САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет прикладной математики – процессов управления

А. П. ИВАНОВ, Ю. В. ОЛЕМСКОЙ

ПРАКТИКУМ ПО ЧИСЛЕННЫМ МЕТОДАМ

МИНИМИЗАЦИЯ КВАДРАТИЧНОЙ ФУНКЦИИ

Методические указания

Санкт-Петербург
2013

ГЛАВА 1. ЗАДАЧА МИНИМИЗАЦИИ КВАДРАТИЧНОЙ ФУНКЦИИ

§1. Постановка задачи.

Пусть X – евклидово n-мерное пространство, обозначаемое далее $\mathbb{R}^n$, $Q = \mathbb{R}^n$. Векторы пространства $\mathbb{R}^n$ будем считать записанными в столбец, обозначая их строчными латинскими буквами; скалярное произведение векторов x и y будем обозначать в зависимости от удобства использования трояко: $x^T y$, либо (x, y), либо $x \cdot y$.

В пространстве
R^n рассмотрим квадратичную функцию

$$f(x) = \frac{1}{2} x^T A x + x^T b, \tag{1.1}$$

где A – положительно определенная матрица, $A = A^T$ и имеют место неравенства

$$m\|x\|^2 \leq (x, Ax) \leq M\|x\|^2, \quad m > 0. \tag{1.2}$$

Отметим, что для такой функции существует единственная точка минимума $\bar{x}$, удовлетворяющая СЛАУ $Ax + b = 0$.

§2. Методы наискорейшего спуска

Поскольку *сходящаяся* минимизирующая последовательность позволяет найти точку минимума рассматриваемой функции с произвольной наперед заданной точностью, то следует рассмотреть вопрос о построении такой последовательности. Поставим следующую задачу: имея точку $x^k \in \mathbb{R}^n$ построить точку $x^{k+1} \in \mathbb{R}^n$ такую, чтобы выполнялось соотношение:

$$f(x^{k+1}) < f(x^k). \tag{2.1}$$

Будем искать точку x^{k+1} в следующем виде:

$$x^{k+1} = x^k + \mu q, \quad q \neq 0, \tag{2.2}$$

где q – заданный вектор из $\mathbb{R}^n$, называемый направлением спуска, а μ – искомый параметр, называемый обычно шагом метода в направлении спуска.

Параметр μ (шаг метода) в формуле (2.2) определим из условия минимума функции $f(\cdot)$ на прямой (2.2). Имеем:

$$\begin{aligned} f(x^{k+1}) = f(x^k + \mu q) \equiv \varphi(\mu) = \\ = \varphi(0) + \mu\dot{\varphi}(0) + \frac{1}{2}\mu^2\ddot{\varphi}(0), \quad \dot{\varphi} = \frac{d\varphi}{d\mu}, \end{aligned} \tag{2.3}$$

где

$$\varphi(0) = f(x^k), \ \dot{\varphi}(0) = q^T(Ax^k + b), \ \ddot{\varphi}(0) = q^T A q. \tag{2.4}$$

Поскольку $\varphi(\mu)$ является квадратным трёхчленом относительно μ с положительным старшим коэффициентом, то у этого трёхчлена существует точка минимума $\bar{\mu}_k$, которая может быть найдена из необходимого условия экстремума $\dot{\varphi}(\mu) = 0$, откуда и получаем:

$$\bar{\mu}_k = -\frac{\dot{\varphi}(0)}{\ddot{\varphi}(0)} = -\frac{q^T(Ax^k + b)}{q \cdot Aq}. \tag{2.5}$$

Очевидно, что если в формуле (2.2) положить $\mu = \bar{\mu}_k$, то построенная по формуле (2.2) точка x^{k+1} будет удовлетворять условию (2.1), причём там будет строгое неравенство в случае $x^k \neq \bar{x}$ и $q^T(Ax^k + b) \neq 0$. Продолжая указанные построения, получим последовательность $\{x^k\}$, которую естественно назвать последовательностью убывания для функции $f(\cdot)$.

Замечание 2.1. Формула

$$\|x^k - \bar{x}\|_2 \leq \frac{1}{m}\|Ax^k + b\|_2, \tag{2.6}$$

позволяет оценить евклидово расстояние от произвольной точки $x^k \in \mathbb{R}^n$ до точки минимума $\bar{x}$ функции $f(\cdot)$.

Замечание 2.2. Формула (2.6) для матриц с диагональным преобладанием величины δ может быть представлена в виде:

$$\|x^k - \bar{x}\|_2 \leq \frac{1}{\delta}\|Ax^k + b\|_2. \tag{2.7}$$

§3. Метод наискорейшего градиентного спуска (МНГС)

Замечание 3.1. Градиентом дифференцируемой скалярной (т.е. $f(x) \in \mathbb{R}^1$) функции $f(\cdot)$ векторного аргумента $x \in \mathbb{R}^n$ называется вектор, составленный из частных производных функции, т.е.

$$\operatorname{grad} f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots \frac{\partial f(x)}{\partial x_n} \right)^T .$$

Замечание 3.2. Для квадратичной функции (1.1) градиент имеет вид $q = \operatorname{grad} f(x) = Ax + b$.

Если в формулах (2.2) считать, что $q = Ax^k + b$ (то есть вектор q является вектором градиента функции $f(\cdot)$ в точке x^k), то соответствующий метод построения последовательности $\{x^k\}$ называют *градиентным методом*. Если к тому же шаг метода μ в формуле (2.2) выбирается по формуле (2.5), то такой метод называют (одношаговым) *методом наискорейшего градиентного спуска* (МНГС). В этом случае формула (2.5) принимает вид

$$\bar{\mu}_k = -\frac{\|Ax^k + b\|^2}{(Ax^k + b)^T A(Ax^k + b)}. \tag{3.1}$$

Замечание 3.3. Метод наискорейшего градиентного спуска сходится для любого начального вектора x^0.

§4. Метод покоординатного спуска (МПС)

В случае выбора направлений спуска q в формуле (2.2) на каждом шаге в виде $q = e^i = (\underbrace{0, 0, \dots, 0, 1}_{i}, 0 \dots, 0)^T$, где e^i – i-ый орт пространства $\mathbb{R}^n$, метод носит название *метода покоординатного спуска*. При выборе шага метода по формуле (2.5) его называют методом *наискорейшего покоординатного спуска* (МНПС). В этом случае формула (2.5) принимает вид

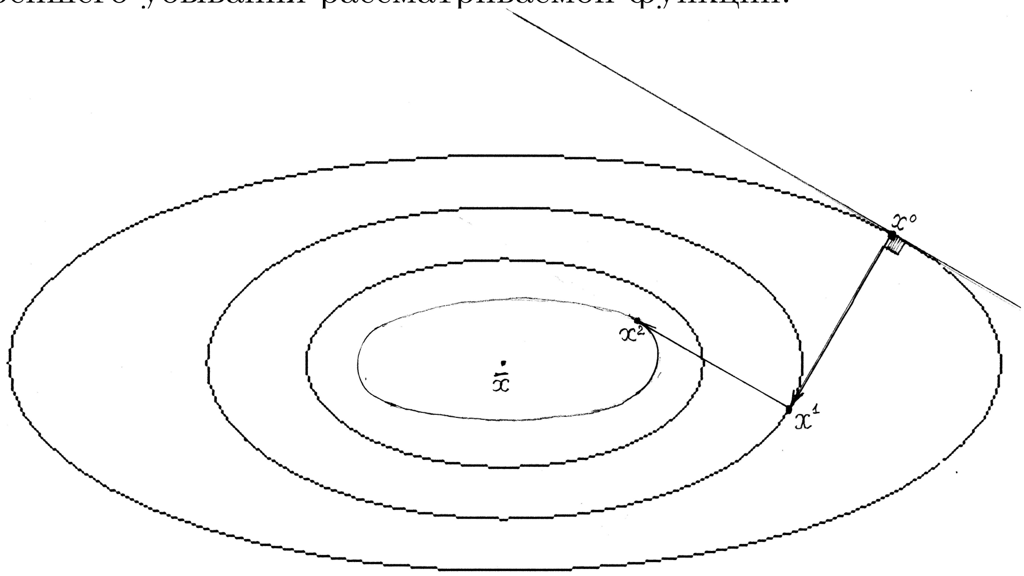
$$\bar{\mu}_k = -\frac{e^i \cdot (Ax^k + b)}{e^i \cdot Ae^i}. \tag{4.1}$$

Замечание 4.1. Метод наискорейшего покоординатного спуска сходится для любого начального вектора x^0.

§5. Геометрическая интерпретация методов

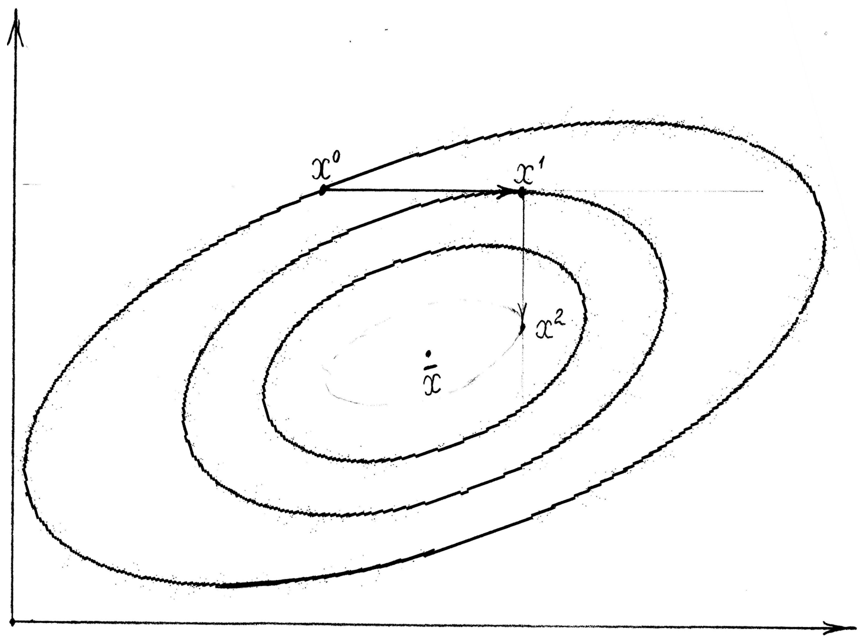
Замечание 5.1. Поверхность (линия) $f(x) = c$, $c = const$ называется *поверхностью (линией) уровня* функции $f(\cdot)$.

Будем далее считать, что $x \in \mathbb{R}^2$. Тогда линиями уровней функции (1.1) являются соосные эллипсы. Кроме того, как известно, градиент функции, вычисленный в точке $\tilde{x}$, ортогонален поверхности уровня $f(x) = f(\tilde{x})$ и направлен сторону наискорейшего возрастания функции, а антиградиент, соответственно, в сторону наискорейшего убывания рассматриваемой функции.



Данный рисунок демонстрирует построение членов минимизирующей последовательности с использованием МНГС. Движение от предыдущей точки к последующей осуществляется по антиградиенту функции, вычисленному в предыдущей точке. При этом градиент ортогонален поверхности уровня функции, на которой расположено предыдущее приближение. Следующая точка последовательности расположена на поверхности уровня функции, для которой прямая, выходящая из предыдущей точки в направлении антиградиента, является касательной.

Следующий рисунок демонстрирует построение минимизирующей последовательности при использовании метода наискорейшего покоординатного спуска. Движение от точки к точке осуществляется по прямым, параллельным координатным осям.



При этом следующее приближение лежит на поверхности уровня минимизируемой функции, для которой прямая, параллельная координатной оси и выходящая из предыдущей точки, является касательной.

ГЛАВА 2. ТИПОВЫЕ ЗАДАЧИ ПО ТЕМЕ.

Задание для самостоятельного выполнения: решить задачи примеров 1–3 для функции

$$f(x_1, x_2, x_3) = x_1^2 + x_1x_3 + x_2^2 + x_3^2 - 2x_1 + x_2 - x_3.$$

§1. Примеры

Ниже приводятся примеры выполнения задач.

Пример 1.1. Записать квадратичную функцию в матричном виде (1.1) и проверить выполнение условия (1.2).
а) $f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 2x_1 + x_2$.
б) $f(x_1, x_2) = x_1^2 + x_1x_2 - x_1 + 2x_2$.
Решение. Для случая а):

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Найдём собственные числа матрицы A, решив уравнение:

$$\det(A - \lambda E) = (2 - \lambda)^2 - 1 = 0$$

где E – единичная матрица, λ – скаляр. Собственные числа этой матрицы $\lambda_1 = 1$, $\lambda_2 = 3$, т.е. условие (1.2) выполнено и функция

$$f(x_1, x_2) = \frac{1}{2}x^T Ax + x^T b$$

удовлетворяет условию соответствующей теоремы и данная функция имеет единственную точку минимума. В данном случае условие (1.2) для $f(\cdot)$ запишется так:

$$\|x\|^2 \leqslant (x, Ax) \leqslant 3\|x\|^2.$$

Для случая б):

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Решив уравнение $\det(A - \lambda E) = 0$, получим $\lambda_{1,2} = \pm\sqrt{5}$, следовательно, функция:

$$f(x_1, x_2) = \frac{1}{2}x^T Ax + x^T b$$

не удовлетворяет условию (1.2).

Пример 1.2. Для случая а): $f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 2x_1 + x_2$. Найти точку минимума $\bar{x}$ функции $f(x_1, x_2)$ с использованием необходимого условия и значение функции $f(\bar{x})$.

Решение. В соответствии с теоремой о существовании и единственности точки минимума для квадратичной функции (1.1) решаем систему линейных алгебраических уравнений:

$$Ax + b = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \end{pmatrix} = 0$$

или же

$$\begin{cases} 2x_1 + x_2 & = & 2 \\ x_1 + 2x_2 & = & -1 \end{cases}$$

Отсюда: $x_1 = \frac{5}{3}$, $x_2 = -\frac{4}{3}$,

$$\bar{x} = \begin{pmatrix} \frac{5}{3} \\ -\frac{4}{3} \end{pmatrix}, \qquad f(\bar{x}) = -\frac{7}{3}.$$

Пример 1.3. Для функции $f(x) = x_1^2 + x_1x_2 + x_2^2 - 2x_1 + x_2$ найти первые три члена последовательности $\{x^k\}$ с использованием метода наискорейшего градиентного спуска.

Решение. Пусть

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Первый член последовательности зададим в виде

$$x^1 = e^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad f(x^1) = -1.$$

Второй и третий члены последовательности получаем по методу наискорейшего градиентного спуска

$$q^1 = Ae^1 + b = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

$$Aq^1 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \qquad \mu_1 = -\frac{q^1 \cdot q^1}{q^1 \cdot Aq^1} = -\frac{1}{2},$$

$$x^2 = x^1 + \mu_1 q^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad f(x^2) = -2;$$

$$q^2 = Ax^2 + b = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \end{pmatrix} == \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$Aq^2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}, \qquad \mu_2 = -\frac{q^2 \cdot q^2}{q^2 \cdot Aq^2} = -\frac{1}{2},$$

$$x^3 = x^2 + \mu_2 q^2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ -1 \end{pmatrix}, \qquad f(x^3) = -\frac{9}{4}.$$

Пример 1.4. Для функции $f(x) = x_1^2 + x_1 x_2 + x_2^2 - 2x_1 + x_2$ найти первые три члена последовательности $\{x^k\}$ с использованием метода наискорейшего покоординатного спуска.

Решение. Здесь, как и ранее

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Первый член последовательности возьмём в виде

$$x^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Второй третий и четвёртый члены последовательности получаем по методу наискорейшего покоординатного спуска

$$x^1 = x^0 + \mu_1 e^1, \quad \mu_1 = -\frac{e^1 \cdot (Ax^0 + b)}{e^1 \cdot Ae^1} = 1, \quad x^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

$$x^2 = x^1 + \mu_2 e^2, \quad \mu_2 = -\frac{e^2 \cdot (Ax^1 + b)}{e^2 \cdot Ae^2} = -1, \quad x^2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

$$x^3 = x^2 + \mu_3 e^1, \quad \mu_3 = -\frac{e^1 \cdot (Ax^2 + b)}{e^1 \cdot Ae^1} = 1/2, \quad x^3 = \begin{pmatrix} 3/2 \\ -1 \end{pmatrix}.$$

Отметим, что выбор направления спуска e^i на каждом шаге находится в руках вычислителя. Имеет смысл после перебора всех координатных ортов от 1 до n в качестве очередного направления спуска назначить тот орт, в направлении которого происходит наибольшее убывание функции.

ЛИТЕРАТУРА

1. *Н. Бахвалов, Н. Жидков, Г. Кобельков.* Численные методы. — М.: Изд. Физматлит, 2006.
2. *В. М. Вержбицкий.* Численные методы. Линейная алгебра и нелинейные уравнения. — М.: Изд. Высшая школа, 2000.