



УДК 051.3

В.В. Аристов, канд. техн. наук
Ин-т проблем моделирования
в энергетике им. Г.Е. Пухова НАН Украины
(Украина, 03164, Киев-164, ул. Генерала Наумова, 15,
тел. (044) 4243251, e-mail: vasily@aristov.com)

Интегро-алгоритмический метод вычисления логарифма матрицы с произвольной точностью*

Предложен интегро-алгоритмический метод аппроксимации и итерационной коррекции для вычисления с высокой точностью логарифмов матрицы. Метод основан на применении линейных многошаговых формул численного интегрирования разностного типа, а также разностно-дифференциальных формул Обрешкова с учетом высших производных. В нем обобщены известные алгоритмы логарифмирования на основе формул Паде. В результате итерационных коррекций повышается их порядок и точность. Предложенные соотношения и программные решения позволяют определять необходимые параметры для организации процессов вычисления логарифмов матрицы с произвольно заданной высокой точностью.

Запропоновано інтегро-алгоритмічний метод апроксимації та ітераційної корекції для обчислення з високою точністю логарифмів матриці. Метод базований на застосуванні лінійних багатокрокових формул чисельного інтегрування різницевого типу, а також різницево-диференціальних формул Обрешкова з урахуванням вищих похідних. В ньому узагальнено відомі алгоритми логарифмування на основі формул Паде. В результаті використання ітераційної корекції підвищується їх порядок і точність. Запропоновані співвідношення і програмні рішення дозволяють визначити необхідні параметри для організації процесів обчислення логарифмів матриці з довільно заданою високою точністю.

К л ю ч е в ы е с л о в а: логарифм матрицы, численное интегрирование, интегро-алгоритмический метод, многошаговые формулы интегрирования, передаточные функции, эквивалентные преобразования.

Итерационно-аппроксимационное матричное логарифмирование на основе разностных ФЧИ. При использовании варианта 1 из формулы (17) с учетом $\mathbf{A} = (\mathbf{C} - \mathbf{E})$ получаем

$$\bar{\mathbf{A}} = - \left(\sum_{v=-n_b}^{n_p} a_v \mathbf{C}^v \right) \left(\sum_{v=-n_b}^{n_p} c_{0v} \mathbf{C}^v \right)^{-1}, \quad (20)$$

* Окончание. Начало статьи см. в № 5, 14.

где a_v и c_{0v} — коэффициенты многошаговой ФЧИ

$$y_{i+1} = \sum_{v=-n_b}^{n_p} a_v y_{i+v} + \sum_{v=-n_b}^{n_p} c_{0v} H f_{i+v}.$$

В формуле (20) степень v матрицы \mathbf{C}^v соответствует $v \in \{-n_b, \dots, n_p\}$. Но поскольку матрица \mathbf{C} должна быть неособенной, то вычисления модифицированной матрицы $\bar{\mathbf{A}}$ удобнее выполнять по формуле

$$\bar{\mathbf{A}} = - \left(\sum_{v=-n_b}^{n_p} a_v \mathbf{C}^{v+n_b} \right) \left(\sum_{v=-n_b}^{n_p} c_{0v} \mathbf{C}^{v+n_b} \right)^{-1} = f_N(\mathbf{C}), \quad (21)$$

не имеющей отрицательных степеней для \mathbf{C} , где индекс N соответствует порядку аппроксимирующей функции (21).

В отличие от аппроксимации логарифмов методом $[m/n]$ Паде $r_{m,n}(\mathbf{C}-\mathbf{E})$ функция $f_N(\mathbf{C})$, во-первых, не обязательно должна иметь суммарный порядок аппроксимации $(m+n)$ и, во-вторых, не имеет смещения на единичную матрицу \mathbf{C} (или на единицу в случае скалярного аргумента). При совпадении порядков аппроксимации погрешности в обоих случаях будут одинаковыми. В работах [3—6] формулы $[m/m]$ Паде использованы для вычисления окончательного значения логарифма матрицы с заданной точностью, поэтому размер радиуса спектра матрицы \mathbf{C} , или ее норма, согласованы с величиной m , чтобы обеспечить требуемую точность. Изменение радиуса спектра достигается масштабированием исходной матрицы \mathbf{C} последовательностью из k операций извлечения квадратного корня с использованием, например, прямого оценочного выражения для нормы матрицы погрешностей $\|r_{m,m}(\mathbf{X}) - \log(\mathbf{E} - \mathbf{X})\| \leq |r_{m,m}(\|\mathbf{X}\|) - \log(1 - \|\mathbf{X}\|)|$, где $\mathbf{X} = \mathbf{E} - \mathbf{C}^{1/2^k}$.

В рассматриваемом интегро-алгоритмическом методе результат аппроксимации логарифма по формуле (21) является промежуточным и на этом этапе не обязательно должен удовлетворять требованиям получения заданной точности. Для достижения любой требуемой точности аппроксимации, при соответствующей минимизации погрешностей округления, могут быть использованы корректирующие уточнения в соответствии с формулой (15).

Если исходная матрица \mathbf{C} плохо обусловлена, т.е. $\text{cond}(\mathbf{C}) = \|\mathbf{C}\| \cdot \|\mathbf{C}^{-1}\| \gg 1$, то, как и в случае аппроксимации Паде, необходимо выполнить ее масштабирование, например k -кратным извлечением квадратного корня (корневое масштабирование):

$$\mathbf{Z} = \mathbf{C}^{1/2^k}. \quad (22)$$

Оценочное соотношение для нормы матрицы погрешностей в этом случае принимает вид

$$\|f_N(\mathbf{Z}) - \log(\mathbf{Z})\| \leq |f_N(1 - \|\mathbf{E} - \mathbf{Z}\|) - \log(1 - \|\mathbf{E} - \mathbf{Z}\|)|.$$

При использовании k -кратного масштабирования (22) вместо матрицы $\bar{\mathbf{A}}$ (21) определяется ее промасштабированный аналог,

$$\bar{\bar{\mathbf{A}}} = - \left(\sum_{v=-n_b}^{n_p} a_v \mathbf{Z}^{v+n_b} \right) \left(\sum_{v=-n_b}^{n_p} c_{0v} \mathbf{Z}^{v+n_b} \right)^{-1} = f_N(\mathbf{Z}), \quad (23)$$

соответствующий при $H = 1$, для системы ОДУ $\mathbf{Y}' = (\mathbf{C}^{1/2^k} - \mathbf{E}) \mathbf{Y}$, $\mathbf{Y}(t_0) = \mathbf{Y}_0$ вместо ФЧИ (14), следующей формуле интегрирования:

$$\mathbf{Y}_{i+1} = \sum_{v=-n_b}^{n_p} a_v \mathbf{Y}_{i+v} + \sum_{s=0}^m \sum_{v=-n_b}^{n_p} c_{sv} \bar{\bar{\mathbf{A}}}^{(s+1)} \mathbf{Y}_{i+v}.$$

Для восстановления масштаба необходимо результат вычисления $\bar{\mathbf{B}} = \ln \mathbf{Z} = \ln(\mathbf{C}^{1/2^k})$ умножить на 2^k . Поэтому запишем $\ln \mathbf{C} = 2^k \bar{\mathbf{B}}$, где

$$\bar{\mathbf{B}} = \sum_{r=0}^{\infty} \rho_r \bar{\bar{\mathbf{A}}}^{r+1} = \sum_{r=0}^{\infty} \rho_r (f_N(\mathbf{Z}))^{r+1}.$$

Сгруппируем бесконечный ряд коэффициентов $\rho_r \in \{\rho_0, \rho_1, \dots, \rho_{\infty}\}$ системной КДФ (3) или (15) следующим образом: $\rho_r \in \{[\rho_0], [\rho_1, \dots, \rho_N], [\rho_{N+1}, \dots, \rho_{N+K}], [\rho_{N+K+1}, \dots, \rho_{\infty}]\}$. Для корректно синтезированных ФЧИ всегда $\rho_0 = 1$. Если эквивалентный порядок аппроксимации ФЧИ равен N , то $\rho_1 = \rho_2 = \dots = \rho_N = 0$. Кроме того, многие ФЧИ (1) не имеют фазовой погрешности [8,10,11], поэтому все нечетные коэффициенты $\rho_{r=2s+1}$ тождественно равны нулю, т.е. $\rho_{2s+1} \equiv 0$. Тогда для ФЧИ с фазовой погрешностью

$$\ln \mathbf{Z} = \bar{\mathbf{B}} = \sum_{r=0}^{\infty} \rho_r \bar{\bar{\mathbf{A}}}^{r+1} = \bar{\bar{\mathbf{A}}} + \sum_{r=N+1}^{N+K} \rho_r \bar{\bar{\mathbf{A}}}^{r+1} + \sum_{r=N+K+1}^{\infty} \rho_r \bar{\bar{\mathbf{A}}}^{r+1},$$

а для ФЧИ без фазовой погрешности, имеющих всегда четное значение порядка аппроксимации N ,

$$\ln \mathbf{Z} = \bar{\mathbf{B}} = \sum_{r=0}^{\infty} \rho_{2r} \bar{\bar{\mathbf{A}}}^{2r+1} = \bar{\bar{\mathbf{A}}} + \sum_{r=N/2+1}^{N/2+K} \rho_{2r} \bar{\bar{\mathbf{A}}}^{2r+1} + \sum_{r=N/2+K+1}^{\infty} \rho_{2r} \bar{\bar{\mathbf{A}}}^{2r+1}. \quad (24)$$

Выражение (24) — точное выражение для матричного логарифма. Его первое приближение $\ln \mathbf{Z} \approx \bar{\mathbf{A}} = f_N(\mathbf{Z})$ вычисляется с помощью коэффициентов ФЧИ по формуле (23). В зависимости от порядка аппроксимации N используемой ФЧИ, величин крайних собственных чисел исходной матрицы \mathbf{C} , числа этапов k масштабирования извлечением квадратного корня и требуемой погрешности аппроксимации логарифма это значение может не соответствовать конечному результату с заданной точностью. Как и в указанных выше вариантах использования Паде аппроксимации [4, 5], более высокая точность может быть достигнута в результате увеличения значения k на этапе масштабирования или в результате применения ФЧИ с большим порядком аппроксимации.

В данном интегро-алгоритмическом методе предлагается использовать вариант, не требующий «сбрасывать» результаты предыдущих вычислений, а продолжать вычисления на основе суммирования корректирующих добавок. Далее будем использовать в основном вариант вычислений (24) для ФЧИ без фазовой погрешности. Запишем

$$\ln \mathbf{Z} \approx f_N(\mathbf{Z}) + \sum_{r=N/2+1}^{N/2+K} \rho_{2r} f_N^{2r+1}(\mathbf{Z}). \quad (25)$$

При этом матрица погрешности аппроксимации $\Delta(\mathbf{Z})$ определяется рядом

$$\Delta(\mathbf{Z}) = \sum_{r=N/2+K+1}^{\infty} \rho_{2r} f_N^{2r+1}(\mathbf{Z}) \approx \rho_{N+2(K+1)} f_N^{N+2K+3}(\mathbf{Z}). \quad (26)$$

Для оптимизации вычислений формулу (25) можно записать по схеме Горнера, подставив $\bar{\mathbf{A}} \approx f_N(\mathbf{Z})$,

$$\ln \mathbf{Z} \approx \bar{\mathbf{A}} + \bar{\mathbf{A}}^{N+3} (\rho_{N+2} + \bar{\mathbf{A}}^2 (\rho_{N+4} + \bar{\mathbf{A}}^2 (\rho_{N+6} + \dots + \bar{\mathbf{A}}^2 (\rho_{N+2K}))))),$$

или в виде итерационного соотношения для K коррекций:

$$\mathbf{W}_i = \mathbf{W}_{i-1} \cdot \bar{\mathbf{Q}}, \quad \mathbf{Q}_i = \mathbf{Q}_{i-1} + \rho_{N+2i} \mathbf{W}_i, \quad i=1, \dots, K, \quad \mathbf{W}_K \approx \ln \mathbf{Z}, \quad (27)$$

где $\mathbf{Q}_0 = f_N(\mathbf{Z})$, $\bar{\mathbf{Q}} = \mathbf{Q}_0^2$, $\mathbf{W}_0 = \mathbf{Q}_0^{N+1}$. Каждая коррекция повышает порядок аппроксимации на два с вычислительными затратами, после предварительного вычисления \mathbf{Q}_0 , $\bar{\mathbf{Q}}$ и \mathbf{W}_0 , в виде одного матричного умножения и операции матричного суммирования с весовым коэффициентом ρ_{N+2i} .

Не каждая многошаговая ФЧИ подходит для использования при вычислениях матричных логарифмов, даже если она имеет достаточно высокий порядок аппроксимации. Вследствие малого уменьшения коэффициента ρ_r в зависимости от индекса r итерационная коррекция (25) для

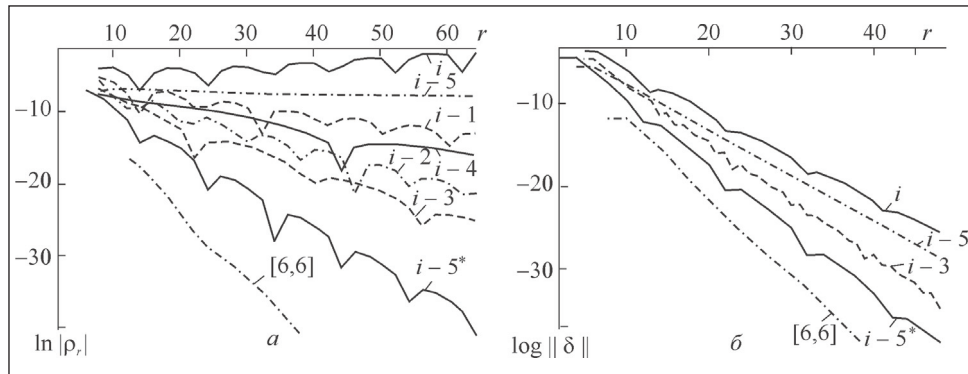


Рис. 2. Максимально точные неявные шестишаговые ФЧИ с одиночными опорными значениями Y_{i-v} : a — графики модулей коэффициентов ρ_r в логарифмическом масштабе; b — графики сходимости итерационных уточнений логарифма матрицы Гильберта

повышения точности может быть слишком медленной, что потребует увеличения значения N , либо окажется необходимым перемасштабирование матрицы Z с большим значением k для обеспечения сходимости.

На рис. 2, a , приведены графики значений для нескольких шестишаговых ФЧИ. С использованием процедуры `SintERDF()` коэффициенты неявной формулы типа Милна формируются следующей программой:

```
m:=0: nb:=5: np:=1:
W0:=[ eC[-1,-5], eC[-1,-4], eC[-1,-3], eC[-1,-2],
eC[-1,-1], eC[-1,1], eC[-1,j] - 1,
eC[-1,1], NULL]: SintERDF(W0):
```

Результатом выполнения программы при $j = -5$ является массив коэффициентов

$$eC_{(i-5)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{41}{140} & \frac{54}{35} & \frac{27}{140} & \frac{68}{35} & \frac{27}{140} & \frac{54}{35} & \frac{41}{140} \end{bmatrix},$$

соответствующий ФЧИ

$$Y_{i+1} = Y_{i-5} + H \left(\frac{41}{140} f_{i-5} + \frac{54}{35} f_{i-4} + \frac{27}{140} f_{i-3} + \frac{68}{35} f_{i-2} + \frac{27}{140} f_{i-1} + \frac{54}{35} f_i + \frac{41}{140} f_{i+1} \right),$$

и график ее коэффициентов ρ_r с первым ненулевым членом $\rho_8 = 3/2800$ (рис. 2, a , $i-5$). Остальные графики ρ_r соответствуют $i-v$, где $v = 0, \dots, 4$.

Все они имеют меньшую точность аппроксимации, так как первым ненулевым членом является ρ_7 , однако сходимость ряда (15) для некоторых ФЧИ выше, чем для $i - 5$. Если список W0 в предыдущей программе синтеза дополнить условием $eC[0,-5] - 3/10$, соответствующим заданию $c_{0,-5} = 3/10$, то будет получена менее точная ФЧИ ($\rho_6 = 1/840$)

$$Y_{i+1} = Y_{i-5} + H \left(\frac{3}{10} f_{i-5} + \frac{3}{2} f_{i-4} + \frac{3}{10} f_{i-3} + \frac{9}{5} f_{i-2} + \frac{3}{10} f_{i-1} + \frac{3}{2} f_i + \frac{3}{10} f_{i+1} \right), \quad (28)$$

однако со значительно более быстрой сходимостью ряда (15), — кривая $i-5^*$ на рис. 2, а. Если задать нулевой список ограничений W0 в указанной выше программе, то будет синтезирована более сложная, но значительно более точная ($\rho_{12} = 1/11099088$) шестишаговая формула с массивом коэффициентов

$$eC_{(6,6)} = \begin{bmatrix} 1 & \frac{132}{7} & \frac{375}{7} & 0 & \frac{375}{7} & \frac{132}{7} & 0 \\ \frac{10}{49} & \frac{360}{49} & \frac{2250}{49} & \frac{4000}{49} & \frac{2250}{49} & \frac{360}{49} & \frac{10}{49} \end{bmatrix}, \quad (29)$$

имеющая в данном случае максимальную сходимость — график [6,6] на рис. 2, а.

На рис. 2, б, приведены графики уменьшения (в зависимости от итераций) нормы матрицы погрешностей вычисления логарифма матрицы Гильберта 12-го порядка с ее предварительным семикратным корневым масштабированием. Из этих графиков видно, что для достижения одинаковой произвольно высокой точности число итераций в случае относительно неточной ФЧИ (28) будет существенно меньше, чем для иных шестишаговых ФЧИ с одной опорной точкой, и лишь незначительно больше, чем для более точной ФЧИ с коэффициентами (29), эквивалентной по коэффициентам ρ_r аппроксимации [6/6] Паде.

Все возможные разностные ФЧИ могут быть синтезированы процедурой SintERDF() при $m:=0$, т.е. без высших производных. По программе $m:=0$: $nb:=j$: $np:=1$: $W0:=[eC[-1,1], NULL]$: SintERDF(W0); при целочисленном значении j синтезируются разностные $(j + 1)$ -шаговые ФЧИ с максимально возможным порядком аппроксимации и минимальным остаточным членом. Такие формулы подпадают под определение аппроксимации Паде, но

в конкретном случае отличаются от известных [3—7] тем, что аппроксимируют не $\log(1 \pm x)$, а $\log(x)$. Поэтому для формального получения, например, из коэффициентов матрицы (29) коэффициентов [6/6] Паде аппроксимации необходимо аппроксимируемый аргумент x в аппроксимирующей функции

$$F(x) = \frac{7}{10} \cdot \frac{-7 - 132x - 375x^2 + 375x^4 + 132x^5 + 7x^6}{1 + 36x + 225x^2 + 400x^3 + 225x^4 + 36x^5 + x^6}$$

увеличить на единицу (или единичную матрицу для матричного логарифма). Тогда для данного случая получим следующую [6/6] Паде функцию:

$$P(x) = \frac{7}{10} \cdot \frac{1320x + 3300x^2 + 2960x^3 + 1140x^4 + 174x^5 + 7x^6}{924 + 2772x + 3150x^2 + 1680x^3 + 420x^4 + 42x^5 + x^6}.$$

Для любой $[m/n]$ Паде аппроксимации $\log(1 \pm x)$ процедурой SintERDF() можно синтезировать соответствующую ФЧИ для аппроксимации скалярного или матричного логарифма $\log(x)$. Организация итерационных уточнений, аналогичных (26), может быть выполнена не только в случае применения ФЧИ, но и при использовании аппроксимации Паде. Коэффициенты r_i системной КПФ (15) вычисляются либо, например, с помощью формулы (7), при разложении ее в степенной ряд, либо с помощью полиномиальных коэффициентов числителя b_v и знаменателя d_v формулы Паде

$$R_0(\bar{p}) = \frac{1}{\bar{p}} \left[\frac{\sum_{v=0}^m d_v (1 - e^{\bar{p}})^v}{\sum_{v=0}^m b_v (1 - e^{\bar{p}})^v} \right]^{-1}.$$

Процедура вычисления таких коэффициентов на языке Maple по исходной матрице формулы Паде следующая:

```
CPFPade:= proc(C0::array, R0m::integer, dd::integer)
  local pp, m1,i,n1,p,Af, Zz, Aa,Tt, rr;
  pp:=sign(dd): Zz:=array(1..R0m,[]): Af := array(1..R0m,[]): m1:=2:
  n1:= ColumnDimension(convert(C0,Matrix)): n1:=n1-1:
  p*sum(P[2,i1]*( pp*(exp(p)-1))^(i1-1),i1=1..n1+1)/
  sum(P[1,i1]*( pp*(exp(p)-1))^(i1-1),i1=1..n1+1);
  Aa:=series(%,p,R0m+2): Tt:=powpoly(convert(Aa, polynom),p):
  for m1 to R0m do Af[m1]:=Tt(m1): Zz[m1]:=Tt(m1) end do:
  for m1 from 1 to R0m-1 do
    for rr from R0m by -1 to m1+1 do
      Zz[rr] := evala(Zz[rr] + sum(Zz[rr - i]*Af[i], i = 1..rr - m1))
    end do
  end do
end proc;
```

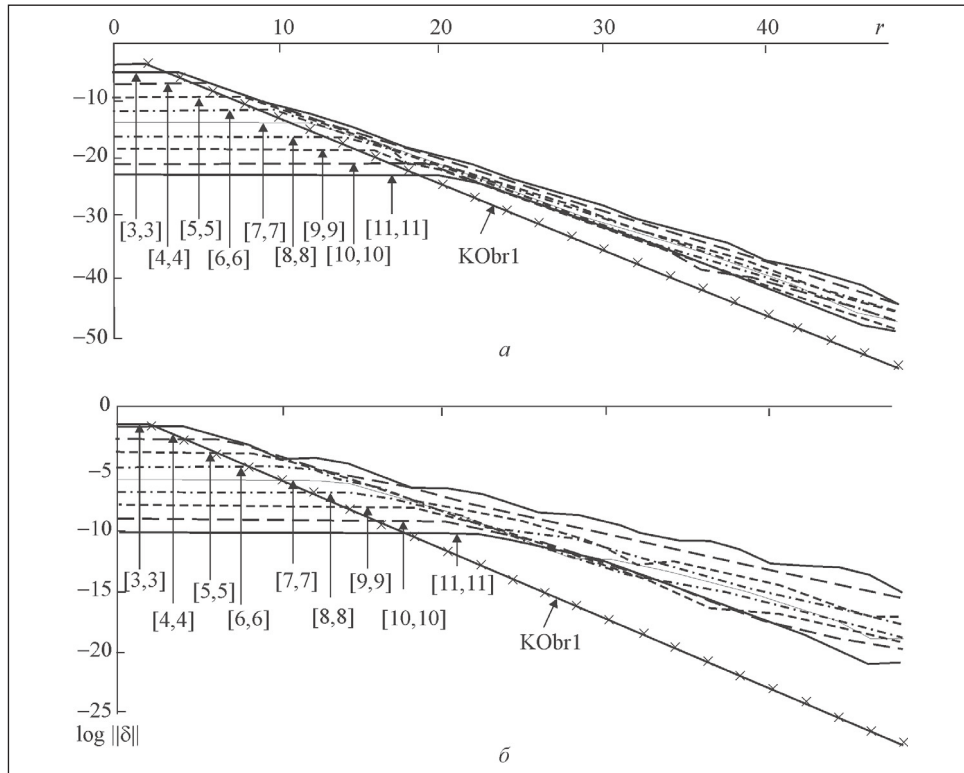



Рис. 3. Нормы погрешностей итерационного вычисления логарифма матрицы Гильберта ($n = 12$) с помощью $[m, m]$ ФЧИ при этапах корневого масштабирования $k = 7$ (а) и $k = 5$ (б): 1kObr1 — модифицированный вариант одношаговой ФЧИ Обрешкова с одной старшей производной

а при $m = 2$ — три корня, нахождение которых в общем случае требует существенных вычислительных затрат. Для $\mathbf{M}_3 \bar{\mathbf{A}}^3 + \mathbf{M}_2 \bar{\mathbf{A}}^2 + \mathbf{M}_1 \bar{\mathbf{A}} + \mathbf{M}_0 = 0$, обозначив

$$\mathbf{W} = (36\mathbf{M}_1\mathbf{M}_2\mathbf{M}_3 - 108\mathbf{M}_0\mathbf{M}_3^2 - 8\mathbf{M}_2^3 + 12\sqrt{3}\sqrt{4\mathbf{M}_1^3\mathbf{M}_3 - \mathbf{M}_1^2\mathbf{M}_2^2 - 18\mathbf{M}_0\mathbf{M}_1\mathbf{M}_2\mathbf{M}_3 + 27\mathbf{M}_0^2\mathbf{M}_3^2 + 4\mathbf{M}_0\mathbf{M}_2^3\mathbf{M}_3})^{1/3},$$

получаем $\bar{\mathbf{A}}_1 = (\mathbf{W} + 4(\mathbf{M}_2^2 - 3\mathbf{M}_1\mathbf{M}_3)\mathbf{W}^{-1} - 2\mathbf{M}_2)\mathbf{M}_3^{-1} / 6$ и два комплексных корня:

$$\bar{\mathbf{A}}_{2,3} = [-(\mathbf{W} + 4(\mathbf{M}_2^2 - 3\mathbf{M}_1\mathbf{M}_3)\mathbf{W}^{-1} + 4\mathbf{M}_2) \pm \pm i\sqrt{3}(\mathbf{W} - 4(\mathbf{M}_2^2 - 3\mathbf{M}_1\mathbf{M}_3)\mathbf{W}^{-1})]\mathbf{M}_3^{-1} / 12,$$

где

$$\mathbf{M}_0 = \sum_{v=-n_b}^{n_p} a_v \mathbf{C}^v, \quad \mathbf{M}_1 = \sum_{v=-n_b}^{n_p} c_{0v} \mathbf{C}^v, \quad \mathbf{M}_2 = \sum_{v=-n_b}^{n_p} c_{1v} \mathbf{C}^v, \quad \mathbf{M}_3 = \sum_{v=-n_b}^{n_p} c_{2v} \mathbf{C}^v.$$

Для упрощения отображения формул перейдем к скалярным переменным, определяющим как изменение погрешности аппроксимации, так и скорость сходимости итераций при вычислении матричных логарифмов. Будем считать, что неособенная матрица \mathbf{C} диагонализируема, т.е. может быть представлена в виде $\mathbf{C} = \mathbf{V}\mathbf{C}_d\mathbf{V}^{-1}$, где $\mathbf{C}_d = \mathbf{E}\lambda$ — диагональная матрица из собственных чисел λ_j . Тогда аппроксимационная матричная функция $F(\mathbf{C})$ принимает вид $F(\mathbf{C}) = F(\mathbf{V}\mathbf{C}_d\mathbf{V}^{-1}) = \mathbf{V}F(\mathbf{C}_d)\mathbf{V}^{-1} = \mathbf{V}(f(\lambda))\mathbf{V}^{-1}$, где $f(\lambda)$ — аппроксимация вектора независимых собственных чисел λ_j . Поэтому в дальнейшем вместо матричной $F(\mathbf{C})$ будем рассматривать скалярную функцию $f(x)$, в которой аргумент x учитывает основные параметры совокупности собственных чисел λ_j . В этом случае вместо (30) для варианта 2 запишем

$$f(x)_{1,2} = \frac{\left[-\frac{1}{2} \sum_{v=-n_b}^{n_p} c_{0v} x^v \pm \sqrt{\left(\sum_{v=-n_b}^{n_p} c_{0v} x^v \right)^2 / 4 - \left(\sum_{v=-n_b}^{n_p} c_{1v} x^v \right) \left(\sum_{v=-n_b}^{n_p} a_v x^v \right)} \right]}{\sum_{v=-n_b}^{n_p} c_{1v} x^v} \quad (31)$$

Для формирования аппроксимирующих функций согласно вариантам 2 и 3 для матричного или скалярного логарифмов необходимо синтезировать соответствующие разностно-дифференциальные ФЧИ, например с помощью приведенной выше процедуры SintERDF(). Задавая целочисленное значение j , программой $m:=m0: nb := j: np:=1: W0:= [eC[-1,1], NULL]: SintERDF(W0)$; синтезируем матрицы коэффициентов $(j+1)$ -шаговых ФЧИ при $m_0 = 1$ и при $m_0 = 2$ с максимально возможным порядком аппроксимации и минимальным остаточным членом.

Пусть $m_0 = 1$ и $j = 0$. В результате выполнения программы получаем коэффициенты

$$eC_{1Obr1} = \begin{bmatrix} 1 & -1 \\ 1/2 & 1/2 \\ 1/12 & -1/12 \end{bmatrix}$$

одношаговой ФЧИ Обрешкова (8) с первой высшей производной, обозначенной 1Obr1. Коэффициенты ρ_r ее системной КПФ определяются по

формуле (9), при этом все нечетные ρ_{2r+1} и ρ_2 равны нулю. Первые ненулевые ρ_{2r} имеют вид

$$\rho_4 \dots \rho_{2r} \dots = \left[\frac{-1}{720}, \frac{-1}{12096}, 0, \frac{1}{2737152}, \frac{1}{38817792}, \dots \right].$$

Подставляя коэффициенты ФЧИ в формулу (31), получаем следующую аппроксимирующую функцию:

$$f_{10br1}(x) = \frac{3+3x-\sqrt{42x-3x^2-3}}{x-1}.$$

Двухшаговой ФЧИ Обрешкова (2Obr1 при $j = 1$) соответствуют матрица коэффициентов

$$eC_{2Obr1} = \begin{bmatrix} -1 & 2 & -1 \\ \frac{3}{8} & 0 & \frac{3}{8} \\ -\frac{1}{24} & \frac{1}{3} & -\frac{1}{24} \end{bmatrix},$$

ряд $\rho_6 \dots \rho_{2r} \dots = \left[\frac{-1}{75600}, 0, \frac{-1}{71280000}, \frac{1}{1263600000}, \frac{-1}{46656000000}, \dots \right]$ и функция

$$f_{2Obr1}(x) = \frac{3}{2} \frac{-3+3x^2 + \frac{1-x}{3} \sqrt{930x-15x^2-15}}{x^2-8x+1}.$$

Четырехшаговой ФЧИ 4Obr1 при $j = 3$ соответствуют матрица

$$eC_{4Obr1} = \begin{bmatrix} -1 & \frac{256}{13} & -\frac{486}{13} & \frac{256}{13} & -1 \\ \frac{15}{52} & \frac{96}{13} & 0 & -\frac{96}{13} & \frac{15}{52} \\ -\frac{3}{130} & \frac{96}{65} & -\frac{324}{65} & \frac{96}{65} & -\frac{3}{130} \end{bmatrix},$$

ряд $\rho_{12} \dots \rho_{2r} \dots = \left[\frac{1}{2427925500}, \frac{1}{19722225600}, \frac{163}{18440280936000}, \dots \right]$ и функция

$$f_{4Obr1}(x) = \frac{5}{12} \frac{15x^4 - 384x^3 + 384x - 15 - \frac{1-x}{5} \sqrt{p(x)}}{x^4 - 64x^3 + 216x^2 - 64x + 1},$$

где $p(x) = -615x^6 + 233010x^5 - 5292405x^4 + 31464420x^3 - 5292405x^2 + 233010x - 615$.

Шестишаговой ФЧИ 6Obr1 при $j = 5$ соответствуют матрица

$$eC_{6Obr1} = \begin{bmatrix} -1 & \frac{205632}{2507} & \frac{1603125}{2507} & \frac{2800000}{2507} & \frac{1603125}{2507} & \frac{205632}{2507} & -1 \\ \frac{630}{2507} & \frac{71280}{2507} & \frac{506250}{2507} & 0 & \frac{506250}{2507} & \frac{71280}{2507} & \frac{630}{2507} \\ \frac{300}{17549} & \frac{64800}{17549} & \frac{1012500}{17549} & \frac{2400000}{17549} & \frac{1012500}{17549} & \frac{64800}{17549} & \frac{300}{17549} \end{bmatrix},$$

ряд

$$\rho_{18 \dots \rho_{2r} \dots} = \left[\frac{-1}{66552040691136}, \frac{-1}{187146339988608}, \frac{-17027}{8780906272265487360}, \dots \right]$$

и функция

$$f_{6Obr1}(x) = \frac{7}{60} \cdot \frac{63x^6 - 7128x^5 + 50625x^4 - 50625x^2 + 7128x - 63 + \frac{1-x}{7} \sqrt{p(x)}}{x^6 - 216x^5 + 3375x^4 - 8000x^3 + 3375x^2 - 216x + 1}$$

при

$$p(x) = -16107x^{10} + 18719610x^9 - 17367591155x^8 + 50447648160x^7 - 416418101190x^6 + 1110670479828x^5 - 416418101190x^4 + 50447648160x^3 - 1736759115x^2 + 18719610x - 16107.$$

В качестве примера для преобразования по варианту 3 возьмем одношаговую ФЧИ Обрешкова (1Obr2) с двумя высшими производными ($m = 2$). В этом случае имеем матрицу

$$eC_{1Obr2} = \begin{bmatrix} 1 & -1 \\ 1/2 & 1/2 \\ 1/10 & -1/10 \\ 1/120 & 1/120 \end{bmatrix},$$

ряд $\rho_6 \dots \rho_{2r} \dots = \left[\frac{1}{100800}, \frac{1}{2592000}, \frac{1}{190080000}, \frac{1}{6739200000}, \dots \right]$ и функцию

$$f_{1\text{Obr}2}(x) = -\frac{1}{5} \cdot \frac{1+18x+x^2+20(1-x)p(x)-100p^2(x)}{(1+x)p(x)}$$

при $p(x) = \frac{2^{2/3}}{20} \left(x^3 - 63x^2 + 63x - 1 + \sqrt{5}(1+x) \sqrt{x^4 + 16x^3 + 1566x^2 + 16x + 1} \right)^{1/3}$.

Практически любая ФЧИ может быть использована для синтеза аппроксимирующей функции и получения произвольно высокой точности логарифмирования в результате предварительного корневого масштабирования и организации корректирующих итераций. Однако для оптимизации вычислительных затрат необходимо учитывать ряд факторов, среди которых — сложность и точность аппроксимирующей формулы (состав, число и сложность операций), сложность масштабирования, скорость сходимости итераций, сложность формирования коэффициентов ρ_r , затраты на дополнительные подготовительные вычисления для организации и оценки итераций. Кроме того, затраты на матричные вычисления можно сократить, применив ортогональные преобразования для получения треугольных матриц, например с помощью декомпозиции Шура. В работах [3, 4, 6, 12] рассмотрены вопросы оптимизации подобных вычислений матричных логарифмов.

Полученные выше формулы синтезированы из условия минимальной погрешности аппроксимации логарифма. Дальнейшее повышение точности может быть достигнуто в результате уточняющих итераций (27). При этом предпочтительно иметь ряд коэффициентов ρ_r с высокой сходимостью к нулю. Но увеличение сходимости обычно приводит к увеличению начальной погрешности, которая при итерациях может быть быстрее скомпенсирована. В итоге общее решение задачи синтеза является компромиссным.

Рассмотрим вариант синтеза одношаговой ФЧИ при $m = 1$, отличный от 1Obr1. Обозначив новую ФЧИ 1kObr1, представим матрицу коэффициентов в виде

$$eC_{1\text{kObr}1} = \begin{bmatrix} 1 & -1 \\ 1/2 & 1/2 \\ w & -w \end{bmatrix}.$$

В этом случае

$$f_{1\text{kObr}1}(x) = \frac{1}{4} \cdot \frac{1+x - \sqrt{1+2x+x^2-16w+32wx-16wx^2}}{w(x-1)}$$

и ряд коэффициентов ρ_r имеет вид

$$\rho_{2\dots\rho_{2r}\dots} = \left[\frac{1}{12} - w, \frac{1}{80} - \frac{w}{4} + w^2, \frac{1}{448} - \frac{w}{16} + \frac{w^2}{2} - w^3, \frac{1}{2304} - \frac{w}{64} + \frac{3w^2}{16} - \frac{5w^3}{6} + w^4, \dots \right].$$

Упростив подкоренное выражение $f_{1kObr1}(x)$ выбором $w = 1/16$, получим

$$f_{1kObr1}(x) = 4 \frac{1+x-2\sqrt{x}}{x-1}, \tag{32}$$

и $\rho_{2\dots\rho_{2r}\dots} = \left[\frac{1}{48}, \frac{1}{1280}, \frac{1}{28672}, \frac{1}{589824}, \frac{1}{11534336}, \dots \right]$, при этом

$$\rho_{2r} = 2^{-4r} / (2r+1). \tag{33}$$

ФЧИ с матрицей eC_{1kObr1} при $w = 1/16$ будем называть одношаговой формулой квази-Обрешкова с первой высшей производной. Достоинствами формулы (32) являются ее простота, более быстрое убывание значений ρ_{2r} по сравнению с $1Obr1$ и формулами Паде любого порядка (см. рис. 1) и связанная с этим более быстрая сходимость итераций (см. рис. 3), а также простая формульная зависимость (33) для определения ρ_{2r} .

На рис. 4 приведены сравнительные графики модулей погрешностей $\delta = \ln(x) - f(x)$ некоторых полученных вариантов аппроксимирующих формул. Графики с обозначением $[m,m]$ соответствуют разностным соотношениям типа (20) с одинаковыми порядками полиномов в числителе и знаменателе, аналогичных аппроксимации $[m/m]$ Паде. Графики с обозначением $1kObr1 + K$, где $K = 1, 2, \dots, 9$, соответствуют формуле квази-Обрешкова (32) с K корректирующими итерациями. Графики отображают погрешность при вычислении логарифма от скалярной переменной x , а также матричного логарифма от j -го собственного числа диагонализированной матрицы C . Поскольку в случае матричного логарифма количество собственных чисел определяется размерностью C , его общая погрешность зависит от композиции диагональной матрицы погрешностей по каждому собственному числу с прямой и инверсной матрицей собственных векторов V .

На рис. 5 приведены относительные значения первых ненулевых коэффициентов ρ_{r0} с указанием градиентных векторов для ряда характерных разностных и разностно-дифференциальных ФЧИ с высшими производными. Определим порядок аппроксимации N для j -шаговой ФЧИ с m высшими производными. Общее число коэффициентов a_v и c_{sv} формулы интегрирования (14) составляет $(j+1)(m+2)$, где $j = n_p + n_b - 1$. Два ограничения накладываются на коэффициенты a_v : $a_1 = 0$ и $\sum_{v=-n_b}^{n_p} a_v = 1$. Еще

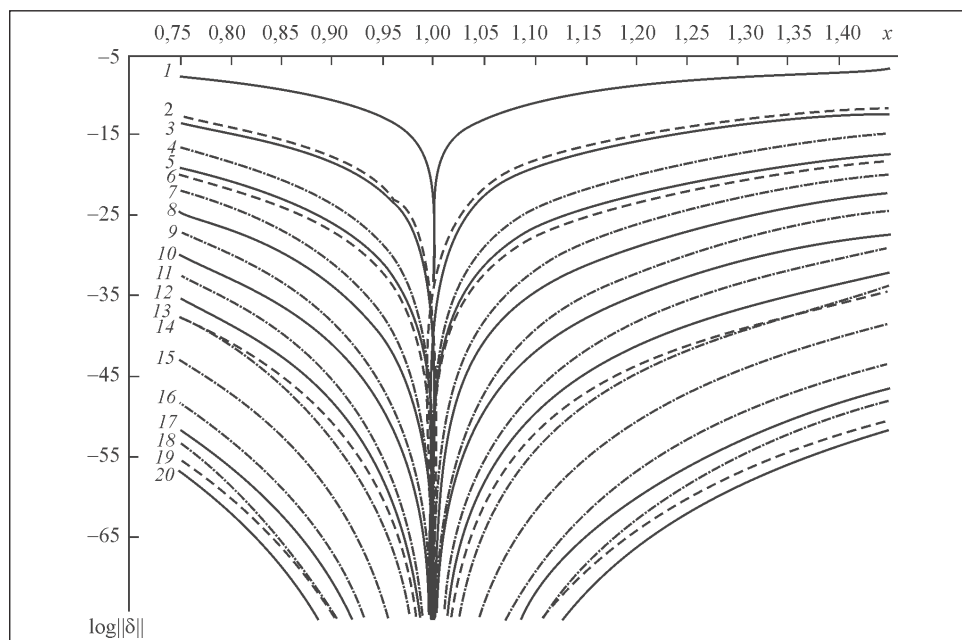


Рис. 4. Сравнительные графики погрешностей аппроксимации логарифма: 1 — 1kObr1; 2 — 1Obr1; 3 — 1kObr1+1; 4 — [3,3]; 5 — 1kObr1+2; 6 — 2Obr1, 1Obr2; 7 — [4,4]; 8 — 1kObr1+3; 9 — [5,5]; 10 — 1kObr1+4; 11 — [6,6]; 12 — 1kObr1+5; 13 — 4Obr1; 14 — [7,7]; 15 — [8,8]; 16 — [9,9]; 17 — 1kObr1+8; 18 — [10,10]; 19 — 6Obr1; 20 — 1kObr1+9

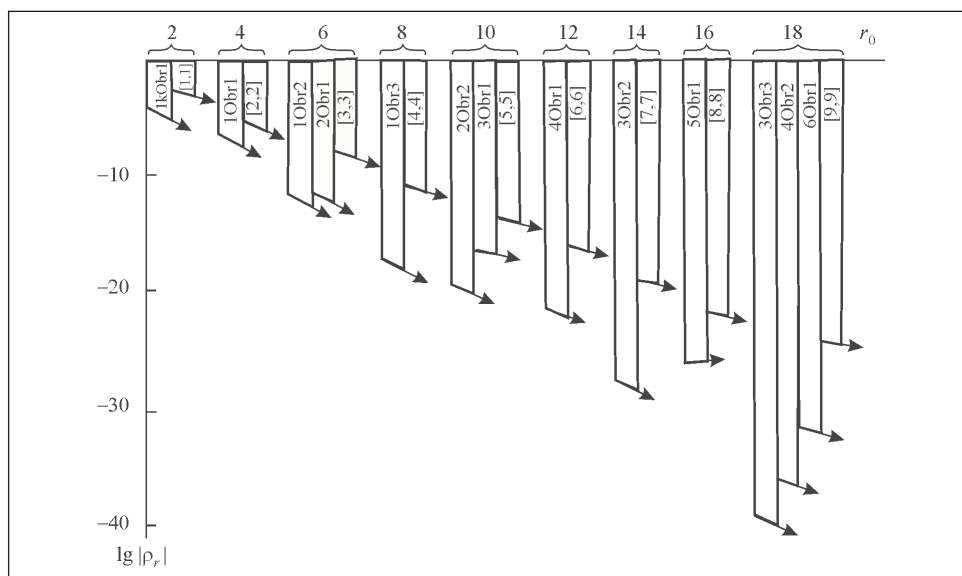


Рис. 5. Начальные коэффициенты ρ_0 итерационной аппроксимации с градиентами убывания коэффициентов ρ_r

два ограничения накладываются на коэффициенты c_{sv} : одно для обнуления фазовой погрешности (необходимо, чтобы все нечетные коэффициенты ρ_{2r-1} были равны нулю), другое — для задания единичного коэффициента передачи результата интегрирования. Тогда максимальный порядок аппроксимации ФЧИ $N = (j+1)(m+2) - 4 + (1 - \text{mod}_2(j+1) \text{mod}_2(m)) = (j+1)(m+2) - 3 - \text{mod}_2(j+1) \text{mod}_2(m)$, где в слагаемом $(1 - \text{mod}_2(j+1) \times \text{mod}_2(m))$ учтен нулевой нечетный коэффициент $\rho_{r,0-1}$ системной КПФ. На рис. 5 индексы первых ненулевых коэффициентов $\rho_{r,0}$ равны $N+1$.

Из анализа графиков, представленных на рис. 4 и 5, можно сделать следующие выводы.

- Точность аппроксимации $\ln(x)$ каждой формулой существенно возрастает в области $x \approx 1$.

- ФЧИ квази-Обрешкова 1kObr1 имеет наибольшую погрешность, однако на каждой итерации погрешность уменьшается приблизительно на 7 десятичных порядков. При вычислительных затратах в виде умножения двух матриц, умножения матрицы на скаляр и суммирования двух матриц порядок аппроксимации на каждой итерации увеличивается на два.

- Применение двух итераций к ФЧИ квази-Обрешкова 1kObr1 позволяет получить результат в 10^3 раз точнее, чем для ФЧИ [3,3] (типа [3/3] Паде), а применение девяти итераций — в 10^4 раз точнее, чем для ФЧИ [10,10] (типа [10/10] Паде).

- Двухшаговая ФЧИ с первой высшей производной имеет такую же точность, как одношаговая ФЧИ с двумя высшими производными.

- Не обнаружено каких-либо преимуществ по точности и скорости сходимости для ФЧИ с несколькими высшими производными, однако сложность вычислений возрастает.

При k -кратном корневом масштабировании (22) матрица погрешностей вычисления логарифма матрицы \mathbf{C} , с учетом выражения (26), оценивается соотношением

$$\Delta(\mathbf{C}) = 2^k \sum_{r=N/2+K+1}^{\infty} \rho_{2r} f_N^{2r+1}(\mathbf{Z}) \approx 2^k \rho_{N+2(K+1)} f_N^{N+2K+3}(\mathbf{Z})$$

и выбором числа коррекций K может быть приведена достаточно близко к нулевой для достижения высокой точности определения матричного логарифма. Естественно, вычисления необходимо выполнять с такой разрядностью, чтобы суммарный эффект от погрешностей округлений не превысил методическую погрешность результирующей аппроксимации.

Для повышения точности при фиксированных значениях k , K и ФЧИ можно дополнительно использовать симметрирование спектра матрицы. Из рис. 4 видно, что для значений собственных чисел x в диапазоне $[0,9 \div 1,1]$

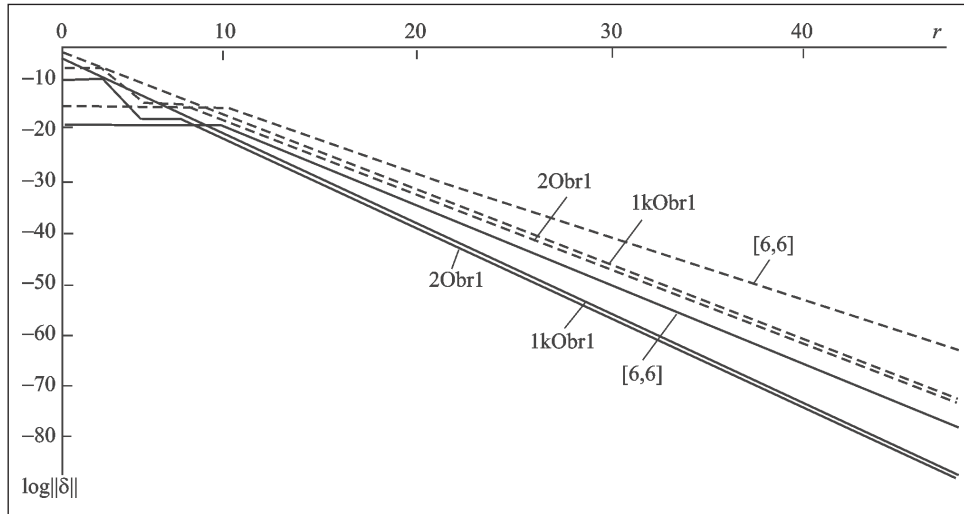


Рис. 6. Норма матрицы погрешности итерационной аппроксимации логарифма матрицы Гильберта 12-го порядка при $k=8$: --- без симметрирования спектра ($S=1$); — с симметрированием спектра ($S \approx \sqrt{\lambda_1 \lambda_{12}} \approx 4,37641 \cdot 10^{-7}$)

наблюдается наибольшее уменьшение погрешности аппроксимации. Поэтому применяется корневое масштабирование (22), позволяющее приблизить спектр \mathbf{Z} к этому диапазону. С помощью корневого масштабирования собственные числа матрицы \mathbf{Z} перемещаются ближе к единице, как в единичной диагональной матрице \mathbf{E} [3—6]. Относительная скорость перемещения при изменении значения k одинакова слева и справа от единицы. Поэтому при исходной асимметрии спектров \mathbf{C} и \mathbf{C}^{-1} спектр матрицы \mathbf{Z} окажется смещенным относительно единицы, что не является оптимальным для получения наибольшей точности.

Для симметрирования введем множитель S такой, чтобы выполнялось условие $\lambda_{\max} / S \approx S / \lambda_{\min}$, т.е. $S \approx \sqrt{\lambda_{\max} \lambda_{\min}}$, где λ_{\min} и λ_{\max} — минимальное и максимальное собственные числа матрицы \mathbf{C} . В этом случае определяется логарифм матрицы \mathbf{C}/S в соответствии с каким-либо описанным выше вариантом и к полученному результату добавляется $(\ln S) \mathbf{E}$.

На рис. 6 приведены результаты вычисления логарифма матрицы Гильберта 12-го порядка тремя методами: двухшаговой ФЧИ Обрешкова 2Obr1, одношаговой 1kObr1 и ФЧИ [6,6] типа Паде. Здесь симметрирование эквивалентно по точности увеличению значения k на единицу. Для точного определения значения S необходимо знать собственные числа, определение которых является непростой задачей. Даже если эти числа известны, то использовать итерационно-аппроксимационные алгоритмы

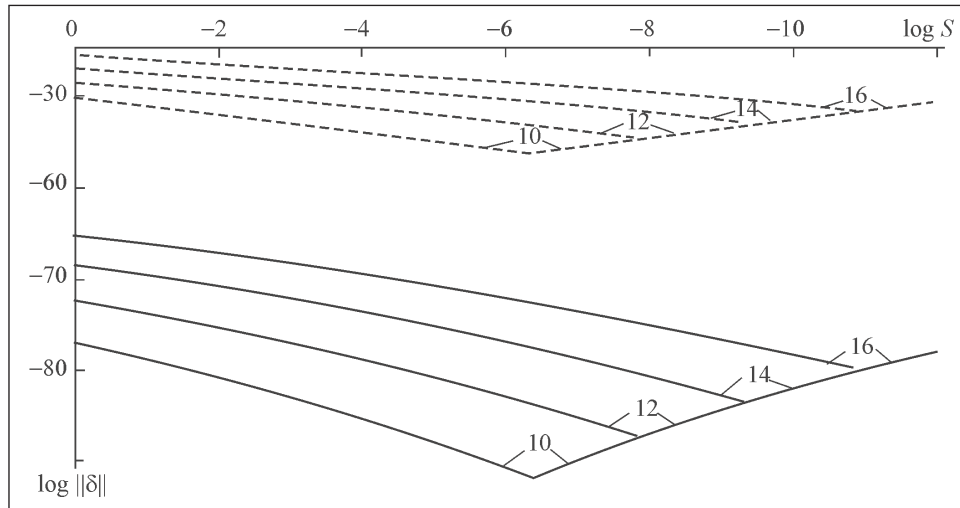


Рис. 7. Зависимости нормы матрицы погрешностей итерационной аппроксимации логарифма матриц Гильберта различного порядка nG в зависимости от симметризирующего множителя S в соответствии с одношаговой аппроксимационной формулой квази-Обрешкова 1kObr1 при $k = 8$: --- $K = 8$; — $K = 23$

нецелесообразно. Однако симметризирующий множитель может быть определен весьма приближенно, что несущественно влияет на изменение точности логарифмирования матрицы.

На рис. 7 приведены графики нормы погрешности логарифмирования матриц Гильберта порядка $nG \in \{10, 12, 14, 16\}$ в зависимости от S . Оптимальные значения S соответствуют точкам изломов кривых и составляют в порядке возрастания nG : $4,376_{10^{-7}}$; $1,427_{10^{-8}}$; $4,252_{10^{-10}}$; $1,308_{10^{-11}}$. Выбор оценочных значений, например $1_{10^{-7}}$, $1_{10^{-8}}$, $1_{10^{-10}}$ и $1_{10^{-11}}$, незначительно снижает точность вычислений, так как они находятся в области минимальных погрешностей. Приближенные величины S можно найти, например, исходя из известных свойств применяемой матрицы или по оценкам спектра, используя нормы $\|C^M\|^{1/M} \approx \lambda_{\max}$ и $\|C^{-M}\|^{-1/M} \approx \lambda_{\min}$ для целого $M \gg 1$.

Выводы

Предложенный интегро-алгоритмический метод аппроксимации и итерационной коррекции для вычисления с высокой точностью логарифмов матрицы основан на применении линейных многошаговых формул численного интегрирования разностного типа, а также разностно-дифференциальных формул Обрешкова с учетом высших производных. Метод также обобщает известные алгоритмы логарифмирования на основе формул

Паде и позволяет повысить их порядок и точность посредством итерационной коррекции. Предложенные соотношения и программные решения позволяют определять необходимые параметры для организации процессов вычисления логарифмов матрицы с произвольно заданной высокой точностью.

СПИСОК ЛИТЕРАТУРЫ

1. Гантмахер Ф.П. Теория матриц. — М. : Наука, 1988. — 552 с.
2. Culver W. J. On the existence and uniqueness of the real logarithm of a matrix // Proc. of the American Mathematical Society. — 1966. — Vol. 17, No 5. — P. 1146 — 1151.
3. Al-Mohy A., Higham N. Improved inverse scaling and squaring algorithms for the matrix logarithm // SIAM J. Sci. Comput. — 2012. — Vol. 34, No 4. — P. C.153 — C.169.
4. Cheng S. H., Higham N. J., Kenney C. S., Laub A. J. Approximating the logarithm of a matrix to specified accuracy // SIAM J. Matrix Anal. Appl. — 2001. — Vol. 22. — P. 1112—1125.
5. Kenney C., Laub A. Condition estimates for matrix functions // Ibid. — 1989. — Vol. 10. — P. 707 — 730.
6. Kenney C., Laub A. A Schur-Frechet algorithm for computing the logarithm and exponential of a matrix // Ibid.—1998. — Vol. 19, No 3. — P. 640 — 663.
7. Higham, Nicholas J. Functions of matrices. Theory and computation// Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. — 425 p. — ISBN: 978-0-89871-646-7 MR2396439 (2009b:15001) MR2396439.
8. Аристов В.В. Функциональные макрооперации: Основы итерационных алгоритмов. — Киев : Наук. думка, 1992. — 280 с.
9. Аристов В.В. Модель многошаговых разностно-дифференциальных формул интегрирования с учетом влияния участка разгона // Электрон. моделирование. — 2013. — 35, No 6. — С. 3—26.
10. Аристов В.В. Интегро-алгоритмические вычисления. — Киев: Наук. думка, 1980. — 192 с.
11. Аристов В.В. Математические модели итерационных соотношений обобщенных CORDIC-алгоритмов // Электрон. моделирование. — 2011. — 33, No 1. — С. 3—29.
12. Higham N.J. Evaluating Pade approximants of the matrix logarithm // SIAM J. Matrix Anal. Appl. — 2001. — Vol. 22, No 4. — P. 1126 — 1135.

V.V. Aristov

INTEGRO-ALGORITHMIC METHOD FOR COMPUTATION THE MATRIX LOGARITHM WITH ARBITRARY ACCURACY

The integro-algorithmic method of approximation and iteration correction for high-accuracy computation of matrix logarithms is proposed. The method is based on the use of linear multistep formulas of numerical integration of the difference type as well as the Obreshkov difference-differential formulas with allowance for higher derivatives. Due to iterations in this case there is not a necessity to choose a high-fidelity primary approximating formula and a basic criterion is a receipt of high-rate of convergence. A method also summarizes the well-known algorithms of taking the logarithm based on the Pade formulas and increases their order and accuracy due to the additional use of iterative correction. The proposed relations and program solutions permit determining necessary parameters for organizing the processes of matrix logarithms computation with arbitrary preset high accuracy.

Keywords: matrix logarithms, numerical integration, integro-algorithmic method, multistep formulas of integration, transfer function, equivalent transformations.

REFERENCES

1. *Gantmacher F.R.* The Theory of Matrices. — Moscow: Nauka, 1988. — 552 p. (in Russian).
2. *Culver W.J.* On the existence and uniqueness of the real logarithm of a matrix // Proc. of the American Mathematical Society. — 1966. — Vol. 17, No 5. — P. 1146—1151.
3. *Al-Mohy A., Higham N.* Improved inverse scaling and squaring algorithms for the matrix logarithm // SIAM J. Sci. Comput. — 2012. — Vol. 34, No 4. — P. C.153—C.169.
4. *Cheng S.H., Higham N.J., Kenney C.S., Laub A.J.* Approximating the logarithm of a matrix to specified accuracy // SIAM J. Matrix Anal. Appl. — 2001. — Vol. 22. — P. 1112—1125.
5. *Kenney C., Laub A.* Condition estimates for matrix functions // Ibid. — 1989. — Vol. 10. — P. 707—730.
6. *Kenney C., Laub A.* A Schur-Frechet algorithm for computing the logarithm and exponential of a matrix // Ibid. — 1998. — Vol. 19, No 3. — P. 640—663.
7. *Higham N.J.* Functions of matrices. Theory and computation // Society for Industrial and Applied Mathematics. SIAM-2008. Philadelphia, 2008. — 425 p.
8. *Aristov V.V.* Functional macrooperation: basics of iterative algorithms. — Kiev: Nauk. Dumka, 1992. — 280 p. (in Russian).
9. *Aristov V.V.* Model multistep difference-differential methods integration with regard to the influence of the starting section // Electronic Modeling. — 2013. — Vol. 35, No 6. — P. 3—26 (in Russian).
10. *Aristov V.V.* Integro-algorithmic computations. — Kiev: Nauk. Dumka, 1980. — 192 p. (in Russian).
11. *Aristov V.V.* Mathematical models of iterative relations generalized CORDIC-algorithms // Electronic Modeling. — 2011. — Vol. 33, No 1. — P. 3—29 (in Russian).
12. *Higham N.J.* Evaluating Pade approximants of the matrix logarithm // SIAM J. Matrix Anal. Appl. — Vol. 22, No 4. — P. 1126—1135.

Поступила 18.06.14

АРИСТОВ Василий Васильевич, канд. техн. наук, вед. науч. сотр. Ин-та проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 1963 г. окончил Харьковский политехнический ин-т. Область научных исследований — теория и реализация итерационных алгоритмов макроопераций, численные методы решения дифференциальных уравнений, схемотехника и архитектура специализированных устройств, геоинформационные системы, радиотехника.