

ЧЕРЕПНЕВ Михаил Алексеевич

**МЕТОД РЕШЕНИЯ БОЛЬШИХ РАЗРЕЖЕННЫХ
СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ**

Москва
2018

Новый Ланцошеподобный блочный алгоритм решения больших разреженных систем линейных уравнений над полем \mathbb{F} .

Пусть требуется решить симметричную систему линейных уравнений:

$$Ax = B, A \in \mathbb{F}^{N \times N}, B \in \mathbb{F}^{N \times n}, \quad (1)$$

где A — симметричная матрица. Пусть n — длина машинного слова (32, 64 или 128), а d — оценка сверху на число ненулевых элементов в каждой строке матрицы A (плотность матрицы).

Во всех Ланцошеподобных методах решения разреженных систем само решение ищется в пространстве Крылова $\langle B, AB, A^2B, \dots \rangle$. Предлагаемый алгоритм в простейшей ситуации использует следующую теорему

Теорема 1 Пусть имеется набор попарно A -ортогональных подпространств $\langle W_i \rangle, i = 0, 1, \dots, m-1$ в \mathbb{F}^N , A -скалярное произведение на которых невырождено. Пусть также матрица A переводит сумму этих подпространств $\langle W \rangle$ в себя и $\langle B \rangle \subseteq \langle W \rangle$. Тогда решение системы (1) в $\langle W \rangle$ единственно и задаётся формулой:

$$X = \sum_{i=0}^{m-1} W_i (W_i^T A W_i)^{-1} W_i^T B. \quad (2)$$

Суть предлагаемого метода состоит в рекуррентном построении приближений Паде к некоторому ряду, коэффициенты которого лежат в кольце квадратных матриц $\mathbb{F}^{sn} \times \mathbb{F}^{sn}$. Из этих приближений получаются блоки W_i , о которых шла речь в предыдущей теореме.

Пусть

$$\alpha(\lambda) = \sum_{i=0}^{\infty} \alpha_i \lambda^{-i}, \alpha_i \in \mathbb{F}^{ns \times ns}, \alpha_i = B^T A^i B,$$

и для некоторых матричных многочленов $Q^{(i)}, P^{(i)}$ выполнено:

$$\begin{aligned} \alpha(\lambda)Q^{(t)}(\lambda) + P^{(t)}(\lambda) &= \sum_{i=t+1}^{\infty} \rho_i \lambda^{-i}, \\ \alpha(\lambda)Q^{(t-1)}(\lambda) + P^{(t-1)}(\lambda) &= \sum_{i=t}^{\infty} \tau_i \lambda^{-i}, \\ \alpha(\lambda)Q^{(t-k-1)}(\lambda) + P^{(t-k-1)}(\lambda) &= \sum_{i=t-k}^{\infty} \sigma_i \lambda^{-i}. \end{aligned} \quad (3)$$

Причём $\deg Q^{(i)}, P^{(i)} = i$.

Такие матричные многочлены принято называть приближениями Паде.

Теорема 2 Следующий алгоритм строит матричные аппроксимации Паде для произвольного ряда по отрицательным степеням формальной переменной с матричными коэффициентами над \mathbb{F} . Этот алгоритм может быть применён для построения пространства Крылова для произвольной матрицы с коэффициентами из \mathbb{F} .

$$Q^{(t+1)}(\lambda) = \sum_{i=0}^{t+1} Q_i^{(t+1)} \lambda^i$$

строится по формуле

$$\begin{pmatrix} O_n & O_n & O_n & \dots & \sigma_{t-k} \\ \dots & \dots & \dots & \dots & \dots \\ O_n & O_n & O_n & \dots & \sigma_{t-1} \\ \rho_{t+1} & O_n & \tau_t & \dots & \sigma_t \\ \rho_{t+2} & \rho_{t+1} & \tau_{t+1} & \dots & \sigma_{t+1} \\ Q_t^{(t)} & O_n & O_n & \dots & O_n \\ Q_{t-1}^{(t)} & Q_t^{(t)} & O_n & \dots & O_n \\ Q_{t-2}^{(t)} & Q_{t-1}^{(t)} & Q_{t-1}^{(t-1)} & \dots & O_n \\ \dots & \dots & \dots & \dots & \dots \\ Q_{t-k-2}^{(t)} & Q_{t-k-1}^{(t)} & Q_{t-k-1}^{(t-1)} & \dots & Q_{t-k-1}^{(t-k-1)} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_0 \\ \nu_{-1} \\ \cdot \\ \cdot \\ \cdot \\ \nu_{-k-1} \end{pmatrix} = \begin{pmatrix} O_n \\ \cdot \\ \cdot \\ \cdot \\ O_n \\ Q_{t+1}^{(t+1)} \\ \cdot \\ \cdot \\ \cdot \\ Q_{t-k-1}^{(t+1)} \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \quad (4)$$

Элементарными преобразованиями над \mathbb{F} со столбцами матричного многочлена $Q^{(i)}(\lambda)$, $i = t - k, \dots, t$, приводящими старший его коэффициент к виду, в котором его первые слева столбцы линейно независимы, а остальные нулевые, а затем умножением на λ столбцов матричного многочлена $Q^{(i)}(\lambda)$, соответствующих нулевым столбцам его старшего коэффициента, итеративно повторяя эту процедуру, добиваемся того, чтобы диагональные блоки нижней части левой матрицы в (4), кроме последнего, были невырождены. Число k должно быть таким, чтобы при $i = t - k, \dots, t$, приведение старшего коэффициента $Q^{(i)}(\lambda)$ к невырожденному виду, использовало не более $i - t + k + 1$ итераций, а при $i = t$, также как и при $i = t - 1$, не более k итераций. В качестве $\bar{\nu}$ берём базис ядра верхней части левой матрицы, содержащей коэффициенты правых частей равенств (3), тогда в нижней части справа получим коэффициенты приближения Паде с номером $t + 1$.

Предложенный в этой теореме алгоритм впервые дал рекуррентные формулы для построения приближений Паде к матричному ряду.

Идея доказательства следующей теоремы была предложена А.М.Зубковым.

Теорема 3 В случае $\mathbb{F} = \mathbb{GF}(2)$, и в предположении, что матричные коэффициенты Паде аппроксимаций случайны и статистически независимы матожидание

$$M(k) \leq 6, 233.$$

Определим A -скалярное произведение двух матричных многочленов $(\varphi(\lambda), \psi(\lambda)) \in \mathbb{F}^{ns \times ns}$ как коэффициент при λ^{-1} в произведении $\varphi(\lambda)^T \alpha(\lambda) \psi(\lambda)$. Обозначим $(C, D) = C^T A D \in \mathbb{F}^{ns \times ns}$ матрицу попарных A -скалярных произведений вектор-столбцов матриц $C, D \in \mathbb{F}^{N \times ns}$. Из этого определения и симметричности матрицы A следует в частности, что $(C, D) = (D, C)^T$.

Далее из матричных приближений Паде строятся A -ортогональные блоки, образующие пространство Крылова. Это делается при помощи формулы:

$$\begin{aligned} Q^{(s)}(\lambda) &= \sum_{i=0}^s Q_i^{(s)} \lambda^i \\ Q^{(s)}(A, B) &= \sum_{i=0}^s A^i B Q_i^{(s)} \end{aligned} \quad (5)$$

с использованием следующей леммы

Лемма 1 $(\varphi(A, B), \psi(A, B)) = (\varphi(\lambda), \psi(\lambda))$.

Отметим также, что

$$\text{если } Q(\lambda) = Q_1(\lambda)\lambda^i a_1 + Q_2(\lambda)\lambda^j a_2, \quad \text{то } Q(A, B) = A^i Q_1(A, B)a_1 + A^j Q_2(A, B)a_2,$$

$$\text{если } R(\lambda) = Q(\lambda)Q^{(k)}(\lambda), \quad \text{то } R(A, B) = Q(A, Q^{(k)}(A, B)),$$

$$\text{если } \alpha(\lambda)Q^{(k)}(\lambda) = \sum_{j=k+1}^{\infty} \mu_j \lambda^{-j}, \quad \text{то } \alpha(\lambda) (\varphi(\lambda)Q^{(k)}(\lambda)) = \sum_j \left(\sum_i \varphi_i \mu_{j+i} \right) \lambda^{-j},$$

Из равенства

$$\alpha(\lambda)Q^{(t)}(\lambda) + P^{(t)}(\lambda) = \sum_{i=t+1}^{\infty} \rho_i \lambda^{-i},$$

имеем, например, что

$$(Q^{(t)}(A, B), Q^{(t)}(A, B)) = (Q_t^{(t)})^T \rho_{t+1}.$$

Для получения линейной по размеру матрицы оценки сложности построения всех необходимых матричных приближений Паде, предложена процедура синхронизации. А именно, поскольку рекуррентные формулы и ортогонализация требуют использования не более 3-х предыдущих векторов, то после вычисления коэффициентов решения часть пространства Крылова можно «забыть» и начать строить приближения Паде снова с первой степени. В результате память не растёт и работа с большими многочленами (Видемае-Кошперсмит) не ведётся (лучшая оценка времени).

Для вычисления коэффициентов рекуррентных зависимостей (принадлежащих $\mathbb{F}^{ns \times ns}$) требуется вычисление базиса пространства Крылова и его ортогонализация (т.е. вычисление скалярных произведений). Для этого предлагается следующая вычислительная система

По сравнению с алгоритмом Монтгомери:

- Возможно распараллеливание высокого уровня (как в алгоритме Видемана-Кошперсмита). Можно вести речь о решении в интернете на вычислителях память которых вмещает саму задачу.

- Линейные комбинации, скалярные произведения, умножения на разреженную матрицу не чередуются, а группируются.
- Точка насыщения больше, чем для Монтгомери с блочным фактором. Меньше линейных комбинаций в 1,5 раза.

По сравнению с алгоритмом Видемана-Копперсмита:

- Требуемая память не растёт в процессе работы ("синхронизация" приводит к тому, что степень приближений не растёт). Есть возможность исключить использование кластера.
- Меньшая трудоёмкость.
- Проще анализ вероятности срабатывания.

Вот результаты замеров времени работы различных программ на 400 и 900 вычислительных узлах современных кластеров.

	400, МВС-1000 ($5,5 \cdot 10^6$)	400, Ломоносов ($33 \cdot 10^6$)	900, Ломоносов ($33 \cdot 10^6$)
Тесты	Р. Montgomery, Алгоритм-К	45 часов	-
	Р. Montgomery, J. Papadopoulos	-	80 часов
	Р. Montgomery, J. Papadopoulos, И.А. Поповян, Ю.В. Нестеренко	-	54 часа
	М.А. Черепнёв, Н.Л. Замарашкин	8 часов	<i>9,4 часов</i>

Наклонным шрифтом указано время, рассчитанное (эксперимент не проводился) исходя из того, что при фиксированном числе вычислительных узлов время растёт не быстрее, чем N^2 , а увеличение в k раз числа используемых вычислительных узлов даёт уменьшение времени как минимум в \sqrt{k} раз. Кроме того узлы "Ломоносова" в 36,8 раз мощнее узлов МВС-1000 (99,5 Гфлопс против 2,7 Гфлопс), отношение соответствующих констант c примерно равно 1,8, а $\sqrt{2,25} = 1,5$.

Разработан способ параллельных вычислений на системе с s узлами, для которой параллельная сложность (без пересылок) примет вид

$$\mathcal{O}\left(\frac{dN^2}{s} + \frac{N^2}{q} + Ns\right), \quad (6)$$

где q - параметр алгоритма, а число пересылок линейно зависит от q .