МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ
Проректор по учебной работе
________ О.Г. Локтионова
«25» ________ 2022 г.

ЧИСЛЕННОЕ РЕШЕНИЕ ЗАДАЧ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ МЕТОДАМИ ПЕРВОГО ПОРЯДКА

Методические указания к лабораторным и практическим занятиям для
студентов направлений подготовки 09.03.01 и 09.04.01
Информатика и вычислительная техника

Курск 2022

УДК 534.1
Составитель Ж.Т. Жусубалиев

Рецензент
Кандидат технических наук, доцент *Т.Н. Конаныхина*

Численное решение задач безусловной минимизации методами первого порядка: методические указания к лабораторным и практическим занятиям для студентов направлений подготовки 09.03.01 и 09.04.01 Информатика и вычислительная техника/ Юго-Зап. гос. ун-т; сост. Ж.Т. Жусубалиев. – Курск, 2022. – 11 с.: ил.1. – Библиогр.: с. 11.

Описывается численные методы безусловной многомерной минимизации методами первого порядка. Описан алгоритм наискорейшего градиентного спуска и сопряженных градиентов. Предназначены для студентов направлений подготовки 09.03.01, 09.04.01 очной и заочной форм обучения.

Текст печатается в авторской редакции

Подписано в печать 2022. Формат $\mathbf{60 \times 84\,{}^1/_{16}}$.
Усл. печ. л. . Уч.-изд. л. . Тираж 50 экз. Заказ . Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

1. Цель работы

Изучение численных методов многомерной безусловной минимизации методами первого порядка.

2. Постановка задачи

Рассмотрим задачу безусловной минимизации дифференцируемой функции многих переменных $\boldsymbol{f}(\boldsymbol{x})$, $\boldsymbol{f} \in \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^n$.

Пусть $\boldsymbol{x}^{(k)}$ – приближение к точке минимума $\boldsymbol{x}_*$ и $\boldsymbol{g}^{(k)} = \boldsymbol{g}(\boldsymbol{x}^{(k)})$. Как мы знаем, что в малой окрестности $\boldsymbol{x}^(k)$ на направление наискорейшего убывания функции $\boldsymbol{f}(\boldsymbol{x})$ указывает антиградиент $-\boldsymbol{g}^{(k)}$.

В градиентных методах направление спуска из точки $\boldsymbol{x}^{(k)}$ выбирается $\boldsymbol{(}p)^(k) = -\boldsymbol{g}^{(k)}$.

Таким образом в градиентных методах

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k\,\boldsymbol{g}^{(k)}, \quad k = 0, 1, 2, ... \tag{1}$$

В зависимости от того, как выбирается шаг α_k, существуют различные модификации метода. Эти методы относятся к методам первого порядка.

3. Метод наискорейшего спуска

Пусть

$$\varphi_k(\alpha) = \boldsymbol{f}(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{g}^{(k)}) \tag{2}$$

есть функция скалярной переменной $\alpha \geqslant 0$.

В методе *наискорейшего градиентного спуска* шаг α_k выбирается из условия

$$\alpha_k = \min_{\alpha > 0} \varphi_k(\alpha).$$

Метод был предложен в 1845 г. французским математиком О.Коши.

Теорема 1. *Пусть $\boldsymbol{f}(\boldsymbol{x})$ дифференцируема. Тогда в итерационном процессе*

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^(k) - \alpha_k\,\boldsymbol{g}^{(k)}, \quad \alpha_k = \min_{\alpha > 0} \varphi_k(\alpha), \quad \varphi_k(\alpha) = \boldsymbol{f}(\boldsymbol{x} - \alpha \boldsymbol{g}^{(k)}), \quad k = 0, 1, 2, ...$$

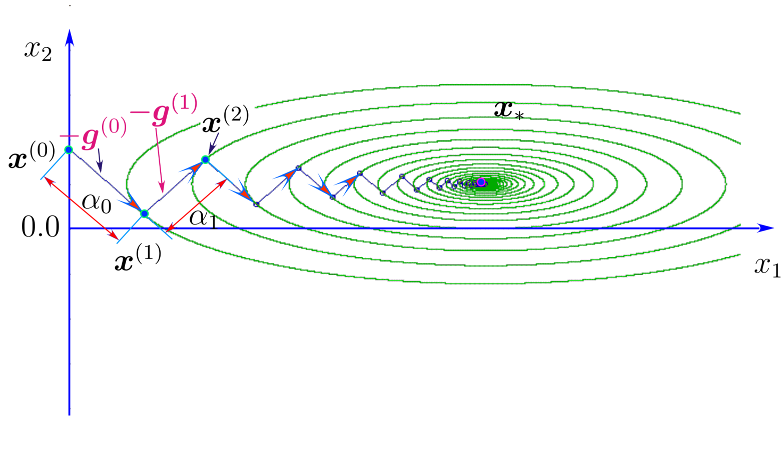

Рис. 1.

для любого $k \geqslant 1$ *справедливо*

$$\left(\nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}), \boldsymbol{p}^{(k)}\right) = 0, \quad \boldsymbol{x} = (x_1, x_2, ..., x_n)^T, \quad \boldsymbol{p} = (p_1, p_2, ..., p_n)^T, \qquad (3)$$

где «T» – знак транспонирования.

Доказательство. Запишем необходимое условие минимума функции $\varphi(\alpha)$ (см. (2))

$$\varphi_k'(\alpha) = 0,$$

используя правило дифференцирования сложной функции

$$\varphi_k'(\alpha) = \frac{d\varphi_k(\alpha)}{d\alpha} = \sum_{i=1}^{n} \frac{\partial \boldsymbol{f}(\boldsymbol{x}_i^{k+1})}{\partial x_i} \cdot \frac{\partial x^{k+1})}{\partial \alpha} = 0.$$

Принимая во внимание, что

$$x_i^{(k+1)} = x_i^{(k)} + \alpha\, p_i^{(k)},$$

получаем условие (3). □

На рис. 1 изображена геометрическая иллюстрация метода наискорейшего градиентного спуска.

Из начальной точки $\boldsymbol{x}^{(0)}$ в направлении $\boldsymbol{p}^{(0)} = -\boldsymbol{g}^{(0)}$ спуск продолжается до тех пор, пока либо не будет достигнута стационарная точка, в которой $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) = 0$, или прямая не коснется в точке $\boldsymbol{x}^{(k+1)})$ некоторой линии уровня. Равенство (3) и есть условие касания.

4. Минимизации квадратичной функции методом наискорейщего градиентного спуска

Решим задачу минимизации квадратичной функции

$$\boldsymbol{f}(\boldsymbol{x}) = \frac{1}{2}\left(A\boldsymbol{x}, \boldsymbol{x}\right) - \left(\boldsymbol{b}, \boldsymbol{x}\right), \quad A \in \mathbb{R}^{n\times n}, \quad \boldsymbol{b}, \boldsymbol{x} \in \mathbb{R}^{n}. \tag{4}$$

Здесь A – симметричная положительно-определенная матрица.

Как мы уже знаем, градиент квадратичной функции

$$\nabla \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}.$$

Тогда формула (1) принимает вид

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \left(A\boldsymbol{x}^{(k)} - \boldsymbol{b}\right), \quad k = 0, 1, 2, ...$$

Найдем функцию $\varphi_k(\alpha)$:

$$\varphi_k(\alpha) = \frac{1}{2}\left(A(\boldsymbol{x}^{(k)} - \alpha\boldsymbol{g}^{(k)}),\ \boldsymbol{x}^{(k)} - \alpha\boldsymbol{g}^{(k)}\right) - (\boldsymbol{b},\ \boldsymbol{x}^{(k)} - \alpha\boldsymbol{g}^{(k)}).$$

Отсюда отсюда получим квадратичную функцию переменной α

$$\varphi_k(\alpha) = \frac{1}{2}\left(A\boldsymbol{g}^{(k)},\ \boldsymbol{g}^{(k)}\right)\cdot\alpha^2 - (\boldsymbol{g}^{(k)},\ \boldsymbol{g}^{(k)})\cdot\alpha + \frac{1}{2}\left(A\boldsymbol{x}^{(k)},\ \boldsymbol{x}^{(k)}\right) - (\boldsymbol{b},\ \boldsymbol{x}^{(k)})$$

Запишем необходимое условие минимума

$$\varphi_k'(\alpha) = 0,$$

где

$$\varphi_k'(\alpha) = \left(A\boldsymbol{g}^{(k)},\ \boldsymbol{g}^{(k)}\right)\cdot\alpha - (\boldsymbol{g}^{(k)},\ \boldsymbol{g}^{(k)}).$$

Таким образом,

$$\alpha_k = \frac{(\boldsymbol{g}^{(k)},\ \boldsymbol{g}^{(k)})}{\left(A\boldsymbol{g}^{(k)},\ \boldsymbol{g}^{(k)}\right)}.$$

Алгоритм

$\varepsilon \leftarrow 10^{-12}$;
$\boldsymbol{x} \leftarrow \boldsymbol{x}_0$;
REPEAT
$\boldsymbol{x}_0 \leftarrow x$;
$\boldsymbol{g}_0 \leftarrow A\boldsymbol{x}_0 - \boldsymbol{b}$;
$\alpha_0 = \dfrac{(\boldsymbol{g}_0, \boldsymbol{g}_0)}{(A\boldsymbol{g}_0, \boldsymbol{g}_0)}$;
$\boldsymbol{x} \leftarrow \boldsymbol{x}_0 - \alpha_0(A\boldsymbol{x}_0 - \boldsymbol{b})$;
UNTIL $||\boldsymbol{g}_0|| < \varepsilon$;

5. Метод наискорейщего градиентного спуска

Итак, подведем итог. Пусть дана функция многих переменных $\boldsymbol{f}(\boldsymbol{x})$, $\boldsymbol{f} \in \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^n$, которая имеет напрерывные частные производные во всех его точках.

Требуется найти локальный минимум функции $\boldsymbol{f}(\boldsymbol{x})$, т.е. такую точку $\boldsymbol{x}_* \in \mathbb{R}^n$, в которой она принимает минимальное значение

$$\boldsymbol{x}_* = \min_{\boldsymbol{x} \in \mathbb{R}^n} \boldsymbol{f}(\boldsymbol{x}).$$

Решение задачи, как мы выяснили выше, состоит в построении последовательности приближений $\boldsymbol{x}^{(0)}$, $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(3)}$, ... к точке $\boldsymbol{x}_*$, таких что $\boldsymbol{f}(\boldsymbol{x}^{k+1}) < \boldsymbol{f}(\boldsymbol{x}^k)$, $k = 0, 1, 2, ...$ (т.е. выполняется условие убывания).

Такая последовательность вычисляется

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \, \boldsymbol{g}^{(k)}, \quad , \boldsymbol{g}^{(k)} = \nabla \boldsymbol{f}(\boldsymbol{x}^k), \quad k = 0, 1, 2, ... \tag{5}$$

Шаг α_k рассчитывается для каждого шага k из условия

$$\alpha_k = \min_{\alpha > 0} \varphi_k(\alpha), \quad \varphi_k(\alpha) = \boldsymbol{f}(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{g}^{(k)}).$$

Построение последовательности приближений $\{\boldsymbol{x}^(k)\}$, $k = 0, 1, 2, ...$ заканчивается в точке $\boldsymbol{x}^(k)$, в которой $||\nabla \boldsymbol{f}(\boldsymbol{x}^k)|| < \varepsilon_1$ или принудительно, когда $k \geqslant m$, где m – предельное число итераций.

Можно заканчивать при одновременном выполнении двух условий $||\boldsymbol{f}(\boldsymbol{x}^{k+1}) - \boldsymbol{f}(\boldsymbol{x}^k)|| < \varepsilon_2$ и $||\boldsymbol{x}^{k+1} - \boldsymbol{x}^k|| < \varepsilon_2$.

С другой стороны, выбор критерия останова – задача не совсем простая.

Алгоритм в общем случае включает

- Шаг 1. Задать $\boldsymbol{x}^{(0)}$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ и m и найти найти градиент функции в произвольной точке $\nabla \boldsymbol{f}(\boldsymbol{x} = \left(\dfrac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1}, ..., \dfrac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_n} \right)$. Это означает, что надо выписать частные производные первого порядка $\boldsymbol{f}(\boldsymbol{x}$ по всем компонентам $x_1, x_2, ..., x_n$ вектора $\boldsymbol{x}$ или составить алгоритм численного нахождения градиента функции $\boldsymbol{f}(\boldsymbol{x})$ методом «конечно-разностной» аппроксимации производных.

- Шаг 2. Положить $k = 0$.

- Шаг 3. Вычислить $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$.

- Шаг 4. Проверить условие останова $||\nabla \boldsymbol{f}(\boldsymbol{x}^k)|| < \varepsilon_1$:

  (а) если условие выполняется, то $\boldsymbol{x}_* = \boldsymbol{x}^{(k)}$;

  (б) если нет, то перейти к шагу 5.

- Шаг 5. Проверить условие $k > m$:

  (а) если условие выполняется, то мы говорим, что принудительно завершаем работу алгоритма (вопрос, насколько близко $\boldsymbol{x}^{(k)}$ к $\boldsymbol{x}_*$, требует дополнительного исследования);

  (б) если нет, то перейти к шагу 6.

- Шаг 6. Вычислить шаг α_k как

  $$\alpha_k = \min_{\alpha > 0} \varphi_k(\alpha), \quad \varphi_k(\alpha) = \boldsymbol{f}(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{g}^{(k)}).$$

  .

- Шаг 7. Вычислить

  $$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \nabla \boldsymbol{f}(\boldsymbol{x}^k).$$

- Шаг 8. Проверить выполнение условий $||\boldsymbol{f}(\boldsymbol{x}^{k+1}) - \boldsymbol{f}(\boldsymbol{x}^k)|| < \varepsilon_2$ и $||\boldsymbol{x}^{k+1} - \boldsymbol{x}^k|| < \varepsilon_2$:

  (а) если условия выполняются то $\boldsymbol{x}_* = \boldsymbol{x}^{(k+1)}$ и закончить поиск.

  (б) если хотя бы одно из условий не выполняется, то положить $k = k + 1$ и перейти к шагу 3.

6. Метод сопряженных градиентов

Рассмотрим задачу минимизации квадратичной функции

$$\boldsymbol{f}(\boldsymbol{x}) = \frac{1}{2} \left(A\boldsymbol{x}, \boldsymbol{x} \right) - \left(\boldsymbol{b}, \boldsymbol{x} \right), \quad A \in \mathbb{R}^{n \times n}, \quad \boldsymbol{b}, \boldsymbol{x} \in \mathbb{R}^n.$$

Ненулевые векторы $\boldsymbol{p}^{(0)}$, $\boldsymbol{p}^{(1)}, ..., \boldsymbol{p}^{(n-1)}$ называются взаимно-сопряженными относительно A, если

$$\left(\boldsymbol{p}^{(k)}), \boldsymbol{p}^{(i)} \right) = 0,$$

для всех $k \neq i$.

В методе сопряженных направлений

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\, \boldsymbol{p}^{(k)}, \quad k = 0, 1, 2, .., n-1, \tag{6}$$

где векторы $\boldsymbol{p}^{(0)}$, $\boldsymbol{p}^{(1)}, ..., \boldsymbol{p}^{(n-1)}$ взаимно сопряжены, а α_k:

$$\alpha_k = -\frac{(\boldsymbol{g}^{(k)},\, \boldsymbol{p}^{(k)})}{\big(A\boldsymbol{p}^{(k)},\, \boldsymbol{p}^{(k)}\big)}$$

находится как решение

$$\alpha_k = \min_{\alpha>0} \varphi_k(\alpha), \quad \varphi_k(\alpha) = \boldsymbol{f}(\boldsymbol{x} + \alpha \boldsymbol{p}^{(k)}).$$

6.1. Метод сопряженных градиентов (метод Флетчера -Ривса)

В этом методе векторы $\boldsymbol{p}^{(0)}$, $\boldsymbol{p}^{(1)}, ..., \boldsymbol{p}^{(n-1)}$ находятся

$$\boldsymbol{p}^{(0)} = -\boldsymbol{g}^{(0)}, \quad \boldsymbol{p}^{(k)} = -\boldsymbol{g}^{(k)} + \beta_{k-1}\boldsymbol{g}^{(k-1)}, \quad k \geqslant 1,$$

где

$$\beta_{k-1} = \frac{(A\boldsymbol{p}^{(k-1)},\, \boldsymbol{p}^{(k)})}{\big(A\boldsymbol{p}^{(k-1)},\, \boldsymbol{p}^{(k-1)}\big)}.$$

Таким образом,

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\, \boldsymbol{p}^{(k)}, \quad k = 0, 1, 2, .., \tag{7}$$

$$\boldsymbol{p}^{(0)} = -\boldsymbol{g}^{(0)}, \quad \boldsymbol{p}^{(k)} = -\boldsymbol{g}^{(k)} + \beta_{k-1}\boldsymbol{g}^{(k-1)}, \quad k \geqslant 1, \tag{8}$$

$$\alpha_k = \min_{\alpha>0} \varphi_k(\alpha), \quad \varphi_k(\alpha) = \boldsymbol{f}(\boldsymbol{x} + \alpha \boldsymbol{p}^{(k)}). \tag{9}$$

Здесь в случае неквадратичных функций

$$\beta_{k-1} = \frac{\big(\boldsymbol{g}^{(k)},\, \boldsymbol{g}^{(k)} - \boldsymbol{g}^{(k-1)}\big)}{|\boldsymbol{p}^{(k-1)}|^2}$$

или

$$\beta_{k-1} = \frac{|\boldsymbol{g}^{(k)}|^2}{|\boldsymbol{p}^{(k-1)}|^2}.$$

ЗАМЕЧАНИЕ:

1. Метод сопряженных градиентов сходится к решених задачи не более чем за n шагов, где n – размерность вектора $\boldsymbol{x}$.

2. Если функция неквадратичная, то метод уже несходится за конечное число итераций. Часто, для уменьшения влияния вычислительной погрешности, при $k = n, 2n, 3n, ...$, коэффициент β_{k-1} обнуляется. Такая процедура называется «обновлением алгоритма».

Алгоритм минимизации квадратичной функции

Задать $\boldsymbol{x}_0$, ε;

1. $\boldsymbol{x} \leftarrow \boldsymbol{x}_0$; $\boldsymbol{g}_0 \leftarrow A\boldsymbol{x}_0 - \boldsymbol{b}$;

$\boldsymbol{p} \leftarrow -\boldsymbol{g}_0$; $k \leftarrow 0$;

REPEAT

2. $\boldsymbol{x}_0 \leftarrow \boldsymbol{x}$;

3. $\boldsymbol{g}_0 \leftarrow A\boldsymbol{x}_0 - \boldsymbol{b}$;

4. $\boldsymbol{p}_0 \leftarrow \boldsymbol{p}$;

5. Выбрать направление $\boldsymbol{p}$ поиска:

If $k = 0$ Then

$$\boldsymbol{p} \leftarrow \boldsymbol{p}_0$$

Else Begin

$$\beta \leftarrow \frac{(A\boldsymbol{p}_0, \boldsymbol{g}_0)}{(A\boldsymbol{p}_0, \boldsymbol{p}_0)};$$

$$\boldsymbol{p} \leftarrow -\boldsymbol{g}_0 + \beta\boldsymbol{p}_0$$

End;

6. Рассчитать шаг α:

$$\alpha \leftarrow -\frac{(\boldsymbol{g}_0, \boldsymbol{p}_0)}{(A\boldsymbol{p}_0, \boldsymbol{p}_0)};$$

7. Рассчитать $\boldsymbol{x}$:

$$\boldsymbol{x} \leftarrow \boldsymbol{x}_0 + \alpha\boldsymbol{p};$$

8. $k \leftarrow k + 1$;

UNTIL $||\boldsymbol{g}_0|| < \varepsilon$;

Задания к лабораторным и практическим занятиям

1. Решить численно задачу минимизации функции

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \mathbf{b}^T\boldsymbol{x}$$

методами наискорейшего градиентного спуска и методом сопряженных градиентов. Решение задачи проиллюстрировать графически.

Здесь

$$A = \begin{bmatrix} N & -N \\ -N & N+1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 3N-1 \\ 5N+3 \end{bmatrix}.$$

Параметр N равен номеру студента в списке группы.

2. Решить аналитически задачу минимизации функции

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \mathbf{b}^T\boldsymbol{x}$$

методом сопряженных градиентов.

Здесь

$$A = \begin{bmatrix} N & -N \\ -N & N+1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 3N-1 \\ 5N+3 \end{bmatrix}.$$

Параметр N равен номеру студента в списке группы.

3. Выполнить 3 итерации методом сопряженных градиентов для

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \mathbf{b}^T\boldsymbol{x}.$$

Здесь

$$A = \begin{bmatrix} N & -N \\ -N & N+1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 3N-1 \\ 5N+3 \end{bmatrix}.$$

Параметр N равен номеру студента в списке группы.

4. Решить задачу минимизации функций

$$f(\boldsymbol{x}) = -x_1 x_2\, \mathrm{e}^{-(x_1+x_2)},\ \boldsymbol{x}^{(0)} = (0,1)^T;$$
$$f(\boldsymbol{x}) = (x_1^2 + x_2^2 - 1)^2 + (x_1 + x_2 - 1)^2, \boldsymbol{x}^{(0)} = (0,3)^T.$$

методами наискорейшего градиентного спуска и Флетчера-Ривса.

Библиографический список

1. *Амосов, А.А.* Вычислительные методы для инженеров: Учебное пособие/А.А. Амосов, Ю.А. Дубинский, Н.В. Копченова – М.: Высш. шк., 1994. – 544 с

2. *Аттетков, А. В.* Введение в методы оптимизации [Текст]: учебное пособие / А. В. Аттетков, В. С. Зарубин, А. Н. Канатников. - М. : Финансы и статистика, 2008. - 272 с.

3. *Гончаров, В. А.* Методы оптимизации [Текст] : учебное пособие / В. А. Гончаров. - М. : Юрайт, 2010. - 191 с.

4. *Пантелеев, А. В.* Методы оптимизации в примерах и задачах [Текст] : учебное пособие / А. В. Пантелеев, Т. А. Летова. - 2-е изд., испр. - М. : Высшая школа, 2005. - 544 с.