О проблемах параллельного решения больших СЛАУ *

В.П.Ильин

Институт вычислительной математики и математической геофизики СО РАН, Новосибирский государственный университет

Аннотация

1 Введение

Как известно, решение больших СЛАУ является узким местом в задачах математического моделирования, поскольку ресурсоемкость данной вычислительной стадии нелинейно растет с увеличением числа степеней свободы. Понятие больших и даже сверхбольших алгебраических систем можно определить как рекордных для своего времени, которые на сегодня можно связывать с размерностями порядка $10^{10} - 10^{12}$, требующих для своей реализации МВС с числом процессоров около $10^5 - 10^7$. Понятно, что для СЛАУ таких размеров реально применять только итерационные методы, так как прямые алгоритмы требуют слишком больших компьютерных ресурсов, особенно памяти.

Главные подходы к решению рассматриваемых задач заключаются в построении предобусловленных итерационных процессов в подпространствах Крылова на основе алгебраической декомпозиции области на подобласти с параметризованными по ширине пересечениями и различными интерфейсными условиями на внутренних границах смежных подобластей. При этом вспомогательные задачи в подобластях решаются синхронно на соответствующих вычислительных устройствах или прямыми, или итерационными приближенными методами.

Основные требования к рассматриваемым алгоритмам являются внутренне противоречивыми. С одной стороны, для увеличения коэффициента ускорения необходимо иметь большое количество подобластей, чтобы задействовать одновременную массовую работу процессоров. Но при этом, в случае применения простейших “стандартных” подходов, значительно растем число внешних итераций по подобластям. Последнее обстоятельство инициирует непрерывное появление большого количества исследований по ускорению этого итерационного процесса, связанных, как правило, с его существенным логическим усложнением.

В целом можно выделить три основных направления развития итерационных методов в последние десятилетия. К первому можно отнести обобщение концепции подпространств Крылова, в которых последовательные приближения конструируются из вариационных, ортогональных и/или проекционных принципов. Здесь можно назвать методы полусопряженных направлений [1] – [3] для решения несимметричных СЛАУ, рациональные крыловские подпространства [4], [5] (в отличие от классических полиномиальных), методы индуцированной редукции размерности IDR(s) [6], [7] (Induced Dimension Reduction), а также

*)Работа поддержана грантами РНФ N 14-11-00485 и РФФИ N 14-07-00128.

алгоритмы дефляционного типа [8], [9] с пополнением крыловского базиса некоторыми дополнительными векторами со специальными свойствами.

Второе направление связано с улучшением обусловленности решаемых СЛАУ за счет использования новых видов предобуславливающих матриц за счет применения методов вложенных сечений и многофронтальных алгоритмов матричных разложений, ассоциированных со специальными упорядоченностями неизвестных. Здесь перспективные подходы основываются на тензорном аппарате [10] и представлении матриц в HSS-формате ([11], Hierarchically Semi-Separable), которые дают эффективные приближенные и/или точные алгоритмы матричных факторизаций. Кроме того, ускорение крыловских итерационных процессов сравнительно недавно предложено осуществлять за счет мультипредобуславливания (см. [12] – [14]), предполагающего подключение нескольких предобуславливающих матриц на каждом итерационном шаге. Данный прием приводит к обобщению известного динамического, или гибкого (flexible), предобуславливания [15], в котором предобуславливатель на текущей итерации только один, но он меняется от шага к шагу.

Третье направление связано с первыми двумя, но оно целевым образом ориентировано на масштабируемый параллелизм с отображением алгоритмов на архитектуру МВС. В данном случае главным орудием является метод декомпозиции сеточных областей, в его геометрической или алгебраической интерпретации. Интуитивно ясно, что многомерную сеточную расчетную область целесообразно разбивать сбалансированным образом на подобласти той же размерности. Конечная же производительность алгоритмов в значительной степени определяется вычислительно-информационными технологиями и качеством программной реализации, так что эффективность распараллеливания необходимо характеризовать не только теоретическими оценками, но и результатами экспериментальных исследований.

Данная работа построена следующим образом. В п. 2 описываются блочные методы полусопряженных направлений с мультипредобуславливанием в абстрактной векторно-матричной форме. В следующем разделе представлена их конкретизация для параллельных алгоритмов декомпозиции подобластей со сравнительным анализом различных структурных подходов типа FETI, BNN, см. [16], [17], и предлагаемых вариантов без подобластей – разделителей, а также некоторых принципов ускорения крыловских процессов на основе агрегации, дефляции и грубосеточной коррекции, осуществляющих разные применения малоранговой аппроксимации матриц, описанных ранее в [18] – [21]. Последний пункт посвящен краткому изложению концепции библиотеки программ KRYLOV [22] как открытого интегрированного инструментария, ориентированного не только на конечных пользователей, заинтересованных в практическом использовании быстрых параллельных алгебраических решателей, но и предоставляющего средства автоматизации построения алгоритмов для разработки и экспериментального исследования новых методов. В заключении приводятся соображения об актуализации цепочки: "рождение идеи и теоретическое исследование методов" – "пробные версии и экспериментальная апробация" – "высокопроизводительная реализация алгоритмов и их практическое внедрение".

2 Блочные методы полусопряженных направлений

Методы полусопряженных градиентов и полусопряженных невязок (SCG, SCR, см. [1, 2]) предназначены для решения вещественных СЛАУ с несимметричными положительно определенными матрицами

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

которые для всех вещественных векторов v удовлетворяют условиям

$$(Av, v) \geq \delta ||v||^2, \ \ \delta > 0, \ \ (v, w) = \sum_{i=1}^{N} v_i w_i, \ \ ||v||^2 = (v, v). \tag{2}$$

В отличие от обобщенных методов минимальных невязок [3], данное семейство алгоритмов полусопряженных направлений обходится без формирования матрицы Хессенберга, что заметно упрощает вычислительную схему и программную реализацию. Как известно, приведенные неравенства обеспечивают положительную определенность симметричной части матрицы $A = \{a_{i,j}\}$:

$$(A^s u, u) \geq \delta(u, u), \ \ A^s = (A + A^T)/2, \ \ A^T = \{a_{j,i}\},$$

а также положительность вещественной части собственных чисел матрицы A (верхний индекс “T” означает транспонирование матрицы).

Для решения уравнения (1) применим итерационный процесс обобщенного, или блочного, крыловского типа:

$$\begin{array}{l} r^0 = f - Au^0, \ \ u^{n+1} = u^n + P_n \bar{\alpha}_n, \ \ n = 0, 1, ... \\ r^{n+1} = r^n - AP_n\bar{\alpha}_n = r^q - AP_q\bar{\alpha}_q - ... - AP_n\bar{\alpha}_n, \ \ 0 \leq q \leq n, \end{array} \tag{3}$$

где $P_n = (p_1^n ... p_{M_n}^n) \in \mathcal{R}^{N,M_n}$ есть составленная из направляющих векторов p_k^n матрица, а $\bar{\alpha}_n = (\alpha_{n,1} ... \alpha_{n,M_n})^T$ – вектор итерационных параметров, которые будем определять из следующих ортогональных свойств:

$$\begin{array}{l} (Ap_k^n, A^\gamma p_{k'}^{n'}) = \rho_{n,k}^{(\gamma)} \delta_{n,n'}^{k,k'}, \ \ \rho_{n,k}^{(\gamma)} = (Ap_k^n, A^\gamma p_k^n), \\ \gamma = 0, 1, ; n' = 0, 1, ..., n-1; \ \ k, k' = 1, 2, ..., M_n. \end{array} \tag{4}$$

Здесь $\delta_{n,n'}^{k,k'}$ – символ Кронекера, равный единице при $n = n'$, $k = k'$ и нулю – в противных случаях, а значения $\gamma = 0, 1$ определяют в дальнейшем метод полусопряженных градиентов или полусопряженных невязок соответственно. Отметим, что в отличие от обычных методов полусопряженных направлений, в формулах (3) на каждой n-й итерации используется не один, а M_n направляющих векторов, количество которых может, вообще говоря, меняться.

Если же в (3) ввести обозначения для “сводных"направляющих векторов

$$\alpha_n p^n = P_n \bar{\alpha}_n, \ \ n = 0, 1, ..., \tag{5}$$

с неопределенными пока скалярными множителями α_n, то вместо (3) для векторов невязки можно дать классическое представление

$$r^{n+1} = r^n - \alpha_n Ap^n = r^q - \alpha_q Ap^q - ... - \alpha_n Ap^n, \ \ 0 \leq q \leq n \tag{6}$$

в некотором подпространстве Крылова

$$\tilde{K}(r^0, A) = \mathrm{span}\{r^0, Ap^0, ..., A^n p^0\}, \tag{7}$$

которое будет конкретизировано после определения $p^0, ..., p^n$.

Очевидно, что выражение (3) может быть записано без использования матриц P_n и векторов $\bar{\alpha}_n$:

$$r^{n+1} = r^n - A\sum_{l=1}^{M_n} \alpha_{n,l} p_l^n = r^q - A\Big(\sum_{l=1}^{M_q} \alpha_{q,l} p_l^q - ... - \sum_{l=1}^{M_n} \alpha_{n,l} p_l^n\Big). \tag{8}$$

Отсюда следует, что при условиях (4) мы можем выписать выражение для функционала

$$\Phi_n^{(\gamma)}(r^{n+1}) \equiv (r^{n+1}, A^{\gamma-1}r^{n+1}) = (r^q, A^{\gamma-1}r^q) - \sum_{l=1}^{M_n}\sum_{k=q}^{n}\alpha_{k,l}[2(r^q, A^\gamma p_q^k) - \alpha_{k,l}(Ap_l^k, A^\gamma p_l^k)]. \tag{9}$$

Таким образом, по условию экстремума

$$\partial\Phi_n^{(\gamma)}/\partial\alpha_{k,l} = 0,$$

для $q = 0, 1, ..., n$ имеем

$$\alpha_{k,l}^{(\gamma)} = (r^q, A^\gamma p_l^k)/\rho_{k,l}^{(\gamma)}, \;\; l = 1, ..., M_n; \;\; k = q, q+1, ..., n. \tag{10}$$

Из соотношений (3), (4), (10) следует, что векторы r^{n+1} и $p_k^{n'}$ удовлетворяют условиям A^γ – полусопряженности в следующем смысле:

$$(r^{n+1}, A^\gamma p_k^{n'}) = 0, \;\; n' = 0, 1, ..., n; \;\; k = 1, ..., M_n, \; \gamma = 0, 1. \tag{11}$$

Замечание 1. *Нетрудно видеть, что соответствующее значение функционала $\Phi_n^{(\gamma)}(r^{n+1})$ принимает значение*

$$\Phi_n^{(\gamma)}(r^{n+1}) = (r^q, r^q) - \sum_{k=q}^{n}\sum_{l=1}^{M_n}(r^q, A^\gamma p_l^k)^2/\rho_{k,l}^{(\gamma)} \tag{12}$$

и достигает своего минимума, если $\gamma = 1$ или если A есть симметричная матрица. Подчеркнем, в частности, что в методе полусопряженных градиентов функционал $\Phi_n^{(0)}(r^{n+1}) = (A^{-1}r^{n+1}, r^{n+1})$ при несимметричности A своего минимума, вообще говоря, не достигает.

Направляющие векторы p_l^n будем определять по условиям ортогональности (5) при $\gamma = 0, 1$ в следующем виде:

$$\begin{aligned} &p_l^0 = B_{0,l}^{-1}r^0, \;\; p_l^{n+1} = B_{n+1,l}^{-1}r^{n+1} + \sum_{k=o}^{n}\sum_{l=1}^{M_k}\beta_{n,k,l}^{(\gamma)}p_l^k, \;\; n = 0, 1, ...; \\ &B_{n,l} \in \mathcal{R}^{N,N}, \;\; l = 1, ..., M_n; \;\; \gamma = 0, 1. \end{aligned} \tag{13}$$

Здесь $\bar{\beta}_{n,k}^{(\gamma)} = \{\beta_{n,k,l}^{(\gamma)}\} = (\beta_{n,k,1}^{(\gamma)} ... \beta_{n,k,M_n}^{(\gamma)})^T \in \mathcal{R}^{M_n}$ суть векторы коэффициентов, а $B_{n,l} \in \mathcal{R}^{N,N}$ – предобуславливающие матрицы, которые выбираются по условиям невырожденности, легкой обратимости и эффективного ускорения конструируемых итерационных процессов. Отметим, что рассматриваемые "предобуславливатели" $B_{n,l}$ являются динамическими, или гибкими (flexible, как в методах FGMRES, [3]), поскольку они для каждого l зависят от номера итерации n.

Замечание 2. *В формулах (12) предполагается, что для всех значений индексов l начальное приближение u^0 и соответствующий вектор невязки – одни и те же. Очевидно, что нетрудно построить блочные версии предлагаемых мультипредобусловленных методов, задавая M_0 различных начальных векторов: $U^0 = (u_1^0 ... u_{M_0}^0)^T$. Более того, данный подход легко переносится на случай решения нескольких СЛАУ с одинаковой матрицей A и разными правыми частями $\bar{f}_m = (f_{1,m} ... f_{N,m})^T, \;\; m \le M_0$. В этом случае можно определить прямоугольные матрицы правых частей и невязок*

$$F = (\bar{f}_1 ... \bar{f}_{M_0}), \;\; R^0 = F - AU^0 \in \mathcal{R}^{N,M_0},$$

причем для $m < M_0$ в матрице F некоторые $M_0 - m$ столбцов будут одинаковыми. В дальнейшем, однако, мы на данных возможных обобщениях останавливаться не будем.

Подставляя выражения (13) в условия ортогональности (4), мы получаем формулы для коэффициентов

$$\begin{aligned}&\beta^{(\gamma)}_{n,k,l} = -(A^\gamma p^k_l, AB^{-1}_{n+1,l}r^{n+1})/\rho^{(\gamma)}_{n,l}, \ \ n = 0, 1, ...;\\ &k = 0, ..., n; \ \ l = 1, ..., M_n.\end{aligned} \tag{14}$$

Теперь мы можем сформулировать следующий результат.

Теорема 1. *Итерационные методы* (3), (11), (13), (14) *для* $\gamma = 0, 1$ *с условиями* (2) *при невырожденности предобуславливающих матриц* $B_{n,l}$, $n = 0, 1, ...;$ $l = 1, ..., M_n$, *обеспечивают условие экстремальности* (10), *соответствующее при* $\gamma = 1$ *минимальности нормы* $||r^{n+1}||$ *в "мультипредобусловленном"подпространстве Крылова*

$$\begin{aligned}&\mathcal{K}_{\sum_{n+1}}(r^0, A) = span\{B^{-1}_{0,1}r^0, ..., B^{-1}_{0,M_0},\\ &AB^{-1}_{1,1}r^1, ..., AB^{-1}_{1,M_1}r^1, ..., A^nB^{-1}_{n,1}r^n, ..., A^nB^{-1}_{n,M_n}r^n\}.\end{aligned} \tag{15}$$

В данных методах векторы невязок являются обобщенно полусопряженными при $\gamma = 0, 1$ *в следующем смысле:*

$$(A^\gamma B^{-1}_{k,l}r^n, r^k) = \begin{cases} 0, & k < n,\\ \sigma^{(\gamma)}_n = (A^\gamma B^{-1}_{n,l}r^n, r^n), & k = n,\end{cases} \tag{16}$$

для $l = 1, ...M_k$. *При этом коэффициенты* $\alpha^{(\gamma)}_{n,l}$ *в* (3), (11) *выражаются формулой*

$$\alpha^{(\gamma)}_{n,l} = (A^\gamma B^{-1}_{n,l}r^n, r^n)/\rho_{n,l}. \tag{17}$$

Доказательство формулы (16) следует при $k < n$ из соотношений

$$\begin{aligned}(A^\gamma B^{-1}_{k,l}r^k, r^n) &= ((A^\gamma p^k_l - \sum_{i=0}^{k-1}\beta_{k,i,l}A^\gamma p^i_l), (r^0 - \sum_{i=0}^{n-1}\alpha^{(\gamma)}_{i,l}Ap^i_l)) =\\ &= (A^\gamma p^k_l, r^0) - \alpha_{k,l}\rho^{(\gamma)}_{k,l} - \sum_{i=0}^{k-1}\beta_{k,i,l}[(A^\gamma p^i_l, r^0) - \alpha^{(\gamma)}_{i,l}\rho^{(\gamma)}_{i,l}] = 0.\end{aligned}$$

Для $k = n$ из аналогичных выражений имеем

$$(A^\gamma B^{-1}_{n,l}r^n, r^n) = (A^\gamma p^n_l, r^0),$$

откуда вследствие (11) получаем формулу (16). □

Отметим, что в частном случае $\gamma = 1$ и $B_{n,l} = I$ (единичная матрица) векторы невязки со свойствами (16) называются правыми A-полусопряженными; в работе [7] соответствующий метод обозначается аббревиатурой GCR (Generalized Conjugate Residual).

Реализация формул (13), (14) последовательно для $n = 0, 1, ...$ при $\gamma = 1$ с алгебраической точки зрения представляет собой преобразование линейно независимых векторов $B^{-1}_{0,l}r^0, ..., B^{-1}_{n,l}r^0$ в обладающие свойством A^TA-ортогональности векторы $p^0_l, ..., p^n_l$ с помощью процесса Грамма–Шмидта. Как известно (см. [8] и цитируемую там литературу), эта вычислительная процедура при больших n может оказываться неустойчивой к погрешностям округлений, и по этой причине рекомендуется использовать модифицированный метод ортогонализации Грама–Шмидта, который в данном случае приводит к следующему алгоритму.

Для вычисления $\beta^{(\gamma)}_{n,k,l}$ вместо (14) предлагается формула

$$\beta^{(\gamma)}_{n,k,l} = -(A^\gamma p^k_l, Ap^{n,k}_l)/\rho^{(\gamma)}_{n,l}, \tag{18}$$

в которой векторы $p^{n,k}_l$ определяются из соотношений

$$\begin{gathered} p^{n,k}_l = p^{n,k-1}_l + \beta^{(\gamma)}_{n,k-1,l} p^{k-1}_l = B^{-1}_{n+1,l} r^{n+1} + \sum_{i=0}^{k-1} \beta^{(\gamma)}_{n,i,l} p^n_l, \;\; l = 1, ..., M_n, \\ k = 0, 1, ..., n+1; \;\; p^{n,0}_l = B^{-1}_{n+1,l} r^{n+1}, \;\; p^{n,n+1}_l = p^{n+1}_l. \end{gathered} \tag{19}$$

При этом в случае $\gamma = 1$ коэффициенты $\beta^{(\gamma)}_{n,i,l}$ в (19) обеспечивают выполнение условий минимальности нормы вектора $p^{n,k}_l$

$$\partial ||p^{n,k}_l|| / \partial \beta_{n,i} = 0, \;\; i = 0, 1, ..., k-1,$$

из которого следует формула, совпадающая с точностью до обозначений с (14). Легко видеть, что в силу справедливых для всех l соотношений ортогональности

$$(Ap^{n,k}_l, A^\gamma p^i_l) = 0, \;\; i = 0, ..., k-1,$$

значения $\beta^{(\gamma)}_{n,k,l}$, вычисляемые по формулам (14) и (18), при точном выполнении арифметических операций совпадают, так как

$$(A^\gamma p^k_l, AB^{-1}_{n+1,l} r^{n+1}) = (A^\gamma p^k_l, Ap^{n,k}_l).$$

Заметим, что в случае $\gamma = 1$ (метод полусопряженных невязок) реализация формул (19) не требует дополнительных матрично-векторных операций, в силу справедливости для всех l рекуррентных соотношений

$$Ap^{n,k}_l = Ap^{n,k-1}_l + \beta^{(\gamma)}_{n,k-1,l} Ap^{k-1}_l.$$

Поскольку методы полусопряженных направлений для решения несимметричных СЛАУ основаны на использовании длинных рекурсий вида (19), их применение для больших n сопряжено с завышенными требованиями оперативной памяти, как и в алгоритме GMRES.

Для устранения данного ограничения будем использовать два общепринятых подхода, один из которых связан с проведением рестартов через заданное число итераций m_{res}, а второй – с лимитированной ортогонализацией, в которой сохраняются направляющие векторы с заранее обусловленного количества последних итераций m_{lim}, см. [1]– [3]. Оба эти вынужденных подхода связаны с неизбежным понижением скорости сходимости итераций. Смягчающим обстоятельством здесь является то, что при удачном выборе предобуславливателей количество итерационных шагов на практике бывает невелико.

3 Алгоритмы декомпозиции областей с грубосеточной коррекцией в подпространствах Крылова

Рассмотренные выше в достаточно абстрактной форме мультипредобусловленные методы полусопряженных направлений в данном разделе конкретизируются в применении к построению естественным образом распараллеливаемых алгебраических методов декомпозиции областей на основе двух типов предобуславливателей. Первый из них обусловлен использованием аддитивного метода Шварца, или блочного метода Якоби, в котором обращение каждого блока означает решение вспомогательной СЛАУ в соответствующей подобласти. Для ускорения получаемого итерационного процесса по подобластям как раз применяется второй тип предобуславливателя, заключающегося в грубосеточной коррекции последовательных приближений, варианты которой также имеют названия агрегации или дефляции, а общий принцип их конструирования состоит в малоранговой аппроксимации обратных матриц. Целью "мультипредобуславливания" в подпространствах Крылова в данном случае является исследование наиболее эффективного комбинирования разных подходов — аддитивного метода Шварца и агрегации.

Изначально методы декомпозиции областей описываются на непрерывном уровне: расчетная область Ω, в которой решается некоторая краевая задача для дифференциального уравнения, разбивается на совокупность P подобластей Ω_s, в каждой из которых ставится соответствующая подзадача. Однако мы рассматривать проблему будем только в дискретном представлении, т.е. в терминах полученной в результате построения сеточной расчетной области, имеющей общее число узлов N и составленной из расчетных непересекающихся подобластей: $\Omega^h = \bigcup\limits_{s=1}^{P} \Omega_s^h$, – в каждой из которых количество узлов равно N_s, $N_1 + N_2 + ... + N_P = N$. Далее верхние индексы "h" ради краткости будем опускать.

Подчеркнем, что мы рассматриваем формирование сеточных подобластей без интерфейсных узлов-разделителей, являющихся общими для двух или более смежных подобластей. Рассмотренная декомпозиция расчетной сеточной области на непересекающиеся подобласти (мы не будем останавливаться на способах их построения и на критериях их качества, что представляет собой отдельный важный вопрос) – только первый шаг к описываемому далее конструированию сеточных подобластей с параметризованными пересечениями.

Для сеточной подобласти Ω_s через $\Gamma_s = \Gamma_s^0$ обозначим ее границу, т.е. совокупность узлов, внешних по отношению к Ω_s, но у которых хотя бы один из соседних узлов лежит в Ω_s ($\bar{\Omega}_s^0 = \Omega_s \cup \Gamma_s^0$ есть замыкание исходной сеточной подобласти Ω_s). Пусть далее Γ_s^1 обозначает первую расширенную границу, или первый внешний фронт $\bar{\Omega}_s$, т.е. множество узлов, не лежащих в $\bar{\Omega}_s$, но имеющих хотя бы один соседний узел из $\bar{\Omega}_s^0$ ($\bar{\Omega}_s^1$ есть первое расширение $\bar{\Omega}_s^0$). Аналогично определим последующие стадии расширения сеточной подобласти, а количество таких стадий Δ будем называть параметром расширенной подобласти $\bar{\Omega}_s = \bar{\Omega}_s^\Delta = \Omega_s^\Delta \bigcup \Gamma_s^\Delta$, где узлы из Γ_s^Δ уже не принадлежат Ω_s^Δ, число узлов которой обозначаем через $\bar{N}_s$.

Далее мы для простоты предполагаем наличие изоморфизма между сеточной и алгебраической постановками задачи в том смысле, что каждому i-му узлу соответствует уравнение и компонента u_i неизвестного вектора СЛАУ. При этом подвекторы u_s и $\bar{u}_s$ с размерностями N_s и $\bar{N}_s$ обозначают совокупности компонент, принадлежащих Ω_s и $\bar{\Omega}_s$.

Для построения итерационного метода декомпозиции областей в подпространствах Крылова мы определим два вида предобуславливателей. Первый из них характеризует "ограничительный" аддитивный метод Шварца (RAS – Restricted Additive Swartz и записывается следующим образом:

$$B_{RAS}^{-1} = R\hat{A}^{-1}W, \;\; \hat{A} = W^T A W = \text{ block-diag } \{\bar{A}_{s,s} \in \mathcal{R}^{\bar{N}_s,\bar{N}_s}\}. \tag{20}$$

Здесь $W = [w_1 ... w_P] \in \mathcal{R}^{N,P}$ есть прямоугольная матрица, каждый столбец w_s которой имеет единичные компоненты в узлах из $\bar{\Omega}_s$ и нулевые – в остальных, а матрица $R \in \mathcal{R}^{N,N}$

состоит из блочных строк $R_s \in \mathcal{R}^{N_s,N}$, каждая из которых представляет собой оператор сужения Ω в Ω_s, т.е. $R_s u = u_s$. Отметим, что даже при симметричности исходной СЛАУ предобуславливающая матрица B_{RAS} из (20) в общем случае симметричной не является. Обращение блоков $A_{s,s}$ матрицы $\hat{A}$ сводится фактически к решению независимых подсистем в соответствующих расширенных подобластях, что и является основой распараллеливания аддитивного метода Шварца, или блочного алгоритма Якоби.

Второй предобуславливатель устанавливает "дальние" связи между подобластями и определяется следующим образом:

$$B_c^{-1} = \Phi \check{A}^{-1} \Phi^T, \;\; \check{A} = \Phi^T A \Phi \in \mathcal{R}^{N_c,N_c}, \tag{21}$$

где $N_c \ll N$ есть размерность (число узлов) некоторой редкой сетки, а $\Phi = [\varphi_1 ... \varphi_{N_c}] \in \mathcal{R}^{N,N_c}$ – это прямоугольная матрица, каждый s-й столбец которой состоит из значений некоторой базисной функции φ_s в узлах исходной сетки Ω. В дальнейшем для простоты будем считать, что $N_c = P$, а i-я компонента вектора-столбца φ_s равна единице, если соответствующий узел принадлежит нерасширенной подобласти Ω_s, и нулю – в противном случае. Из (21) видно, что умножение на предобуславливающую матрицу B_c^{-1}, к чему и сводится фактически метод грубосеточной коррекции, заключается главным образом к решению вспомогательной СЛАУ с "агрегированной" матрицей $\check{A}$ малого порядка, устанавливающей на каждой итерации связи между всеми подобластями. Очевидно, что если Φ есть матрица полного ранга, что мы и предполагаем, а матрица A – невырождена, то также невырожденной будет и матрица $\check{A}$, которая называется малоранговой аппроксимацией (точнее говоря, так называется $\Phi^T \check{A} \Phi \in \mathcal{R}^{N_c,N_c}$) исходной матрицы A.

Представленные в формулах (20), (21) предобуславливающие матрицы, после введения обозначений

$$B_{RAS}^{-1} = B_{n,1}^{-1}, \;\; B_c^{-1} = B_{n,2}^{-1}$$

могут использоваться в методах полусопряженных направлений (13)–(19). Если при этом вспомогательные СЛАУ с матрицами $\hat{A}$ и $\check{A}$ решаются прямыми методами, то данные предобуславливатели фактически не зависят от номера итерации n. В случае же использования "внутренних" итерационных процессов мы приходим к динамическому, или гибкому, предобуславливанию.

На методах декомпозиции областей основаны различные вычислительные технологии распараллеливания алгоритмов. Этим вопросам посвящено огромное количество литературы на сайте [23], и мы на данных проблемах останавливаться не будем.

4 О концепции библиотеки алгоритмов KRYLOV

Теоретические оценки скорости сходимости итерационного решения СЛАУ, а также более сравнительный анализ эффективности различных алгоритмов и выработка рекомендаций по выбору наилучшего из них для конкретной алгебраической системы или класса задач представляют собой далеко не простую проблему. Тем более это относится к оценке реальной производительности программной реализации методов на МВС гетерогенной архитектуры со сложной иерархией общей и распределенной памяти. Поэтому непременным условием развития новых численных методов является их экспериментальное исследование, требующее проведение массовых систематических расчетов на представительное серии методических и практических задач. Такая трудоемкая проблема требует разработки специализированных средств автоматизации тестирования и верификации программных реализаций, в том числе создания каталогизированных наборов характерных примеров.

Надо сказать, что в вычислительной алгебре такого типа работы ведутся уже много лет, в результате чего существуют достаточно широко используемые специалистами доступные в интернете матричные коллекции, например, MatrixMarket, Florida и Boeing.

Надо еще сказать, что в мире к настоящему времени накоплено большое количество программного обеспечения по решению задач линейной алгебры, в том числе для разреженных СЛАУ, которое существует в виде или специализированных программных библиотек, или в составе каких-либо прикладных систем, как коммерческих, так и общедоступных (Open Source). В качестве примеров можно назвать MATLAB, NETLIB, PETSc, Hypre, Trilinos, MKL, SparseKit и т.д., этот список можно значительно продолжить. Важно отметить, что в данной области уже сформировались основные стандарты на матричные структуры данных и имеется достаточно представительный набор эффективно реализованных матрично-векторных операций в широко используемых системах BLAS и SparseBlas.

На фоне наличия такого "зоопарка" мы рассмотрим концепцию формирования интегрированного программного окружения, включающего вычислительные инструментарии не только для эффективного решения алгебраических задач на современных МВС, но и для оперативной разработки новых алгоритмов. Некоторые аспекты подобной разработки (библиотеки KRYLOV) изложены в [22].

Мотивировкой к организации такого суперпроекта являются следующие предпосылки. Во-первых, существует очень широкий и непрерывно пополняемый класс актуальных задач линейной алгебры, который с одной стороны можно характеризовать по типу матриц: вещественные и комплексные, эрмитовые и неэрмитовые, симметричные и несимметричные, положительно определенные и знаконеопределенные, а также множество специальных. С другой стороны, многообразные СЛАУ появляются из старых и новых методов решения дифференциальных и/или интегральных уравнений: Максвелла, Ламе, Дарси, Навье–Стокса и т.п.,– структурные и спектральные особенности которых обязательно необходимо учитывать.

Вторая предпосылка заключается в постоянном и активном развитии новых методов вычислительной алгебры, для которых производительные технологии экспериментальных исследований — это своя фундаментальная проблема (по аналогии – теоретическая физика не может развиваться без экспериментальной). И наконец, третья предпосылка: прикладное программное обеспечение должно разрабатываться адаптируемым к бурной эволюции архитектур и платформ суперкомпьютеров. Характерное требование к таким разработкам — это отсутствие программных ограничений на число степеней свободы в решаемой задаче и на количество используемых вычислительных процессов и/или ядер.

С учетом изложенных соображений предлагается концепция библиотеки KRYLOV как интегрированного инструментального окружения для решения широкого класса задач вычислительной алгебры, открытого для согласованного развития его различными группами разработчиков, а также предусматривающие разные интерфейсы и режимы использования конечными пользователями.

Функциональное наполнение библиотеки предполагает автоматизированную сбалансированную геометрическую или алгебраическую декомпозицию задачи, использование различных типов итерационных процессов в подпространствах Крылова с применением разных видов предобуславливающих матриц и критериев точности расчета, а также программы – генераторы матриц для характерных классов задач.

Программная реализация основывается на унифицированных форматах данных с возможной их конвертацией и с формированием MPI-процессов и многопотоковых вычислений на гетерогенных кластерах с распределенной и общей памятью CPU или GPGPU. Внутренние и внешние интерфейсы конструируются по компонентным технологиям CCA (Common

Component Architecture, [24]), поддерживающим многоязыковость и кросс-платформенность разработок, а также возможность переиспользования внешних программных продуктов. Мы не будем вдаваться в детали такого большого проекта, так как эта актуальная проблема представляет тему для отдельных исследований.

5 Заключение

Мы привели краткий обзор современных подходов к актуальной проблеме решения больших СЛАУ, с изложением некоторых авторских результатов, целью чего была демонстрация многообразия существующих алгоритмов, активно развивающихся и теоретически, и в плане реализационных технологий. Следует сказать, что различные аспекты рассматриваемых направлений составляют неразрывные части вычислительной алгебры, которые в целом формируют научно-технологическую цепочку: "рождение идеи и теоретическое исследование методов" – "пробные версии и экспериментальная апробация" – "высокопроизводительная реализация алгоритмов и их практическое внедрение".

Список литературы

[1] Eisenstat S.C., Elman H.C., Schultz M.H. Variational iterative methods for nonsymmetric systems of linear equations // SIAM J. Numer. Anal. 1983, vol. 20, №3, 345-357.

[2] Yuan J.Y., Golub G.H., Plemmons R.J., Cecilio W.A. Semi-conjugate direction methods for real positive definite systems//BIT, v. 44, N 1, 2004, 189-207.

[3] Ильин В.П., Ицкович Е.А. О методах полусопряженных направлений с динамическим предобуславливанием.– Новосибирск, СибЖИМ. 2007, т. 10, № 4, 41-54.

[4] Ruhe A. Rational Krylov sequence methods for eigenvalue computation/ Linear Algebra Appl., vol.58, 1984, 391-405.

[5] Druskin V., Knizhnerman L., Simoncini V. Analysis of the rational Krylov subspace and adi methods for solving the Lyapunov equation.-SIAM J. Numer. Anal., vol. 49, N 5, 2011, 1875-1898.

[6] Gijzen Van M., Sonneveld P. Algorithm 913: An Elegant IDR(s) Variant that Efficiently Exploits Biorthogonality Properties. ACM Transactions on Mathematical Software, vol. 38, N 1, Article, 5, 2011.

[7] Martin B., van Gijzen, Gerard L.G., Sleijpen and Jens-Peter M. Zemke. Flexible and multi-shift induced dimension reduction algorithvs for solving large sparse linear systems.–Numer. Linear Algebra Appl.,vol. 22, 2015, 1-25.

[8] Chapman A., Saad Y. Deflated and augmented Krylov subspace techniques // Numer. Linear Algebra Applic. 1997, vol. 4, № 1, 43-66.

[9] Gurieva Y.L., Il'in V.P. Parallel Aproaches and Technologies of Domain Decomposition Methods.– Jounal of Mathematical Sciences, vol. 207, N 5, 2015, 724-735.

[10] Oseledets I.V., Tyrtyshnikov E.E. Breaking the Curse of Dimensionality. Or How to Use SVD in Many Dimensions.-SIAM J. Sci. Comput., v. 31, Iss. 5, 2009, 3744-3759.

[11] Hackbusch W. Hierarchische Matrizen: Algorithmen and Analysis, Springer, 2009.

[12] Bridson R., Greif C. A multipreconditioned conjugate gradient algorithm.//SIAM J. Matrix Anal. Appl., vol. 27, N 4, 1056-1068.

[13] Greif C., Rees T., Szyld D.B. MPGMRES: a generalized minimum residual method with multiple preconditioners. Tech. Rep. 11-12-23, Department of Mathematics, Temple University (2011).

[14] Greif C., Rees T., Szyld D.B. Additive Schwarz with Variable Weights. Proc. of 21-th Internat. Confer. on DDMs. Lecture Notes in Comp. Sci. and Eng., vol. 98, Springer, Berlin, 2014, 661-668.

[15] Saad Y. Iterative Methods for Sparse Linear Systems, 2nd ed. Society for Industrial and Applied Mathematics, Philadelphia, 2003, 547 p.

[16] Dolean V., Jolivet P., Natan F. An Introduction to Domain Decomposition Methods: algorithms, theory and parallel implementation,https:archives-ouvertes.fr/cel-01100932, Lecture Notes to appear in SIAM collection, 2015.

[17] Toselli A., Widlund O. Domain Decomposition Methods – Algorithms and Theory. Springer Series in Comput. Math. 2005, vol. 34, 450 p.

[18] Chu M.T., Funderlic R.E., Golub G.H. A rank-one reduction formula and its application to matrix factorization.-SIAM REv., v. 37, 1995,512-530.

[19] Kolotilina L.Yu. Eigenvalue Bounds and Inequalities Using Vector Aggregation of Matrices.– Linear Algebra and its Applications, vol. 271, 1998, 139-167.

[20] Dubois O., Gander M.J., St-Cyr A., Loisel S., Szyld D. The Optimized Schwarz Method with a Coarse Grid Correction // SIAM Journal on Scientific Computing. 2012, vol. 34, № 1, 421-458.

[21] Efendiev Y., Galvis J., Lazarov R., Willems J. Robust domain decomposition preconditioners for abstract symmetric positive definite bilinear forms. Esaim Math. Model. Anal., 46(5): 1175-1199, 2012. ISSN: 0764-583X.DOI:10.1051/m2an/2011073. URL:http://dx.doi.org/10.1051/m2an/2011073.

[22] Бутюгин Д.С., Гурьева Я.Л., Ильин В.П., Перевозкин Д.В., Петухов А.В., Скопин И.Н. Функциональность и технологии алгебраических решателей в библиотеке Krylov // Вестник ЮУрГУ. Серия “Вычислительная математика и информатика”. 2013, т. 2, № 3, 92-105.

[23] URL: http://www.ddm.org

[24] CCA: The Common Component Architecture Forum. www.cca-forum.org/.