Универсальный блочный метод Ланцоша—Паде для систем линейных уравнений над большими простыми полями

М. А. ЧЕРЕПНЁВ
Московский государственный университет им. М. В. Ломоносова
e-mail: cherepniov@gmail.com

Н. Л. ЗАМАРАШКИН
Институт вычислительной математики РАН
e-mail: nikolai.zamarashkin@gmail.com

УДК 519.61

Ключевые слова: блочный метод Ланцоша, матричные аппроксимации Паде, конечные поля, пространства Крылова, параллельные вычисления.

Аннотация

В работе предложен универсальный алгоритм, предназначенный для решения больших разреженных систем линейных уравнений над конечными полями с большим простым числом элементов. Такие системы возникают при решении задачи дискретного логарифмирования по модулю простого числа. Алгоритм разработан для использования на параллельных вычислительных системах с разнообразной параллельной архитектурой и характеристиками. Новый метод наследует структурные свойства метода Ланцоша, позволяя, однако, гибко управлять сложностью параллельных вычислений и количеством обменов.

Abstract

M. A. Cherepniov, N. L. Zamarashkin, The universal block Lanczos–Padé method for linear systems over large prime fields, Fundamentalnaya i prikladnaya matematika, vol. 19 (2014), no. 6, pp. 225—249.

In this paper, we propose a universal algorithm designed for solving large sparse linear systems over finite fields with a large prime number of elements. Such systems arise in the solution of the discrete logarithm problem modulo a prime number. The algorithm has been developed for parallel computing systems with various parallel architectures and properties. The new method inherits the structural properties of the Lanczos method however providing flexible control over the complexity of parallel computations and the intensity of exchanges.

1. Введение

В настоящее время среди методов, которые используются для решения больших разреженных систем над конечными полями, одним из наиболее часто

Фундаментальная и прикладная математика, 2014, том 19, № 6, с. 225—249.

упоминаемых является метод Ланцоша (см., например, [1—5, 7—9, 11, 12]). Простота алгоритма — главное достоинство метода. Однако по своей природе метод Ланцоша является последовательным алгоритмом, на каждой итерации которого происходит большой обмен данными между различными вычислительными узлами.

Несмотря на то что использование блочных версий алгоритма Ланцоша несколько улучшает масштабируемость параллельных программ (см., например, [3]), оно так или иначе не позволяет эффективно применять вычислительные системы с производительностью более 100 терафлопс, причём недостатки всех программных реализаций метода Ланцоша становятся тем более заметными, чем ниже скорость обмена данными в коммуникационной сети вычислительной системы.

В настоящей работе предлагается подход к построению *универсального* блочного метода Ланцоша—Паде для решения больших разреженных систем линейных уравнений над конечными полями с большим (на практике порядка 10^{150}) числом элементов. Термин «универсальный» используется в работе как замена выражению «эффективный на параллельных вычислительных системах с различной архитектурой и характеристиками». В предлагаемом подходе имеется возможность гибко управлять сложностью вычислений и размером обмениваемых данных. При построении универсального метода Ланцоша—Паде мы следуем [5, 6, 10].

2. A-ортогональный базис пространства Крылова и метод Ланцоша

Пусть $A \in \mathbb{F}^{n\times n}$ — симметричная матрица, а $B \in \mathbb{F}^{n\times K}$ — блок, составленный из K векторов с элементами из «большого» конечного простого поля $\mathbb{F}$. Рассмотрим систему линейных уравнений с матрицей A и правой частью B

$$AX = B. \tag{1}$$

Для решения системы (1) можно использовать блочный метод Ланцоша, $(k+1)$-я итерация которого состоит из следующих операций:

1) вычислить произведение AQ_k матрицы A на блок $Q_k \in \mathbb{F}^{n\times K}$;
2) построить новый блок направлений $Q_{k+1} \in \mathbb{F}^{n\times K}$ при помощи коротких рекуррентных соотношений

$$Q_{k+1} = AQ_k - Q_kC_k - Q_{k-1}C_{k-1} \tag{2}$$

с матрицами коэффициентов C_k и C_{k-1} размера $K\times K$, вычисляемыми по формулам

$$C_k = (Q_k^{\mathrm{T}}AQ_k)^{-1}Q_k^{\mathrm{T}}A^2Q_k, \quad C_{k-1} = (Q_{k-1}^{\mathrm{T}}AQ_{k-1})^{-1}Q_{k-1}^{\mathrm{T}}A^2Q_k.$$

Для блоков направлений Q_s справедливы соотношения A-ортогональности (т. е. $Q_l^{\mathrm{T}} A Q_r = 0$ для $l \neq r$) [11]. Кроме того, в дальнейшем будем предполагать, что все матрицы вида $Q_s^{\mathrm{T}} A Q_s$, $s \in \{1, \ldots, n/K - 1\}$ являются невырожденными. Это справедливо с вероятностью, близкой к 1, в предположении, что матрицы $Q_s^{\mathrm{T}} A Q_s$ являются независимыми случайными симметричными матрицами с элементами из поля $\mathbb{F}$. Справедливость такого предположения для задачи дискретного логарифмирования вытекает из того, что размер исходной матрицы существенно меньше характеристики поля, и подтверждается многочисленными численными экспериментами (см., например, [1, 2]);

3) вычислить новое «приближённое» решение X_{k+1} в виде

$$X_{k+1} = \sum_{i=1}^{k+1} Q_i (Q_i^{\mathrm{T}} A Q_i)^{-1} Q_i^{\mathrm{T}} B = X_k + Q_{k+1} (Q_{k+1}^{\mathrm{T}} A Q_{k+1})^{-1} Q_{k+1}^{\mathrm{T}} B.$$

Из сделанного предположения, что матрицы $Q_s^{\mathrm{T}} A Q_s$ невырожденные, следует, что каждый блок направлений Q_s имеет полный ранг K (т. е. что столбцы, образующие Q_s, линейно независимы), а линейная оболочка столбцов, из которых составлены блоки $Q_1, Q_2, \ldots, Q_s$, совпадает с линейной оболочкой столбцов, входящих в блоки $B, AB, \ldots, A^{s-1}B$.

В случае конечных полей метод Ланцоша является итерационной процедурой лишь условно. Для того чтобы найти решение задачи (1), требуется построить пространство Крылова (блочное пространство Крылова) полной размерности. Таким образом, метод Ланцоша — это существенно последовательная процедура построения A-ортогонального базиса $Q_1, Q_2, \ldots, Q_{n/K}$ пространства Крылова $\mathcal{K}(A, B)$,

$$\mathcal{K}(A, B) = \mathrm{Span}\{B, AB, A^2B, \ldots, A^{n/K-1}B\},$$

и одновременного вычисления коэффициентов разложения решения X в построенном базисе. Каждый новый блок направлений Q_{s+1} может быть вычислен только при условии, что известны предыдущие блоки $Q_1, Q_2, \ldots, Q_s$.

Последовательная структура алгоритма Ланцоша подразумевает, что ресурс независимых вычислений, необходимый при реализации метода на параллельных вычислительных системах, весьма ограничен. Более того, явное (непосредственное) вычисление векторов направлений Q_s, как правило (см., например, [3]), сопровождается существенным обменом данными между различными вычислительными узлами, что увеличивает время работы алгоритма на вычислительных системах с медленными межузловыми связями.

Идея универсального алгоритма состоит в том, чтобы не вычислять блоки Q_s явно, а использовать отрезки степенного базиса пространства Крылова. Относительно большой обмен данными в универсальном алгоритме будет иметь место только в моменты инициализации новых отрезков степенного базиса. Частота таких инициализаций является параметром алгоритма. Именно возможность управлять размером обмениваемых данных является главной особенностью нового алгоритма.

3. A-ортогональный базис пространства Крылова и матричные приближения Паде

Нам потребуется ряд конструкций, чтобы установить связь между матричными приближениями Паде некоторого специального ряда и A-ортогональными базисами пространства Крылова $\mathcal{K}(A, B)$.

Итак, пусть $\alpha(x)$ — формальный матричный ряд

$$\alpha(x) = \sum_{i=0}^{\infty} \alpha_i x^{-i}, \tag{3}$$

коэффициентами которого являются матрицы $\alpha_i \in \mathbb{F}^{K\times K}$. Будем говорить, что пара матричных полиномов $\mathcal{Q}^{(s)}(x)$ и $\mathcal{P}^{(s)}(x)$ степени s

$$\mathcal{Q}^{(s)}(x) = \mathcal{Q}_0^{(s)} + \mathcal{Q}_1^{(s)} x + \ldots + \mathcal{Q}_s^{(s)} x^s,$$
$$\mathcal{P}^{(s)}(x) = \mathcal{P}_0^{(s)} + \mathcal{P}_1^{(s)} x + \ldots + \mathcal{P}_s^{(s)} x^s$$

с коэффициентами $\mathcal{Q}_i^{(s)}, \mathcal{P}_j^{(s)} \in \mathbb{F}^{K\times K}$ задаёт *матричное приближение Паде степени* s для ряда (3), если выполняется соотношение

$$\alpha(x)\mathcal{Q}^{(s)}(x) - \mathcal{P}^{(s)}(x) = \sum_{i=s+1}^{\infty} \rho_i^{(s)} x^{-i}. \tag{4}$$

Ряд в правой части (4) будем называть *остаточным рядом* приближения Паде $\mathcal{P}^{(s)}(x)$, $\mathcal{Q}^{(s)}(x)$ степени s и обозначать через $\mathcal{R}^{(s)}(x)$.

Для ряда (3) и пары произвольных матричных полиномов определим *билинейное отображение*. А именно, пусть $\mathcal{Q}_1(x)$ и $\mathcal{Q}_2(x)$ — два матричных полинома. Рассмотрим ряд, полученный формальным произведением $\mathcal{Q}_1^{\mathrm{T}}(x)$, $\alpha(x)$ и $\mathcal{Q}_2(x)$,

$$\mathcal{Q}_1^{\mathrm{T}}(x)\alpha(x)\mathcal{Q}_2(x) = \sum_{i=-t}^{\infty} \gamma_i x^{-i}. \tag{5}$$

Тогда значение билинейного отображения $\langle \mathcal{Q}_1, \mathcal{Q}_2 \rangle$ для пары полиномов $\mathcal{Q}_1$ и $\mathcal{Q}_2$ определяется как значение коэффициента γ_1 формального ряда (5):

$$\langle \mathcal{Q}_1, \mathcal{Q}_2 \rangle = \gamma_1 \in \mathbb{F}^{K\times K}. \tag{6}$$

Наконец, установим способ построения блока Q размера $n \times K$ по заданному матричному полиному $\mathcal{Q}(x)$, квадратной $(n \times n)$-матрице A и $(n \times K)$-блоку B. Для этого рассмотрим полином $\mathcal{Q}(x)$ степени s с матричными коэффициентами $\mathcal{Q}_0, \mathcal{Q}_1, \ldots, \mathcal{Q}_s$:

$$\mathcal{Q}(x) = \mathcal{Q}_0 + \mathcal{Q}_1 x + \ldots + \mathcal{Q}_s x^s.$$

Определим блок $Q = \mathcal{Q}(A, B) \in \mathbb{F}^{n\times K}$ по следующему правилу (см., например, [5]):

$$Q = \mathcal{Q}(A, B) = B\mathcal{Q}_0 + AB\mathcal{Q}_1 + \ldots + A^s B\mathcal{Q}_s. \tag{7}$$

В дальнейшем нас будут интересовать только те ряды $\alpha(x)$ вида (3), коэффициенты α_i которых задаются некоторым специальным образом, а именно

$$\alpha_i = \alpha_i^{\mathrm{T}} = B^{\mathrm{T}} A^i B$$

с симметричной матрицей A и некоторым блоком B. Таким образом, начиная с этого момента и всюду далее будем считать, что введённый нами ранее ряд $\alpha(x)$ имеет вид

$$\alpha(x) = \sum_{i=0}^{\infty} (B^{\mathrm{T}} A^i B) x^{-i}. \tag{8}$$

Следующая лемма связывает значение билинейного отображения $\langle \mathcal{Q}_1, \mathcal{Q}_2 \rangle$ для матричных полиномов $\mathcal{Q}_1$ и $\mathcal{Q}_2$ с матричным произведением блоков $Q_1 = \mathcal{Q}_1(A, B)$ и $Q_2 = \mathcal{Q}_2(A, B)$, порождённых теми же самыми полиномами.

Лемма 1 [5]. *Справедливо равенство*

$$\langle \mathcal{Q}_1, \mathcal{Q}_2 \rangle = Q_1^{\mathrm{T}} A Q_2. \tag{9}$$

С помощью леммы 1 докажем утверждение, которое связывает матричные приближения Паде различных степеней с некоторым A-ортогональным базисом пространства $\mathcal{K}(A, B)$.

Утверждение 1. *Пусть $A \in \mathbb{F}^{n \times n}$ — симметричная матрица, $B \in \mathbb{F}^{n \times K}$ — блок из K линейно независимых векторов, $\alpha_i \in \mathbb{F}^{K \times K}$ — квадратные матрицы вида*

$$\alpha_i = \alpha_i^{\mathrm{T}} = B^{\mathrm{T}} A^i B.$$

Рассмотрим приближения Паде $\mathcal{P}^{(s)}(x)$, $\mathcal{Q}^{(s)}(x)$ ряда (8) *степеней $0, \ldots, l$:*

$$\begin{aligned}
\alpha(x)\mathcal{Q}^{(0)}(x) - \mathcal{P}^{(0)}(x) &= \sum_{i=1}^{\infty} \rho_i^{(0)} x^{-i}, \\
\alpha(x)\mathcal{Q}^{(1)}(x) - \mathcal{P}^{(1)}(x) &= \sum_{i=2}^{\infty} \rho_i^{(1)} x^{-i}, \\
&\ldots \\
\alpha(x)\mathcal{Q}^{(l)}(x) - \mathcal{P}^{(l)}(x) &= \sum_{i=l+1}^{\infty} \rho_i^{(l)} x^{-i},
\end{aligned} \tag{10}$$

предполагая, что матрицы $Q^{(s)}(A, B)^{\mathrm{T}} A Q^{(s)}(A, B)$ являются невырожденными. Тогда для любого s, $0 \leqslant s \leqslant l$, блоки $Q_0 = \mathcal{Q}^{(0)}(A, B), \ldots, Q_s = \mathcal{Q}^{(s)}(A, B)$, построенные по $\mathcal{Q}$-полиномам приближений Паде, задают A-ортогональный базис пространства Крылова $\mathcal{K}_l(A, B)$:

$$\mathcal{K}_l(A, B) = \mathrm{Span}(B, AB, \ldots, A^l B).$$

Кроме того,

$$Q^{(s)}(A, B)^{\mathrm{T}} A Q^{(s)}(A, B) = Q_s^{(s)T} \rho_{s+1}^{(s)}$$

и старшие коэффициенты матричных $\mathcal{Q}$-полиномов и старшие коэффициенты соответствующих остаточных рядов, стоящие в последнем равенстве справа, являются невырожденными матрицами.

Доказательство. Рассмотрим два $\mathcal{Q}$-полинома $\mathcal{Q}^{(m)}(x)$ и $\mathcal{Q}^{(k)}(x)$ степени m и k соответственно, причём $m \neq k$. Не теряя общности, считаем, что $m > k$. Рассмотрим остаточный ряд $\mathcal{R}^{(m)}(x)$ для приближения $\mathcal{Q}^{(m)}(x)$, $\mathcal{P}^{(m)}(x)$ степени m:

$$\alpha(x)Q^{(m)}(x) - P^{(m)}(x) = \sum_{i=m+1}^{\infty} \rho_i^{(m)} x^{-i}. \tag{11}$$

Старший ненулевой член ряда (11) имеет степень не выше $-(m+1)$, следовательно, старший член произведения $\mathcal{R}^{(m)}(x)$ на матричный полином $\left(\mathcal{Q}^{(k)}(x)\right)^{\mathrm{T}}$ будет иметь степень не выше $k-m-1$. Принимая во внимание, что $k < m$, а следовательно, $k-m-1 < -1$, заключаем, что

$$\langle \mathcal{Q}^{(m)}, \mathcal{Q}^{(k)} \rangle = 0.$$

Теперь непосредственно из леммы 1 вытекает A-ортогональность блоков $\mathcal{Q}^{(m)}(A,B)$ и $\mathcal{Q}^{(k)}(A,B)$. Осталось заметить, что размерности векторных пространств, образованных векторами в блоках $\mathcal{Q}^{(0)}(A,B), \ldots, \mathcal{Q}^{(l)}(A,B)$ и $B, AB, \ldots, A^lB$, будут совпадать, если только старшие коэффициенты $\mathcal{Q}$-полиномов приближений Паде являются невырожденными.

Остальное доказывается аналогично. □

Утверждение 1 предоставляет иной («неявный») способ описания A-ортогональных базисов пространства Крылова $\mathcal{K}(A,B)$. А именно, чтобы задать A-ортогональный базис, достаточно знать соответствующие $\mathcal{Q}$-полиномы приближений Паде. Вопросы вычисления $\mathcal{Q}$-полиномов с помощью коротких рекуррентных соотношений рассматриваются в разделе 4.

4. Рекуррентные формулы для приближений Паде

Предположим, что уже известны приближения Паде $\mathcal{P}^{(s-1)}$, $\mathcal{Q}^{(s-1)}$ и $\mathcal{P}^{(s)}$, $\mathcal{Q}^{(s)}$ порядков $s-1$ и s (здесь и везде далее верхний индекс в круглых скобках говорит о том, к какому порядку приближения Паде относится данная величина) соответственно с остаточными рядами $\mathcal{R}^{(s-1)}(x)$ и $\mathcal{R}^{(s)}(x)$:

$$\mathcal{R}^{(s-1)}(x) = \alpha(x)\mathcal{Q}^{(s-1)}(x) - \mathcal{P}^{(s-1)}(x) = \sum_{i=s}^{\infty} \rho_i^{(s-1)} x^{-i},$$

$$\mathcal{R}^{(s)}(x) = \alpha(x)\mathcal{Q}^{(s)}(x) - \mathcal{P}^{(s)}(x) = \sum_{i=s+1}^{\infty} \rho_i^{(s)} x^{-i}.$$

Новое приближение Паде $\mathcal{Q}^{(s+1)}(x)$, $\mathcal{P}^{(s+1)}(x)$ степени $s+1$ будем искать в виде

$$\mathcal{Q}^{(s+1)}(x) = x\mathcal{Q}^{(s)}(x) + \mathcal{Q}^{(s)}(x)\nu_0 + \mathcal{Q}^{(s-1)}(x)\nu_1, \tag{12}$$

$$\mathcal{P}^{(s+1)}(x) = x\mathcal{P}^{(s)}(x) + \mathcal{P}^{(s)}(x)\nu_0 + \mathcal{P}^{(s-1)}(x)\nu_1, \tag{13}$$

$$\mathcal{R}^{(s+1)}(x) = x\mathcal{R}^{(s)}(x) + \mathcal{R}^{(s)}(x)\nu_0 + \mathcal{R}^{(s-1)}(x)\nu_1, \tag{14}$$

где ν_0 и ν_1 — $(K \times K)$-матрицы, требующие определения. Из условия равенства нулю коэффициентов $\rho_s^{(s+1)}$ и $\rho_{s+1}^{(s+1)}$ остатка ряда Паде $\mathcal{R}^{(s+1)}(x)$ следует, что матрицы ν_0 и ν_1 удовлетворяют системе линейных уравнений

$$\rho_{s+1}^{(s)} + \rho_s^{(s-1)}\nu_1 = \rho_s^{(s+1)} = 0, \tag{15}$$

$$\rho_{s+2}^{(s)} + \rho_{s+1}^{(s)}\nu_0 + \rho_{s+1}^{(s-1)}\nu_1 = \rho_{s+1}^{(s+1)} = 0. \tag{16}$$

Оставшиеся коэффициенты $\rho_{s+k+1}^{(s+1)}$ остатка $\mathcal{R}^{(s+1)}(x)$ с $k > 0$ получаются по следующей формуле:

$$\rho_{s+k+1}^{(s+1)} = \rho_{s+k+2}^{(s)} + \rho_{s+k+1}^{(s)}\nu_0 + \rho_{s+k+1}^{(s-1)}\nu_1.$$

Следуя (12), запишем для полинома $\mathcal{Q}^{(s+1)}(x)$

$$\mathcal{Q}^{(s+1)}(x) = \mathcal{Q}^{(s)}(x)(Ix + \nu_0) + \mathcal{Q}^{(s-1)}(x)\nu_1. \tag{17}$$

Применяя данное преобразование t раз, получаем представление для полинома $\mathcal{Q}^{(s+t)}(x)$ в виде

$$\mathcal{Q}^{(s+t)}(x) = \mathcal{Q}^{(s)}(x)\mathcal{H}_{1s}^{(t)}(x) + \mathcal{Q}^{(s-1)}(x)\mathcal{H}_{0s}^{(t)}(x), \tag{18}$$

где $\mathcal{H}_{1s}^{(t)}(x)$ — матричный полином степени t, старший коэффициент которого — единичная $(K \times K)$-матрица, а $\mathcal{H}_{0s}^{(t)}(x)$ — некоторый матричный полином степени не выше $t-1$.

Наконец, несложно проверить, что согласно (12) явное вычисление матричных коэффициентов полинома $\mathcal{Q}^{(s+1)}(x)$ даётся рекуррентной формулой

$$\begin{bmatrix} \mathcal{Q}_{s+1}^{(s+1)} \\ \mathcal{Q}_{s}^{(s+1)} \\ \mathcal{Q}_{s-1}^{(s+1)} \\ \mathcal{Q}_{s-2}^{(s+1)} \\ \dots \\ \mathcal{Q}_{1}^{(s+1)} \\ \mathcal{Q}_{0}^{(s+1)} \end{bmatrix} = \begin{bmatrix} \mathcal{Q}_{s}^{(s)} & 0 & 0 \\ \mathcal{Q}_{s-1}^{(s)} & \mathcal{Q}_{s}^{(s)} & 0 \\ \mathcal{Q}_{s-2}^{(s)} & \mathcal{Q}_{s-1}^{(s)} & \mathcal{Q}_{s-1}^{(s-1)} \\ \mathcal{Q}_{s-3}^{(s)} & \mathcal{Q}_{s-2}^{(s)} & \mathcal{Q}_{s-2}^{(s-1)} \\ \dots & \dots & \dots \\ \mathcal{Q}_{0}^{(s)} & \mathcal{Q}_{1}^{(s)} & \mathcal{Q}_{1}^{(s-1)} \\ 0 & \mathcal{Q}_{0}^{(s)} & \mathcal{Q}_{0}^{(s-1)} \end{bmatrix} \begin{bmatrix} I \\ \nu_0 \\ \nu_1 \end{bmatrix}. \tag{19}$$

Заметим, что, зная коэффициенты матричного $\mathcal{Q}$-полинома $\mathcal{Q}^{(s+1)}(x)$, можно формально построить блок $Q_{s+1} = \mathcal{Q}^{(s+1)}(A, B)$, который в силу доказанного ранее будет A-ортогонален предыдущим $\mathcal{Q}$-блокам $Q_0, \dots, Q_s$. Таким образом, использование коротких рекуррентных соотношений для вычисления Паде приближений матричного ряда даёт альтернативный алгоритм построения A-ортогонального базиса пространства $\mathcal{K}(A, B)$, причём в этом случае блоки $\mathcal{Q}(A, B)$

A-ортогонального базиса задаются неявно, через коэффициенты соответствующих $\mathcal{Q}$-полиномов.

Подход к вычислению A-ортогонального базиса, основанный на приближениях Паде, имеет видимые преимущества с точки зрения параллельных вычислений. В классическом методе Ланцоша самая «тяжёлая» операция алгоритма — это операция умножения матрицы A на блоки Q_i. Однако, как мы увидим, в алгоритме, основанном на вычислении приближений Паде (для краткости будем говорить об алгоритме Ланцоша—Паде), эта операция присутствует лишь при вычислении коэффициентов $\alpha_i = B^{\mathrm{T}}A^iB$ ряда (3). Вычисление $(K \times K)$-матриц α_i несложно реализовать с использованием K вычислительных узлов, т. е. когда количество узлов совпадает с размером блока. Общий обмен данными при таком вычислении не превосходит cnK (где n — размерность системы, K — размер блока, а c — константа, обозначающая количество машинных слов, необходимое для хранения одного элемента поля $\mathbb{F}$), что существенно меньше, чем в случае стандартного метода Ланцоша.

Проблема, однако, в том, что мы ещё не предъявили способа получения решения X системы $AX = B$. В следующем разделе будет показано, как провести оставшиеся вычисления, не вычисляя явным образом A-ортогональный базис пространства $\mathcal{K}(A, B)$. Новый метод мы будем называть *блочным методом Ланцоша—Паде*.

5. Блочный метод Ланцоша—Паде

Будем строить алгоритм типа Ланцоша для решения системы уравнений, не вычисляя A-ортогональных блоков Q явно, работая только с коэффициентами $\mathcal{Q}$-полиномов Паде.

Несложно проверить, что полиномы $\mathcal{Q}^{(0)}(x)$ и $\mathcal{Q}^{(1)}(x)$, взятые в виде $\mathcal{Q}^{(0)}(x) = I$ и $\mathcal{Q}^{(1)}(x) = Ix - \alpha_1^{-1}\alpha_2$ с $\alpha_1 = B^{\mathrm{T}}AB$ и $\alpha_2 = B^{\mathrm{T}}A^2B$, являются $\mathcal{Q}$-полиномами Паде нулевой и первой степени соответственно. Используя частный случай соотношения (18), представим $\mathcal{Q}$-полином Паде степени $s+1$ в виде

$$\mathcal{Q}^{(s+1)}(x) = \mathcal{Q}^{(1)}(x)\mathcal{H}_1^{(s+1)}(x) + \mathcal{Q}^{(0)}(x)\mathcal{H}_0^{(s+1)}(x) \tag{20}$$

с матричным полиномом $\mathcal{H}_1^{(s+1)}(x)$ степени s и старшим коэффициентом, равным единичной матрице, и некоторым матричным полином $\mathcal{H}_0^{(s+1)}(x)$ степени не выше $s-1$. Из (17) и (20) следует, что $\mathcal{H}_1^{(s+1)}(x)$ и $\mathcal{H}_0^{(s+1)}(x)$ удовлетворяют одному и тому же рекуррентному соотношению

$$\mathcal{H}_1^{(s+1)}(x) = \mathcal{H}_1^{(s)}(x)(Ix + \nu_0) + \mathcal{H}_1^{(s-1)}(x)\nu_1, \tag{21}$$

$$\mathcal{H}_0^{(s+1)}(x) = \mathcal{H}_0^{(s)}(x)(Ix + \nu_0) + \mathcal{H}_0^{(s-1)}(x)\nu_1, \tag{22}$$

отличаясь только выбором начальных условий. А именно, начальные полиномы выбираются в следующем виде: $\mathcal{H}_1^{(0)}(x) = 0$, $\mathcal{H}_1^{(1)}(x) = I$, $\mathcal{H}_0^{(0)}(x) = I$, $\mathcal{H}_0^{(1)}(x) = 0$.

5.1. Матричные обозначения

Для удобства дальнейшей записи алгоритмов определим специальные матричные обозначения. Составим из коэффициентов полиномов $\mathcal{H}_1^{(s)}(x)$ и $\mathcal{H}_0^{(s)}(x)$ блоки $H_1^{(s)}$ и $H_0^{(s)}$ размера $n \times K$, расположив коэффициенты меньших степеней в строках с большими номерами:

$$H_1^{(s)} = \begin{bmatrix} 0 \\ \dots \\ 0 \\ \mathcal{H}^{(s)}_{1(s-1)} \\ \mathcal{H}^{(s)}_{1(s-2)} \\ \dots \\ \mathcal{H}^{(s)}_{11} \\ \mathcal{H}^{(s)}_{10} \end{bmatrix}, \quad H_0^{(s)} = \begin{bmatrix} 0 \\ \dots \\ 0 \\ 0 \\ \mathcal{H}^{(s)}_{0(s-2)} \\ \dots \\ \mathcal{H}^{(s)}_{01} \\ \mathcal{H}^{(s)}_{00} \end{bmatrix}. \tag{23}$$

В таком случае рекуррентные соотношения (21) и (22) эквивалентны матричным равенствам

$$H_i^{(s+1)} = \begin{bmatrix} ZH_i^{(s)} & H_i^{(s)} & H_i^{(s-1)} \end{bmatrix} \begin{bmatrix} I \\ \nu_0 \\ \nu_1 \end{bmatrix}, \quad i = 0, 1, \tag{24}$$

где $Z \in \mathbb{F}^{n \times n}$ — матрица блочного сдвига вида

$$Z = \begin{bmatrix} 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & I_k & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & I_k \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Наряду с блоками $H_0^{(s)}$ и $H_1^{(s)}$ определим блок $R^{(s)}$ размера $2n \times K$, который соответствует остаточному ряду Паде степени s (заметим, что размер блоков R вдвое превосходит размер блоков H). Блок $R^{(s)}$ зададим следующим образом:

$$R^{(s)} = \begin{bmatrix} \star \\ \star \\ \rho^{(s)}_{2n/K-s} \\ \dots \\ \rho^{(s)}_{s+1} \\ 0 \\ \dots \\ 0 \end{bmatrix}, \tag{25}$$

через $\star$ обозначаются те элементы, которые не являются существенными (другими словами, значения этих элементов не будут влиять на вычисления в алгоритме). Число нулевых матриц в $R^{(s)}$ в точности равно $s + 1$, поскольку $R^{(s)}$

соответствует остаточному ряду Паде степени s. По аналогии с формулами (24) запишем

$$R^{(s+1)} = \begin{bmatrix} Z^{\mathrm{T}} R^{(s)} & R^{(s)} & R^{(s-1)} \end{bmatrix} \begin{bmatrix} I \\ \nu_0 \\ \nu_1 \end{bmatrix}. \tag{26}$$

Заметим, что в соответствии с определением число «несущественных» элементов в $R^{(s+1)}$ увеличивается на 1 по сравнению с числом «несущественных» элементов в $R^{(s)}$. Таким образом, формула (26) может рассматриваться как равенство только для «существенных» компонент $R^{(s+1)}$.

Обозначим через Q_1 и Q_0 блоки $\mathcal{Q}_1(A,B)$ и $\mathcal{Q}_0(A,B)$ соответственно. По определению $Q_0 = B$, $Q_1 = AB - B\alpha_1^{-1}\alpha_2$. Составим матрицы $K_0, K_1 \in \mathbb{F}^{n\times n}$ из блоков вида $A^j Q_0$ и $A^j Q_1$ с j от 0 до $n/K-1$, записанных в обратном порядке:

$$\begin{aligned} K_0 &= \begin{bmatrix} A^{n/K-1}Q_0 & A^{n/K-2}Q_0 & \dots & AQ_0 & Q_0 \end{bmatrix}, \\ K_1 &= \begin{bmatrix} A^{n/K-1}Q_1 & A^{n/K-2}Q_1 & \dots & AQ_1 & Q_1 \end{bmatrix}. \end{aligned} \tag{27}$$

Из (18) следует, что для блока $Q_{s+1} = \mathcal{Q}^{(s+1)}(A,B)$ справедливо представление

$$Q_{s+1} = \sum_{i=0}^{s} A^i Q_1 \mathcal{H}_{1i}^{(s+1)} + \sum_{i=0}^{s-1} A^i Q_0 \mathcal{H}_{0i}^{(s+1)}, \tag{28}$$

которое удобно записать в виде

$$Q_{s+1} = K_0 H_0^{(s+1)} + K_1 H_1^{(s+1)}. \tag{29}$$

По построению (см. утверждение 1) блоки Q_s образуют A-ортогональный базис всего пространства, следовательно, решение системы $X_{n/K}$ записывается в виде линейной комбинации блоков Q_i:

$$X_{n/K} = \sum_{i=0}^{n/K-1} Q_i (Q_i^{\mathrm{T}} A Q_i)^{-1} Q_i^{\mathrm{T}} B.$$

Вводя обозначение Z_i для матриц,

$$Z_i = (Q_i A Q_i^{\mathrm{T}})^{-1} Q_i^{\mathrm{T}} B,$$

придём к краткой записи для $X_{n/K}$:

$$X_{n/K} = \sum_{i=0}^{n/K-1} Q_i Z_i. \tag{30}$$

Наконец, подставляя в (30) выражение для Q_i из (29), найдём, что

$$\begin{aligned} X_{n/K} = \sum_{i=0}^{n/K-1} Q_i Z_i = \sum_{i=0}^{n/K-1} \bigl(K_0 H_0^{(i)} + K_1 H_1^{(i)}\bigr) Z_i = \\ = K_0 \Biggl(\sum_{i=0}^{n/K-1} H_0^{(i)} Z_i \Biggr) + K_1 \Biggl(\sum_{i=0}^{n/K-1} H_1^{(i)} Z_i \Biggr). \end{aligned}$$

Чтобы ещё немного упростить запись и получить матричное соотношение для $X_{n/K}$, определим систему блоков $G_0^{(s)}$ и $G_1^{(s)}$ размера $n \times K$

$$G_0^{(s)} = \sum_{i=0}^{s} H_0^{(i)} Z_i = G_0^{(s-1)} + H_0^{(s)} Z_s, \tag{31}$$

$$G_1^{(s)} = \sum_{i=0}^{s} H_1^{(i)} Z_i = G_1^{(s-1)} + H_1^{(s)} Z_s, \tag{32}$$

где s меняется в пределах от 1 до $n/K-1$. Используя (31) и (32), окончательно запишем $X_{n/K}$ как

$$X_{n/K} = K_0 G_0^{n/K-1} + K_1 G_1^{n/K-1}. \tag{33}$$

Преимущество полученной записи для $X_{n/K}$ состоит в том, что в ней нет явной зависимости от блоков Q_i, а имеется зависимость лишь от систем K_0 и K_1 блочно-крыловского типа. Вычисление K_0 и K_1 легко производить на параллельных вычислительных системах. Для этого каждый из $2K$ векторов, входящих в блок Q_0 или Q_1, можно умножать на матрицу A независимо, тем самым достигая максимально возможного эффекта. Чтобы завершить описание метода, необходимо предъявить способы вычисления блоков $G_0^{(n/K)}$ и $G_1^{(n/K)}$, которые не использовали бы явный вид блоков Q_i.

5.2. Вычисление $Q_i^{\mathrm{T}} B$

Используя для блока Q_i выражение

$$Q_i = K_0 H_0^{(i)} + K_1 H_1^{(i)},$$

получаем, что

$$Q_i^{\mathrm{T}} B = \left(K_0 H_0^{(i)} + K_1 H_1^{(i)}\right)^{\mathrm{T}} B = \left(H_0^{(i)}\right)^{\mathrm{T}} K_0^{\mathrm{T}} B + \left(H_1^{(i)}\right)^{\mathrm{T}} K_1^{\mathrm{T}} B.$$

Рассмотрим матрицы $K_0^{\mathrm{T}} B$ и $K_1^{\mathrm{T}} B$. Из (27) следует, что $K_0^{\mathrm{T}} B$ является блоком (т. е. матрицей размера $n \times K$), составленным из $(K \times K)$-матриц $\alpha_i = B^{\mathrm{T}} A^i B$, а именно

$$K_0^{\mathrm{T}} B = \begin{bmatrix} \alpha_{n/K} \\ \cdots \\ \alpha_1 \\ \alpha_0 \end{bmatrix}.$$

Если дополнительно определить матрицы $\beta_i = Q_1^{\mathrm{T}} A^i B$, то для блока $K_1^{\mathrm{T}} B$ будет справедливо аналогичное равенство

$$K_1^{\mathrm{T}} B = \begin{bmatrix} \beta_{n/K} \\ \cdots \\ \beta_1 \\ \beta_0 \end{bmatrix}.$$

Заметим, что матрицы α_i и β_i можно вычислять независимо, не изменяя тем самым параллельной сложности алгоритма. Окончательно для $Q_i^{\mathrm{T}}B$ получаем следующее выражение:

$$Q_i^{\mathrm{T}}B = \sum_{j=0}^{i-2}\bigl(\mathcal{H}_{0j}^{(i)}\bigr)^{\mathrm{T}}\alpha_j + \sum_{j=0}^{i-1}\bigl(\mathcal{H}_{1j}^{(i)}\bigr)^{\mathrm{T}}\beta_j. \tag{34}$$

Сложность такого вычисления, как нетрудно проверить, оценивается величиной $\mathcal{O}\bigl((2i-1)K^3\bigr)$. Следовательно, сложность вычисления всех матриц $Q_i^{\mathrm{T}}B$ оценивается как

$$\{\text{сложность вычисления всех } Q_i^{\mathrm{T}}B\} = \sum_{i=0}^{n/K}\Bigl(\mathcal{O}\bigl((2i-1)K^3\bigr)\Bigr) \approx \mathcal{O}(n^2K).$$

С другой стороны, матрицы $Q_i^{\mathrm{T}}B$ можно вычислять, используя рекуррентные соотношения (17):

$$\mathcal{Q}^{(i)}(x) = \mathcal{Q}^{(i-1)}(x)(Ix+\nu_0) + \mathcal{Q}^{(i-2)}(x)\nu_1.$$

Для блоков Q_i, Q_{i-1} и Q_{i-2} и последнее равенство эквивалентно равенству

$$Q_i = AQ_{i-1} + Q_{i-1}\nu_0 + \tilde{Q}_{i-2}\nu_1.$$

Тогда

$$Q_i^{\mathrm{T}}B = Q_{i-1}^{\mathrm{T}}AB + \nu_0^{\mathrm{T}}Q_{i-1}^{\mathrm{T}}B + \nu_1^{\mathrm{T}}Q_{i-2}^{\mathrm{T}}B. \tag{35}$$

В силу свойства A-ортогональности для $i > 2$ первое слагаемое в (35) всегда равно нулю. Окончательно получаем, что

$$Q_i^{\mathrm{T}}B = \sum_{j=0}^{1}\nu_j^{\mathrm{T}}(Q_{i-j-1}^{\mathrm{T}}B). \tag{36}$$

Сложность (36) не выше $\mathcal{O}(K^3)$, следовательно, сложность вычисления всех $Q_i^{\mathrm{T}}B$ не выше $\mathcal{O}(nK^2)$.

5.3. Вычисление $Q_i^{\mathrm{T}}AQ_i$

Чтобы вычислить $Q_i^{\mathrm{T}}AQ_i$, воспользуемся утверждением 1:

$$Q_i^{\mathrm{T}}AQ_i = \bigl(\mathcal{Q}_i^{(i)}\bigr)^{\mathrm{T}}\rho_{i+1}^{(i)}. \tag{37}$$

Поскольку старшие коэффициенты $\mathcal{Q}_i^{(i)}$ у $\mathcal{Q}$-полиномов не изменяются с номером i, то сложность вычисления матриц $Q_i^{\mathrm{T}}AQ_i$ не превосходит

$$\{\text{сложность вычисления всех } Q_i^{\mathrm{T}}AQ_i\} = \sum_{i=0}^{n/K-1}\bigl(\mathcal{O}(K^3)\bigr) = \mathcal{O}(nK^2).$$

Более того, можно считать, что коэффициенты $\mathcal{Q}_i^{(i)}$ тождественно равны I, и в таком случае вычисление $Q_i^{\mathrm{T}}AQ_i$ не требует дополнительных затрат.

5.4. Вычисление Z_i

Поскольку

$$Z_i = (Q_i^{\mathrm{T}} A Q_i)^{-1} Q_i^{\mathrm{T}} B,$$

то сложность нахождения всех матриц Z_i складывается из сложности вычисления всех $(K \times K)$-матриц $Q_i^{\mathrm{T}} B$, матриц $Q_i^{\mathrm{T}} A Q_i$, а также из сложности вычисления матриц, обратных к $Q_i^{\mathrm{T}} A Q_i$. Суммируя всё сказанное выше, получаем, что

$$\{\text{сложность вычисления всех } Z_i\} = \mathcal{O}(n^2 K) + \mathcal{O}(nK^2), \tag{38}$$

где второе слагаемое отвечает за операцию обращения уже вычисленных матриц $Q_i^{\mathrm{T}} A Q_i$ и операцию произведения матриц $(Q_i^{\mathrm{T}} A Q_i)^{-1}$ и $Q_i^{\mathrm{T}} B$.

5.5. Алгоритм Ланцоша—Паде

Выпишем алгоритм Ланцоша—Паде.

Шаг 0:

а) вычислить $\alpha_0 = B^{\mathrm{T}} B$, $\alpha_1 B^{\mathrm{T}} A B$, $\alpha_2 = B^{\mathrm{T}} A^2 B$;

б) определить $(n \times K)$-блоки Q_0 и Q_1:

$$Q_0 = B, \quad Q_1 = AB - B\alpha_1^{-1}\alpha_2;$$

в) определить $(n \times K)$-блоки $H_1^{(0)}$ и $H_0^{(0)}$:

$$H_1^{(0)} = H_0^{(1)} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ 0 \end{bmatrix}, \quad H_0^{(0)} = H_1^{(1)} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ I_k \end{bmatrix}.$$

Шаг 1:

а) построить расширенный набор векторов Крылова $Q_0, AQ_0, \ldots, A^{2n/K} Q_0$;

б) вычислить $\alpha_i = Q_0^{\mathrm{T}} A^i Q_0$ для всех i от 0 до $2n/K$;

в) присвоить элементам блоков $R^{(0)}$ и $R^{(1)}$ значения в соответствии со следующими формулами:

$$R^{(0)} = \begin{bmatrix} \alpha_{2n/K} \\ \alpha_{2n/K-1} \\ \cdots \\ \alpha_2 \\ \alpha_1 \\ 0 \end{bmatrix}, \quad R^{(1)} = \begin{bmatrix} 0 \\ \alpha_{2n/K} - \alpha_{2n/K-1}\alpha_1^{-1}\alpha_2 \\ \cdots \\ \alpha_3 - \alpha_2\alpha_1^{-1}\alpha_2 \\ 0 \\ 0 \end{bmatrix}.$$

Шаг 2: для всех s от 2 до $n/K + 1$ повторять а)—д):

а) найти $(K \times K)$-матрицы ν_0, ν_1, решая систему уравнений (15), (16);

б) вычислить $H_0^{(s)}$ и $H_1^{(s)}$ по формуле (24);

в) вычислить остаток ряда Паде $R^{(s)}$ для приближения Паде степени s по формуле (26);

г) вычислить $Z_s = (Q_s^{\mathrm{T}} A Q_s)^{-1} Q_s^{\mathrm{T}} B$:

iv) вычислить $Q_s^{\mathrm{T}} B$ по формуле (34) или по формуле (36);

iv) вычислить $Q_s^{\mathrm{T}} A Q_s$ по формуле (37);

д) вычислить $G_0^{(s)}$ и $G_1^{(s)}$ по рекуррентным формулам:

$$G_0^{(s)} = G_0^{(s-1)} + H_0^{(s)} Z_s, \quad G_1^{(s)} = G_1^{(s-1)} + H_1^{(s)} Z_s.$$

Шаг 3: вычислить $X_{n/K}$ по формуле (33).

5.6. Алгоритмическая и параллельная сложность метода Ланцоша—Паде

Вычислительная сложность алгоритма Ланцоша—Паде складывается из сложности нескольких групп вычислений. Первая такая группа вычислений (шаг 1 алгоритма Ланцоша—Паде) связана с вычислениями квадратных $(K \times K)$-матриц α_i и β_i, где индекс i изменяется от 0 до $2n/K$. Прямой путь к получению этих матриц состоит в том, чтобы сначала последовательно вычислить блоки $Q_0, AQ_0, \ldots, A^{2n/K} Q_0$ и блоки and $Q_1, AQ_1, \ldots, A^{2n/K} Q_1$, затратив $\mathcal{O}(4pn^2)$ операций (p — оценка сверху на число ненулей в строках матрицы A), а затем определить сами матрицы, затратив ещё не более $\mathcal{O}(n^2 K)$ операций с элементами поля $\mathbb{F}$.

Несложно понять, что описанные вычисления удобны для реализации на параллельных вычислительных системах.

Рассмотрим систему с K вычислительными узлами. Будем считать, что перед началом вычислений на каждом узле имеются как матрица системы A, так и блоки Q_0 и Q_1 размера $n \times K$. Пусть вычислительный узел с номером J строит последовательности столбцов

$$(Q_0)_J, A(Q_0)_J, \ldots, A^{2n/K}(Q_0)_J, \quad (Q_1)_J, A(Q_1)_J, \ldots, A^{2n/K}(Q_1)_J$$

размера n. Алгоритмическая сложность такого вычисления не выше $\mathcal{O}\big(2p(n^2/K)\big)$. Рассматривая выражения вида $Q_0^{\mathrm{T}}(A^i(Q_0)_J)$, $Q_0^{\mathrm{T}}(A^i(Q_1)_J)$, $Q_1^{\mathrm{T}}(A^i(Q_1)_J)$, заметим, что в результате очевидных вычислений на J-м узле получаются J-е столбцы матриц α_i и β_i, причём сложность вычислений на каждом узле не превосходит $\mathcal{O}(n^2/K)$. Таким образом, с использованием K вычислительных узлов количество вычислений для шага 1 на каждом узле сократилось в K раз.

Рассмотрим шаг 2. Сложность решения одной системы линейных уравнений относительно матриц ν_0, ν_1 не превосходит $\mathcal{O}(K^3)$ операций в поле $\mathbb{F}$. Следовательно, решение всех таких систем требует $\mathcal{O}(nK^2)$ операций.

Далее, для фиксированного s сложность вычисления блоков $H_0^{(s)}$ и $H_1^{(s)}$ не более $\mathcal{O}(nK^2)$. Следовательно, нахождение всех блоков H_0 и H_1 потребует не

более $\mathcal{O}(n^2K)$ действий с элементами поля $\mathbb{F}$. Аналогичную сложность имеет вычисление последовательности блоков $R^{(s)}$.

Наконец, сложность получения матриц Z_s рассматривалась нами в разделе 5.4, а сложность получения блоков $G^{(s)}$, как нетрудно проверить, оценивается сверху величиной $\mathcal{O}(n^2K)$. Отметим, не обсуждая деталей, что, как и на шаге 1, на шаге 2 все вычисления идеально масштабируются на системах с K независимыми узлами, а параллельная сложность шага 2 в K раз меньше алгоритмической сложности.

В соответствии с формулой (33) шаг 3 является ничем иным, как двукратным произведением плотных $(n\times n)$-матриц K_0 и K_1 на $(n\times K)$-блоки G_0 и G_1. Таким образом, сложность такого вычисления не выше $\mathcal{O}(n^2K)$. Можно показать, что и в этом случае на системе с K узлами вычисления можно производить в K раз быстрее.

Собирая вместе сказанное выше, получаем, что сложность блочного метода Ланцоша—Паде с размером блока K не превосходит

$$\{\text{сложность}\} = \mathcal{O}(pn^2 + n^2K + nK^2), \tag{39}$$

причём существует реализация такого метода на вычислительной системе с K узлами, для которой сложность не выше

$$\{\text{параллельная сложность}\} = \mathcal{O}\left(p\frac{n^2}{K} + n^2 + nK\right). \tag{40}$$

Заметим, что, как следует из (40), параллельная сложность метода Ланцоша—Паде содержит слагаемые, одно из которых (n^2) не зависит от K, а другое (nK) растёт с размером блока K. На практике при малых K эти слагаемые малы по сравнению с первым слагаемым. Однако с ростом K ситуация меняется. Последнее говорит о том, что предлагаемый подход имеет определённые ограничения при его реализации на параллельных вычислительных системах. Однако основное преимущество рассматриваемой схемы вычислений состоит в сравнительно небольшом обмене данными между узлами (поскольку нет явных вычислений с векторами направлений). По этой причине метод Ланцоша—Паде особенно эффективен для вычислительных систем с медленными межузловыми коммуникациями. Однако если вычислительный комплекс обладает высокоскоростной коммуникационной сетью, то организацию вычислений в методе Ланцоша—Паде можно изменить, уменьшив параллельную сложность алгоритма за счёт определённого увеличения количества обменов. В этом состоит идея *универсального метода Ланцоша—Паде*.

6. Универсальный блочный метод Ланцоша—Паде

Относительно высокая алгоритмическая сложность блочного метода Ланцоша—Паде определяется необходимостью проводить вычисления с блоками $H_0^{(s)}$, $H_1^{(s)}$, $R_0^{(s)}$, $R_1^{(s)}$ размера $n \times K$ и $2n \times K$ (см. раздел 5.1). Идея универсального

метода состоит в переходе к более коротким блокам. Для этого весь процесс вычислений разбивается на несколько последовательных шагов.

Итак, фиксируем натуральные t и q, такие что $tq = n/K$. Для произвольного целого $p \leqslant n/K$ запишем его в виде $p = jt+s$, где s — остаток от деления p на t, и следовательно, $s < t$. Принимая во внимание (18), полином $\mathcal{Q}^{(p)}(x)$ степени p будем представлять в виде

$$\mathcal{Q}^{(p)}(x) = \mathcal{Q}^{(jt+s)}(x) = \mathcal{Q}^{(jt+1)}(x)\mathcal{H}_{1j}^{(s)}(x) + \mathcal{Q}^{(jt)}(x)\mathcal{H}_{0j}^{(t)}(x), \tag{41}$$

где, как и следует, степень полинома $\mathcal{H}_{1j}^{(s)}(x)$ не выше $s-1$, а степень полинома $\mathcal{H}_{0j}^{(s)}$ не выше $s-2$. Будем строить вычисления в алгоритме таким образом, чтобы использовать полиномы $\mathcal{H}_{1j}^{(s)}(x)$, $\mathcal{H}_{0j}^{(s)}$. Очевидно, что вычисления с полиномами малых степеней имеют меньшую вычислительную сложность. Это наблюдение и составляет основу универсального метода.

Несложно проверить, что для фиксированного j полиномы $\mathcal{H}_{0j}^{(s)}$ и $\mathcal{H}_{1j}^{(s)}$ удовлетворяют одному и тому же рекуррентному соотношению

$$\mathcal{H}_{1j}^{(s+1)}(x) = \mathcal{H}_{1j}^{(s)}(x)(Ix + \nu_0) + \mathcal{H}_{1j}^{(s-1)}(x)\nu_1, \tag{42}$$

$$\mathcal{H}_{0j}^{(s+1)}(x) = \mathcal{H}_{0j}^{(s)}(x)\,(Ix + \nu_0) + \mathcal{H}_{0j}^{(s-1)}(x)\nu_1, \tag{43}$$

отличие состоит только в выборе начальных условий. А именно, начальные полиномы выбираются в виде

$$\mathcal{H}_{1j}^{(0)}(x) = 0, \quad \mathcal{H}_{1j}^{(1)}(x) = I, \quad \mathcal{H}_{0j}^{(0)}(x) = I, \quad \mathcal{H}_{0j}^{(1)}(x) = 0.$$

Конечно, чтобы пользоваться формулами (42), (43), необходимо ещё предоставить способ нахождения для каждого s матриц ν_0 и ν_1, зависящих от s. Заметим, что для этих целей мы не можем воспользоваться формулами (25), (26). Их использование привело бы к тому, что сложность алгоритма осталась бы неизменной. Для того чтобы избавиться от необходимости работать с большими блоками $R^{(p)}$, нам потребуется следующее вспомогательное утверждение.

Лемма 2. *Пусть $l > t > 0$, и пусть среди всех коэффициентов $\mathcal{Q}$-полинома Паде $\mathcal{Q}^{(l)}(x)$ известны только первые t. Тогда коэффициенты $\rho_{l+1}^{(l)}, \rho_{l+2}^{(l)}, \ldots, \rho_{l+t}^{(l)}$ остаточного ряда Паде $\mathcal{R}^{(l)}(x)$ могут быть определены как решения системы линейных уравнений с нижней треугольной блочной теплицевой матрицей*

$$\begin{bmatrix} \mathcal{Q}_l^{(l)T} & & & & \\ \mathcal{Q}_{l-1}^{(l)T} & \mathcal{Q}_l^{(l)T} & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathcal{Q}_{l-t+2}^{(l)T} & \mathcal{Q}_{l-t+3}^{(l)T} & & \mathcal{Q}_l^{(l)T} & \\ \mathcal{Q}_{l-t+1}^{(l)T} & \mathcal{Q}_{l-t+2}^{(l)T} & & \mathcal{Q}_{l-1}^{(l)T} & \mathcal{Q}_l^{(l)T} \end{bmatrix} \begin{bmatrix} \rho_{l+1}^{(l)} \\ \rho_{l+2}^{(l)} \\ \cdots \\ \rho_{l+t-1}^{(l)} \\ \rho_{l+t}^{(l)} \end{bmatrix} = \begin{bmatrix} Q_l^{\mathrm{T}} A Q_l \\ Q_l^{\mathrm{T}} A^2 Q_l \\ \cdots \\ Q_l^{\mathrm{T}} A^{t-1} Q_l \\ Q_l^{\mathrm{T}} A^t Q_l \end{bmatrix}, \tag{44}$$

где Q_l — блок, соответствующий полиному $\mathcal{Q}^{(l)}(x)$, $Q_l = \mathcal{Q}^{(l)}(A, B)$.

Доказательство. Пусть $\mathcal{Q}^{(l)}(x)$ — $\mathcal{Q}$-полином Паде степени l:

$$\mathcal{Q}^{(l)}(x) = \mathcal{Q}_l^{(l)} x^l + \mathcal{Q}_{l-1}^{(l)} x^{l-1} + \ldots + \mathcal{Q}_1^{(l)} x + \mathcal{Q}_0^{(l)}.$$

По определению (см. раздел 3) значение билинейного отображения $\langle \mathcal{Q}^{(l)}(x), \mathcal{Q}^{(l)}(x) \rangle$ является коэффициентом при x^{-1} формального ряда $\big(\mathcal{Q}^{(l)}(x)\big)^{\mathrm{T}} \alpha(x) \mathcal{Q}^{(l)}(x)$. Так как $\mathcal{Q}^{(l)}$ — это $\mathcal{Q}$-полином Паде, то

$$\begin{aligned}
\langle \mathcal{Q}^{(l)}(x), \mathcal{Q}^{(l)}(x) \rangle &= \operatorname{coeff}_{-1}\Big(\big(\mathcal{Q}^{(l)}(x)\big)^{\mathrm{T}} \alpha(x) \mathcal{Q}^{(l)}(x)\Big) = \\
&= \operatorname{coeff}_{-1}\Big(\big(\mathcal{Q}^{(l)}(x)\big)^{\mathrm{T}} \big(\alpha(x) \mathcal{Q}^{(l)}(x) - \mathcal{P}^{(l)}(x)\big)\Big) = \\
&= \operatorname{coeff}_{-1}\left(\big(\mathcal{Q}^{(l)}(x)\big)^{\mathrm{T}} \Big(\sum_{i=l+1}^{\infty} \rho_i^{(l)} x^{-i}\Big)\right) = \big(\mathcal{Q}_l^{(l)}\big)^{\mathrm{T}} \rho_{l+1}^{(l)}.
\end{aligned}$$

С другой стороны, в соответствии с леммой 1

$$\langle \mathcal{Q}^{(l)}(x), \mathcal{Q}^{(l)}(x) \rangle = Q_l^{\mathrm{T}} A Q_l.$$

Сопоставляя последние равенства, получаем первую строку в уравнении (44). Чтобы получить k-ю строку в уравнении (44), рассмотрим билинейное отображение на полиномах $x^k \mathcal{Q}^{(l)}(x)$ и $\mathcal{Q}^{(l)}(x)$. Используя те же преобразования, получим

$$\begin{aligned}
\big\langle x^k \mathcal{Q}^{(l)}(x), \mathcal{Q}^{(l)}(x) \big\rangle &= \operatorname{coeff}_{-1}\Big(\big(x^k \mathcal{Q}^{(l)}(x)\big)^{\mathrm{T}} \alpha(x) \mathcal{Q}^{(l)}(x)\Big) = \\
&= \operatorname{coeff}_{-1}\left(\big(x^k \mathcal{Q}^{(l)}(x)\big)^{\mathrm{T}} \Big(\sum_{i=l+1}^{\infty} \rho_i^{(l)} x^{-i}\Big)\right) = \\
&= \big(\mathcal{Q}_{l-k}^{(l)}\big)^{\mathrm{T}} \rho_{l+1}^{(l)} + \big(\mathcal{Q}_l^{(l-k+1)}\big)^{\mathrm{T}} \rho_{l+2}^{(l)} + \ldots + \big(\mathcal{Q}_l^{(l)}\big)^{\mathrm{T}} \rho_{l+k}^{(l)}.
\end{aligned}$$

Остаётся только заметить, что в соответствии с леммой 1 из раздела 3

$$\big\langle x^k \mathcal{Q}^{(l)}(x), \mathcal{Q}^{(l)}(x) \big\rangle = Q_l^{\mathrm{T}} A^{k+1} Q_l. \qquad \square$$

С помощью леммы 2 вычисления полиномов $\mathcal{H}_{0j}^{(s)}$ и $\mathcal{H}_{1j}^{(s)}$ будем производить по следующей схеме. Будем предполагать, что для некоторого фиксированного j одновременно известны блоки

$$Q_{jt} = \mathcal{Q}^{(jt)}(A, B), \quad Q_{jt+1} = \mathcal{Q}^{(jt+1)}(A, B),$$

а также $2(t+1)$ старших коэффициентов матричных полиномов $\mathcal{Q}^{(jt)}(x)$ и $\mathcal{Q}^{(jt+1)}(x)$. Поскольку блок Q_{jt} известен явно, то можно найти матрицы $\alpha_{jt,s} = Q_{jt}^{\mathrm{T}} A^s Q_{jt}$, где s изменяется от 0 до $2(t+1)$. Решая систему линейных уравнений (44), находим $2(t+1)$ старших коэффициентов $\rho_{jt+s}^{(jt)}$ остатка ряда Паде приближения Паде степени jt. Используя найденную часть остатка ряда Паде, вычисляем матричные полиномы $\mathcal{H}_{0j}^{(s)}(x)$ и $\mathcal{H}_{1j}^{(s)}(x)$ для всех s от 0 до $t+1$.

6.1. Матричные обозначения

Фиксируем натуральные t и q, такие что $tq = n/K$. Определим квадратные $(n \times n)$-матрицы K_0 и K_1 следующего вида:

$$K_0 = \begin{bmatrix} K_{q-1}^0 & K_{q-2}^0 & \dots & K_0^0 \end{bmatrix}, \quad K_1 = \begin{bmatrix} K_{q-1}^1 & K_{q-2}^1 & \dots & K_0^1 \end{bmatrix}, \tag{45}$$

где $(n \times tK)$-матрицы K_j^0 и K_j^1, $j = 0, \dots, q-1$, являются цепочками из крыловских блоков:

$$K_j^0 = \begin{bmatrix} A^{t-1}Q_{tj} & A^{t-2}Q_{tj} & \dots & AQ_{tj} & Q_{tj} \end{bmatrix}, \tag{46}$$

$$K_j^1 = \begin{bmatrix} A^{t-1}Q_{tj+1} & A^{t-2}Q_{tj+1} & \dots & AQ_{tj+1} & Q_{tj+1} \end{bmatrix}. \tag{47}$$

Несложно понять, что формулы (45), (46) и (47) задают базис пространства Крылова $\mathcal{K}(A, B)$. Блоки K_j^0 и K_j^1 сами являются цепочками из блоков крыловского типа. Можно сказать, что K_0 и K_1 — это кусочно крыловские базисы пространства $\mathcal{K}(A, B)$.

Помимо матриц K_j^0 и K_j^1, определим $\big(n \times (t+1)\big)$-матрицы $\hat{K}_j^0$ и $\hat{K}_j^1$ вида

$$\hat{K}_j^0 = \begin{bmatrix} A^tQ_{tj} & A^{t-1}Q_{tj} & \dots & AQ_{tj} & Q_{tj} \end{bmatrix} \tag{48}$$

$$\hat{K}_j^1 = \begin{bmatrix} A^tQ_{tj+1} & A^{t-1}Q_{tj+1} & \dots & AQ_{tj+1} & Q_{tj+1} \end{bmatrix}. \tag{49}$$

Представим произвольное $p \leqslant n/K$ в виде $p = jt + s$, где s — остаток от деления p на t, $s < t$. Тогда, принимая во внимание (18), получаем, что

$$\mathcal{Q}^{(p)}(x) = \mathcal{Q}^{(jt+s)}(x) = \mathcal{Q}^{(jt+1)}(x)\mathcal{H}_{1j}^{(s)}(x) + \mathcal{Q}^{(jt)}(x)\mathcal{H}_{0j}^{(s)}(x), \tag{50}$$

причём степень полиномов $\mathcal{H}_{1j}^{(s)}(x)$ не выше $s-1$, а степень полинома $\mathcal{H}_{0j}^{(s)}$ не выше $s-2$. Определим $\big((t+1)K \times K\big)$-блоки $H_{0j}^{(s)}$ и $H_{1j}^{(s)}$, элементами которых будут $(K \times K)$-матрицы коэффициентов полиномов $\mathcal{H}_{0j}^{(s)}$ и $\mathcal{H}_{1j}^{(s)}$:

$$H_{1j}^{(s)} = \begin{bmatrix} 0 \\ \dots \\ 0 \\ \mathcal{H}_{1j(s-1)}^{(s)} \\ \mathcal{H}_{1j(s-2)}^{(s)} \\ \dots \\ \mathcal{H}_{1j1}^{(s)} \\ \mathcal{H}_{1j0}^{(s)} \end{bmatrix}, \quad H_{0j}^{(s)} = \begin{bmatrix} 0 \\ \dots \\ 0 \\ 0 \\ \mathcal{H}_{0j(s-2)}^{(s)} \\ \dots \\ \mathcal{H}_{0j1}^{(s)} \\ \mathcal{H}_{00}^{(s)} \end{bmatrix}. \tag{51}$$

В таком случае для блока $Q_p = \mathcal{Q}^{(p)}(A, B)$ справедливо представление

$$Q_p = Q_{jt+s} = \hat{K}_j^0 H_{0j}^{(s)} + \hat{K}_j^1 H_{1j}^{(s)}. \tag{52}$$

Существенным отличием (52) от (29) является размер входящих в них матриц.

Замечание. Размер блоков $H_{0j}^{(s)}$ и $H_{1j}^{(s)}$ был выбран таким образом, чтобы иметь возможность представления блоков $Q_{t(j+1)}$ и $Q_{t(j+1)+1}$ через блоки Q_{tj} и Q_{tj+1}:

$$Q_{t(j+1)} = \hat{K}_j^0 H_{0j}^{(t)} + \hat{K}_j^1 H_{1j}^{(t)},$$
$$Q_{t(j+1)+1} = \hat{K}_j^0 H_{0j}^{(t+1)} + \hat{K}_j^1 H_{1j}^{(t+1)}.$$

Вслед за блоками $H_{0j}^{(s)}$, $H_{1j}^{(s)}$ определим блок $R_j^{(s)}$ размера $2(t+1)K \times K$:

$$R_j^{(s)} = \begin{bmatrix} \star \\ \star \\ \rho_{2(t+1)-s}^{(jt+s)} \\ \dots \\ \rho_{s+1}^{(jt+s)} \\ 0 \\ \dots \\ 0 \end{bmatrix}, \tag{53}$$

где через $\star$ обозначаются элементы, которые не являются существенными (другими словами, значения этих элементов не будут влиять на вычисления в алгоритме). Число нулевых матриц в блоке $R_j^{(s)}$ равно s. Если для некоторого s известны матрицы ν_0 и ν_1, то следствием формул (42) и (43) будут следующие соотношения для блоков $H_{0j}^{(s+1)}$ и $H_{1j}^{(s+1)}$:

$$H_{ij}^{(s+1)} = \begin{bmatrix} ZH_{ij}^{(s)} & H_{ij}^{(s)} & H_{ij}^{(s-1)} \end{bmatrix} \begin{bmatrix} I \\ \nu_0 \\ \nu_1 \end{bmatrix}, \quad i = 0, 1, \tag{54}$$

с матрицей блочного сдвига $Z \in \mathbb{F}^{K(t+1)\times K(t+1)}$ вида

$$Z = \begin{bmatrix} 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & I_k & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & I_k \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

В то же время блок $R_j^{(s+1)}$ определяется соотношением

$$R^{(s+1)} = \begin{bmatrix} Z^{\mathrm{T}} R^{(s)} & R^{(s)} & R^{(s-1)} \end{bmatrix} \begin{bmatrix} I \\ \nu_0 \\ \nu_1 \end{bmatrix}. \tag{55}$$

Заметим, что в соответствии с определением число «несущественных» элементов в $R_j^{(s+1)}$ увеличивается на 1 по сравнению с числом «несущественных» элементов в $R_j^{(s)}$. Таким образом, формула (55) может рассматриваться как равенство только для «существенных» компонент $R_j^{(s+1)}$.

Так как для решения системы $X_{n/K}$ справедливо представление

$$X_{n/K} = \sum_{i=0}^{n/K-1} Q_i Z_i \tag{56}$$

с матрицами $Z_i = (Q_i A Q_i^{\mathrm{T}})^{-1} Q_i^{\mathrm{T}} B$, то, подставив в (56) выражение для Q_i из (52), найдём, что

$$\begin{aligned} X_{n/K} &= \sum_{i=0}^{n/K-1} Q_i Z_i = \sum_{j=0}^{n/(Kt)-1} \sum_{s=0}^{t-1} \left(\hat{K}_j^0 H_{0j}^{(s)} + \hat{K}_j^1 H_{1j}^{(s)} \right) Z_{jt+s} = \\ &= \sum_{j=0}^{n/(Kt)-1} \hat{K}_j^0 \left(\sum_{s=0}^{t-1} H_{0j}^{(s)} Z_{jt+s} \right) + \sum_{j=0}^{n/(Kt)-1} \hat{K}_j^1 \left(\sum_{s=0}^{t-1} H_{1j}^{(s)} Z_{jt+s} \right) = \\ &= \sum_{j=0}^{n/(Kt)-1} \left(\hat{K}_j^0 G_{0j}^{(t-1)} + \hat{K}_j^1 G_{1j}^{(t-1)} \right) = \sum_{j=0}^{n/(Kt)-1} \hat{X}_j, \end{aligned}$$

где $(t+1)K \times K$ блоки $G_{0j}^{(s)}$ и $G_{1j}^{(s)}$ определяются с помощью соотношений

$$G_{0j}^{(s)} = \sum_{i=0}^{s} H_{0j}^{(i)} Z_{jt+s} = G_{0j}^{(s-1)} + H_{0j}^{(s)} Z_{jt+s}, \tag{57}$$

$$G_{1j}^{(s)} = \sum_{i=0}^{s} H_{1j}^{(i)} Z_{jt+s} = G_{1j}^{(s-1)} + H_{1j}^{(s)} Z_{jt+s} \tag{58}$$

с параметром s, меняющимся в пределах от 1 до $t-1$, а $(n \times K)$-блоки $\hat{X}_j$ — это ничто иное, как

$$\hat{X}_j = \hat{K}_j^0 G_{0j}^{(t-1)} + \hat{K}_j^1 G_{1j}^{(t-1)}.$$

Чтобы завершить описание универсального метода Ланцоша—Паде и оценить его сложность (параллельную сложность), требуется предъявить способы вычисления основных величин алгоритма, как это было сделано в случае метода Ланцоша—Паде из раздела 5.

6.2. Вычисление $Q_i^{\mathrm{T}} B$

Для вычисления $Q_i^{\mathrm{T}} B$ достаточно воспользоваться рекуррентной формулой (36) из раздела 5.2. Сложность такого вычисления останется неизменной и будет не выше

$$\{\text{сложность вычисления всех } Q_i^{\mathrm{T}} B\} = \mathcal{O}(nK^2). \tag{59}$$

6.3. Вычисление $Q_i^{\mathrm{T}} A Q_i$

Вычисление $Q_i^{\mathrm{T}} A Q_i$ в случае универсального алгоритма ничем не отличается от вычисления, представленного ранее в разделе 5.3. Следовательно, сохраняется и сложность такого вычисления:

$$\{\text{сложность вычисления всех } Q_i^{\mathrm{T}} A Q_i\} = \mathcal{O}(nK^2).$$

6.4. Вычисление Z_i

Поскольку

$$Z_i = (Q_i^{\mathrm{T}} A Q_i)^{-1} Q_i^{\mathrm{T}} B,$$

то сложность нахождения всех матриц Z_i складывается из сложности вычисления всех $(K \times K)$-матриц $Q_i^{\mathrm{T}} B$, матриц $Q_i^{\mathrm{T}} A Q_i$, а также из сложности вычисления матриц, обратных к $Q_i^{\mathrm{T}} A Q_i$. Таким образом,

$$\{\text{сложность вычисления всех } Z_i\} = \mathcal{O}(nK^2), \tag{60}$$

где второе слагаемое отвечает за операцию обращения уже вычисленных матриц $Q_i^{\mathrm{T}} A Q_i$ и операцию произведения матриц $(Q_i^{\mathrm{T}} A Q_i)^{-1}$ и $Q_i^{\mathrm{T}} B$.

6.5. Описание универсального алгоритма Ланцоша—Паде

Фиксируем положительные целые t и K. Будем считать, что n делится на произведение tK без остатка и найдётся такое целое q, что $q = n/(tK)$. Выпишем *универсальный алгоритм Ланцоша—Паде*.

Шаг 0:

а) положить $j = 0$;

б) вычислить α_0, α_1, α_2;

в) положить

$$Q_0 = B, \quad Q_1 = AB - B\alpha_1^{-1}\alpha_2;$$

г) положить $(t+1) \times K$-блоки $H_{10}^{(0)}$, $H_{00}^{(0)}$, $H_{10}^{(1)}$, $H_{00}^{(1)}$ равными

$$H_{10}^{(0)} = H_{00}^{(1)} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ 0 \end{bmatrix}, \quad H_{00}^{(0)} = H_{10}^{(1)} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ I_k \end{bmatrix};$$

д) определить блоки $Q^{0,0}$, $Q^{1,0}$ размера $2(t+1) \times K$, компоненты которых — старшие $2t+1$ коэффициентов полиномов $\mathcal{Q}^{(0)}(x)$ и $\mathcal{Q}^{(1)}(x)$, в виде

$$Q^{1,0} = \begin{bmatrix} 0 \\ \cdots \\ I_k \\ -\alpha_1^{-1}\alpha_2 \end{bmatrix}, \quad Q^{0,0} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ I_k \end{bmatrix}.$$

Шаг 1:

а) вычислить $\hat{K}_0^0$ и $\hat{K}_0^1$;

б) вычислить $\alpha_i = Q_0^{\mathrm{T}} A^i Q_0$, $\beta_i = Q_1^{\mathrm{T}} A^i Q_0$ для всех i от 0 до $2(t+1)$;

в) присвоить элементам блоков $R_0^{(0)}$ и $R_0^{(1)}$ значения в соответствии со следующими формулами:

$$R_0^{(0)} = \begin{bmatrix} \alpha_{0,2(t+1)} \\ \alpha_{0,2t+1} \\ \cdots \\ \alpha_{0,2} \\ \alpha_{0,1} \\ 0 \end{bmatrix}, \quad R_0^{(1)} = \begin{bmatrix} 0 \\ \alpha_{0,2(t+1)} - \alpha_{0,2t+1}\alpha_{0,1}^{-1}\alpha_{0,2} \\ \cdots \\ \alpha_{0,3} - \alpha_{0,2}\alpha_{0,1}^{-1}\alpha_{0,2} \\ 0 \\ 0 \end{bmatrix}. \qquad (61)$$

Шаг 2: для $s := 2, \ldots, t+1$ повторять следующие операции:

а) найти $(K \times K)$-матрицы ν_0, ν_1, решая систему уравнений (15), (16);

б) вычислить $H_{00}^{(s)}$ и $H_{10}^{(s)}$;

в) вычислить остаток ряда Паде $R_0^{(s)}$ для приближения Паде степени s по формуле (55);

г) вычислить блок $Q^{s,0}$, состоящий из $2(t+1)$ старших коэффициентов полинома $\mathcal{Q}^{(s)}$;

д) вычислить $(K \times K)$-матрицы Z_s, используя соотношение $Z_s = (Q_s^{\mathrm{T}} A Q_s)^{-1} Q_s^{\mathrm{T}} B$:

i) вычислить $Q_s^{\mathrm{T}} B$ по формуле (36);

ii) вычислить $Q_s^{\mathrm{T}} A Q_s$ по формуле (37);

е) вычислить $G_0^{(s)}$ и $G_1^{(s)}$:

$$G_{00}^{(s)} = G_{00}^{(s-1)} + H_{00}^{(s)} Z_s, \quad G_{10}^{(s)} = G_{10}^{(s-1)} + H_{10}^{(s)} Z_s. \qquad (62)$$

Шаг 3: вычислить $\hat{X}_0$ по формуле

$$\hat{X}_0 = \hat{K}_0^0 G_{00}^{(t-1)} + \hat{K}_0^1 G_{10}^{(t-1)}.$$

Шаг 4: Для $j := 1, \ldots, n/(Kt) - 1$ повторять следующие операции:

а) вычислить $(n \times K)$-блоки $Q_{j(t+1)}$ и $Q_{j(t+1)+1}$ по формуле

$$Q_{j(t+1)} = \hat{K}_j^0 H_{0j}^{(t)} + \hat{K}_j^1 H_{1j}^{(t)}, \quad Q_{j(t+1)+1} = \hat{K}_j^0 H_{0j}^{(t+1)} + \hat{K}_j^1 H_{1j}^{(t+1)};$$

б) вычислить $\hat{K}_j^0$ и $\hat{K}_j^1$;

в) вычислить $\alpha_i = Q_{jt}^{\mathrm{T}} A^i Q_{jt}$, $\beta_i = Q_{jt+1}^{\mathrm{T}} A^i Q_{jt}$ для всех i от 0 до $2(t+1)$;

г) вычислить $2(t+1) \times K$ блоки $R_j^{(0)}$ и $R_j^{(1)}$, используя лемму 2 из раздела 6;

д) для $s := 2, \ldots, t+1$ повторять следующие операции:

i) найти $(K \times K)$-матрицы ν_0, ν_1, решая систему уравнений (15), (16);

ii) вычислить $H_{0j}^{(s)}$ и $H_{1j}^{(s)}$ по формуле (54);

iii) вычислить остаток ряда Паде $R_j^{(s)}$ для приближения Паде степени s по формуле (55);

iv) вычислить блок $Q^{s,j}$, состоящий из $2(t+1)$ старших коэффициентов полинома $\mathcal{Q}^{(jt+s)}$;

v) вычислить $(K \times K)$-матрицы Z_{jt+s}, используя соотношение $Z_{jt+s} = (Q_{jt+s}^{\mathrm{T}} A Q_{jt+s})^{-1} Q_{jt+s}^{\mathrm{T}} B$:
   — вычислить $Q_{jt+s}^{\mathrm{T}} B$;
   — вычислить $Q_{jt+s}^{\mathrm{T}} A Q_{jt+s}$ по формуле (37);

vi) вычислить $G_{0j}^{(s)}$ и $G_{1j}^{(s)}$:

$$G_{0j}^{(s)} = G_{0j}^{(s-1)} + H_{0j}^{(s)} Z_{jt+s}, \quad G_{1j}^{(s)} = G_{1j}^{(s-1)} + H_{1j}^{(s)} Z_{jt+s}; \qquad (63)$$

e) вычислить $\hat{X}_j$ по формуле

$$\hat{X}_j = \hat{K}_j^0 G_{0j}^{(t-1)} + \hat{K}_j^1 G_{1j}^{(t-1)}.$$

Шаг 5: вычислить $X_{n/K}$ по формуле

$$X_{n/K} = \hat{X}_0 + \hat{X}_1 + \ldots + \hat{X}_{q-1}.$$

6.6. Сложность универсального метода Ланцоша—Паде

Анализ сложности универсального метода Ланцоша—Паде мало чем отличается от анализа сложности простого метода Ланцоша—Паде. Бóльшая часть вычислений в универсальном методе Ланцоша—Паде является калькой с вычислений в простом методе Ланцоша—Паде. Тем не менее существуют определённые отличия.

Во-первых, вычисления с блоками $H_{0j}^{(s)}$, $H_{1j}^{(s)}$ и $R_{0j}^{(s)}$, $R_{1j}^{(s)}$ становятся проще, так как блоки становятся короче. Действительно, для фиксированного j сложность вычислений с такими блоками для s от 0 до $t+1$ не превосходит $\mathcal{O}(t^2K^3)$, а поскольку сам параметр j изменяется в пределах от 0 до $n/(Kt)$, то сложность общей работы с блоками не превосходит $\mathcal{O}(ntK^2) = \mathcal{O}(n^2K/q)$, что в q раз меньше, чем в случае простого метода Ланцоша—Паде.

Во-вторых, по сравнению с методом Ланцоша—Паде в универсальный алгоритм добавляется специализированная процедура вычисления коэффициентов остатка ряда Паде длины $2(t+1)$ для полиномов вида $\mathcal{Q}^{(jt)}(x)$ с j от 0 до $q-1$. Сложность этой процедуры определяется сложностью решения системы линейных уравнений с нижней треугольной блочно-теплицевой матрицей размера $2(t+1) \times 2(t+1)$ (см. лемму 2). Сложность одного такого вычисления не выше $\mathcal{O}(t^2K^3)$. Всего же таких вычислений $q = n/(Kt)$. Следовательно, общая сложность всех таких вычислений $\mathcal{O}(ntK^2) = \mathcal{O}(n^2K/q)$.

Кроме того, в универсальном алгоритме требуется вычислять $2(t+1)$ старших коэффициентов всех $\mathcal{Q}$-полиномов. Сложность этой группы вычислений, как нетрудно проверить, также оценивается выражением $\mathcal{O}(n^2K/q)$.

Наконец, остальные вычисления и их сложность рассматривались нами в разделах 6.2 и 6.3. Собирая вместе оценки сложности, приходим к утверждению, что для сложности универсального метода справедливо соотношение

$$\{\text{сложность универсального метода}\} = \mathcal{O}\left(pn^2 + \frac{n^2K}{q} + nK^2\right). \qquad (64)$$

Не вдаваясь в детали, укажем, что несложно предъявить способ параллельных вычислений на системе с K узлами, для которой параллельная сложность универсального алгоритма Ланцоша—Паде примет вид

$$\{\text{параллельная сложность метода}\} = \mathcal{O}\left(\frac{pn^2}{K} + \frac{n^2}{q} + nK\right). \qquad (65)$$

Сравнивая (65) с (40), отметим, что в универсальном методе слагаемое вида n^2 переходит в слагаемое вида n^2/q, что и делает новый метод более эффективным.

Чем же мы платим? Напомним, что универсальный метод предполагает явное вычисление блоков вида Q_{jt} и Q_{jt+1}. Каждое такое вычисление связано с большим обменом данными между вычислительными узлами. Чем меньше t, тем чаще происходит такой обмен, и время на обмен увеличивается. С другой стороны, чем меньше t, тем больше q и тем меньше параллельная сложность алгоритма.

Работа выполнена при финансовой поддержке Минобрнауки России, соглашение от 17 июня 2014 г. № 14.604.21.0034 (идентификатор соглашения RFMEFI60414X0034) в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно—технологического комплекса России на 2014—2020 годы».

Литература

[1] Дорофеев А. Я. Вычисление логарифмов в конечном простом поле методом линейного решета // Тр. по дискр. матем. — 2002. — Т. 5. — С. 29—50.

[2] Дорофеев А. Я. Решение систем линейных уравнений при вычислении логарифмов в конечном простом поле // Матем. вопр. криптогр. — 2012. — Т. 3, № 1. — С. 5—51.

[3] Замарашкин Н. Л. Алгоритмы для разреженных систем линейных уравнений в $GF(2)$: Учебное пособие. — М.: Изд-во Моск. ун-та, 2013.

[4] Поповян И. А., Нестеренко Ю. В., Гречников Е. А. Вычислительно сложные задачи теории чисел: Учебное пособие. — М.: Изд-во Моск. ун-та, 2012.

[5] Черепнёв М. А. Блочный алгоритм типа Ланцоша решений разреженных систем линейных уравнений // Дискрет. матем. — 2008. — Т. 20, № 1. — С. 145—150.

[6] Cherepnev M. A. Version of block Lanczos-type algorithm for solving sparse linear systems // Bull. Math. Soc. Sci. Math. Roumanie. — 2010. — Vol. 53 (101), no. 3. — P. 225—230.

[7] Coppersmith D. Solving linear equations over $GF(2)$: Block Lanczos algorithm // Linear Algebra Appl. — 1993. — Vol. 193. — P. 33—60.

[8] Gutknecht M. H. A completed theory of the unsymmetric Lanczos process and related algorithms. I // SIAM J. Matrix Anal. Appl. — 1992. — Vol. 13, no. 2. — P. 594—639.

[9] Gutknecht M. H. A completed theory of the unsymmetric Lanczos process and related algorithms. II // SIAM J. Matrix Anal. Appl. — 1994. — Vol. 15, no. 1. — P. 15—58.

[10] Lanczos C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators // J. Res. Nat. Bur. Standards. — 1950. — Vol. 45. — P. 255—282.

[11] Montgomery P. A block Lanczos algorithm for finding dependencies over $GF(2)$ // Advances in Cryptology — EUROCRYPT '95. — Berlin: Springer, 1995. — (Lect. Notes Comp. Sci.; Vol 921). — P. 106—120.

[12] Peterson M., Monico C. $\mathbb{F}_2$ Lanczos revisited // Linear Algebra Appl. — 2008. — Vol. 428. — P. 1135—1150.