, $\mathbf{B} \in \mathbb{R}^{K \times K}$, $\mathbf{C} \in \mathbb{R}^{N \times N}$:

$$c_{i,j} = \sum_{i_1+i_2-1=i} \sum_{j_1+j_2-1=j} a_{i_1,j_1} b_{i_2,j_2},$$

где $\mathbf{A} = (a_{i_1,j_1})$, $\mathbf{B} = (b_{i_2,j_2})$, $\mathbf{C} = (c_{i,j})$.

Известно соотношение, связывающее матричные веса с весовой матрицей.

Теорема 1 [8]. $\|\mathbf{Z}\|_{\mathbf{W}} = \|\mathbf{Z}\|_{\mathbf{L},\mathbf{R}}$ для любого ряда $\mathbf{Z} \in \mathbb{R}^N$ тогда и только тогда, когда $\mathbf{W} = \mathbf{W}(\mathbf{L}, \mathbf{R}) = \mathbf{L} * \mathbf{R}$.

Рассмотрим интересующий нас случай диагональной матрицы \mathbf{R} . При этом пусть p обозначает количество ненулевых главных диагоналей матрицы \mathbf{L} в ее верхнем или нижнем треугольнике, то есть \mathbf{L} является $(2p+1)$ -диагональной.

Определим матрицы $\mathbf{M}_i = \mathbf{L} * \mathbf{E}_{i,i}$, где $\mathbf{E}_{i,i} \in \mathbb{R}^{K \times K}$ — матрица, содержащая единицу по индексу (i, i) и ноль по всем остальным индексам. Заметим, что $\mathbf{M}_i \in \mathbb{R}^{N \times N}$ содержит матрицу \mathbf{L} как подматрицу в строчках и столбцах с i -го по $(i+L-1)$ -й, при этом все остальные элементы нулевые. Представим матрицу \mathbf{R} как $\mathbf{R} = \sum_{i=1}^K r_i \mathbf{E}_{i,i}$.

Следствие. Рассмотрим $(2p+1)$ -диагональную матрицу \mathbf{L} и диагональную $\mathbf{R} = \text{diag}(\mathbf{r})$, $\mathbf{r} = (r_1, \dots, r_K)^T$. Тогда весовая матрица \mathbf{W} является $(2p+1)$ -диагональной, $\mathbf{W} = \sum_{i=1}^K r_i \mathbf{M}_i$, где $\mathbf{M}_i = \mathbf{L} * \mathbf{E}_{i,i}$.

3. Эквивалентные формулировки задачи поиска весов. 3.1. Постановка задачи. Сформулируем задачу поиска матричных весов \mathbf{R} при фиксированной матрице \mathbf{L} и ограничении на вырожденность матрицы \mathbf{R} , таких что $\mathbf{W}(\mathbf{L}, \mathbf{R})$ равна заданной обратной матрице ковариаций $\mathbf{W}_0 = \Sigma^{-1}$ или приближена к ней.

Во введении было упомянуто, что уже для авторегрессионного случая $\mathbf{W}_0 = \Sigma_N(\mathbf{b})$ не существует таких невырожденных \mathbf{L} , \mathbf{R} , что $\mathbf{W}(\mathbf{L}, \mathbf{R}) = \mathbf{L} * \mathbf{R} = \mathbf{W}_0$ [8]. Поэтому необходимо поставить задачу поиска \mathbf{L} и \mathbf{R} , которые бы дали матрицу $\mathbf{W}(\mathbf{L}, \mathbf{R})$, схожую с \mathbf{W}_0 по некоей мере близости этих матриц.

Рассмотрим разумную функцию различия между обратной ковариационной матрицей $\mathbf{W}_0 = \Sigma^{-1}$ и весовой матрицей \mathbf{W} . В качестве такой меры выберем дивергенцию Кульбака — Лейблера [15] между многомерными нормальными распределениями с нулевыми средними и ковариационными матрицами $\Sigma = \mathbf{W}_0^{-1}$ (соответствующей модели реального шума в предположении гауссовости) и $\mathbf{W}^{-1} = (\mathbf{W}(\mathbf{L}, \mathbf{R}))^{-1}$ (соответствующей модели с ковариационной матрицей, обратной весовой матрице). По определению дивергенция выглядит следующим образом [16]:

$$D_{\text{KL}}(N(\mathbf{0}_N, \mathbf{W}_0^{-1}) \| N(\mathbf{0}_N, \mathbf{W}^{-1})) = \frac{1}{2} (\text{tr}(\mathbf{W}\mathbf{W}_0^{-1}) - N - \log(\det(\mathbf{W}\mathbf{W}_0^{-1}))).$$

С учетом условия на невырожденность матрицы \mathbf{R} , ограничим ее число обусловленности: $\text{cond } \mathbf{R} \leq 1/\alpha$, $0 < \alpha \leq 1$. Получим оптимизационную задачу в следующем виде:

$$\mathbf{R}^* = \arg \min_{\substack{\text{cond } \mathbf{R} \leq 1/\alpha \\ \mathbf{W} = \mathbf{L} * \mathbf{R} \\ \mathbf{R} = \text{diag}(\mathbf{r})}} D_{\text{KL}}(N(\mathbf{0}_N, \mathbf{W}_0^{-1}) \| N(\mathbf{0}_N, \mathbf{W}^{-1})). \quad (7)$$

Введем функцию, которая представляет собой упрощение дивергенции Кульбака — Лейблера:

$$D_{\mathbf{W}_0}^*(\mathbf{W}) = \text{tr}(\mathbf{W}\mathbf{W}_0^{-1}) - \log(\det(\mathbf{W})), \quad (8)$$

после чего сформулируем следующую задачу:

$$\mathbf{R}^* = \arg \min_{\substack{\text{cond } \mathbf{R} \leq 1/\alpha \\ \mathbf{W} = \mathbf{L} * \mathbf{R} \\ \mathbf{R} = \text{diag}(\mathbf{r})}} D_{\mathbf{W}_0}^*(\mathbf{W}). \quad (9)$$

Предложение 1. Задачи (7) и (9) эквивалентны.

ДОКАЗАТЕЛЬСТВО. Целевые функции задач отличаются только на константные множитель и слагаемое. \square

3.2. Выбор левой весовой матрицы \mathbf{L} . В общем случае выбор подходящей матрицы \mathbf{L} по заданной матрице Σ (или Σ^{-1}) — нетривиальная задача. Ниже мы приведем подход, основанный на статье [11] и позволяющий выбрать \mathbf{L} в некоторых частных случаях.

Используя разложение Холецкого, любую положительно определенную ковариационную матрицу Σ , чья обратная является $(2p+1)$ -диагональной, можно записать как

$$\Sigma = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{D} ((\mathbf{I} - \mathbf{A})^{-1})^T,$$

или через ее обращение:

$$\Sigma^{-1} = (\mathbf{I} - \mathbf{A})^T \mathbf{D}^{-1} (\mathbf{I} - \mathbf{A}), \quad (10)$$

где \mathbf{A} — нижнетреугольная $(p+1)$ -диагональная матрица с нулевой главной диагональю, \mathbf{D} — положительно определенная диагональная матрица. Пусть \mathbf{X} — случайный вектор с нулевым средним и ковариационной матрицей Σ . Тогда его можно представить в виде $\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{V} = \mathbf{A}\mathbf{X} + \mathbf{V}$, где \mathbf{V} — случайный вектор с нулевым средним

и ковариационной матрицей \mathbf{D} , независимый с \mathbf{X} [11]. Добавление в представление ненулевой матрицы \mathbf{A} дает обобщение процесса авторегрессии на «динамический» случай, \mathbf{V} — вектор некоррелированных ошибок.

В случае, когда Σ соответствует классическому стационарному процессу авторегрессии $\text{AR}(\mathbf{b})$, матрица $\mathbf{A} = \mathbf{A}(\mathbf{b})$ является теплицевой, значения на диагоналях соответствуют значениям вектора коэффициентов $\mathbf{b} = (b_1, \dots, b_p)^T$:

$$\mathbf{a}_i = (b_i, \dots, b_i)^T \in \mathbb{R}^{N-i}, \quad i = 1, \dots, p, \quad (11)$$

где \mathbf{a}_i обозначает p -ю главную диагональ матрицы \mathbf{A} . Это следует из определения $x_k = \sum_{i=1}^p b_i x_{k-i} + \varepsilon_k$, ε_k — независимые одинаково распределенные величины с нулевым средним и дисперсией σ^2 , $\mathbf{X} = (x_1, \dots, x_N)^T$, при этом добавляется требование к нахождению всех корней характеристического полинома вне единичной окружности. Заметим, что все элементы главной диагонали матрицы \mathbf{D} , за исключением первых p величин, также равны σ^2 , что тоже следует из определения.

Исходя из этого, мы можем подбирать \mathbf{L} в представлении, аналогичном (10):

$$\mathbf{L} = (\mathbf{I}_L - \mathbf{A}_L)^T \mathbf{D}_L^{-1} (\mathbf{I}_L - \mathbf{A}_L), \quad (12)$$

где $\mathbf{A}_L, \mathbf{D}_L$ — матрицы порядка $L \times L$, имеющие ту же структуру, что и соответствующие матрицы \mathbf{A} и \mathbf{D} .

В частности, если матрица \mathbf{A} теплицева, то матрицу \mathbf{A}_L можно тоже выбрать $(p+1)$ -диагональной нижнетреугольной с теми же значениями на диагоналях. Заметим, что случай теплицевой матрицы \mathbf{A} может соответствовать и нестационарному процессу авторегрессии, если хотя бы один корень характеристического полинома с коэффициентами \mathbf{b} лежит внутри единичного круга. Случай (возможно, нестационарного) белого шума соответствует нулевой матрице \mathbf{A} . Если матрица \mathbf{D} содержит все одинаковые элементы, кроме, возможно, первых p , то в качестве \mathbf{D}_L можно взять диагональную матрицу, у которой на диагонали находятся первые L элементов диагонали матрицы \mathbf{D} .

Замечание 1. Такой выбор матриц \mathbf{A}_L и \mathbf{D}_L для случая стационарного процесса авторегрессии эквивалентен выбору $\mathbf{L} = \Sigma_L^{-1}(\mathbf{b})$ — обратной автоковариационной матрицы процесса авторегрессии длины L с теми же коэффициентами авторегрессии \mathbf{b} , что и у Σ , где $\Sigma_L^{-1}(\mathbf{b})$ можно напрямую вычислить по формулам, приведенным в [12].

Таким образом, мы обобщили выбор матрицы \mathbf{L} на класс ковариационных матриц Σ , которым соответствует теплицева матрица \mathbf{A} и для которых диагональ матрицы \mathbf{D} содержит равные элементы после первых p элементов. Помимо стационарных, такой класс включает и нестационарные процессы авторегрессии, когда нестационарность достигнута за счет выбора коэффициентов полинома \mathbf{b} .

Остается непокрытым случай, когда матрица \mathbf{A} не является теплицевой или диагональ матрицы \mathbf{D} содержит разные элементы и вне первых p (например, случай нестационарного белого шума). Без дополнительной информации можно лишь предложить эвристики, которые точно дадут положительно определенную матрицу \mathbf{L} . К примеру, можно выбрать \mathbf{A}_L теплицевой, где элементы на диагоналях равны усредненным значениям элементов диагонали \mathbf{A} , а матрицу \mathbf{D}_L выбрать соответствующей таковой при стационарном варианте, если матрица \mathbf{A} теплицева и коэффициенты \mathbf{b} отвечают стационарному случаю, либо единичной, если подобные условия не выполнены.

4. Задача поиска весов как задача минимизации гладкой функции.

4.1. Подход к оптимизации. Основная трудность при решении задачи (7) и эквивалентной ей задачи (9) — набор ограничений на элементы матрицы \mathbf{R} . Сформулируем эквивалентную форму задачи (9), в которой вместо вектора \mathbf{r} , $\mathbf{R} = \text{diag}(\mathbf{r})$, используется вектор $\tilde{\mathbf{r}}$, оптимизация по которому происходит в многомерном кубе.

Рассмотрим следующую задачу:

$$\tilde{\mathbf{r}}^* = \arg \min_{\substack{\tilde{\mathbf{r}} = (\tilde{r}_1, \dots, \tilde{r}_K)^T \\ \tilde{r}_i \leq 1, i=1, \dots, K \\ \tilde{r}_i \geq \alpha, i=1, \dots, K}} D_2^*(\tilde{\mathbf{r}}), \text{ где } D_2^*(\tilde{\mathbf{r}}) = -\log(\det(\mathbf{L} * \text{diag}(\tilde{\mathbf{r}}))) + N \log \langle \tilde{\mathbf{r}}, \mathbf{q}(\mathbf{W}_0^{-1}) \rangle, \quad (13)$$

$\mathbf{q}(\mathbf{Z}) = (q_1(\mathbf{Z}), \dots, q_K(\mathbf{Z}))^T : q_i(\mathbf{Z}) = \text{tr}(\mathbf{M}_i \mathbf{Z}) = \langle \mathbf{M}_i, \mathbf{Z} \rangle_F$ — вектор, определенный для симметричной положительно определенной матрицы $\mathbf{Z} \in \mathbb{R}^{N \times N}$, $\langle \cdot, \cdot \rangle_F$ и $\langle \cdot, \cdot \rangle$ обозначают соответственно фробениусово и обычное евклидово скалярные произведения.

Теорема 2. *Задачи (9) и (13) эквивалентны.*

Доказательство. Параметризуем вектор $\mathbf{r} = c\tilde{\mathbf{r}}$ с учетом ограничения на число обусловленности матрицы \mathbf{R} , где $c > 0$ — свободный множитель, при этом $\tilde{r}_i \leq 1$, $\tilde{r}_i \geq \alpha$, $i = 1, \dots, K$, $\tilde{\mathbf{r}} = (\tilde{r}_1, \dots, \tilde{r}_K)^T$. Тогда задачу (9) с учетом следствия теоремы 1 можно переписать следующим образом:

$$\tilde{\mathbf{r}}^* = \arg \min_{\substack{\tilde{r}_i \leq 1, i=1, \dots, K \\ \tilde{r}_i \geq \alpha, i=1, \dots, K \\ c > 0}} \sum_{i=1}^K c\tilde{r}_i \text{tr}(\mathbf{M}_i \mathbf{W}_0^{-1}) - \log \left(\det \left(\sum_{i=1}^K c\tilde{r}_i \mathbf{M}_i \right) \right). \quad (14)$$

Преобразуем целевую функцию из задачи (14):

$$\begin{aligned} \sum_{i=1}^K c\tilde{r}_i \text{tr}(\mathbf{M}_i \mathbf{W}_0^{-1}) - \log \left(\det \left(\sum_{i=1}^K c\tilde{r}_i \mathbf{M}_i \right) \right) = \\ = c \langle \tilde{\mathbf{r}}, \mathbf{q}(\mathbf{W}_0^{-1}) \rangle - N \log c - \log(\det(\mathbf{L} * \text{diag}(\tilde{\mathbf{r}}))). \end{aligned}$$

Заметим, что легко выполнить оптимизацию по множителю c , если найти выражение, при котором целевая функция достигает минимума. Достаточно взять у выражения производную по c и приравнять ее к нулю. Получим оптимальное значение $c^* = \frac{N}{\langle \tilde{\mathbf{r}}, \mathbf{q}(\mathbf{W}_0^{-1}) \rangle}$. Подстановкой $c = c^*$ в (14) и сокращением константного слагаемого получим выражение (13). \square

Предложение 2. *Функция $D_2^*(\tilde{\mathbf{r}})$ из (13) является гладкой, ее градиент определяется равенством*

$$(D_2^*(\tilde{\mathbf{r}}))' = -\mathbf{q}((\mathbf{L} * \text{diag}(\tilde{\mathbf{r}}))^{-1}) + \frac{N \mathbf{q}(\mathbf{W}_0^{-1})}{\langle \tilde{\mathbf{r}}, \mathbf{q}(\mathbf{W}_0^{-1}) \rangle}.$$

Доказательство. Достаточно продифференцировать выражение (13) по $\tilde{\mathbf{r}}$, используя известную формулу $\log \det \mathbf{Z} = \text{tr} \log \mathbf{Z}$, где $\mathbf{Z} = \mathbf{L} * \text{diag}(\tilde{\mathbf{r}}) \in \mathbb{R}^{N \times N}$. \square

Обоснуем необходимость введения задачи поиска весов в форме (13). Согласно утверждению теоремы 2 и предложений 1 и 2, исходная задача минимизации (7)

может быть записана в эквивалентной форме оптимизации гладкой функции в многомерном кубе. Поэтому для численного решения задачи (13) можно использовать, например, достаточно эффективный численный метод L-BFGS-B [14], разработанный для функций такого вида.

4.2. Исследование трудоемкости. Для работы метода требуется уметь вычислять как значение целевой функции $D_2^*(\tilde{\mathbf{r}})$ в заданной точке, так и ее градиент.

Пусть \mathbf{W}_0 и \mathbf{L} являются $(2p+1)$ -диагональными матрицами, где p фиксировано. Рассмотрим асимптотическую сложность вычисления функции $D_2^*(\tilde{\mathbf{r}})$ и ее градиента по N . Применим следующую вычислительную схему.

Предварительно вычислим $\mathbf{q}(\mathbf{W}_0^{-1})$, что сводится к обращению матрицы \mathbf{W}_0 и вычислению K фробениусовских скалярных произведений. Это потребует временной сложности $O(N^2)$, так как $(2p+1)$ -диагональную матрицу можно обратить за время $O(N^2)$ с использованием разложения Холецкого [17]. Дальнейшее вычисление K скалярных произведений с $\mathbf{M}_1, \dots, \mathbf{M}_K$ потребует $O(N^2)$ времени, так как каждая из матриц \mathbf{M}_i содержит $O(N)$ ненулевых элементов, и $K \leq N$.

Вычисление свертки $\mathbf{L} * \text{diag}(\tilde{\mathbf{r}}) = \sum_{i=1}^K \tilde{r}_i \mathbf{M}_i$, которая является $(2p+1)$ -диагональной матрицей, тоже занимает $O(N^2)$ времени (сложение K матриц, каждая из которых содержит $O(N)$ ненулевых элементов). Взятие логарифма ее определителя делается через разложение Холецкого и занимает $O(N)$ времени. Трудоемкость вычисления градиента, требующего нахождения $-\mathbf{q}((\mathbf{L} * \text{diag}(\tilde{\mathbf{r}}))^{-1})$, эквивалентна трудоемкости вычисления $\mathbf{q}(\mathbf{W}_0^{-1})$.

Таким образом, временная сложность вычисления функции $D_2^*(\tilde{\mathbf{r}})$ и ее градиента в случае $(2p+1)$ -диагональных матриц весов составляет $O(N^2)$, что допустимо для многих практических приложений.

Следующее замечание устанавливает порядок трудоемкости для частного случая $p = 0$.

Замечание 2. В случае диагональных матриц \mathbf{W}_0 и \mathbf{L} асимптотическая сложность вычисления функции $D_2^*(\tilde{\mathbf{r}})$ и ее градиента может быть снижена до $O(N \log N)$ за счет применения преобразования Фурье для вычисления $\mathbf{L} * \text{diag}(\tilde{\mathbf{r}})$. В случае, когда $\mathbf{L} = \mathbf{I}_L$, асимптотика может быть снижена до $O(N)$ за счет вычисления сверток с использованием накопленных сумм.

Таким образом, был получен метод поиска весов, эффективный с вычислительной точки зрения и учитывающий вид ковариационной матрицы шума.

5. Численный эксперимент. Ниже приведем численное сравнение эффективности полученных с помощью решения задачи (13) весов и метода нахождения весов, упомянутых в работе [10]. Сравнение было проведено на примере синусоидального сигнала.

Предложенный в [10, алгоритм 3] метод (в дальнейшем назовем его «квадратичный метод») решал иную задачу. Во-первых, там использовалась минимизация квадратичного функционала вместо дивергенции Кульбака — Лейблера. Во-вторых, метод никак не учитывал вид матриц \mathbf{L} и \mathbf{W}_0 — можно считать, что веса искались в предположении, что $\mathbf{L} = \mathbf{I}_L$ и $\mathbf{W}_0 = \mathbf{I}_N$ являются единичными матрицами соответствующих размеров. Новый подход учитывает форму данных матриц.

Для того чтобы изучить разницу между полученными весами, был взят сигнал $S = (s_1, \dots, s_N)$ длины $N = 40$ и ранга $r = 2$, имеющий вид

$$s_k = 5 \sin \frac{2\pi k}{6}, \quad k = 1, \dots, N,$$

и рассматривались три ряда $X_i = S + R_i$, R_i — случайный гауссовский шум с нулевым средним, $i = 1, 2, 3$. В качестве первой модели шума R_1 была взята стационарная авторегрессия с нулевым средним, $p = 1$, $\mathbf{b} = (0.9)^T$, $\sigma = 0.1$. Варианты R_2 и R_3 используют нестационарную авторегрессию. Согласно представлению ковариационной матрицы (10), при теплицевости матрицы \mathbf{A} нестационарности можно добиться как выбором коэффициентов \mathbf{b} характеристического полинома \mathbf{b} , где \mathbf{A} и \mathbf{b} связаны соотношением (11), так и выбором матрицы \mathbf{D} с неравными коэффициентами. Для варианта R_2 возьмем $\mathbf{b} = (0.9)^T$, $\mathbf{D} = 0.01 \cdot \text{diag}(1^2, 2^2, \dots, 40^2)$, нестационарность достигается за счет матрицы \mathbf{D} . Для шума R_3 возьмем $\mathbf{b} = (1)^T$, $\mathbf{D} = \text{diag}(0.01, \dots, 0.01)$ в представлении (10) матрицы Σ , что соответствует модели броуновского шума, который представляет из себя нестационарную авторегрессию за счет выбора коэффициентов \mathbf{b} .

Точность оценки сигнала \hat{S} измерялась с помощью среднего по точкам ряда и по $I = 10000$ реализациям ряда квадрата отклонения от сигнала S :

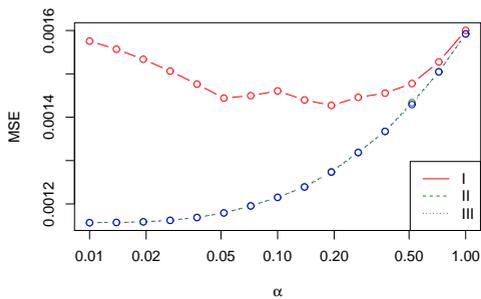
$$\text{MSE}_i = \frac{1}{I} \frac{1}{N} \sum_{j=1}^I \|\hat{S}_j - S\|^2,$$

где \hat{S}_j получена с помощью применения алгоритма Кэдзоу (6) к j -й реализации ряда X_i . Эту меру оценки сигнала будем называть MSE (mean-squared error). Длина окна зафиксирована равной $L = 20$. В каждом из вариантов у метода Кэдзоу было произведено 40 итераций.

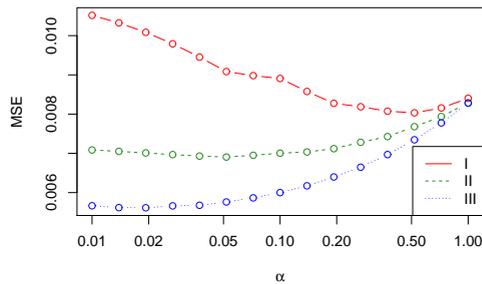
Рассматривались следующие три типа весов. Тип I соответствует методу Oblique Cadzow из работы [10], в нем $\mathbf{L} = \mathbf{I}_L$, \mathbf{R} получена квадратичным методом, то есть информация о виде матрицы Σ никак не используется. Тип II — промежуточный вариант, в котором \mathbf{R} также получена квадратичным методом без использования матрицы Σ , но для метода Кэдзоу матрица \mathbf{L} выбрана уже не единичной, а двудиагональной согласно подразделу 3.2. Тип III соответствует этой работе, в нем матрица \mathbf{L} выбрана согласно подразделу 3.2, для выбора \mathbf{R} используется решение задачи (13).

Результаты сравнения различных типов весов для рядов X_1 и X_2 представлены на рис. 1, для ряда X_3 — на рис. 2.

Для всех трех моделей шума тип весов I провалился и показал большую ошибку (в случае рядов X_2 и X_3 даже возрастающую при уменьшении α). Это связано с тем, что в нем никак не учитывается авторегрессионная природа шума. Тип II показал очень близкий к типу III результат в случае стационарной авторегрессии из-за схожести полученных в результате поиска матричных весов, что видно на рис. 1, а. Однако в нестационарных случаях X_2 и X_3 тип II отличается от типа III. На рис. 1, б видно, что полученные им оценки хуже, чем у типа III при всех $\alpha < 1$, а на рис. 2, б — что ошибка начинает возрастать при уменьшении α на промежутке $\alpha < 0.05$ и становится значительно хуже, чем у метода III. Это объясняется тем, что квадратичный метод поиска весов никак не учитывает информацию о виде ковариационной матрицы Σ .

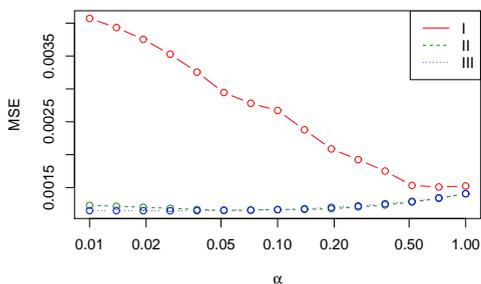


а

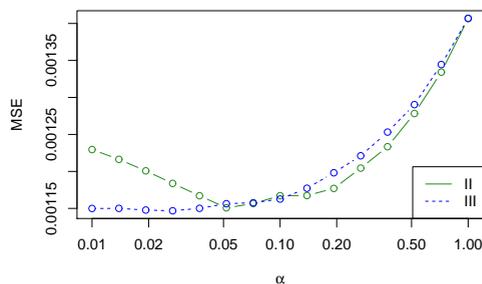


б

Рис. 1. Сравнение методов по MSE при различных α и различных алгоритмах получения весов (I, II, III) для рядов X_1 (а) и X_2 (б).



а



б

Рис. 2. Сравнение методов по MSE при различных α и различных алгоритмах получения весов для ряда X_3 .

6. Заключение. Таким образом, в работе удалось улучшить метод, предлагаемый в [10], тем самым еще больше увеличив преимущество взвешенных вариантов метода Кэдзоу по сравнению со стандартно используемым невзвешенным вариантом [13, 18]. А именно, с помощью нового подхода к поиску весов, состоящего в решении задачи (7) по заданной обратной ковариационной матрице шума Σ^{-1} , удалось получить более точные, чем в работе [10], оценки сигнала конечного ранга по зашумленному ряду, что подтверждено численным экспериментом. Также удалось расширить область применения метода Кэдзоу на случай модели авторегрессионного шума, в том числе и на нестационарный случай.

Для решения задачи поиска весов был применен обобщенный подход, сводящийся к поиску весовой матрицы, близкой к обратной от требуемой ковариационной матрицы шума в смысле дивергенции Кульбака — Лейблера между соответствующими распределениями. Принципиальное отличие от предыдущего подхода, изложенного в [10], состоит в том, что вид ковариационной матрицы учитывается при поиске весов. Для исходной задачи (7) построена эквивалентная задача (13), которая является задачей оптимизации гладкой функции в многомерном кубе, а для решения подобных задач существуют эффективные методы оптимизации. Одна итерация оптимизационного метода имеет временную трудоемкость $O(N^2)$ в слу-

чае $(2p + 1)$ -диагональной матрицы Σ^{-1} , но может быть сокращена до $O(N \log N)$ и $O(N)$ в случае диагональной и единичной матриц соответственно.

Литература

1. Golyandina N., Nekrutkin V., Zhigljavsky A. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall/CRC (2001).
2. Golyandina N., Zhigljavsky A. *Singular spectrum analysis for time series*. 2nd ed. Berlin, Heidelberg, Springer (2020).
3. Van Der Veen A.-J., Deprettere E. F., Swindlehurst A. L. Subspace-based signal analysis using singular value decomposition. *Proceedings of the IEEE* **81** (9), 1277–1308 (1993).
4. Gonen E., Mendel J.M. Subspace-based direction finding methods. In: Madisetti V.K., Williams D.B. (eds.) *The Digital Signal Processing Handbook*, vol. 62. Boca Raton, CRC Press LLC (1999).
5. Heinig G., Rost K. *Algebraic Methods for Toeplitz-like Matrices and Operators*. In: Operator Theory: Advances and Applications. Birkhäuser, Verlag (1985).
6. Adachi K. *Matrix-based introduction to multivariate data analysis*. Springer (2016).
7. Allen G. I., Grosenick L., Taylor J. A generalized least-square matrix decomposition. *Journal of the American Statistical Association* **109** (505), 145–159 (2014).
8. Golyandina N., Zhigljavsky A. Blind deconvolution of covariance matrix inverses for autoregressive processes. *Linear Algebra and its Applications* **593**, 188–211 (2020).
9. Zvonarev N., Golyandina N. Iterative algorithms for weighted and unweighted finite-rank time-series approximations. *Statistics and Its Interface* **10** (1), 5–18 (2017).
10. Звонарев Н. К. Поиск весов в задаче взвешенной аппроксимации временным рядом конечного ранга. *Вестник Санкт-Петербургского университета. Серия 1. Математика. Механика. Астрономия* **3** (61), вып. 4, 570–581 (2016). <https://doi.org/10.21638/11701/spbu01.2016.406>
11. Wiesel A., Globerson A. Covariance estimation in time varying ARMA processes. *IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM) / IEEE*, 357–360 (2012).
12. Verbyla A. A note on the inverse covariance matrix of the autoregressive process. *Australian & New Zealand Journal of Statistics* **27** (2), 221–224 (1985).
13. Cadzow J. A. Signal enhancement: a composite property mapping algorithm. *IEEE Trans. Acoust.* **36** (1), 49–62 (1988).
14. Byrd R. H., Lu P., Nocedal J., Zhu C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* **16** (5), 1190–1208 (1995).
15. Kullback S., Leibler R. A. On information and sufficiency. *The annals of mathematical statistics* **22** (1), 79–86 (1951).
16. Duchi J. *Derivations for linear algebra and optimization* (2007). Доступно на: https://web.stanford.edu/~jduchi/projects/general_notes.pdf (дата обращения: 01.04.2021).
17. Golub G. H., Van Loan C. F. *Matrix computations*. 3rd ed. Baltimore, MD, USA, Johns Hopkins University Press (1996).
18. Gillard J. Cadzow's basic algorithm, alternating projections and singular spectrum analysis. *Stat. Interface* **3** (3), 335–343 (2010).

Статья поступила в редакцию 11 декабря 2020 г.;
после доработки 11 января 2021 г.;
рекомендована в печать 19 марта 2021 г.

Контактная информация:

Звонарев Никита Константинович — канд. физ.-мат. наук; nikitazvonaev@gmail.com

Search of weights in the problem of finite-rank signal estimation in presence of noise*

N. K. Zvonarev

St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Zvonarev N. K. Search of weights in the problem of finite-rank signal estimation in presence of noise. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2021, vol. 8 (66), issue 3, pp. 417–429. <https://doi.org/10.21638/spbu01.2021.304> (In Russian)

The problem of weighted finite-rank time-series approximation is considered for signal estimation in “signal plus noise” model, where the inverse covariance matrix of noise is $(2p + 1)$ -diagonal. Finding of weights, which improve the estimation accuracy, is examined. An effective method for the numerical search of the weights is constructed and proved. Numerical simulations are performed to study the improvement of the estimation accuracy for several noise models.

Keywords: time series, Cadzow’s iterative method, time series of finite rank, weighted low-rank approximation, Kullback – Leibler divergence, optimization methods.

References

1. Golyandina N., Nekrutkin V., Zhigljavsky A. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall/CRC (2001).
2. Golyandina N., Zhigljavsky A. *Singular spectrum analysis for time series*. 2nd ed. Berlin, Heidelberg, Springer (2020).
3. Van Der Veen A.-J., Deprettere E. F., Swindlehurst A. L. Subspace-based signal analysis using singular value decomposition. *Proceedings of the IEEE* **81** (9), 1277–1308 (1993).
4. Gonen E., Mendel J.M. Subspace-based direction finding methods. In: Madisetti V.K., Williams D.B. (eds.) *The Digital Signal Processing Handbook*, vol. 62. Boca Raton, CRC Press LLC (1999).
5. Heinig G., Rost K. *Algebraic Methods for Toeplitz-like Matrices and Operators*. In: Operator Theory: Advances and Applications. Birkhäuser, Verlag (1985).
6. Adachi K. *Matrix-based introduction to multivariate data analysis*. Springer (2016).
7. Allen G. I., Grosenick L., Taylor J. A generalized least-square matrix decomposition. *Journal of the American Statistical Association* **109** (505), 145–159 (2014).
8. Golyandina N., Zhigljavsky A. Blind deconvolution of covariance matrix inverses for autoregressive processes. *Linear Algebra and its Applications* **593**, 188–211 (2020).
9. Zvonarev N., Golyandina N. Iterative algorithms for weighted and unweighted finite-rank time-series approximations. *Statistics and Its Interface* **10** (1), 5–18 (2017).
10. Zvonarev N. K. Search of weights in problem of weighted finite-rank time-series approximation. *Vestnik of Saint Petersburg University. Series 1. Mathematics. Mechanics. Astronomy* **3** (61), iss. 4, 570–581 (2016). <https://doi.org/10.21638/11701/spbu01.2016.406> (In Russian)
11. Wiesel A., Globerson A. Covariance estimation in time varying ARMA processes. *IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM) / IEEE*, 357–360 (2012).
12. Verbyla A. A note on the inverse covariance matrix of the autoregressive process. *Australian & New Zealand Journal of Statistics* **27** (2), 221–224 (1985).
13. Cadzow J. A. Signal enhancement: a composite property mapping algorithm. *IEEE Trans. Acoust.* **36** (1), 49–62 (1988).
14. Byrd R. H., Lu P., Nocedal J., Zhu C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* **16** (5), 1190–1208 (1995).
15. Kullback S., Leibler R. A. On information and sufficiency. *The annals of mathematical statistics* **22** (1), 79–86 (1951).
16. Duchi J. *Derivations for linear algebra and optimization* (2007). Available at: https://web.stanford.edu/~jduchi/projects/general_notes.pdf (accessed: April 1, 2021).

*The work is supported by Russian Foundation for Basic Research (grant No. 20-01-00067).

17. Golub G. H., Van Loan C. F. *Matrix computations*. 3rd ed. Baltimore, MD, USA, Johns Hopkins University Press (1996).

18. Gillard J. Cadzow's basic algorithm, alternating projections and singular spectrum analysis. *Stat. Interface* **3** (3), 335–343 (2010).

Received: December 11, 2020

Revised: January 11, 2021

Accepted: March 19, 2021

Author's information:

Nikita K. Zvonarev — nikitazvonarev@gmail.com