

# Крылов: библиотека алгоритмов и программ для решения СЛАУ\*

Д.С.Бутюгин, В.П.Ильин, Е.А.Ицкович, Д.В.Кныш, А.В.Петухов

*Институт вычислительной математики и  
математической геофизики СО РАН, Новосибирск*

Рассматривается концепция, архитектура и технологические принципы библиотеки алгоритмов и программ Крылов для решения СЛАУ с разреженными матрицами высокого порядка, возникающими при сеточных аппроксимациях многомерных краевых задач. Библиотека включает прямые и итерационные предобусловленные методы в подпространствах Крылова для симметричных и несимметричных, положительно определенных и знаконеопределенных алгебраических систем. Программные реализации на языке Фортран-90 активно используют вычислительные инструментарии библиотеки Intel MKL. Распараллеливание алгоритмов осуществляется средствами систем OpenMP и MPI на МВС с общей и распределенной памятью.

## 1 Введение

Современные итерационные алгоритмы для решения больших разреженных систем линейных алгебраических уравнений (СЛАУ) с матрицами достаточно общего вида: комплексными (эрмитовыми и неэрмитовыми) и вещественными, симметричными и несимметричными, положительно определенными и знаконеопределенными, основаны на вариационных методах в подпространствах Крылова и на предобуславливании СЛАУ. По этим вопросам имеется огромное количество журнальных статей и трудов конференций, но мы ограничимся ссылками на опубликованные за последние 15 лет монографии [1]–[4]. Эффективность и надежность (робастность) алгоритмов в значительной степени также зависит от масштабирования матриц, переупорядочивания неизвестных и различных приемов повышения вычислительной устойчивости, что формально также можно отнести к вариантам предобуславливания, см. [5] и цитируемую там литературу.

Ввиду особой актуальности алгебраических проблем в серьезных задачах наукоемкого математического моделирования программное обеспечение по решению СЛАУ насчитывает большое количество свободно распространяемых и коммерческих продуктов в форме библиотек или прикладных пакетов, в том числе распараллеливаемых на многопроцессорных вычислительных системах (МВС) с общей и/или распределенной памятью, среди которых можно назвать NETLIB, PETSc, SPARSEKIT, pARMS, hYPRE, SuperLU, SparSol, см. [6]–[12].

Концепция описываемой нами библиотеки Крылов заключается в единообразной модульной реализации представительного набора крыловских методов, как классических, так и разработанных в последние годы алгоритмов, основанных на  $A$ -ортогонализации Арнольди и Ланцоша, полусопряженных и бисопряженных направлений, использующих различные эффективные предобуславливатели и ориентированных на эффективное решение широкого класса СЛАУ, см. [4], а также работы [13]–[17], в которых приведен обширный экспериментальный материал.

---

\*Работа поддержана грантом РФФИ N 08-01-00526 и грантом ОМН РАН N 1.3.4.

Мы рассматриваем три группы алгоритмов, применяемых при решении сеточных СЛАУ, получаемых при аппроксимации многомерных краевых задач для систем дифференциальных уравнений (СДУ) или эквивалентных вариационных постановок методами конечных разностей, конечных объемов или конечных элементов (МКР, МКО, МЭ) различных порядков. Первая группа связана с решением задач с полностью или частично разделяющимися переменными и основана на кратных быстрых преобразованиях Фурье (БПФ), экономичных алгоритмах декомпозиции, редукции и комбинированных подходах, ориентированных на достижение рекордной производительности в актуальных характерных приложениях. Такие алгебраические системы и методы имеют самостоятельное значение во многих практических случаях, но могут также зачастую использоваться в качестве экономичных вспомогательных подзадач при решении более сложных проблем. Второй класс представляет семейство универсальных предобусловленных итерационных процессов в подпространствах Крылова, в том числе оригинальные методы полусопряженных и бисопряженных направлений для разного типа матриц (симметричных и несимметричных, положительно определенных и знаконеопределенных, полного и неполного ранга), представляемых в общепринятых сжатых форматах. Третья группа методов посвящена решению крупноблочных СЛАУ, включая алгебраические задачи с седловой точкой, получаемые при дискретизации векторных уравнений электромагнетизма, гидрогазодинамики и др. Здесь подходы основаны на использовании комбинации прямых и итерационных решателей с учетом специальных структурных и спектральных матричных свойств. Блочные алгоритмы лежат также в основе распространенных и хорошо распараллеливаемых методов декомпозиции, для которых типичными являются двухуровневые итерационные процессы.

Описываемые нами реализации алгебраических методов на многопроцессорных и многоядерных вычислительных системах с разделенной, общей и неоднородной многоуровневой памятью основываются на алгебраических принципах декомпозиции областей, минимизации коммуникационных потерь и на использовании высокопроизводительных инструментальных средств библиотеки MKL компании Intel. Распараллеливание алгоритмов осуществляется средствами MPI, OpenMP и гибридного программирования. Проблеме распараллеливания численных решений математического моделирования посвящено большое количество публикаций, см. [18, 20] и цитируемые там работы.

Рассматриваемые алгоритмы излагаются для простоты в применении к вещественной квадратной СЛАУ вида

$$Au = f, \quad A = \{a_{i,j}\} \in \mathcal{R}^{N,N}, \quad u = \{u_i\}, \quad f = \{f_i\} \in \mathcal{R}^N, \quad (1)$$

хотя они могут использоваться и для решения комплексных алгебраических систем.

Библиотека Krylov рассчитана на быстрое решение больших практических СЛАУ, с порядками до десятков и сотен миллионов, а также на возможное широкое использование в образовательном процессе по естественно-научным дисциплинам.

Данная работа построена следующим образом. Пункты 2–4 содержат краткое описание и классификацию основных методов Крылова, с акцентом на вычислительные, но не теоретические аспекты. Остальные пункты посвящены технологическим вопросам: модульная архитектура библиотеки, позволяющая легко разрабатывать и включать новые алгоритмы или их программные версии, типовые форматы матричных данных, поддерживаемые имеющимися конверторами и реализациями, принципы распараллеливания и оптимизации кода с использованием средств SPARSE BLAS библиотеки MKL, организация пользовательского интерфейса, тестирования алгоритмов и программ, и т.д.

## 2 Методы в подпространствах Крылова

Мы приводим краткие описания крыловских методов для решения несимметричных СЛАУ, классифицируя их по способам ортогонализации базисных векторов и по использованию вариационных принципов Галеркина или минимизации функционалов. Алгоритмы для симметричных систем следуют из них как частные случаи и хорошо освещены в литературе. Отметим, что реально методы Крылова применяются к предобусловленным СЛАУ, описываемым в следующем разделе. Для изложения крыловских алгоритмов это роли не играет, и ниже можно считать, что  $A$  и  $f$  – это не матрица и вектор правой части исходной системы (1), а полученные в результате предобуславливания.

### 2.1 Методы минимальных итераций

Общая идея построения итерационного решения СЛАУ с помощью ортогонализации или  $A$ -ортогонализации Арнольди заключается в представлении очередной итерации в виде линейной комбинации ортогональных векторов с помощью следующих единообразных выражений при  $s = 0, 1$ :

$$u^n = u^0 + y_1 v^1 + \dots + y_n v^n, \quad (v^n, A^s v^k) = d_n^{(s)} \delta_{k,n}, \quad d_n^{(s)} = (v^n, A^s v^n) \equiv \|v^n\|_s^2. \quad (2)$$

В применении к решению СЛАУ (1) при заданном начальном приближении  $u^0$  процесс ортогонализации определяется рекуррентными соотношениями

$$v^{n+1} = Av^n - \sum_{k=1}^n h_{k,n}^{(s)} v^k, \quad v^1 = r^0 = f - Au^0, \quad h_{k,n}^{(s)} = (Av^n, A^s v^k) / d_n^{(s)}. \quad (3)$$

Традиционно векторы  $v^k$  нормализуются, что можно представить формулами (3) с заменой векторов  $v^{n+1}$  и  $v^1$  на  $\tilde{v}^{n+1}$  и  $\tilde{v}^1$ , полагая дополнительно  $v^{n+1} = \tilde{v}^{n+1} / \|\tilde{v}^{n+1}\|$ ,  $v^1 = \tilde{v}^1 / \|\tilde{v}^1\|$  и  $d_n^{(s)} = 1$ . Однако эта процедура является необязательной и приводит к излишним вычислениям, в силу чего она в дальнейшем не используется. Векторы  $v^k$  составляют базис пространства Крылова

$$\mathcal{K}_{n+1}(r^0, A) = \text{span}\{v^1, \dots, v^{n+1}\} = \text{span}\{r^0, Ar^0, \dots, A^n r^0\}. \quad (4)$$

Введя матрицы (здесь и далее  $s = 0, 1$ , а сами индексы “ $s$ ” для краткости опускаем)

$$\bar{H}_n = \{h_{k,n}\} = \begin{bmatrix} H_n \\ e_n^t \end{bmatrix} \in \mathcal{R}^{n+1,n}, \quad H_n \in \mathcal{R}^{n,n}, \quad V_{n+1} = (v^1, \dots, v^{n+1}) \in \mathcal{R}^{N,n+1} \quad (5)$$

мы можем записать соотношения (индекс “ $t$ ” означает транспонирование)

$$V_n^t (A^s)^t V_n = D_n, \quad V_n^t (A^s)^t A V_n = D_n H_n, \quad D_n = \text{diag}\{d_i\} \in \mathcal{R}^{n,n}. \quad (6)$$

Здесь  $\bar{H}_n$  есть матрица Хессенберга ( $h_{k+1,k} = 1, h_{k,n} = 0$  для  $k > n+1$ ), а  $e_n^t$  – вектор-строка с единственным ненулевым (равным единице) элементом в последней  $n$ -й позиции. Из (2) имеем

$$u^n = u^0 + V_n y^n, \quad y^n = (y_1, \dots, y_n)^t, \quad r^n = r^0 - A V_n y^n, \quad (7)$$

где  $r^n = f - A u^n$  есть невязка  $n$ -й итераций. Для нахождения вектора коэффициентов  $y^n$  используются условия Галеркина (ортогональность или  $A$ -ортогональность  $r^n$  базисным векторам  $v^1, \dots, v^n$ ):

$$V_n^t (A^s)^t r^n = V_n^t (A^s)^t r^0 - V_n^t (A^s)^t A V_n y^n = d_1 e_1 - D_n H_n y^n = 0, \quad (8)$$

откуда получаем уравнение  $H_n y^n = e_1$ , решение которого находится методом Гаусса:

$$H_n = L_n U_n, \quad L_n z^n = e_1, \quad U_n y^n = z^n, \quad (9)$$

где  $U_n$  есть верхняя треугольная, а  $L_n$  – двухдиагональная нижняя треугольная матрица с единичной главной диагональю. Отметим, что в силу выражений

$$r^n = v^1 - V_{n+1} \bar{H}_n y^n = v^1 - V_n H_n y^n - v^{n+1} e_n^t y^n = -v^{n+1} y_n \quad (10)$$

норма  $r^n$  определяется выражением  $\|r^n\|_2^2 \equiv (r^n, r^n) = (v^{n+1}, v^{n+1}) y_n^2$ , которое позволяет контролировать условие окончания итераций  $\|r^n\|_2^2 \leq (f, f) \varepsilon^2$ ,  $\varepsilon \ll 1$ , без вычисления векторов невязки.

Рассмотренные алгоритмы при  $s = 0, 1$  мы будем называть методами минимальных и  $A$ -минимальных итераций соответственно (при  $s = 0$  и ортогонализации векторов  $v^k$  получаем известный метод Арнольди, или FOM – Full Orthogonalization Method). Отметим, что вычисляемые итерационные приближения  $u^n$  в (2), (7) не меняются от того, применяется ли процесс нормализации векторов  $v^n$  в (3) или нет, поскольку это приводит только к взаимно сокращающимся масштабированиям множителей в матрично-векторном произведении  $V_n y^n$ .

## 2.2 Методы минимальных невязок

Одним из самых популярных алгоритмов решения несимметричных СЛАУ является метод обобщенных минимальных невязок (GMRES), который основан на тех же формулах (2)–(7) при  $s = 0$ . Алгоритм, получаемый из них при  $s = 1$ , будем называть по аналогии методом обобщенных  $A$ -минимальных невязок (A-GMRES). При этом в обоих последних алгоритмах базисные векторы  $v^k$  строим без их нормализации.

В отличие от предыдущих, методы GMRES и A-GMRES вместо принципа Галеркина для нахождения вектора  $y^n$  используют условие минимизации функционалов  $\Phi_s(r^n) = (r^n, A^s r^n)$ , для которых в силу соотношения  $r^n = V_{n+1}(e_1 - \bar{H}_n y^n)$  имеем:

$$\begin{aligned} \Phi_s(r^n) &= (V_{n+1}^t (A^s)^t V_{n+1} (e_1 - \bar{H}_n y^n), (e_1 - \bar{H}_n y^n)) = \\ &= \|d_1^{1/2} e_1 - D_{n+1}^{1/2} \bar{H}_n y^n\|_2^2 = \|\bar{g}^n - \bar{R}_n y^n\|_2^2, \quad \bar{g}^n = (g^n, g_{n+1})^t \in \mathcal{R}^{n+1}, \quad g^n \in \mathcal{R}^n. \end{aligned} \quad (11)$$

Здесь предполагается положительная определенность матрицы  $D_{n+1}$ , а также введены обозначения

$$\bar{g}^n = Q_{n+1} d_1^{1/2} e_1, \quad \bar{R}_n = Q_{n+1} D_{n+1}^{1/2} \bar{H}_n, \quad Q_{n+1} = G_n \cdots G_1, \quad (12)$$

где  $G_k \in \mathcal{R}^{n+1, n+1}$  суть элементарные ортогональные матрицы вращения Гивенса, с помощью которых хессенбергова матрица  $D_{n+1/2}^{1/2} \bar{H}_n$  преобразуется в верхнюю треугольную  $\bar{R}_n = \begin{bmatrix} R_n \\ e_n^t \end{bmatrix}$ ,  $R_n \in \mathcal{R}^{n, n}$ .

Очевидно, что метод наименьших квадратов обеспечивает минимум функционала

$$\min_{y^n} \{\Phi_s(r^n)\} = \min_{y^n} \{\|g^n - R_n y^n\| + g_{n+1}^2\} = g_{n+1}^2,$$

который дает однозначное решение  $y^n = R_n^{-1} g^n$ .

## 2.3 Методы квазiminимальных невязок

Альтернативой к ортогонализации Арнольди с “длинными” рекурсиями для построения базисов подпространств Крылова является процесс биортогонализации Ланцоша, который описывается следующими трехчленными соотношениями при  $s = 0$  и  $n = 1, 2, \dots$ :

$$v^{n+1} = A v^n - \alpha_n^{(s)} v^n - \beta_n^{(s)} v^{n-1}, \quad w^{n+1} = A^t w^n - \alpha_n^{(s)} w^n - \beta_n^{(s)} w^{n-1}, \quad (13)$$

где  $\beta_1^{(s)} = 0$  и для простоты полагается  $v^1 = w^1 = r^0 \equiv f - Au^0$ , а коэффициенты  $\alpha_n^{(s)}$  и  $\beta_n^{(s)}$  определяются из условий  $(v^n, A^s w^k) = d_n^{(s)} \delta_{n,k}$ ,  $d_n^{(s)} = (v^n, A^s w^n)$  с помощью формул

$$\alpha_n^{(s)} = (Av^n, A^s w^n)/d_n^{(s)}, \quad \beta_n^{(s)} = (Av^n, A^s w^{n-1})/d_{n-1}^{(s)}. \quad (14)$$

Из соотношений (13) для  $s = 0, 1$  следуют выражения

$$\begin{aligned} AV_n &= V_n T_n + v^{n+1} e_n^t = V_{n+1} \bar{T}_n, & \bar{T}_n &\in \mathcal{R}^{n+1,n}, \\ A^t W_n &= W_n T_n + w^{n+1} e_n^t = W_{n+1} \bar{T}_n, & W_n &\in \mathcal{R}^{N,n}, \end{aligned} \quad (15)$$

где  $T_n = \{t_{k,n}\}$  есть квадратная трехдиагональная матрица с элементами  $t_{k+1,k} = 1$ ,  $t_{k,k} = \alpha_k$ ,  $t_{k,k+1} = \beta_k$  и  $\bar{T}_n = \begin{bmatrix} T_n \\ e_n^t \end{bmatrix}$ ,  $W_n = [w^1, \dots, w^n]$ , причем для введенных матриц выполняются равенства

$$W_n^t A^s AV_n = T_n, \quad W_n^t A^s V_n = D_n = \text{diag}\{d_n^{(s)}\}. \quad (16)$$

Здесь и далее индексы  $s$  для краткости опускаются, а рекуррентные формулы (13) при  $s = 1$  будем называть процессом  $A$ -биортогонализации.

Построение итерационных приближений, как и ранее, проводится с помощью формул (7). При этом для векторов невязки справедливо соотношение

$$r^n = v^1 - V_{n+1} \bar{T}_n y^n = V_{n+1} (e_1 - \bar{T}_n y^n) \equiv V_{n+1} z^n, \quad (17)$$

причем матрица  $V_{n+1}$ , в отличие от предыдущих случаев, не является ортогональной, что затрудняет решение проблемы минимизации нормы невязки  $r^n$ .

Определяя вектор коэффициентов  $y^n$  по условию минимума нормы вектора  $z^n$  из (17), мы при  $s=0$  приходим к известному методу квазимиимальных невязок QMR (Quasi - Minimal Residual), если только векторы  $v^n$  и  $w^n$  из (13) подвергнуть дополнительному (но необязательному) процессу нормализации. Алгоритм, получаемый из аналогичных выражений при  $s = 1$ , будем называть методом  $A$ -квазимиимальных невязок и обозначать A-QMR. В обоих случаях методом наименьших квадратов ищется решение задачи

$$y^n = \arg\{\min_y \|z^n\|\}, \quad (18)$$

которое, как и в методах GMRES, A-GMRES, находится с помощью ортогональных матриц вращения Гивенса из (12):

$$\begin{aligned} \|z^n\|_2^2 &= \|\bar{g}^n - \bar{R}_n y^n\|_2^2, & \bar{R}_n &= Q_{n+1} \bar{T}_n = \begin{bmatrix} R_n \\ 0 \end{bmatrix} \in \mathcal{R}^{n+1,n}, \\ R_n &\in \mathcal{R}^{n,n}, & \bar{g}^n &= Q_{n+1} e_1 = (g^n g_{n+1})^t \in \mathcal{R}^{n+1}, & g^n &\in \mathcal{R}^n, \end{aligned} \quad (19)$$

где  $Q_{n+1} = G_n \cdots G_1 \in \mathcal{R}^{n+1,n+1}$ ,  $\bar{R}_n$  и  $R_n$  суть треугольные трехдиагональные матрицы, а  $G_k$  – элементарные матрицы вращения Гивенса (почти единичные, с четырьмя отличными от нуля и единицы элементами:  $g_{k,k} = g_{k+1,k+1} = c_k$ ,  $g_{k,k+1} = -g_{k+1,k} = s_k$ ,  $c_k^2 + s_k^2 = 1$ ) элементы которых  $c_k, s_k$  подбираются по условию “зануления”  $t_{k+1,k}$ .

Искомый результат получается следующим образом:

$$y^n = R_n^{-1} g^n, \quad \min\{\|z\|_2^2\} = g_{n+1}^2. \quad (20)$$

## 2.4 Методы бисопряженных направлений

Метод биортогонализации Ланцоша, в несколько отличной от (13) реализации, порождает метод бисопряженных градиентов (BiCG), описываемый следующими формулами при  $s = 0$ :

$$\begin{aligned} u^{n+1} &= u_n + \alpha_n^{(s)} p^n, & r^0 &= \tilde{r}^0 = p^0 = \tilde{p}^0 = f - Au^0, \\ r^{n+1} &= r^n - \alpha_n^{(s)} Ap^n, & \tilde{r}^{n+1} &= \tilde{r}^n - \alpha_n^{(s)} A^t \tilde{p}^n, \\ p^{n+1} &= r^{n+1} + \beta_n^{(s)} p^n, & \tilde{p}^{n+1} &= \tilde{r}^{n+1} + \beta_n^{(s)} \tilde{p}^n, \end{aligned} \quad (21)$$

где  $\tilde{r}^n$  и  $\tilde{p}^n$  суть двойственные векторы и коррекции, а коэффициенты  $\alpha_n^{(s)}, \beta_n^{(s)}$  по условиям биортогонализации векторов

$$\begin{aligned} (A^s p^n, A^t \tilde{p}^k) &= \rho_n^{(s)} \delta_{k,n}, & \rho_n^{(s)} &= (A^s p^n, A^t \tilde{p}^n), \\ (A^s r^n, \tilde{p}^k) &= (A^s r^n, \tilde{p}^n) \delta_{k,n}, & (A^s p^k, \tilde{r}^n) &= (A^s p^n, \tilde{r}^n) \delta_{k,n}, \\ (A^s r^n, \tilde{r}^k) &= \sigma_n^{(s)} \delta_{k,n}, & \sigma_n^{(s)} &= (A^s r^n, \tilde{r}^n), \end{aligned} \quad (22)$$

определяются формулами

$$\alpha_n^{(s)} = \sigma_n^{(s)} / \rho_n^{(s)}, \quad \beta_n^{(s)} = \sigma_n^{(s)} / \sigma_n^{(s)}. \quad (23)$$

При  $s = 1$  приведенные соотношения определяют метод бисопряженных невязок (BiCR, см. [14]).

Данные итерационные процессы могут вырождаться при обращении  $\rho_n^{(s)}$  в нуль (или “почти нуль”), во избежание чего применяется процедура “заглядывания вперед” (look ahead, см. [2], [3]).

В методах BiCG и BiCR векторы невязок могут быть представлены в виде

$$r^n = \varphi_n^{(s)}(A) r^0, \quad \tilde{r}^n = \varphi_n^{(s)}(A^t) \tilde{r}^0,$$

где  $\varphi_n^{(s)}(t)$  – некоторый многочлен степени  $n$ . Отсюда могут быть построены сдвоенные (squared) методы бисопряженных градиентов и бисопряженных невязок (SCG и SCR), реализующие векторы невязок

$$r^n = \Phi_n^{(s)}(A) r^0, \quad \Phi_n^{(s)}(t) = (\text{varphi}_n^{(s)}(t))^2$$

и дающие в “хороших” случаях двукратное ускорение сходимости. Реализация этих методов вместо умножения на транспонированную матрицу  $A^t$  на каждой итерации использует повторное умножение на  $A$ .

Отметим еще стабилизированные методы бисопряженных градиентов и бисопряженных невязок (BiCGStab и BiCRStab), обладающие повышенной устойчивостью за счет введения дополнительного итерационного параметра, вычисляемого из условия локальной (но глобальной!) минимизации нормы вектора невязки, см. [4], [14].

Отметим, что формулы (21) для симметричных СЛАУ не требуют вычисления векторов  $\tilde{r}^n, \tilde{p}^n$  и переходят в методы сопряженных градиентов и сопряженных невязок (CG и CR), обеспечивающие на каждой итерации минимизацию функционалов  $(A^{s-1} r^n, r^n)$  в подпространствах Крылова.

## 2.5 Вопросы эффективности и надежности алгоритмов

С точки зрения практической эффективной применимости рассмотренных методов в подпространствах Крылова, существенное значение имеют различные алгоритмические аспекты

- а. Оценки скорости сходимости и числа итераций, необходимого для достижения необходимой точности итерационного приближения, можно получить только для методов, обладающих

свойством минимизации функционалов (GMRES, A-GMRES, QMR, A-QMR) и только для СЛАУ с диагонализруемыми матрицами, т.е. представимыми в виде

$$A = X\Lambda X^{-1}, \quad \Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\},$$

где  $\lambda_i$  суть собственные числа  $A$ , а  $X$  – квадратная матрица, столбцы которой являются соответствующими собственными векторами. При этом количество итераций  $n(\varepsilon)$ , обеспечивающее выполнение неравенства  $\Phi_s(r^n) \leq \varepsilon^2 \Phi_s(r^0)$ , где  $\Phi_s(r^n) = (A^s r^n, r^n)$  для алгоритмов GMRES, A-GMRES и  $\Phi_s(r^n, \bar{r}^n) = (A^s r^n, \bar{r}^n)$  для QMR и A-QMR, определяется неравенством

$$n(\varepsilon) \leq |\ln(\varepsilon)/\text{cond}(X)|/|\ln(\gamma_1/\gamma_2)|.$$

Здесь  $\text{cond}(X)$  – число обусловленности матрицы  $X$ , а  $\gamma_1$  и  $\gamma_2$  – параметры эллипса в комплексной плоскости, содержащего все собственные значения  $\lambda_i$ , см. подробнее [2], [4]. Отметим, что наряду с данной “классической” оценкой, имеются и другие подходы к исследованию сходимости вариационных методов, основанных, в частности, на понятии  $K$ -обусловленности матриц, см. [5] и цитируемые там работы.

- б. Рассмотренные выше методы с “длинными” рекурсиями при большом числе итераций  $n$  требуют слишком много вычислительных ресурсов для решения СЛАУ сверхвысоких порядков, в первую очередь из-за требований к оперативной памяти для хранения базисных векторов  $v^k$ . В этих случаях применяется два подхода – периодические рестарты или ограниченная ортогонализация, – которые формально можно объединить следующим образом. Первый прием заключается в том, что на итерациях с номерами  $n = n_r \equiv \left[ \frac{n}{m_r} \right]$  ( $m_r$  – заданный период рестартов, а  $[m]$  – целая часть  $m$ ) последовательность векторов  $v^k$  обрывается и начинается снова с переопределением  $v^1 = r^{n_r} = f - Au^{n_r}$ . Ограничение ортогонализации состоит в использовании в разложении (2) только  $m_0$  последних базисных векторов и “забывании” остальных, т.е.  $u^n = u^0 + \sum_{k=n-m_0+1}^n y_k v^k$ . Объединение данных подходов при  $m_r > m_0$  основано и на отбрасывании “слишком старых” векторов, и на проведении рестартов.
- в. Известно, что процесс ортогонализации в конечнозначной арифметике приводит к эффекту “разортогонализации” векторов, который оказывается критическим при вычислении длинных последовательностей. Основные способы для борьбы с такой неустойчивостью – это проведение переортогонализации векторов или использование матричных операций отражения Хаусхолдера вместо вращений Гивенса [2]. Отметим, что в приведенных выше формулах используется модифицированная ортогонализация Грама–Шмидта, которая гораздо устойчивей его исходного варианта.

### 3 Алгоритмы предобуславливания СЛАУ

Итерационные методы Крылова мы рассматриваем в сочетании с предобуславливанием СЛАУ, значительно повышающим эффективность алгоритмов. Такое предобуславливание можно реализовывать несколькими способами. Пусть, например,  $B$  есть легко обрабатываемая матрица со свойством  $\text{cond}(B^{-1}A) \ll \text{cond}(A)$ . Одностороннее предобуславливание осуществляется путем умножения исходной системы (1) на  $B^{-1}$ , в результате чего получаем предобусловленную СЛАУ

$$\bar{A}u \equiv B^{-1}Au = B^{-1}f \equiv \bar{f}. \quad (24)$$

Применение рассмотренных выше методов к ее решению сводится всего лишь к замене  $A$  и  $f$  в приведенных формулах на  $\bar{A}$  и  $\bar{f}$  соответственно. Если для предобуславливающей матрицы существует экономичная факторизация  $B = L_B U_B$ , то зачастую удобнее использовать двустороннее предобуславливание:

$$\bar{A}\bar{u} \equiv L_B^{-1} A U_B^{-1} \bar{u} = L_B^{-1} f \equiv \bar{f}, \quad \bar{u} = U_B u. \quad (25)$$

Пусть, наконец, для решения (1) имеется некоторый сходящийся итерационный процесс

$$u^{n+1} = T u^n + g, \quad u^n \xrightarrow{n \rightarrow \infty} u. \quad (26)$$

Тогда мы фактически имеем предобусловленную СЛАУ

$$\bar{A}u \equiv (I - T)u = g \equiv \bar{f}, \quad (27)$$

к решению которой также эффективно применимы методы в подпространствах Крылова. При этом явное представление матрицы  $T$  и вектора  $g$  не требуется, а на каждом шаге крыловского метода нужно фактически проведение одной итерации вида (26).

В качестве примера двустороннего предобуславливания рассмотрим метод неполной факторизации в экономичной модификации Айзенштата (ниже предполагается  $A = D - L - U$ , где  $D$ ,  $L$  и  $U$  – диагональная, нижняя и верхняя треугольные матрицы):

$$L_B = (G - L)G^{-1/2}(G - U), \quad G = \frac{1}{\omega}D - \theta S, \quad S e = \left( \frac{1 - \omega}{\omega} D + L G^{-1} U \right) e, \quad (28)$$

где  $G$  и  $S$  – диагональные матрицы,  $e$  – вектор из единичных компонент, а  $\omega$  и  $\theta$  – итерационные параметры. В этом случае матрица  $\bar{A}$  из (25) может быть представлена в виде

$$\begin{aligned} \bar{A} &= (I - \bar{L})^{-1} + (I - \bar{U})^{-1} - (I - \bar{L})^{-1}(2I - \bar{D})(I - \bar{U})^{-1}, \\ \bar{L} &= G^{-1/2} L G^{-1/2}, \quad \bar{U} = G^{-1/2} U G^{-1/2}, \quad \bar{D} = G^{-1/2} D G^{-1/2}, \quad \bar{u} = (I - \bar{U})G^{-1/2} u, \end{aligned} \quad (29)$$

и умножение на нее требует практически столько же арифметических операций ( $I$  – единичная матрица), что и на исходную матрицу  $A$ .

Рассмотрим теперь методы полусопряженных градиентов и полусопряженных невязок (SCG и SCR), которые по вариационным свойствам в определенном смысле эквивалентны описанным в пп. 2.1, 2.2 алгоритмам, но допускают простое введение динамических предобуславливающих матриц  $B_n$ , которые могут меняться на каждой итерации. Данные методы описываются следующими формулами при  $s = 0, 1$  (см. подробнее [4], [16]):

$$\begin{aligned} r^0 &= f - A u^0, \quad u^{n+1} = u^n + \alpha_n^{(s)} p^n, \\ r^{n+1} &= r^n - \alpha_n^{(s)} A p^n, \quad p^0 = B_0^{-1} r^0, \end{aligned} \quad (30)$$

где коэффициенты  $\alpha_n^{(s)}$  по условиям полусопряженности векторов

$$(A p^n, A^s p^k) = \rho_n^{(s)} \delta_{n,k}, \quad \rho_n^{(s)} = (A p^n, A^s p^n), \quad k = 0, 1, \dots, n-1,$$

определяются соотношениями  $\alpha_n^{(s)} = (A^s B_n^{-1} r^n, r^n) / \rho_n^{(s)}$ .

Корректирующие векторы  $p^n$  вычисляются из “длинных” рекурсий

$$\begin{aligned} p^{n+1} &= r^{n+1} + \sum_{k=0}^n \beta_{n,k}^{(s)} p^k = p^{n+1,k} + \sum_{j=k}^n \beta_{n,j} p^j, \\ p^{n+1,0} &= B_{n+1}^{-1} r^{n+1}, \quad p^{n+1,k} = p^{n+1,k-1} + \beta_{n,k-1} p^{k+1}, \quad p^{n+1} = p^{n+1,n+1}. \end{aligned} \quad (31)$$



Здесь коэффициенты  $\beta_{n,j}^{(s)}$  из условий обобщенной полусопряженности векторов невязок

$$(A^s B_k^{-1} r^k, r^n) = \sigma_n^{(s)} \delta_{n,k}, \quad \sigma_n^{(s)} = (A^s B_n^{-1} r^n, r^n), \quad k < n,$$

находятся с помощью формул

$$\beta_{n,j}^{(s)} = -(A^s p^j, A p^{n,j}) / \rho_n^{(s)}, \quad j = 0, 1, \dots, n.$$

#### 4 Решение задач с разделяющимися переменными

$$\bar{A}\bar{u} = \bar{f} \equiv \{f_{i,j,k}\}, \quad \bar{u} = \{u_{i,j,k}\}, \quad \bar{A} = \bar{A}^x + \bar{A}^y + \bar{A}^z, \quad (32)$$

$$\bar{A}^z \bar{A}^{xy} = \bar{A}^{xy} \bar{A}^z, \quad \bar{A}^{xy} = \bar{A}^x + \bar{A}^y \quad (33)$$

$$\begin{aligned} (A_r^x + A_r^y) v^r &= f^r, \quad e = 1, \dots, N, \\ v^r &= \{v_{i,j}^r\}, \quad f^r = \{f_{i,j}^r\}, \quad i, j = 1, \dots, N, \\ A_r^x &= A^x + 0.5\lambda_r^z I, \quad A_r^y = A^y + 0.5\lambda_r^z I, \end{aligned} \quad (34)$$

$$\begin{aligned} (\gamma_n I + A_r^x) v_r^{-1/2} &= (\gamma_n I + A_r^y) v_r^{n-1} + f^r, \\ (\gamma_n I + A_r^y) w_r^n &= 2(v_r^{n-1/2} - v_r^n), \\ v_r^n &= v_r^{n-1} + w_r^n, \quad n = 1, 2, \dots \end{aligned} \quad (35)$$

$$(Au)_p = -A_p u_{p-1} + B_p u_p - C_p u_{p+1} = f_p, \quad p = 1, 2, \dots, P, \quad A_1 = C_p = 0, \quad (36)$$

$$\begin{aligned} u_{p+1} &= B_{p+1}^{-1} (f_{p+1} - A_{p+1} u_p - C_{p+1} u_{p+2}) = w_{p+1} - \hat{v}_{p+1} u_p - \check{v}_{p+1} u_{p+2}, \\ B_{p+1} \hat{v}_{p+1} &= [1, \dots, 0]^T a, \quad B_{p+1} \check{v}_{p+1} = [0, \dots, 1]^T c \end{aligned} \quad (37)$$

$$\begin{aligned} \bar{a}_p u_{p-2} + \bar{b}_p u_p + \bar{c}_p u_{p+2} &= f_p, \\ \bar{a}_p &= A_p B_{p-1}^{-1} A_{p-1}, \quad \bar{b}_p = B_p - A_p B_{p-1}^{-1} C_{p-1} - C_p B_{p+1}^{-1} A_{p+1}, \quad \bar{c}_p = C_p B_{p+1}^{-1} C_{p+1} \end{aligned} \quad (38)$$

#### 5 Модульная структура алгоритмов

#### 6 Матричные форматы данных

#### 7 Распараллеливание алгоритмов

#### 8 Пользовательский интерфейс

#### 9 Технологии программирования и тестирования

#### Список литературы

- [1] Axelsson O. Iterative solution methods.-N.Y.: Cambridge Univ. Press., 1994.
- [2] Saad Y. Iterative methods for sparse linear systems.-NY: PWS Publish., 1996.
- [3] Greenbaum A. Iterative methods for solving linear systems.-SIAM Pull., Philadelphia, 1997.

- [4] Ильин В.П. Методы и технологии конечных элементов.-Новосибирск: Изд. ИВМиМГ СО РАН, 2007.
- [5] Kaporin I.E. Scaling, reordering and diagonal pivoting in ILU preconditioning.– Rus. J. Num. Anal. Math. Mod., v. 22, N 4, 2007, 341-376.
- [6] Barrett R., Berry M., Chan T., Demmel J., Donatto J., Dongarra J., Eijkhout V., Pozo R., Romine Ch., Van der Vorst H. Templates for the solution of linear systems: Building Blocks for Iterative Methods (electronic version).
- [7] PETSc Home Portable, Extensible Toolkit for Scientific Computation. <http://www-unix.mcs.anl.gov/petsc/>
- [8] Saad Y. SPARSEKIT: A basic tool for sparse matrix computations.– Tech. Report N 90-20, Research Institute for Advanced Computer Science. NASA Ames Research Center, Moffet field, CA, 1990.
- [9] Saad Y., Sosonkina M. pARMS: A package for the parallel iterative solution of general large sparse linear systems user's guide.–Report UMSI 2004-8, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, MN, 2004.
- [10] Xiaoye S.Li. An overview of SuperLU: algorithms, implementation and user interface.–ACM Trans. on Math. Software, v. 31, N 2, 2005, 302-325.
- [11] Falgout R.D., Yang U.M. hypre: A library of high performance preconditioners. Lecture Notes in Computer Science, v. 2331, 2002, 632-641.
- [12] Diyankov O.V., Koshelev S.V., Kotegov, et al. SparSol: sparselinear systems solver.–Rus. J. Num. Anal. Math. Mod., v.22, N 4, 2007, 325-339.
- [13] П'ин В.П. Methods of semiconjugate directions.–Russ. J. Numer. Anal. Math. Modelling., v. 23, N 4, 2008, 369-387.
- [14] Ильин В.П. Методы бисопряженных направлений в подпространствах Крылова.–СибЖИМ, т.11, N 4, 2008, 47-60.
- [15] Ильин В.П. О методах сопряженных и полусопряженных направлений с преобуславливающими проекторами.–ДАН, т. 419, N 3, 2008, 303-306.
- [16] Ильин В.П., Ицкович Е.А. О методах полусопряженных направлений с динамическим преобуславливанием.–СибЖИМ, т. 10, N 4, 2007, 41-54.
- [17] Ильин В.П. Об итерационном методе Качмажа и его обобщениях.– СибЖИМ, т. 9, N 3, 2006, 39-49.
- [18] Ильин В.П. Параллельные алгоритмы для больших прикладных задач: проблемы и технологии.//Автометрия, N 2, 2007, 3-21.
- [19] Ильин В.П. Проблемы высокопроизводительных технологий решения больших разреженных СЛАУ.–Вычислительные методы и программирование, т. 10, 2009, 141-147.
- [20] Ильин В.П., Кныш Д.В. Параллельные алгоритмы решения разделяющихся краевых задач.– Труды конфер. ПАВТ-2008, Санкт-Петербург, изд. Политехн. ун-та (СПбПУ), 2008, 107-118.