

МИНОБРНАУКИ РОССИИ
Ярославский государственный университет им. П.Г. Демидова

Кафедра нелинейной динамики

УТВЕРЖДАЮ

Декан математического факультета



Нестеров П.Н.

20 июня 2023 г.

Рабочая программа дисциплины
Суперкомпьютерные технологии и основы искусственного интеллекта

Направление подготовки (специальности)
01.03.02 Прикладная математика и информатика

Направленность (профиль)
«Прикладное программирование и информационные технологии»

Форма обучения очная

Программа рассмотрена
на заседании кафедры
от 12 апреля 2023 г., протокол № 8

Программа одобрена НМК
математического факультета
протокол № 9 от 3 мая 2023 г.

1. Цели освоения дисциплины

Целями преподавания дисциплины «Суперкомпьютерные технологии и основы искусственного интеллекта» являются:

знакомство слушателей с теоретическими основами и современными программно-аппаратными средствами для решения прикладных задач в следующих областях:

1. Классические высокопроизводительные вычисления (HPC) на примерах математического моделирования, комбинаторики и статистических методов
2. Машинное обучение с приложениями в области машинного зрения, обработки естественных языков, задач биоинформатики
3. Введение в искусственный интеллект и взаимодействие интеллектуальных агентов

2. Место дисциплины в структуре образовательной программы

Дисциплина «Суперкомпьютерные технологии и основы искусственного интеллекта» относится к вариативной части Блока 1.

3. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Процесс изучения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО, ООП ВО и приобретения следующих знаний, умений, навыков и (или) опыта деятельности:

Формируемая компетенция (код и формулировка)	Индикатор достижения компетенции (код и формулировка)	Перечень планируемых результатов обучения
Общепрофессиональные компетенции		
ОПК-4 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ИД-ОПК-4.1 Понимает принципы работы современных информационных технологий ИД-ОПК-4.2 Способен использовать современные информационные технологии для решения задач профессиональной деятельности	<p>Знать: архитектуру кластерных вычислительных систем и архитектуру сопроцессоров для задач машинного обучения</p> <p>Уметь: оценивать применимость вычислительных платформ для прикладных задач</p> <p>Знать: методы решения задач математического моделирования и машинного обучения на современных вычислительных платформах</p> <p>Уметь: выбирать эффективные алгоритмы и/или их библиотечные реализации для решения прикладных задач</p> <p>Владеть навыками: разработки программного обеспечения для задач математического моделирования и машинного обучения</p>

4. Объем, структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетные единицы, 108 акад. часов.

№ п/п	Темы (разделы) дисциплины, их содержание	Семестр	Виды учебных занятий, включая самостоятельную работу студентов, и их трудоемкость (в академических часах)						Формы текущего контроля успеваемости Форма промежуточной аттестации (по семестрам) Формы ЭО и ДОТ (при наличии)
			Контактная работа					самостоятельная работа	
			лекции	практические	лабораторные	консультации	аттестационные испытания		
1	Введение	15	4	6		1		4	Задания для самостоятельной работы
	<i>в том числе с ЭО и ДОТ</i>	0	0	0		0		0	
2	Высокопроизводительны е вычисления	37	12	18		3		4	Практическое задание 1
	<i>в том числе с ЭО и ДОТ</i>	0	0	0		0		0	
3	Машинное обучение	27	8	12		3		4	Задания для самостоятельной работы
	<i>в том числе с ЭО и ДОТ</i>	0	0	0		0		0	
4	Искусственный интеллект	27	8	12		3		4	Практическое задание 2
	<i>в том числе с ЭО и ДОТ</i>	0	0	0		0		0	
		2					0.3	1.7	зачет
	ИТОГО	108	32	48		10	0.3	17.7	
	<i>в том числе с ЭО и ДОТ</i>	0	0	0		0	0	0	

Содержание разделов дисциплины:

1. Введение

1.1. Обзор курса и его мотивация. Работа с вычислительным кластером. Введение в Python notebooks / JupyterLab

1.2. Классические высокопроизводительные вычисления. Машинное обучение. Глубокое обучение. Суррогатные модели - синтез классического НРС и машинного обучения.

2. Высокопроизводительные вычисления

2.1. Современные архитектуры и программные модели (многоядерность, grgri, кластеры, - openmp, cuda, mpi), high performance vs high throughput/availability. OpenMP и параллелизм циклов

2.2. Параллелизм задач, виды и ограничения

2.3. Классические задачи НРС: дифференциальные уравнения - математическое моделирование. Решение уравнения теплопроводности.

2.4. Классические задачи НРС: комбинаторика, графы — рендеринг. Динамический параллелизм для деревьев

2.5. MPI и крупнозернистый параллелизм. CUDA и линейная алгебра. Быстрое

преобразование Фурье

2.6 Классические задачи НРС: Монте-Карло - статистические методы. Параллельные генераторы случайных чисел

3. Машинное обучение

3.1. Формальная модель обучения, равномерная сходимость. Обзор фреймворков для машинного обучения / искусственного интеллекта.

3.2. Линейные предикторы. Дилемма смещения-сложности. VC-размерность. Фреймворки для ML/AI: Tensorflow, PyTorch

3.3. Бустинг. Метод опорных векторов. Кластеризация

3.4. Распознавание речи. Задачи геологии и сейсмологии.

4. Искусственный интеллект

4.1. Нейросети. Полносвязные и рекуррентные слои. Обработка естественных языков.

4.2. Сверточные нейросети. Задачи машинного зрения, детекция, трекинг.

4.3. Генеративно-состязательные сети.

4.4. Метрики качества детекторов и классификаторов.

5. Образовательные технологии, в том числе технологии электронного обучения и дистанционные образовательные технологии, используемые при осуществлении образовательного процесса по дисциплине

В процессе обучения используются следующие образовательные технологии:

Вводная лекция – дает первое целостное представление о дисциплине и ориентирует студента в системе изучения данной дисциплины. Студенты знакомятся с назначением и задачами курса, его ролью и местом в системе учебных дисциплин и в системе подготовки в целом. Дается краткий обзор курса, история развития науки и практики, достижения в этой сфере, имена известных ученых, излагаются перспективные направления исследований. На этой лекции высказываются методические и организационные особенности работы в рамках данной дисциплины, а также дается анализ рекомендуемой учебно-методической литературы.

Академическая лекция с элементами лекции-беседы – последовательное изложение материала, осуществляемое преимущественно в виде монолога преподавателя. Элементы лекции-беседы обеспечивают контакт преподавателя с аудиторией, что позволяет привлекать внимание студентов к наиболее важным темам дисциплины, активно вовлекать их в учебный процесс, контролировать темп изложения учебного материала в зависимости от уровня его восприятия.

Практическое занятие – занятие, посвященное освоению конкретных умений и навыков по закреплению полученных на лекции знаний.

Консультации – вид учебных занятий, являющийся одной из форм контроля самостоятельной работы студентов. На консультациях по просьбе студентов рассматриваются наиболее сложные моменты при освоении материала дисциплины, преподаватель отвечает на вопросы студентов, которые возникают у них в процессе самостоятельной работы.

В процессе обучения используются следующие технологии электронного обучения

и дистанционные образовательные технологии:

Jupyter Notebooks, JupyterHub, Google Colab

6. Перечень лицензионного и (или) свободно распространяемого программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине

Используется следующее свободно распространяемое программное обеспечение:

Nvidia CUDA, Nvidia cuDNN, OpenMP, OpenMPI, Python, NumPy, SciPy, Pandas, OpenCV, PyTorch, YOLOv5

7. Перечень современных профессиональных баз данных и информационных справочных систем, используемых при осуществлении образовательного процесса по дисциплине (при необходимости)

При изучении дисциплины используются:

- справочные системы используемых программных пакетов и библиотек, предоставляемые их разработчиками.

- автоматизированная библиотечно-информационная система «БУКИ-NEXT»

http://www.lib.uniyar.ac.ru/opac/bk_cat_find.php

8. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет» (при необходимости), рекомендуемых для освоения дисциплины

а) основная литература

Параллельные вычисления на GPU: архитектура и программная модель CUDA: учеб. пособие для вузов / [А.В. Боресков и др.]; УМО по классическому унив. образованию; МГУ. – М.: МГУ, 2012. – 333 с.

Программируем на Python / М. Доусон; [пер. с англ. В.Порицкого]. – СПб.: Питер, 2016. – 414 с.

б) дополнительная литература

Д. Уоткинс **Основы матричных вычислений**. М.:Бином, 2006.

Коршунов Д.А. , Чернова Н.И. **Сборник задач и упражнений по математической статистике**. Учебное пособие. Новосиб. Гос. Унив. 2004.

Shai Shalev-Shwartz and Shai Ben-David **Understanding Machine Learning: From Theory to Algorithms**. Cambridge University press, 2014

<https://www.cs.huji.ac.il/w~shais/UnderstandingMachineLearning/>

В.В.Воеводин, Вл.В.Воеводин **Параллельные вычисления** -СПб.: БХВ-Петербург, 2002.- 608 с. ISBN 5-94157-160-7

в) ресурсы сети «Интернет» (при необходимости)

MIT 6.S191 Introduction to Deep Learning <http://introtodeeplearning.com/>

PyTorch <https://pytorch.org/>

YOLOv5 <https://github.com/ultralytics/yolov5>

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине включает в свой состав специальные помещения:

- учебные аудитории для проведения занятий лекционного типа;
- компьютерные классы для проведения практических занятий (семинаров);
- учебные аудитории для проведения групповых и индивидуальных консультаций;
- учебные аудитории для проведения текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы;
- помещения для хранения и профилактического обслуживания технических средств обучения.
- вычислительный кластер ЯрГУ с программным обеспечением для параллельных вычислений и машинного обучения.

Специальные помещения укомплектованы средствами обучения, служащими для представления учебной информации большой аудитории.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде ЯрГУ.

Автор(ы):
научный сотрудник, к.ф.-м.н, Глызин Д.С.

должность, ученая степень

подпись

И.О. Фамилия

Приложение № 1 к рабочей программе дисциплины
«Суперкомпьютерные технологии и основы искусственного интеллекта»
наименование дисциплины

Фонд оценочных средств
для проведения текущего контроля успеваемости
и промежуточной аттестации студентов
по дисциплине

4. Типовые контрольные задания и иные материалы,
используемые в процессе текущего контроля успеваемости

Задания для самостоятельной работы (преподавателем не проверяются)

1. По теме «Введение»

1. Найти оценку максимального правдоподобия параметра b равномерного распределения на отрезке $[0, b]$
2. Найти оценку максимального правдоподобия параметра $b > 0$ равномерного распределения на отрезке $[-b, 0]$
3. Найти оценку максимального правдоподобия параметра $b > 0$ равномерного распределения на отрезке $[-b, b]$
4. Найти оценку максимального правдоподобия параметра $b > 0$ равномерного распределения на отрезке $[b, b+2]$
5. Найти оценку максимального правдоподобия параметра $b > 0$ равномерного распределения на отрезке $[b, 2b]$
6. Пусть X_1, \dots, X_n — выборка из равномерного распределения на отрезке $[0, b]$. С помощью статистики $X_{(n)}$ постройте точный доверительный интервал уровня $1-\epsilon$ для параметра b .

2. По теме «Машинное обучение»

Рассмотрим классификаторы, возвращающие 0 для всех точек внутри прямоугольной области со сторонами, параллельными осям координат. Предположим также, что выполнено условие реализуемости.

1. Покажите, что алгоритм, возвращающий минимальный прямоугольник, содержащий позитивные метки из обучающей выборки, минимизирует эмпирический риск.
2. Покажите, что если размер обучающей выборки больше или равен $4 \log(4/\delta) / \epsilon$ то с вероятностью по меньшей мере $1-\delta$ этот алгоритм возвращает гипотезу с ошибкой не более ϵ
3. Распространите результат из п.2 на пространство R^d .
4. Покажите, что трудоемкость данного алгоритма полиномиальна по размерности d , по $1/\epsilon$, а также по $\log(1/\delta)$.

Практическое задание 1 (по теме «Высокопроизводительные вычисления»)

- Вариант 1. Решите численно с применением технологии OpenMP уравнение теплопроводности в квадратной области с заданными граничными условиями.
- Вариант 2. Решите численно с применением технологии MPI уравнение теплопроводности в квадратной области с заданными граничными условиями.
- Вариант 3. Решите численно с применением технологии CUDA уравнение теплопроводности в квадратной области с заданными граничными условиями.

Практическое задание 2 (по теме «Искусственный интеллект»)

Выбрав один из открытых наборов данных для тренировки машинного зрения (<https://public.roboflow.com/>) произведите тренировку в сервисе Google Colab (<https://colab.research.google.com>) нейросети YOLOv5s (<https://github.com/ultralytics/yolov5>) Проанализируйте полученные результаты с точки зрения качества детекции и классификации.

5. Список вопросов и (или) заданий для проведения промежуточной аттестации
На зачете проверяется сформированность компетенции ОПК-4 (индикаторы ИД-ОПК-4_1, ИД-ОПК-4_2)

Список вопросов к зачету (проведение зачета подразумевает наличие у студента разработанных программ для практических заданий 1 и 2):

по практическому заданию 1:

1. Объясните выбор цикла для распараллеливания в решении задачи математического моделирования
2. Измените направление распараллеливания (по строкам или по столбцам) расчета матрицы состояния
3. Измените расположение граничного условия в программе
4. Проинтерпретируйте результаты ускорения работы параллельной программы

по практическому заданию 2:

5. Проинтерпретируйте кривую precision в результатах обучения нейросети
6. Проинтерпретируйте кривую recall в результатах обучения нейросети
7. Проинтерпретируйте кривую mAP в результатах обучения нейросети
8. Объясните различия кривых mAP для обучающей и валидационной выборок.

Приложение № 2 к рабочей программе дисциплины
«Суперкомпьютерные технологии и основы искусственного интеллекта»
наименование дисциплины

Методические указания для студентов по освоению дисциплины

Курс состоит из лекций и практических занятий. Лекции обеспечивают теоретическую подготовку студентов, создают представление о наборе решаемых современными вычислительными средствами задач, о том, какие результаты могут быть сформулированы на вероятностном языке строго. Практические занятия требуют от студентов базовых навыков программирования и проходят путем демонстрации решения типовых задач средствами языков C++ и Python с использованием соответствующих библиотек. Крайне важной является самостоятельная работа над программным кодом и наработка навыков создания воспроизводимых расчетов с анализом полученных характеристик.

Методические указания по разработке программы для параллельного решения задачи теплопроводности (практическое задание 1):

Решение двумерного уравнения
теплопроводности
с использованием OpenMP

Глызин Дмитрий Сергеевич
glyzin@gmail.com
<http://parallel.ncycle.org>

ЯрГУ им. П.Г. Демидова

I семестр 2011/2012 учебного года

Уравнение теплопроводности

Рассмотрим уравнение

$$\frac{\partial u(x, y, t)}{\partial t} = \frac{\partial^2 u(x, y, t)}{\partial x^2} + \frac{\partial^2 u(x, y, t)}{\partial y^2} \quad (1)$$

в квадрате $x, y \in [0, 1]$, при $t \geq 0$, с граничными условиями

$$u(x, 0, t) = u(x, 1, t) = u(1, y, t) = 0, \quad u(0, y, t) = 100, \quad (2)$$

и начальным условием

$$u(x, y, t) = 0, \quad 0 < x < 1, \quad 0 < y < 1. \quad (3)$$

Задача (1)-(3) является моделью распространения тепла по квадратной пластине, одна сторона которой постоянно нагревается до 100 градусов, а остальные — охлаждаются до нуля.

Рассмотрим уравнение

$$\frac{\partial u(x, y, t)}{\partial t} = \frac{\partial^2 u(x, y, t)}{\partial x^2} + \frac{\partial^2 u(x, y, t)}{\partial y^2} \quad (1)$$

в квадрате $x, y \in [0, 1]$, при $t \geq 0$, с граничными условиями

$$u(x, 0, t) = u(x, 1, t) = u(1, y, t) = 0, \quad u(0, y, t) = 100, \quad (2)$$

и начальным условием

$$u(x, y, t) = 0, \quad 0 < x < 1, \quad 0 < y < 1. \quad (3)$$

Задача (1)-(3) является моделью распространения тепла по квадратной пластине, одна сторона которой постоянно нагревается до 100 градусов, а остальные — охлаждаются до нуля.



Метод

Будем решать задачу (1)-(3) численно, разбив квадрат равномерной сеткой из N точек по каждому направлению и заменяя производные на конечные разности по этой сетке. Обозначим $\Delta x = 1/(N - 1)$, $\Delta y = 1/(N - 1)$ и фиксируем шаг по времени Δt , тогда

$$\frac{\partial u(x, y, t)}{\partial t} \approx \frac{u(x, y, t + \Delta t) - u(x, y, t)}{\Delta t}, \quad (4)$$

$$\frac{\partial^2 u(x, y, t)}{\partial x^2} \approx \frac{u(x + \Delta x, y, t) + u(x - \Delta x, y, t) - 2u(x, y, t)}{\Delta x^2}, \quad (5)$$

$$\frac{\partial^2 u(x, y, t)}{\partial y^2} \approx \frac{u(x, y + \Delta y, t) + u(x, y - \Delta y, t) - 2u(x, y, t)}{\Delta y^2}. \quad (6)$$



Метод

Теперь введем обозначения для узлов сетки: $x_i = i\Delta x$, $y_j = j\Delta y$, $t_k = k\Delta t$. Значение функции в узле обозначим $u_{i,j,k} = u(x_i, y_j, t_k)$. Как известно, при выполнении условия

$$\left(\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} \right) \Delta t \leq \frac{1}{2} \quad (7)$$

решение конечно-разностной аппроксимации

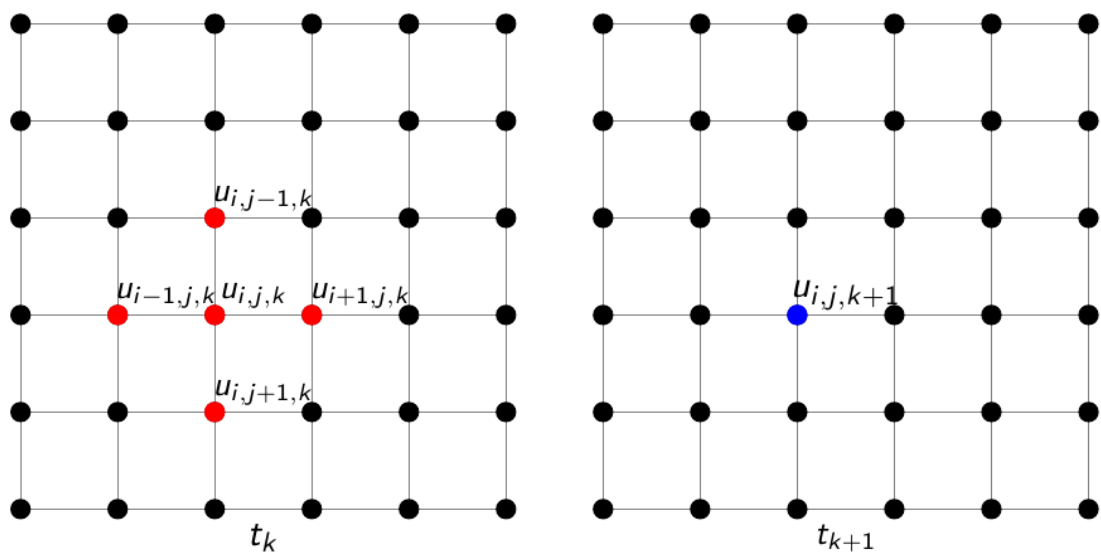
$$\frac{u_{i,j,k+1} - u_{i,j,k}}{\Delta t} = \frac{u_{i+1,j,k} + u_{i-1,j,k} - 2u_{i,j,k}}{\Delta x^2} + \frac{u_{i,j+1,k} + u_{i,j-1,k} - 2u_{i,j,k}}{\Delta y^2} \quad (8)$$

сходится с измельчением шагов к решению задачи (1)-(3). Из уравнения (8) можно выразить $u_{i,j,k+1}$ через значения в предыдущий момент времени.



Метод

Для того, чтобы найти значение $u_{i,j,k+1}$ по формуле (8) в точке, отмеченной синим, потребуется пять значений в точках, отмеченных красным:



Параллелизм

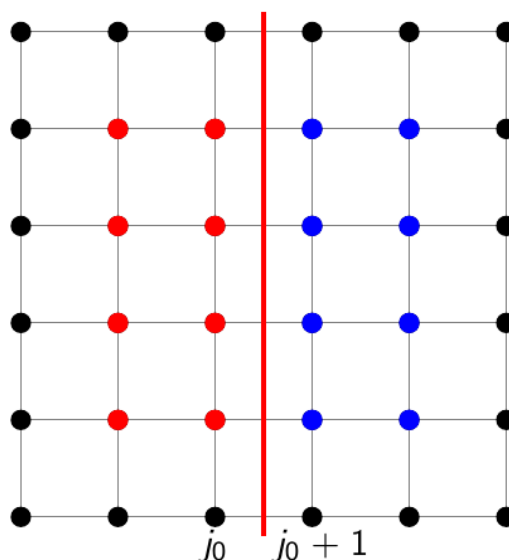
Для параллельного решения задачи воспользуемся декомпозицией области. Пусть у нас имеется p процессоров, разделим пространственную сетку по одной из координат на p прямоугольных блоков одинаковой величины. На каждой итерации по времени вычисление значений элементов блока не зависит от значений в других блоках, за исключением пограничных рядов.

Таким образом, итерации каждого блока можно выполнять на отдельном процессоре, обмениваясь в конце итерации с "соседями" значениями из пограничных рядов.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ≡ ↺ ↻

Параллелизм

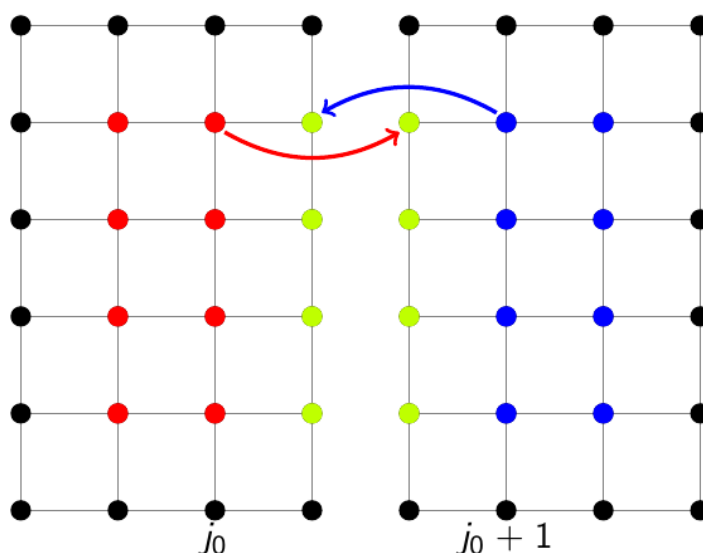
Разделим область по вертикали между столбцами j_0 и $j_0 + 1$. Черными обозначены граничные точки, не требующие пересчета, красными и синими — точки, обрабатываемые первым и вторым процессорами.



◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ≡ ↺ ↻

Параллелизм

Дополним матрицы каждого из процессоров фиктивными столбцами с нужной стороны, и после каждой итерации будем копировать в них крайний вычисленный столбец соседнего процессора:



Требования к программе

Программа должна компилироваться либо в Visual Studio 2008 (необходимо прислать проект целиком), либо с помощью TDM-GCC (или gcc в Linux), в этих случаях нужно прислать make-файл.

Ввод параметров осуществляется через файл "input.txt", лежащий в той же директории, что исполняемый файл. Каждая строка файла содержит по одному числу в следующем порядке:

- ① p , количество нитей, на которое распараллеливается вычисление ($1 \leq p \leq 4$)
- ② N , количество точек сетки по каждому направлению
- ③ Δt , шаг по времени
- ④ T , длина промежутка времени, на котором решается задача
- ⑤ filename, имя файла для вывода данных

Требования к программе

Параметры должны удовлетворять условию $4(N - 1)^2 \Delta t \leq 1$, в противном случае программа должна выводить сообщение об ошибке и останавливаться.

Время вычислений нужно измерить и вывести в файл filename первой строкой, в начале которой нужно поставить знак # (знак комментария в gnuplot). Это время не должно учитывать ввод или вывод.



Требования к программе

Выводить приближение $u(T, x, y)$ в файл filename нужно в соответствии с требованиями gnuplot для хранения трехмерных графиков, т.е. в виде:

$x_0 \quad y_0 \quad u(T, x_0, y_0)$

$x_0 \quad y_1 \quad u(T, x_0, y_1)$

...

$x_0 \quad y_{N-1} \quad u(T, x_0, y_{N-1})$

пустая строка

$x_1 \quad y_0 \quad u(T, x_1, y_0)$

$x_1 \quad y_1 \quad u(T, x_1, y_1)$

...

$x_1 \quad y_{N-1} \quad u(T, x_1, y_{N-1})$

пустая строка

...

$x_{N-1} \quad y_0 \quad u(T, x_{N-1}, y_0)$

$x_{N-1} \quad y_1 \quad u(T, x_{N-1}, y_1)$

...

$x_{N-1} \quad y_{N-1} \quad u(T, x_{N-1}, y_{N-1})$



Полезные советы

- Сравнивайте результаты, которые выводит программа при вычислении в одну нить с вычислениями в несколько нитей. Результаты должны быть идентичными, это простой способ проверить корректность работы параллельной программы.
- Поскольку параллелизм задачи при большом размере достаточно высок, ускорение на двухядерном процессоре при запуске в одну и в две нити должно быть близко к 1.5. Это хороший признак того, что программа работает эффективно.
- При работе с массивами избегайте лишнего копирования больших объемов данных. В случае, когда требуется поменять массивы ролями, можно обойтись обменом указателей.
- Во избежание лишних проверок можно считать, что $N - 2$ делится на p нацело.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ↺ ↻

