

Коньшин Игорь Николаевич

ПАРАЛЛЕЛЬНЫЕ МЕТОДЫ РЕШЕНИЯ
СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ С СИММЕТРИЧНЫМИ
ПОЛОЖИТЕЛЬНО-ОПРЕДЕЛЕННЫМИ МАТРИЦАМИ
НА ОСНОВЕ АДДИТИВНОГО РАЗЛОЖЕНИЯ С ПЕРЕКРЫТИЯМИ

01.01.07 — Вычислительная математика

Автореферат

диссертации на соискание учёной степени
кандидата физико-математических наук

Работа выполнена в Учреждении Российской академии наук Вычислительный центр им. А. А. Дородницына РАН.

Научный руководитель: кандидат физико-математических наук
Капорин И. Е.

Официальные оппоненты: доктор физико-математических наук,
профессор
Кобельков Г. М.

доктор физико-математических наук,
доцент
Василевский Ю. В.

Ведущая организация: Научно-исследовательский вычислительный
центр МГУ

Защита состоится 10 декабря 2009 г. в 15 ч. 00 мин. на заседании Диссертационного совета Д 002.017.01 при Учреждении Российской академии наук Вычислительный центр им. А. А. Дородницына РАН по адресу: 119333, г. Москва, ул. Вавилова, дом. 40, тел: (499) 135-24-89.

С диссертацией можно ознакомиться в библиотеке ВЦ РАН.

Автореферат разослан 05 ноября 2009 г.

Учёный секретарь Диссертационного совета
доктор физико-математических наук, профессор

Зубов В. И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Объект исследования и актуальность темы. Актуальность темы диссертационной работы обусловлена необходимостью эффективного численного решения систем линейных алгебраических уравнений с разреженными симметричными положительно-определенными матрицами большой размерности. Численное решение систем линейных уравнений является одной из наиболее часто встречающихся задач в научно-технических исследованиях и практических приложениях. Подобные задачи возникают, например, в математической физике при численном решении дифференциальных и интегральных уравнений. При этом прикладные задачи часто требуют решения систем линейных уравнений с большим числом неизвестных. К таким системам, например, приводит численное решение двумерных и особенно трехмерных задач математической физики, в которых условия физической и геометрической аппроксимации двумерной и трехмерной области диктуют использование достаточно мелкой расчетной сетки с большим числом расчетных узлов. Для возникающих систем линейных уравнений применение прямых методов решения оказывается неприемлемым как по причине большого времени решения, так и недостаточности объема оперативной памяти для хранения данных задачи. Итерационные методы решения систем линейных уравнений намного экономичнее с точки зрения использования оперативной памяти, но для ускорения их сходимости требуется строить эффективные предобуславливания, особенно при решении систем с плохо обусловленными матрицами.

Существенные затруднения связаны с тем, что наиболее актуальной задачей является решение систем линейных уравнений настолько большой размерности, что их решение возможно только на современных параллельных ЭВМ с распределенной памятью, что подразумевает разбиение данных на блоки, каждый из которых обрабатывается отдельным процессором. Поэтому для таких многопроцессорных вычислительных машин необходимо построение специальных параллельных предобуславливаний, использующих разбиение на блоки матрицы решаемой системы. Причем для универсальности метода требуется построение разбиения на блоки без при-

влечения информации о пространственном расположении узлов расчетной сетки.

Эффективность использования каждого процессора многопроцессорной ЭВМ в конечном итоге определяет эффективность решения задачи в целом, поэтому регулирование загруженности процессоров также является актуальной задачей. Использование специальных методов балансировки вычислений на различных процессорах призвано решить задачу повышения эффективности параллельных вычислений.

Целью диссертационной работы является разработка, исследование и реализация на параллельных ЭВМ высокопроизводительного итерационного метода решения жестких систем линейных алгебраических уравнений с разреженными симметричными положительно-определенными матрицами, а также численные исследования работоспособности предложенного метода для различных типов плохо обусловленных систем линейных уравнений.

В соответствии с целью исследования были поставлены следующие **задачи**:

1. Разработка и исследование параллельного итерационного предобусловленного метода решения жестких систем линейных алгебраических уравнений с разреженными симметричными положительно-определенными матрицами.
2. Разработка эффективного способа балансировки вычислений для подзадач на различных процессорах.
3. Исследование эффективности метода в зависимости от параметров предобусловливания и количества процессоров.
4. Подтверждение эффективности метода для различных типов решаемых задач и на различных вычислительных системах.

Методы исследования. В диссертационной работе используется математический аппарат численных методов, теории итерационных алгоритмов, линейной алгебры, теории графов, методов параллельного программирования.

Научная новизна:

1. Установлено, что при однопроцессорной реализации методов решения систем линейных уравнений с симметричными положительно-определенными матрицами, метод приближенного треугольного разложения второго порядка существенно превосходит по качеству предобусловливания аналогичные методы первого порядка.
2. Показано, что предложенная схема построения параллельного предобусловливания имеет более высокую скорость сходимости по сравнению с некоторыми известными схемами типа декомпозиции области.
3. Предложена комбинация приближенного треугольного разложения второго порядка и неполного обратного треугольного разложения, позволяющая получить высокое качество параллельного предобусловливания.
4. Разработана новая методика балансировки вычислений на основе восполнения треугольных сомножителей элементами матрицы ошибок и показана надежность и эффективность ее применения.
5. Проведены численные эксперименты, показавшие, что как общие арифметические затраты, так и количество итераций для предложенного метода слабо зависят от количества используемых процессоров по сравнению с известными параллельными предобусловливаниями.
6. Численными экспериментами подтверждена теоретически обоснованная высокая параллельная эффективность и вычислительная надежность разработанного метода.

Научная и практическая ценность. Разработан параллельный метод решения жестких систем линейных алгебраических уравнений с симметричными положительно-определенными матрицами, который может быть использован во многих приложениях. Предложенный метод показал высокую надежность и эффективность при решении задач из самых различных научных и практических областей применения. Выполненная работа вносит теоретический и практический вклад в развитие высокопроизводительных параллельных вычислений, так как она призвана не только

решить проблему экономии памяти и сокращения времени вычислений при решении систем линейных уравнений, но и сделать возможным решение задач, которые ранее в принципе не могли быть решены на однопроцессорных компьютерах.

Положения, выносимые на защиту:

1. Теоретические основы построения эффективного параллельного предобусловливания для решения систем линейных алгебраических уравнений с симметричными положительно-определенными матрицами.
2. Методика балансировки вычислений на основе восполнения треугольных сомножителей элементами матрицы ошибок.
3. Результаты практического исследования эффективности предложенного алгоритма на тестовых и прикладных задачах, его вычислительных и коммуникационных затрат, а также времени счета при изменении параметров предобусловливания и количества процессоров.

Обоснованность и достоверность результатов основывается на строгих математических доказательствах и проверке теории в численных экспериментах. Эффективность построения параллельного предобусловливания обосновывается использованием широкого ряда тестовых задач для верификации полученных численных результатов, а также прямым сравнением результатов расчетов с результатами, полученными при использовании других предобусловливающих семейства блочных методов Якоби. Достоверность того, что предложенный метод является эффективным, подтверждена результатами расчетов и сравнением с результатами других авторов на примере задач из коллекции университета Флориды.

Апробация работы. Результаты работы докладывались и обсуждались на международной конференции “PaCT99” (Санкт-Петербург, Россия, 1999 г.), Всемирном 16-ом Конгрессе “IMACS 2000” по научным вычислениям, прикладной математике и моделированию (Лозанна, Швейцария, 2000), Голландско-российском симпозиуме NWO (МГУ, Москва, июнь 2000), Симпозиуме NWO (Амстердам-Ниймеген, Голландия, ноябрь 2000),

Втором Международном научно-практическом Семинаре и Всероссийской молодежной школе “Высокопроизводительные параллельные вычисления на кластерных системах” (Нижний Новгород, Россия, 2002), Международной конференции “Parallel CFD 2003” (Москва, 2003), Международной конференции “VIII Забабахинские научные чтения” (РФЯЦ-ВНИИТФ, Снежинск, Россия, 2005), Международной конференции по численной геометрии, построению расчетных сеток и научным вычислениям “NUM-GRID2008” (ВЦ РАН, Москва, 2008), на научно-исследовательских семинарах Вычислительного центра РАН, Московского физико-технического института, Института автоматизации проектирования РАН, а также на семинарах Департамента информационных технологий Республики Индия (C-DAC, Пуна, Индия), Institut Français du Pétrol (Париж, Франция), Exxon-Mobil Upstream Research Co. (Хьюстон, США).

Личный вклад соискателя. В работах соискателя соавтором является научный руководитель Капорин И. Е., ему принадлежит постановка задач и выбор математических методов исследований. Соискателем проведены теоретические исследования, разработаны параллельные предобуславливания, выполнены параллельные реализации предложенных алгоритмов и исследованы границы применения предложенных методов.

Публикации. Основные публикации по теме диссертации включают 11 работ, из них 3 работы [6, 7, 10] в ведущих отечественных и международных рецензированных научных изданиях, рекомендованных ВАК, 3 в других международных рецензируемых изданиях, 1 в российском рецензируемом издании, 1 в трудах международной конференции, 3 публикации в других научных изданиях.

Кроме того, по теме диссертации имеются публикации в трудах и тезисах 7 всероссийских и 16 международных конференций.

Структура и объем работы. Диссертационная работа состоит из введения, 5 глав, заключения, списка литературы, содержит 18 таблиц и 23 рисунка. Объем работы 138 страницы. Список литературы включает в себя 112 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Введение содержит обоснование актуальности исследуемой проблемы, формулировку цели и задач диссертационной работы, краткое изложение полученных результатов, их практической ценности, а также положений, выносимых на защиту, и описание структуры диссертации.

Первая глава посвящена описанию метода сопряженных градиентов и теории его сходимости. В начале главы дан обзор итерационных методов решения систем линейных уравнений, особенностей их применения, и указано, когда их использование может быть более эффективно по сравнению с прямыми методами решения. Показана необходимость построения предобусловливания для ускорения сходимости итерационных методов, особенно для достаточно жестких задач.

На примере решения системы линейных алгебраических уравнений

$$Ax = b \quad (1)$$

с симметричной положительно-определенной (с.п.о.) разреженной $n \times n$ матрицей A рассматривается предобусловленный метод сопряженных градиентов (МСГ), в стандартной двучленной форме:

$$\begin{aligned} r_0 &= b - Ax_0, \quad p_0 = Hr_0, \quad i = 0, 1, \dots, \\ \alpha_i &= \frac{r_i^T Hr_i}{p_i^T Ap_i}, \quad x_{i+1} = x_i + p_i \alpha_i, \quad r_{i+1} = r_i - Ap_i \alpha_i, \\ \beta_i &= \frac{r_{i+1}^T Hr_{i+1}}{r_i^T Hr_i}, \quad p_{i+1} = Hr_{i+1} + p_i \beta_i. \end{aligned} \quad (2)$$

Здесь $H \sim A^{-1}$ – предобусловливатель для матрицы A , x_i и $r_i = b - Ax_i$ есть приближение к решению и невязка на i -й итерации, соответственно. Хорошо известная верхняя граница количества итераций, необходимых для уменьшения A^{-1} -нормы начальной невязки в ε^{-1} раз, т. е. $\sqrt{r_0^T A^{-1} r_0} \leq \varepsilon \sqrt{r_0^T A^{-1} r_0}$, выражается приближенной формулой

$$i_C(\varepsilon) = \left\lceil \frac{1}{2} \sqrt{C} \log \frac{2}{\varepsilon} \right\rceil,$$

где

$$C = C(HA) = \lambda_{\max}(HA) / \lambda_{\min}(HA)$$

есть спектральное число обусловленности предобусловленной матрицы HA . Величину $C(HA)$ принято рассматривать как критерий качества предобусловливания. Однако основными недостатками такого критерия является как проблематичность его непосредственного практического использования в достаточно общих случаях (например, при построении приближенного разложения Холецкого для произвольной с.п.о. матрицы), так и его недостаточная согласованность с реальной скоростью сходимости МСГ, которая существенно зависит не только от границ спектра, но и от характера распределения собственных значений предобусловленной матрицы внутри этих границ.

С целью нахождения более конструктивного критерия качества предобусловливания была разработана альтернативная теория сходимости МСГ. И. Е. Капориным показано, что справедлива неумлучшаемая оценка, измеряемая H -нормой невязки, вида

$$\sqrt{r_i^T H r_i} \leqslant \left(K^{1/i} - 1 \right)^{i/2} \sqrt{r_0^T H r_0}, \quad (3)$$

где

$$K = K(HA) = \left(\frac{1}{n} \operatorname{tr}(HA) \right)^n / \det(HA)$$

есть так называемое K -число обусловленности предобусловленной матрицы HA . Из неравенства (3) можно получить упрощенную, но наглядную верхнюю оценку количества итераций, необходимых для уменьшения H -нормы начальной невязки в ε^{-1} раз:

$$i_K(\varepsilon) = \lceil \log K + \log \varepsilon^{-1} \rceil. \quad (4)$$

В заключение формулируется основной вывод о возможности построить предобусловливатель H путем минимизации $K(HA)$ (или соответственной верхней оценки) при ограничениях, накладываемых структурой предобусловливателя H .

Вторая глава содержит описание приближенных симметричных треугольных разложений. В начале главы дан обзор различных способов построения приближенных треугольных разложений. Обсуждаются особенности построения неполных симметричных треугольных разложений с фиксированной структурой разреженности (разложение “по позициям”), а

также приближенных треугольных разложений с отсечением малых элементов по значениям. Показаны преимущества и недостатки этих подходов, а также делается вывод о необходимости применения приближенного треугольного разложения второго порядка.

Далее для удобства предполагается, что матрица A симметрично отмасштабирована таким образом, что на главной диагонали стоят единичные элементы. Приводится матричное соотношение, лежащее в основе предобусловливания с использованием приближенного треугольного разложения второго порядка (Incomplete Cholesky 2-nd order, IC2), которое имеет вид

$$A = U^T U + U^T R + R^T U, \quad (5)$$

где U – невырожденная верхнетреугольная матрица (используемая как приближение правого множителя точного разложения Холецкого для матрицы A) и R – строго верхнетреугольная матрица ошибок с “малыми” элементами.

Важными свойствами матрицы

$$M = U^{-T} A U^{-1} \equiv I + R U^{-1} + U^{-T} R^T,$$

полученной в результате двустороннего IC2-предобусловливания, являются следующие:

$$\text{diag} M = I, \quad K(M) = \det(I + R A^{-1} R^T), \quad (6)$$

$$K(HA) = K(U^{-T} A U^{-1}) = K(M),$$

где $H = (U^T U)^{-1}$.

Рекуррентные соотношения для вычисления IC2-разложения могут быть легко получены из (5), например, если потребовать, чтобы в структуре разреженности матриц U и R не было совпадающих позиций ненулевых элементов, и чтобы их ненулевые элементы удовлетворяли соотношениям $|u_{ij}| \geq \tau$ и $|r_{ij}| < \tau$ соответственно. Здесь $0 < \tau \ll 1$ представляет собой малый параметр, определяющий качество разложения (будем называть также этот параметр параметром отсечения малых элементов в IC2 разложении). В частности, из второго уравнения (6) легко получается оценка

$$\log K(M) \leq c_0(A) \tau^2, \quad (7)$$

сопоставление которой с (4) объясняет, почему это разложение называется разложением 2-го порядка.

Однако как вычисление, так и применение приближенного разложения Холецкого для всей матрицы A не может быть эффективно реализовано на параллельных архитектурах с распределенной памятью, которые имеют значительные накладные расходы на организацию пересылок данных. Поэтому в заключение делается вывод о необходимости рассмотрения специальных блочных алгоритмов, требующих малого количества инициализаций обменов при их параллельной реализации. При построении таких параллельных предобусловливателей может оказаться эффективным применение приближенного разложения Холецкого 2-го порядка для решения некоторых подзадач на различных процессорах.

Третья глава содержит описание аддитивных приближенных разложений. В начале главы дан обзор различных схем построения параллельных предобусловливателей. Перечислены их основные свойства и особенности их реализации. Рассматриваются некоторые варианты аддитивных предобусловливателей типа декомпозиции области.

Приведено описание блочной версии алгоритма неполного обратного треугольного разложения (Block Incomplete Inverse Cholesky, ВИС). Пусть матрица A переупорядочена и разбита на блоки тем же образом, что и в случае применения блочно-диагонального предобусловливания, т. е. t -й диагональный блок симметрично переупорядоченной матрицы имеет размер n_t , где $n_1 + \dots + n_p = n$. Здесь $t = 1, 2, \dots, p$ и p – блочная размерность матрицы A . Для t -го диагонального блока определим “базисное” множество индексов как $\{k_{t-1} + 1, \dots, k_t\}$, где $k_{t-1} = n_1 + \dots + n_{t-1}$ ($k_0 = 0$, $k_p = n$), и введем “перекрывающиеся” множества индексов $\{j_t(1), \dots, j_t(m_t - n_t)\}$, $j_t(p) \leq k_{t-1}$. Для каждого t такое индексное множество, как правило, включает те индексы, не превосходящие k_t , которые оказываются наиболее существенно связанными с t -м базисным множеством индексов, например, в смысле графа связности разреженной матрицы A (или матрицы, полученной из A заменой относительно малых элементов на нули). Здесь $m_t \geq n_t$ и $m_1 = n_1$, т. е. первое такое множество является пустым.

В этих обозначениях ВИС-предобусловливатель может быть записан в следующем аддитивном виде:

$$H = H_{\text{ВИС}} = \sum_{t=1}^p V_t U_t^{-1} \begin{bmatrix} 0 & 0 \\ 0 & I_{n_t} \end{bmatrix} U_t^{-T} V_t^T, \quad (8)$$

где V_t выбираются в виде прямоугольных матриц, столбцы которых являются единичными n -векторами с определенными выше индексами,

$$V_t = [e_{j_t(1)} | \cdots | e_{j_t(m_t - n_t)} | e_{k_{t-1}+1} | \cdots | e_{k_t}], \quad t = 1, 2, \dots, p, \quad (9)$$

а каждая верхнетреугольная матрица U_t является правым множителем Холецкого t -й “расширенной” диагональной $(m_t \times m_t)$ -подматрицы $V_t^T A V_t$, т. е.

$$V_t^T A V_t = U_t^T U_t, \quad t = 1, 2, \dots, p. \quad (10)$$

Далее рассматривается использование ИС2-разложения в ВПС-предобусловливании. Для каждого $t = 1, 2, \dots, p$ заменим точное разложение Холецкого (10) соответствующим приближенным ИС2-разложением, аналогичным (5):

$$V_t^T A V_t = \tilde{U}_t^T \tilde{U}_t + \tilde{U}_t^T R_t + R_t^T \tilde{U}_t. \quad (11)$$

где R_t – локальные матрицы ошибок с элементами $|r_{ij}| < \tau$. Обозначим

$$\tilde{H} = \sum_{t=1}^p V_t \tilde{U}_t^{-1} \begin{bmatrix} 0 & 0 \\ 0 & I_{n_t} \end{bmatrix} \tilde{U}_t^{-T} V_t^T$$

и будем для простоты предполагать, что стратегия отсечения малых элементов выбрана таким образом, что первые $m_t - n_t$ диагональных элементов \tilde{U}_t совпадают с соответствующими элементами U_t . Используя это предположение и изложенные выше свойства ИС2- и ВПС-разложений, доказывается следующая теорема.

Теорема 1. Пусть A – с.п.о. матрица, тогда из свойств разложений ИС2 с порогом τ отсечения малых элементов и ВПС с выделением p блоков следует оценка отношения K -чисел обусловленности матриц $\tilde{H}A$ и HA :

$$\frac{K(\tilde{H}A)}{K(HA)} = \prod_{t=1}^p \det(I_{m_t} + R_t(V_t^T A V_t)^{-1} R_t^T) \leq e^{c_0 \tau^2}. \quad (12)$$

Используя (7) и (12), получаем

$$\log K(\tilde{H}A) \leq \log K(HA) + c_0 \tau^2,$$

и из (4) непосредственно следует оценка

$$i_{\tilde{H}}(\varepsilon) \lesssim i_H(\varepsilon) + c_0 \tau^2.$$

Результаты численных экспериментов подтверждают эту зависимость от τ для набора из 5-и тестовых задач, подробно описанных далее. Они также показывают, что при $\tau = 10^{-3}$ количество итераций незначительно отличается от количества итераций для “точного” ВПС-предобусловливания (8), отвечающего значению $\tau = 0$, и остается практически неизменным при $\tau < 10^{-4}$.

В заключение делается вывод, что для некоторого достаточно малого значения параметра отсечения τ такое комбинированное ВПС-IC2-разложение будет давать предобусловливатель \tilde{H} , позволяющий достичь почти такой же скорости сходимости итераций МСГ, как и K -оптимальное ВПС-предобусловливание (8). В то же время ВПС-IC2-разложение требует значительно меньших затрат для вычисления, хранения и применения предобусловливателя по сравнению с “точным” ВПС-предобусловливанием (8).

Четвертая глава содержит описание параллельной реализации и методик балансировки вычислений. В начале главы дан обзор современных параллельных ЭВМ, моделей параллельных вычислений и способов оценки эффективности вычислений на таких ЭВМ. Для повышения параллельной эффективности применяется балансировка вычислений на различных этапах решения. Приведены различные варианты проведения балансировки на этапе выполнения итераций МСГ.

Далее описана параллельная реализация предобусловленного МСГ с использованием аддитивного приближенного разложения с перекрытиями. Данный способ построения эффективных предобусловливателей был реализован с использованием обменов данными на основе стандарта MPI. Так была достигнута переносимость программного обеспечения на широкий класс ЭВМ с параллельной архитектурой.

Предположим, что система линейных уравнений (1) решается на параллельной ЭВМ с распределенной памятью, имеющей p процессорных элементов. Будем также предполагать, что количество блоков совпадает с количеством процессоров, на которые разбита исходная матрица A , а t -й блок соответствует t -му процессору, $t = 1, 2, \dots, p$.

Предложенный алгоритм может быть реализован следующим образом. На этапе построения ВПС-IC2 предобусловливания выполняются IC2-разложения (11) локальных подматриц $V_t^T A V_t$ на локальном t -м процессоре. При этом никаких обменов данными не происходит.

На этапе итераций МСГ (2) с выполнением обменов данными происходят следующие операции: а) умножение матрицы A на вектор, б) умножение предобусловливателя \tilde{H} на вектор, в) скалярные произведения.

Выполнение умножения матрицы A на вектор является часто используемой и хорошо исследованной операцией. Ее параллельная эффективность зависит от графа связности матрицы и топологии разбиения на подобласти. Эффективность умножения предобусловливателя \tilde{H} на вектор зависит кроме этого еще и от геометрии перекрытия подобластей. Параллельная эффективность выполнения скалярных произведений определяется только количеством неизвестных в подобластях. Балансировки вычислений на всех описанных этапах взаимосвязаны, причем наиболее сложной является балансировка заполненности предобусловливателя. Основное внимание уделено именно этой проблеме.

Из вышеизложенного следуют свойства параллельной реализации метода ВИС-ИС2-МСГ:

1. Общее количество инициализаций обменов не зависит от количества блоков и размерности системы линейных уравнений и равно 5 на каждой итерации МСГ.
2. Коммуникационные затраты невелики по сравнению с арифметическими затратами.
3. Все вычисления хорошо сбалансированы, если размеры локальных подматриц и размеры локальных треугольных множителей приблизительно одинаковы.

Далее описывается построение разбиения матрицы на блоки, которое выполняется без привлечения информации о пространственной топологии исходной физической постановки задачи, т. е. без явного использования пространственной геометрии сетки, которая в большинстве случаев бывает не известна. Для этого применяется широко известный открытый пакет METIS, разбивающий граф связности разреженной матрицы на p примерно равных по размеру “подобластей” (или блоков).

Перекрытие подобластей строится с использованием структуры разреженности матрицы A^q , где q – целое число, $q \geq 0$, называемое в дальнейшем параметром перекрытия.

Приводится теоретический анализ стратегий *постфильтрации* построенного предобусловливателя, которые применяются для балансиров-

ки его заполненности. Отдельно рассматриваются отсечения элементов в традиционном IC-разложении первого порядка и два варианта отсечения элементов в IC2-разложении: по треугольным сомножителям и по матрице ошибок IC2-разложения.

Анализ параллельной эффективности предобусловленного МСГ показывает, что несбалансированность независимых подзадач на стадии применения предобусловливания (вследствие неодинаковой плотности треугольных сомножителей для различных подматриц) в наибольшей степени отвечает за общую потерю параллельной эффективности.

Применение параллельного ВИС-IC2-предобусловливания сводится к независимому решению p нижне- и верхнетреугольных систем линейных уравнений. Даже если исходная матрица A была разбита на блоки одинакового размера, и в используемых перекрывающихся блоках было приблизительно равное количество узлов, величины $NZ(\tilde{U}_t)$, из-за использования отсечения элементов, по модулю меньших τ , могут, в общем случае, сильно отличаться. При этом на этапе итераций процедуры решения систем линейных уравнений с треугольными сомножителями становятся несбалансированными, что может существенно сказаться на параллельной эффективности.

Рассматривается постфильтрация множителей \tilde{U}_t по традиционно применяемой стратегии. Главным недостатком традиционной фильтрации является то, что при ее применении сходимость итераций может ухудшиться. Это может произойти, даже если используется *стабилизированное* IC2-разложение, которое запишем в виде:

$$A = U^T U + U^T R + R^T U - S,$$

где R – строго верхнетреугольная матрица ошибок с элементами r_{ij} , $|r_{ij}| < \tau$, S – симметричная матрица ошибок $S = S^T > 0$ с элементами s_{ij} , $|s_{ij}| < \tau^2$, $i \neq j$, $|s_{ii}| < c_1 \tau^2$ для некоторого $c_1 > 1$, и τ – параметр отсечения, $0 < \tau \ll 1$. Известно, что для предобусловленной матрицы $M = U^{-T} A U^{-1}$ при выполнении неравенства $R^T R \leq \gamma S$ для всех собственных значений матрицы M верна оценка

$$\lambda_i(M) \leq 1 + \gamma.$$

В случае, когда γ является величиной порядка 1, можно рассчитывать на хорошую устойчивость и быструю сходимость МСГ.

В связи с этим рассматривается расщепление матрицы U , используемое в традиционной стратегии фильтрации:

$$U = \tilde{U} + \hat{R},$$

где \hat{R} включает некоторое количество наименьших по величине ненулевых элементов матрицы U , а также оценивается качество предобусловливания, полученного с использованием постфильтрации треугольного сомножителя \tilde{U} .

Теорема 2. *Для всех собственных значений предобусловленной матрицы $\tilde{M} = \tilde{U}^{-T}A\tilde{U}^{-1}$, при выполнении условий $R^T R \leq \gamma S$ и $\hat{R}^T \hat{R} \leq \hat{\gamma} S$, где $\hat{\gamma} < (\sqrt{\gamma} + \sqrt{1 + \gamma})^{-2}$ верна оценка*

$$\lambda_i(\tilde{M}) \leq \frac{1 + \gamma}{1 - \hat{\gamma} - 2\sqrt{\gamma\hat{\gamma}}}, \quad i = 1, \dots, n. \quad (13)$$

Таким образом, получены условия, показывающие, что при использовании традиционной фильтрации сходимость МСГ может ухудшиться. Результаты приведенных ниже численных экспериментов подтверждают, что примерно в 1/4 случаев происходит резкое ухудшение сходимости более чем в 2 раза, вызванного именно возрастанием $\lambda_{\max}(\tilde{M})$.

Рассматривается постфильтрация, выполняемая по принципиально новой стратегии. Для приближенного треугольного разложения второго порядка IC2 предложено переносить наибольшие по величине элементы из матрицы ошибок \tilde{R}_t в вычисленные треугольные множители. Это не только обеспечивает сбалансированность вычислений, но и не ухудшает скорость сходимости МСГ.

Для этого IC2-разложение переписывается в эквивалентном виде:

$$A + R^T R + S = (U + R)^T (U + R),$$

которое позволяет рассмотреть расщепление матрицы ошибок IC2-разложения

$$R = \hat{R} + \tilde{R},$$

где \hat{R} включает некоторое количество относительно больших по величине элементов матрицы R , и оценить качество предобусловливания, полученного с использованием восполнения треугольного сомножителя IC2-разложения

$$\tilde{U} = U + \hat{R}.$$

Теорема 3. Для предобусловленной матрицы

$$\tilde{M} = \tilde{U}^{-T} A \tilde{U}^{-1}$$

при выполнении условия $\tilde{R}^T \tilde{R} \leq S + R^T R$ верна оценка

$$K(\tilde{M}) \leq (\det U)^2 / \det A.$$

Таким образом, если норма матрицы \tilde{R} достаточно мала, то K -число обусловленности предобусловленной матрицы \tilde{M} имеет верхнюю границу, которая не зависит от возмущения, вносимого описанной постфильтрацией. Это дает основания полагать, что сходимость МСГ при этом, по крайней мере, не ухудшится.

Результаты приведенных численных экспериментов показывают, что для рассмотренных задач более чем в 70% случаев было зафиксировано уменьшение времени счета, обусловленное улучшением сходимости МСГ, а в остальных случаях результаты оказались сравнимы.

В заключение делается вывод о надежности и эффективности применения предложенной стратегии постфильтрации, заключающейся в восполнении элементов предобусловливателя из матрицы ошибок.

Пятая глава содержит описание результатов численных экспериментов для различных тестовых задач, и также анализ сходимости предложенного метода в зависимости от параметров предобусловливания. В начале главы дается обзор приложений, где требуется решать жесткие системы линейных уравнений с разреженными с.п.о. матрицами. Дано описание нескольких наиболее представительных коллекций тестовых матриц, которые используются исследователями для проверки своих алгоритмов. Описаны также используемые для сравнения методы предобусловливания и изложена методика проведения численных экспериментов.

Для сравнения предложенного метода с некоторыми другими стратегиями предобусловливания, также требующими малого количества инициализаций обменов на этапе итераций, приводятся данные численных экспериментов для точечного метода Якоби (Point Jacobi, PJ), приближенного блочного метода Якоби (Block Jacobi, BJ), а также обобщения метода Якоби, использующее взвешенное перекрытие блоков (Overlapping Block Jacobi, OBJ).

Необходимо заметить, что в последнем случае перекрытие подобластей (или блоков) строится “во все стороны” (ОВЖ-тип перекрытия), а не только в сторону меньших индексов, как это происходит в случае ВПС-предобусловливания. Второй особенностью метода ОВЖ является то, что до и после обращения диагональных подматриц применяется масштабирование весовой диагональной матрицей $D^{-1/2}$, где j -й элемент D равен числу взаимных перекрытий подобластей для j -го узла. Построение ОВЖ-предобусловливателя может быть представлено в виде

$$H_{\text{ОВЖ}} = D^{-1/2} \left(\sum_{t=1}^p V_t (V_t^T A V_t)^{-1} V_t^T \right) D^{-1/2}, \quad (14)$$

где V_t определены выше в (9).

Заметим, что блочно-диагональный предобусловливатель, эквивалентный применению блочного метода Якоби без перекрытия (где $m_t = n_t$, $t = 1, 2, \dots, p$), может быть записан как

$$H_{\text{ВЖ}} = \sum_{t=1}^p E_t (E_t^T A E_t)^{-1} E_t^T, \quad (15)$$

где $E_t = [e_{k_{t-1}+1} | \dots | e_{k_t}]$, $t = 1, 2, \dots, p$, так что $E_t^T A E_t$ в точности является t -м диагональным блоком матрицы A . Рассмотренные методы ВЖ и ОВЖ называют также методами типа декомпозиции области (Domain Decomposition, DD).

Предобусловливание по точечному методу Якоби (PJ) соответствует матрица-предобусловливатель

$$H_{\text{PJ}} = D_A^{-1},$$

где D_A^{-1} – диагональ матрицы A . Если же исходная матрица A уже отмасштабирована по Якоби, то $H_{\text{PJ}} = I$.

Во всех рассматриваемых методах (кроме точечного метода Якоби) использовался одинаковый алгоритм приближенного IC2-разложения для диагональных блоков.

Приведены результаты численных экспериментов для задач из коллекции университета Флориды¹. Численные эксперименты выполнялись на высокопроизводительном вычислительном комплексе MVS6000IM, использующем процессоры Dual Core Intel Itanium 2 с частотой 1.6 ГГц.

¹<http://www.cise.ufl.edu/research/sparse/matrices>

В качестве основной модельной задачи (обозначаемой далее DCD) была рассмотрена конечно-разностная дискретизация задачи минимизации квадратичного функционала $\min \int c|\Delta u|^2 dx_1 dx_2$ на квадрате $[0, 1] \times [0, 1]$, с кусочно-постоянными коэффициентами в виде

$$c(x_1, x_2) = \begin{cases} c^*, & 1/3 < x_1, x_2 < 2/3 \\ 1, & \text{иначе,} \end{cases}$$

с граничными условиями $u = 0$ и $\partial u / \partial n = 0$, с точным решением

$$u^* = u(x_1, x_2) = x_1 e^{x_1 x_2} \sin(\pi x_1) \sin(\pi x_2)$$

и начальным приближением $u_0 = 0$. Было выбрано значение $c^* = 10^3$ и использовалось 300 точек дискретизации вдоль каждого из направлений.

Также были рассмотрены несколько задач с плохо обусловленными разреженными матрицами из коллекции, собранной в университете Флориды. Характеристики тестовых матриц представлены в табл. 1, где использованы следующие обозначения: n – размерность матрицы A ; $\text{NZ}(A)$ – количество ненулевых элементов матрицы A ; $\text{Cond}(A_S)$ – нижняя оценка спектрального числа обусловленности матрицы $A_S = D_A^{-1/2} A D_A^{-1/2}$, т. е. исходной матрицы, симметрично отмасштабированной к единичной главной диагонали. Следует заметить, что только одна из рассмотренных матриц является М-матрицей. Необходимо также обратить внимание на широкий круг областей потенциального применения рассматриваемого метода решения систем линейных уравнений.

При построении предобусловливателя по умолчанию использовался порог отсечения малых элементов τ , равный 10^{-3} . Параметр, определяющий размер перекрытия q , был выбран равным 10 во всех рассмотренных случаях.

Во всех экспериментах правая часть выбиралась равной $b = Ax^*$, используя $x^* = [1 \ 1 \ \dots \ 1]^T$ в качестве точного решения и $x_0 = [0 \ 0 \ \dots \ 0]^T$ в качестве начального приближения. Итерации МСГ проводились до достижения критерия остановки

$$\|r_i\| \leq \varepsilon \|r_0\|, \quad \varepsilon = 10^{-8}. \quad (16)$$

В табл. 2 использованы следующие обозначения: p – количество процессоров (блоков); Dens – плотность предобусловливателя $\text{NZ}(U)/\text{NZ}(U_A)$

Таблица 1. Характеристики матриц из коллекции университета Флориды.

Матрица	n	$NZ(A)$	$Cond(A_S)$	Описание
DCD	90 000	1 164 004	9.61×10^{10}	расчет конструкций (пластина с утолщением)
bcsstk25	15 439	252 241	3.55×10^6	расчет конструкций (небоскреб)
msc23052	23 052	1 154 814	7.57×10^7	расчет конструкций
gridgena	48 962	512 084	1.35×10^5	оптимизация вычислительных сеток
cvxbqp1	50 000	349 968	1.31×10^6	выпуклое квадратичное программирование
oilpan	73 752	3 597 188	1.03×10^8	расчет конструкций
s3dkq4m2	90 449	4 820 891	1.74×10^{10}	МКЭ для цилиндрической оболочки
s3dkt3m2	90 449	3 753 461	3.12×10^{10}	МКЭ для цилиндрической оболочки
x104	108 384	10 167 624	2.12×10^9	расчет конструкций (SESAM)
bmw7st_1	141 347	7 339 667	1.03×10^7	расчет конструкций (матрица жесткости)
g2_circuit	150 102	726 674	2.34×10^5	моделирование СБИС (AMD, M-матрица)
pwtk	217 918	11 634 424	4.25×10^7	расчет конструкций (воздуховод)
hood	220 542	10 768 436	5.42×10^5	расчет конструкций (автомобилестроение)
BenElechi1	245 874	13 150 496	1.84×10^9	пространственно-плоскостная задача
msdoor	415 863	20 240 935	1.95×10^8	расчет конструкций (автомобилестроение)
af_1_k101	503 625	17 550 675	4.43×10^7	расчет конструкций (листовая штамповка)
parabolic_fem	525 825	3 674 625	2.10×10^5	МКЭ для конвекции-диффузии

по отношению к плотности верхнего треугольника исходной матрицы A ; I_{iter} – количество итераций, необходимых для достижения требуемой точности ε (16); T_{tot} – общее время решения задачи в секундах. Приведенные здесь данные свидетельствуют, что рассматриваемая задача DCD является достаточно сложной для решения. Это видно как из нижней оценки ее числа обусловленности (см. табл. 1), так и из приведенных данных о сходимости традиционного применяемого предобусловливания первого порядка $IC(\tau)$ (эквивалентного использованию $IC2(\tau, \tau)$), $\tau = 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$, а также сходимости применяемого предобусловливания второго порядка $IC2(\tau, \tau^2)$. Количество итераций для $IC(\tau)$, при использовании разумных значений τ , достаточно велико. Результаты, приведенные в табл. 2, показывают значительно бóльшую эффективность предобусловливания второго порядка точности $IC2$, сочетающего скорость сходимости $IC(\tau^2)$ с заполненностью $IC(\tau)$.

В табл. 2 также представлены результаты сравнения стратегий предобусловливания: $IC2$ -разложение с описанным выше ВИС предобусловливанием и с предобусловливанием по блочному методу Якоби (BJ) (который эквивалентен ВИС с параметром $q = 0$). Можно видеть, что сходимость

Таблица 2. Сравнение предобусловливающих для задачи DCD ($\tau = 10^{-3}$).

Метод	p	Dens	Iter	T_{tot}
IC(10^{-3})	1	2.44	9013	513.80
IC(10^{-4})	1	21.34	5014	1273.97
IC(10^{-5})	1	45.65	2113	1058.28
IC(10^{-6})	1	51.79	601	377.20
IC2(τ, τ^2)	1	4.47	626	94.38
IC2(τ, τ^2)-ВПС	8	5.37	528	8.17
IC2(τ, τ^2)-ВЖ	8	4.02	7033	55.13

для ВЖ достаточно медленная по сравнению с предобусловливанием ВПС, где применяются специальным образом построенные перекрытия.

Еще более наглядно сравнение с другими методами предобусловливания представлено на рис. 1а. Приведены данные времени счета для предобусловливающих по методам РЖ, ВЖ, ОВЖ, а также используемого метода ВПС. Для аппроксимации задач в подобластях во всех случаях использовалось разложение второго порядка IC2. Видно, что время решения для точечного метода Якоби РЖ неприемлемо велико во всех случаях. При использовании предобусловливания ОВЖ обычно происходит несущественное изменение количества итераций по сравнению с предобусловливанием без перекрытия ВЖ, но из-за значительно более плотного предобусловливателя выигрыша во времени решения не происходит. Результаты проведенных расчетов показывают, что для предложенного ВПС-IC2-предобусловливания получаются наилучшие результаты как по количеству итераций, так по общему времени решения.

Будем в дальнейшем использовать следующие обозначения: STD – описанное выше ВПС-IC2-предобусловливание, без использования какой-либо постфильтрации и балансировки; MIN – традиционно применяемая постфильтрация каждой матрицы \tilde{U}_t , которая уменьшает все $NZ(\tilde{U}_t)$ до их минимального значения среди всех p блоков; MAX – предложенный способ дополнения в каждую \tilde{U}_t максимальных по модулю элементов из матрицы ошибок R_t , чтобы увеличить все $NZ(\tilde{U}_t)$ до их максимального значения.

На рис. 1б–1г представлены результаты, иллюстрирующие применение различных стратегий балансировки вычислений. Для решения всех 17 задач из рассматриваемой коллекции приводятся графики ускорения вре-

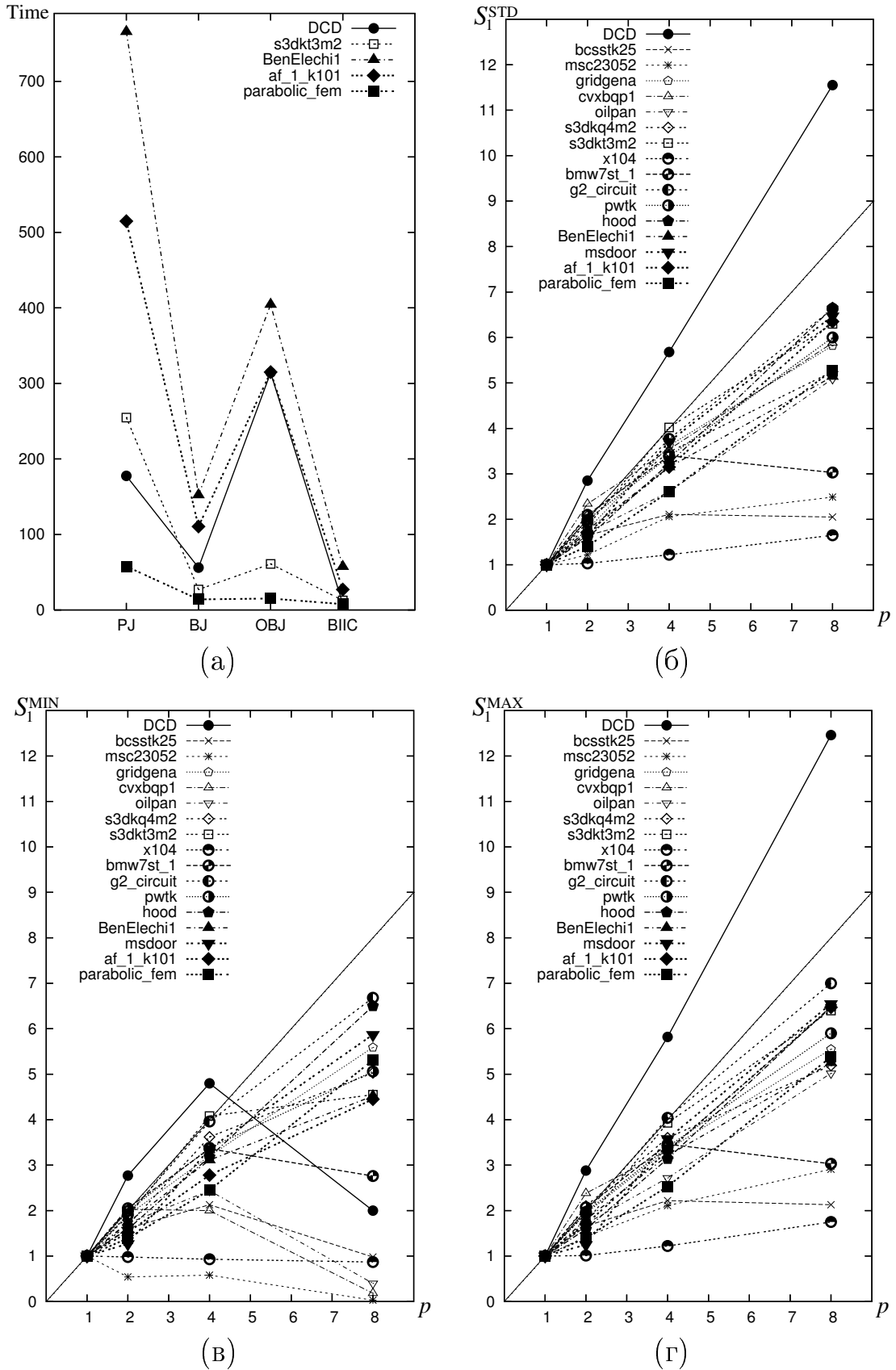


Рис. 1. Задачи из коллекции университета Флориды: (а) сравнение различных методов для $p = 8$; (б) ускорение без балансировки для ВИIC-IC2-МСТ; (в) ускорение для балансировки по традиционной стратегии MIN; (г) ускорение для балансировки по предложенной стратегии MAX.

мени счета

$$S_1 = \frac{T_{\text{tot}}(1)}{T_{\text{tot}}(p)},$$

где $T_{\text{tot}}(p)$ – общее время решения задачи на p процессорах, $p = 1, 2, 4, 8$.

Данные, приведенные на рис. 1б, показывают, что предложенное ВПС-IC2 предобусловливание даже без применения какой-либо балансировки показывает достаточно устойчивые высокие результаты ускорения. Если исключить из рассмотрения задачи небольшой размерности, то достигнутое ускорение составит более 5 на 8 процессорах.

На рис. 1в представлены результаты, показывающие, что наиболее очевидная и традиционно применяемая балансировка, основанная на стратегии MIN, дает медленную скорость сходимости. Так, резкое возрастание количества итераций для задачи DCD для ВПС-IC2 предобусловливания с 8 процессорами и традиционно используемой стратегией балансировки MIN объясняется чрезмерно большим значением $\lambda_{\max}(\tilde{M})$, где $\tilde{M} = HA$ – предобусловленная матрица (см. (13)). В этом случае оцененная граница для этой величины получилась равной 101.55, в то время как типичное значение, которое гарантирует численную устойчивость итераций МСГ, ниже 2. По этой же причине для задачи “msc23052” сходимость стала хуже почти в 100 раз. Здесь соответствующие спектральные характеристики оценены как $\lambda_{\max}(\tilde{M}) > 5 \cdot 10^4$ и $\text{Cond}(\tilde{M}) > 6 \cdot 10^8$. Примерно в половине рассмотренных случаев количество итераций при применении стратегии MIN возросло более чем в 2 раза, а в 17% случаев более чем в 10 раз. Наблюдаемые отрицательные эффекты могут быть связаны с нарушением сформулированного в теореме 2 условия на верхнюю границу нормы отбрасываемой части U_t .

С другой стороны, рис. 1г показывает, что предложенная балансировка на основе стратегии MAX достаточно работоспособна и эффективна. Улучшение достигнуто более чем в 70% случаев, в остальных случаях время счета отличалось не более чем на 5%, что можно считать лежащим в области ошибок измерения.

Отметим, что применение стратегий STD и MAX позволило получить сверхлинейное ускорение в 14% случаев. Причем для стратегии MAX выигрыш был более значителен. Помимо возможных эффектов, зависящих от оборудования (улучшение работы кэш-памяти при уменьшении размерности подзадачи на каждом из процессоров) указанное явление может иметь

чисто алгебраический характер. Это связано с более точной аппроксимацией задач в подобластях меньшего размера, что приводит к уменьшению общих арифметических затрат по сравнению с однопроцессорной реализацией без выделения блоков.

Проводилось также сравнение с опубликованными результатами других исследователей, которыми рассматривался алгебраический многосеточный метод (AMG) и алгебраический рекурсивный многоуровневый метод (ARMS) на примере решения 5 задач из рассмотренной коллекции. Исходя из опубликованных ими результатов, сравнения проводились для решения на одном процессоре. Практически во всех случаях предложенный метод дал лучшие результаты (иногда до 100 раз по времени счета), что свидетельствует о его более широкой области применимости при решении жестких задач.

Следует также заметить, что исследования не были ограничены только описанной коллекцией матриц. Были рассмотрены и другие области применения предложенного метода, а именно, задачи упругости тонкостенных оболочек и задачи линейной теории упругости. Кроме этого исследовалась эффективность предложенного метода при использовании до 64 блоков в ВПС-IC2-предобусловливания. Результаты этих расчетов, также приведенные в диссертационной работе, еще раз продемонстрировали надежность и эффективность предложенной методики построения параллельных предобусловливаний.

Выводы

Как теоретические исследования, так и экспериментальные результаты параллельной реализации ВПС-IC2-предобусловливания для МСГ показали его высокую эффективность для решения систем линейных алгебраических уравнений с разреженными с.п.о. плохо обусловленными матрицами. Результаты проведенных численных экспериментов свидетельствуют, что чем “сложнее” для решения выбирается задача, чем больше число обусловленности матрицы системы линейных уравнений, тем больше оказывается эффективность предложенного метода предобусловливания по сравнению с традиционно используемыми параллельными методами.

Численная реализация предложенного метода показала значительное ускорение по сравнению с расчетами на одном процессоре. Чем большей

размерности берется решаемая задача, тем большего ускорения удается достигнуть. Отсутствие замедления даже для задач малой размерности показывает высокую эффективность использования ресурсов параллельной ЭВМ. Повышение эффективности использования каждого процессора было достигнуто за счет новой методики балансировки вычислений. Таким образом, параллельные свойства предложенного алгоритма ВПС-IC2-МСГ при использовании умеренного количества процессоров оказались достаточно хорошими, как и ожидалось из теоретических построений.

Основные результаты

1. Разработан, исследован и реализован новый параллельный метод ВПС-IC2-МСГ решения жестких систем линейных уравнений с разреженными с.п.о. матрицами.
2. Разработана, исследована и реализована новая методика балансировки вычислений для подзадач на различных процессорах.
3. В численных экспериментах показана надежность предложенного метода, в частности, относительно изменения таких параметров предобусловливания, как параметр отсечения малых элементов, параметр перекрытия и количество блоков.
4. Показана высокая эффективность предложенного параллельного метода, вплоть до сверхлинейного ускорения.

Основное содержание диссертационной работы изложено в следующих публикациях:

- [1] *Коньшин И.Н.* О реализации методов разбиения области на вычислительных системах с матричными процессорами // Методы математической физики. – М.: ОВМ АН СССР. 1988. 23 с.
- [2] *Коньшин И.Н.* Оптимизация многосеточных методов декомпозиции области // Численные методы и программное обеспечение. – М.: ОВМ АН СССР. 1990. 22 с.
- [3] *Ibragimov I.V., Kolesnikov P.A., Konshin I.N., et al.* Numerical experiments with industrial iterative solvers on massively parallel computers. I: Numerical experiments with the A_SPARSE solver on CRAY T3D //

- In: Proc. of the High Performance Computing Symposium'95. Phoenix, Arizona, USA. 1995. P. 283–289.
- [4] *Kaporin I.E., Konshin I.N.* Parallel solution of large sparse SPD linear systems based on overlapping domain decomposition // Lecture Notes in Computer Science, Vol.1662. Springer-Verlag, Berlin–Heidelberg–New-York. 1999. P. 436–445.
- [5] *Axelsson O., Kaporin I., Konshin I., et al.* Comparison of algebraic solution methods on a set of benchmark problems in linear elasticity // Tech. Report of Department of Mathematics, University of Nijmegen, The Netherlands. 2000. 89 p.
- [6] *Капорин И.Е., Коньшин И.Н.* Параллельное решение симметричных положительно-определенных систем на основе перекрывающегося разбиения на блоки // Ж. Вычисл. Матем. и Матем. Физ. 2001. Т. 41. № 4. С. 515–528.
- [7] *Kaporin I.E., Konshin I.N.* A parallel block overlap preconditioning with inexact submatrix inversion for linear elasticity problems // J. Numer. Lin. Alg. Appl. 2002. V. 9. № 2. P. 141–162.
- [8] *Капорин И.Е., Коньшин И.Н.* Параллельное решение линейных систем при использовании приближенной факторизации перекрывающихся блоков // В кн.: “Математическое моделирование. Проблемы и результаты.” (Ред. О.М.Белоцерковский, В.А.Гуцин) – М.: Наука. 2003. С. 308–319.
- [9] *Kaporin I.E., Konshin I.N.* Post-filtering of IC factors for load balancing in parallel preconditioned CG solvers // Numerical geometry, grid generation and high performance computing (NUMGRID2008). (Eds. V.A.Garanzha, Yu.G.Evtushenko, B.K.Soni, N.P.Weatherill) – М.: Folium, 2008. P. 158–164.
- [10] *Капорин И.Е., Коньшин И.Н.* Постфильтрация множителей IC2-разложения для балансировки параллельного предобусловливания // Ж. Вычисл. Матем. и Матем. Физ. 2009. Т. 49. № 6. С. 940–957.
- [11] *Kaporin I.E., Konshin I.N.* Load balancing of parallel block overlapped incomplete Cholesky preconditioning // Lecture Notes in Computer Science, Vol. 5698. Springer-Verlag, Berlin–Heidelberg. 2009. P. 304–315.