Владикавказский математический журнал
октябрь–декабрь, 2006, Том 8, Выпуск 4

УДК **519.3**

НЕКОТОРЫЕ МЕТОДЫ МИНИМИЗАЦИИ МАКСИМУМА КВАДРАТИЧНЫХ ФУНКЦИЙ[1]

Л. Н. Полякова

Светлой памяти Александра Моисеевича посвящается

В работе рассматриваются несколько алгоритмов минимизации функции максимума от квадратичных функций в евклидовом пространстве $\mathbb{R}^n$. Показывается, что данную задачу можно свести к нахождению точки с наименьшей евклидовой нормой, принадлежащей пересечению квадрик. Описывается метод минимизации функции максимума на $\mathbb{R}^n$ с постоянным шагом, аналогичный градиентному методу минимизации с постоянным шагом сильно выпуклой функции. Доказывается геометрическая скорость сходимости генерируемой последовательности к точке минимума.

1. Введение

Класс функций максимума от непрерывно дифференцируемых функций является одним из наиболее исследованных среди негладких функций (см., например, [1, 2]). Напомним некоторые свойства таких функций. Пусть

$$f(x) = \max_{i \in I_1} f_i(x), \quad f_i(x) = \frac{1}{2}\langle A_i x, x\rangle + \langle b_i, x\rangle + c_i, \quad i \in I_1 = 0 \cup I,\ I = 1, \dots, m,\ x \in \mathbb{R}^n,$$

где A_i — симметричные матрицы размера $n \times n$, $b_i \in \mathbb{R}^n$, $c_i \in \mathbb{R}$, $i \in I_1$.

В каждой точке $x \in \mathbb{R}^n$ функция f непрерывна и дифференцируема по любому направлению $g \in \mathbb{R}^n$, при этом ее производная по направлению имеет вид (см, например, [1])

$$f'(x, g) = \max_{v \in \partial f(x)} \langle v, g\rangle,\ \partial f(x) = \mathrm{co}\left\{ \bigcup_{i \in R_f(x)} (A_i x + b_i) \right\},\ R_f(x) = \{ i \in I_1 \mid f_i(x) = f(x)\}.$$

Здесь через $\mathrm{co}\{X\}$ обозначена выпуклая оболочка множества X. Множество $\partial f(x^*)$ называется *субдифференциалом* функции f в точке x^*.

Рассмотрим оптимизационную задачу: найти

$$\inf_{x \in \mathbb{R}^n} f(x). \tag{1}$$

[1] Работа выполнена при поддержке Российского Фонда фундаментальных Исследований (Грант РФФИ № 06-01-00276).

Для нее сформулируем необходимые условия минимума.

Теорема 1. [1] 1) *Для того чтобы точка $x^* \in \mathbb{R}^n$ была точкой минимума функции f на $\mathbb{R}^n$, необходимо, чтобы выполнялось включение*

$$0_n \in \partial f(x^*). \tag{2}$$

2) *Для того чтобы точка $x^* \in \mathbb{R}^n$ была точкой минимума функции f на $\mathbb{R}^n$, необходимо, чтобы нашлись такие неотрицательные коэффициенты μ_i^*, $i \in I_1$, $\sum\limits_{i=0}^{m} \mu_i^* = 1$, чтобы выполнялись условия*

$$\sum_{i=0}^{m} \mu_i^* A_i x^* = -\sum_{i=0}^{m} \mu_i^* b_i, \tag{3}$$

$$\mu_i^* \left(f_i(x) - f(x) \right) = 0, \quad i \in I_1. \tag{4}$$

Точка $x^* \in \mathbb{R}^n$, для которой выполняется условие (2) или условия (3), (4) называется *стационарной* точкой функции f на $\mathbb{R}^n$. Из условия (4) следует, что если в стационарной точке $i \notin R_f(x^*)$, то $\mu_i = 0$.

Пусть точка $x^* \in \mathbb{R}^n$ — стационарная точка функции f на $\mathbb{R}^n$ и числа μ_i^*, $i \in I_1$, таковы, что для них в x^* выполнены соотношения (3), (4). Введем множество

$$M_f(x^*) = \{\mu = (\mu_0^*, \mu_1^*, \ldots, \mu_m^*) \in \mathbb{R}^{m+1} \mid \mu_i^* \geqslant 0,\ i \in I_1,\ \sum_{i=0}^{m} \mu_i^* = 1\}.$$

Если все матрицы A_i, $i \in I$, — неотрицательно определенные, то функции f_i, $i \in I_1$, являются выпуклыми, и тогда необходимые условия минимума, сформулированные в теореме 1, являются также достаточными условиями. Если все матрицы A_i, $i \in I_1$, — положительно определенные, то решение задачи (1) единственно.

Рассмотрим невырожденное преобразование $x = Dy$. В результате такого преобразования квадратичные формы $\langle A_i x, x\rangle$, $i \in I_1$, будут приведены к квадратичным формам $\langle \bar{A}_i y, y\rangle$, $i \in I_1$. Причем матрицы A_i, $\bar{A}_i$ связаны соотношениями

$$A_i = \left(D^T\right)^{-1} \bar{A}_i D^{-1}, \quad \bar{A}_i = D^T A_i D, \quad i \in I_1,$$

где через D^T обозначена транспонированная матрица.

Определим функцию

$$\varphi(y) = \max_{i \in I} \varphi_i(y), \quad y \in \mathbb{R}^n,$$

где

$$\varphi_i(y) = \frac{1}{2}\langle \bar{A}_i y, y\rangle + \langle \bar{b}_i, y\rangle + c_i, \quad \bar{b}_i = D^T b_i \in \mathbb{R}^n, \quad i \in I_1.$$

Рассмотрим оптимизационную задачу: найти

$$\inf_{y \in \mathbb{R}^n} \varphi(y). \tag{5}$$

Множество $\partial\varphi(y) = \mathrm{co}\left\{ \bigcup\limits_{i \in R_\varphi(y)} \left(\bar{A}_i y + \bar{b}_i\right) \right\}$ есть субдифференциал функции φ в точке $y \in \mathbb{R}^n$. Нетрудно видеть, что при таком преобразовании $R_f(x) = R_\varphi(y)$.

Пусть точка $y^* \in \mathbb{R}^n$ является решением задачи (5), тогда необходимо, чтобы

$$0_n \in \operatorname{co} \left\{ \bigcup_{i \in R_\varphi(y^*)} \left(\bar{A}_i y^* + \bar{b}_i \right) \right\}.$$

Лемма 1. 1. *Если точка $x^* \in \mathbb{R}^n$ является стационарной точкой функции f на $\mathbb{R}^n$, то $y^* = D^{-1}x^*$ есть стационарная точка функции φ на $\mathbb{R}^n$.*

2. *Если точка $y^* \in \mathbb{R}^n$ является стационарной точкой функции φ на $\mathbb{R}^n$, то $x^* = Dy^*$ есть стационарная точка функции f на $\mathbb{R}^n$.*

3. *В обоих случаях справедливо равенство $M_f(x^*) = M_\varphi(y^*)$.*

$\lhd$ Докажем первое утверждение леммы. Пусть $x^* \in \mathbb{R}^n$ — стационарная точка функции f на $\mathbb{R}^n$. Тогда выполнено включение (2), т. е.

$$0_n \in \operatorname{co} \left\{ \bigcup_{i \in R_f(x^*)} (A_i x^* + b_i) \right\} = \operatorname{co} \left\{ \bigcup_{i \in R_\varphi(y^*)} (A_i D y^* + b_i) \right\}$$
$$= \operatorname{co} \left\{ \bigcup_{i \in R_\varphi(y^*)} \left(D^T\right)^{-1} \left(\bar{A}_i y^* + \bar{b}_i \right) \right\}.$$

Отсюда следует, что $0_n \in \operatorname{co} \left\{ \bigcup_{i \in R_\varphi(y^*)} \left(\bar{A}_i y^* + \bar{b}_i \right) \right\} = \partial\varphi(y^*)$.

Второе утверждение леммы доказывается аналогично.

Из доказательства также следует справедливость равенства $M_f(x^*) = M_\varphi(y^*)$. $\rhd$

Из линейной алгебры известно (см., например, [3, 4]), что любую квадратичную форму с положительно определенной матрицей можно привести невырожденным преобразованием к нормальному виду.

В дальнейшем будем предполагать, что матрица A_0 — положительно определенная. Тогда инфимум в задаче (1) достигается. Предположим, что невырожденное преобразование $x = Dy$ приводит квадратичную форму $\langle A_0 x, x \rangle$ к нормальному виду. В этом случае, $\bar{A}_0 = E$, где через E обозначена единичная матрица размера $[n \times n]$. Затем рассмотрим преобразование $y = z - \bar{b}_0$ и квадратичные функции

$$\psi_0(z) = \frac{1}{2}\langle z, z \rangle + \tilde{c}_0, \quad \psi_i(z) = \frac{1}{2}\langle \bar{A}_i z, z \rangle + \langle \tilde{b}_i, z \rangle + \tilde{c}_i, \quad i \in I,$$

где

$$\tilde{c}_0 = c_0 - \frac{1}{2}\|\bar{b}_0\|^2, \quad \tilde{b}_i = \langle \bar{b}_i, z \rangle - \frac{1}{2}\langle \bar{A}_i z, \bar{b}_i \rangle,$$
$$\tilde{c}_i = c_i + \frac{1}{2}\langle \bar{A}_i \bar{b}_0, \bar{b}_0 \rangle - \langle \bar{b}_i, \bar{b}_0 \rangle, \quad i \in I.$$

Обозначим $\psi(z) = \max\limits_{i \in I_1} \psi_i(z)$. Рассмотрим оптимизационную задачу: найти

$$\min_{z \in \mathbb{R}^n} \psi(z).$$

Если точка $z^* \in \mathbb{R}^n$ является стационарной точкой функции ψ на $\mathbb{R}^n$, то несложно показать, что $y^* = z^* - \bar{b}_0$ есть стационарная точка функции φ на $\mathbb{R}^n$, а $x^* = D^{-1}y^*$ — стационарная точка функции f на $\mathbb{R}^n$.

Таким образом, для того чтобы решить задачу (1), необходимо уметь минимизировать функцию

$$f(x) = \max_{i \in I_1} f_i(x),$$

где

$$f_0(x) = \frac{1}{2}\langle x, x\rangle,\ \ f_i(x) = \frac{1}{2}\langle A_i x, x\rangle + \langle b_i, x\rangle + c_i, \quad x \in \mathbb{R}^n,$$

A_i, — симметричные матрицы размера $[n \times n]$, $b_i \in \mathbb{R}^n$, $c_i \in \mathbb{R}$, $i \in I$.

Если матрицы A_i, $i \in I$, — положительно определенные, то решение задачи (1) существует и единственно. В дальнейшем будем предполагать, что в точке минимума x^* функции f на $\mathbb{R}^n$ все функции f_i, $i \in I_1$, активны, т. е. выполнено равенство $R_f(x^*) = I_1$. Тогда задача (1) эквивалентна задаче условной оптимизации: найти

$$\min_{x \in X} \frac{1}{2}\langle x, x\rangle,$$

где

$$X = \left\{ x \in \mathbb{R}^n \mid \frac{1}{2}\langle A_i x, x\rangle + \langle b_i, x\rangle + c_i = \frac{1}{2}\langle x, x\rangle,\ i \in I \right\}.$$

Множество X может быть записано в виде

$$X = \{x \in \mathbb{R}^n \bigm| f_i(x) - f_0(x) = 0,\ i \in I\}.$$

Пусть задана квадратичная функция

$$q(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c,$$

множество $X(q) = \{x \in \mathbb{R}^n \bigm| q(x) = 0\}$ называется *квадрикой*. Если обозначить через

$$q_i(x) = f_i(x) - f_0(x), \quad i \in I,$$

то множество X есть пересечение квадрик, определяемых квадратичными функциями q_i, $i \in I$. Поэтому задача минимизации функции f на $\mathbb{R}^n$ сводится к нахождению точки из множества X, имеющей наименьшую евклидову норму.

2. Минимизации максимума двух квадратичных функций

Рассмотрим случай минимизации функции максимума от двух квадратичных функций в предположении, что матрица A_0 положительно определена. Из линейной алгебры известно, что существует невырожденное преобразование $x = Dy$, приводящее квадратичную форму $\langle A_1 x, x\rangle$ к каноническому виду, а форму $\langle A_0 x, x\rangle$ — к нормальному виду. Таким образом, можно считать, что функции f_0 и f_1 имеют вид

$$f_0(x) = \frac{1}{2}\langle x, x\rangle, \quad f_1(x) = \frac{1}{2}\langle \Theta x, x\rangle + \langle b, x\rangle + c, \quad x, b \in \mathbb{R}^n,\ c \in \mathbb{R},$$

где матрица Θ — диагональная матрица ($\Theta = \mathrm{diag}\,\{\theta_j\}$, $j \in J = 1..n$), полученная в результате соответствующего невырожденного преобразования, и $f(x) = \max\{f_0(x), f_1(x)\}$.

Отметим тот факт, что если в результате преобразования оказалось, что число $c \leqslant 0$, то решение задачи (1) есть нулевая точка. Поэтому будем считать $c > 0$. Рассмотрим множество точек x^*, которые являются стационарными для функции f_1 на $\mathbb{R}^n$, т. е. точки, для которых

$$\Theta x^* + b = 0_n, \tag{6}$$

и выполнено неравенство $f_0(x^*) < f_1(x^*)$. Если все числа $\theta_j > 0$, $j \in J$, то функция f_1 сильно выпукла и точка $x^* = -\Theta^{-1} b_1$ будет единственным решением задачи (1). Если среди чисел θ_j есть отрицательные, то может оказаться, что точка $x^* = -\Theta^{-1} b$ являющаяся решением системы (6) не есть точка локального минимума функции f на $\mathbb{R}^n$.

Предположим, что числа $\theta_j \geqslant 0$ при любом $j \in J$. Нас будет интересовать случай, когда $R_f(x^*) = \{0, 1\}$. Тогда если точка $x^* \in \mathbb{R}^n$ является решением задачи (1), то необходимо, чтобы нашлось положительное число μ^*, для которого справедливо равенство

$$\Theta x^* + b = -\mu^* x^*. \tag{7}$$

Уравнение (7) можно переписать в виде $(\Theta + \mu^* E)\, x^* = -b$. Отсюда имеем

$$x^* = -\left(\Theta + \mu^* E\right)^{-1} b. \tag{8}$$

Так как матрица $\Theta + \mu^* E$ диагональная, то и обратная к ней также будет диагональной,

$$(\Theta + \mu^* E)^{-1} = \mathrm{diag}\ \left\{\frac{1}{\theta_j + \mu^*}\right\}, \quad j \in J.$$

Следовательно, j-ая координата вектора x^* будет равна величине $x_j^* = -\dfrac{b_j}{\theta_j + \mu^*}$, где b_j — j-ая координата вектора b. Покажем, что нахождение числа μ^* сводится к нахождению корня многочлена с действительными коэффициентами степени $2n$.

Зафиксируем произвольное число $\mu \in \mathbb{R}$, $x_j^* = -\dfrac{b_j}{\theta_j + \mu}$. Подставляя в выражение (8) значение x_j^*, получим уравнение

$$\sum_{j=1}^{n} \left[\frac{(\theta_j - 1) b_j^2}{2(\theta_j + \mu)^2} - \frac{b_j^2}{\theta_j + \mu}\right] + c = 0. \tag{9}$$

Если обозначить через $p_j(\mu) = 2\mu + \theta_j + 1$, $j \in J$, то уравнение (9) можно записать в виде

$$\sum_{j=1}^{n} \frac{p_j(\mu) b_j^2}{2(\theta_j + \mu)^2} - c = 0.$$

Пусть

$$P(\mu) = \sum_{j=1}^{n} p_j(\mu) b_j^2 \prod_{k=1, k\neq j}^{n} (\theta_k + \mu)^2 - 2c \prod_{k=1}^{n} (\theta_k + \mu)^2.$$

Очевидно, что многочлен $P(\mu)$ является многочленом степени $2n$, и, в силу нашего предположения, имеет единственный положительный корень μ^*.

Заметим также, что если точка $x^* \in \mathbb{R}^n$ является решением задачи (1), то необходимо и достаточно, чтобы нашлось такое положительное число ν^*, чтобы выполнялись равенство

$$\nu^* \left(\Theta x^* + b\right) = -x^*.$$

Из этого уравнения следует, что $x^* = -\nu^* \left(\nu^* \Theta + E\right)^{-1} b$. (В нашем случае, $\nu^* = \dfrac{1}{\mu^*}$).

Пусть $\nu \in \mathbb{R}$ и j-ая координата вектора x^* равна величине $x_j^* = -\dfrac{\nu b_j}{\nu\theta_j + 1}$. Подставляя в выражение (8) значение x^*, получим уравнение

$$\sum_{j=1}^{n} \left[\frac{(\theta_j - 1)\nu^2 b_j^2}{2(\nu\theta_j + 1)^2} - \frac{\nu b_j^2}{\nu\theta_j + 1} \right] + c = 0. \tag{10}$$

Если обозначить через $s_j(\nu) = \nu^2(1+\theta_j)+2\nu,\ j \in J$, то уравнение (10) можно записать в виде

$$\sum_{j=1}^{n} \frac{s_j(\nu) b_j^2}{2(\nu\theta_j + 1)^2} - c = 0.$$

Нахождение нужного нам числа ν^* также сводится к определению единственного положительного корня многочлена $S(\nu)$ степени $2n$ следующего вида

$$S(\nu) = \sum_{j=1}^{n} s_j(\nu) \prod_{k=1, k\neq j}^{n} (\nu\theta_j + 1)^2 - 2c \prod_{k=1}^{n} (\nu\theta_k + 1)^2.$$

Остановимся на одном варианте определения преобразования D (см., например, [3]). Рассмотрим спектральное разложение матрицы A_0: $A_0 = Q_0 \Lambda Q_0^T$, где Q_0 — ортогональная матрица, матрица Λ — диагональная матрица, на главной диагонали которой стоят собственные числа λ_j матрицы A_0, расположенные в порядке невозрастания, т. е.

$$\Lambda = \mathrm{diag}\ \{\lambda_j\}, \quad \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0.$$

Невырожденное преобразование $x = S(\Lambda) Q_0^T y$, где $S(\Lambda) = \left(\sqrt{\Lambda}\right)^{-1} = \mathrm{diag}\ \left\{\dfrac{1}{\sqrt{\lambda_j}}\right\}$, приводит квадратичную форму $\langle A_0 x, x\rangle$ к нормальному виду $\langle y, y\rangle$, а квадратичную форму $\langle A_1 x, x\rangle$ — к виду $\langle \bar{A}_1 y, y\rangle$. Заметим, что

$$\bar{A}_1 = Q_0 S(\Lambda) A_1 S(\Lambda) Q_0^T.$$

Рассмотрим спектральное разложение матрицы $\bar{A}_1$: $\bar{A}_1 = Q_1 \Theta Q_1^T$, где матрица Q_1 — ортогональная, а Θ — диагональная матрица, на главной диагонали которой находятся собственные числа $\theta_j,\ j \in J$, матрицы $\bar{A}_1$, расположенные в порядке невозрастания, т. е.

$$\Theta = \mathrm{diag}\ \{\theta_j\}, \quad \theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta_n > 0.$$

Тогда невырожденное преобразование $y = Q_1^T z$, приводит квадратичную форму $\langle y, y\rangle$ к виду $\langle z, z\rangle$, а квадратичную форму $\langle \bar{A}_1 y, y\rangle$ к виду $\langle \Theta z, z\rangle$.

Таким образом, невырожденное преобразование $y = Dx$, где $D = S(\Lambda) Q_0^T Q_1^T$ приводит к нормальному виду квадратичную форму $\langle A_0 x, x\rangle$, а форму $\langle A_1 x, x\rangle$ — к каноническому виду.

3. Минимизация функции максимума на плоскости

В двумерном случае при минимизации функции максимума мы сталкиваемся с тремя вариантами:

1) точка минимума функции максимума совпадает с точкой минимума одной из функций;

2) в точке минимума активны две функции;

3) в точке минимума активны три функции.

Если оказалось, что в точке минимума активны более трех функций, то нулевая точка должна содержаться в выпуклом многоугольнике, натянутом на градиенты активных функций в этой точке. Но в двумерном случае, как следует из теоремы Каратеодори, для представления любой точки из выпуклой оболочки, натянутой на конечный набор векторов, в виде выпуклой комбинации этих векторов достаточно не более трех точек. Стало быть, в точке минимума можно рассматривать не более трех активных функций.

Если в точке минимума функции максимума активны две функции, то задача минимизации может быть сведена к нахождению положительного корня многочлена четвертой степени. Рассмотрим

Пример 1. Пусть

$$f(x) = \max\{f_0(x), f_1(x)\}, \quad x = (x_1, x_2) \in \mathbb{R}^2,$$

где

$$f_0(x) = \frac{1}{2}\langle x, x\rangle, \quad f_1(x) = \frac{1}{2}\langle \Theta x, x\rangle + \langle b, x\rangle + c.$$

$$\Theta = \operatorname{diag}\{4, 6\}, \quad b = (-3, -4) \in \mathbb{R}^2, \quad c = 2.5.$$

Найдем точку минимума функции $f(x)$. Так как коэффициент $c > 0$, то нулевая точка не является точкой минимума функции f на $\mathbb{R}^2$. Точка $x_1^* = \left(\frac{3}{4}, \frac{3}{4}\right) \in \mathbb{R}^2$ есть точка минимума функции f_1 на $\mathbb{R}^2$. В ней $f_0(x_1^*) = \frac{145}{288}$, $f_1(x_1^*) = \frac{1}{24}$. Следовательно, точка x_1^* также не является точкой минимума функции f. Таким образом, функция f достигает минимума в точке x^*, для которой $R_f(x^*) = \{0, 1\}$. Тогда

$$p_1(\mu) = 2\mu + 5, \quad p_1(\mu) = 2\mu + 7,$$

$$P(\mu) = 9(2\mu + 5)(6 + \mu)^2 + 16(2\mu + 7)(4 + \mu)^2 - 5(4 + \mu)^2(6 + \mu)^2$$

$$= 5\mu^4 + 50\mu^3 + 111\mu^2 - 196\mu - 532 = (\mu - 2)(5\mu^3 + 60\mu^2 + 231\mu + 266).$$

Многочлен $5\mu^3+60\mu^2+231\mu+266$ не имеет положительных корней, поэтому число $\mu^* = 2$ является единственным положительным корнем многочлена $P(\mu)$, и точкой минимума функции f на $\mathbb{R}^2$ будет точка $x^* = \left(\frac{1}{2}, \frac{1}{2}\right)$.

Рассмотрим задачу минимизации на плоскости функции максимума от трех квадратичных функций:

$$f_0(x) = \frac{1}{2}\langle x, x\rangle, \quad f_1(x) = \frac{1}{2}\langle \Lambda x, x\rangle + \langle b_1, x\rangle + c_1, \quad f_2(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b_2, x\rangle + c_2,$$

где матрица Λ — диагональная, матрица A — симметричная, x, b_1 $b_2 \in \mathbb{R}^2$, c_1, $c_2 \in \mathbb{R}$. Предположим, что в точке минимума x^* активны три функции. Тогда точку минимума следует искать среди решений системы нелинейных уравнений:

$$\begin{cases} \frac{1}{2}\langle \Lambda x, x\rangle + \langle b_1, x\rangle + c_1 = \frac{1}{2}\langle x, x\rangle, \\ \frac{1}{2}\langle Ax, x\rangle + \langle b_2, x\rangle + c_2 = \frac{1}{2}\langle x, x\rangle. \end{cases}$$

Среди этих точек выбирается точка x^*, для которой нулевая точка содержится в треугольнике, натянутом на векторы x^*, $\Lambda x^* + b_1$, $Ax^* + b_2$.

Следует отметить тот факт, что если найденная точка минимума не является точкой минимума ни одной из функций f_p, $p = 0, 1, 2$, то она будет точкой строгого локального минимума функции f.

4. Минимизация функции максимума квадратичных функций с постоянным шагом

Пусть

$$f(x) = \max_{i \in I} f_i(x),\ f_i(x) = \frac{1}{2}\langle A_i x, x\rangle + \langle b_i, x\rangle + c_i,\ i \in I = \{1, \ldots, s\},$$

где матрицы A_i — положительнно определенные, x, $b_i \in \mathbb{R}^n$, $c_i \in \mathbb{R}$, $i \in I$. Известно, что функция f является сильно выпуклой на $\mathbb{R}^n$. В качестве константы сильной выпуклости можно взять константу $m = \min\limits_{i \in I} \frac{m_i}{2}$, где m_i — наименьшее собственное число матрицы A_i. Т. е.,

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2) - m\lambda(1-\lambda)\|x_1 - x_2\|^2,$$
$$\forall\, \lambda \in [0,1], \quad \forall\, x_1, x_2 \in \mathbb{R}^n, \quad i \in I.$$

Рассмотрим задачу: найти

$$\min_{x \in \mathbb{R}^n} f(x). \tag{11}$$

Решение задачи (11) существует и точка минимума функции f на $\mathbb{R}^n$ единственна.

Обозначим через $M = \max\{1, \max\limits_{i \in I} M_i\}$, где M_i — наибольшее собственное число матрицы A_i. Тогда справедливо неравенство

$$\langle A_i x, x\rangle \leqslant M\|x\|^2 \quad \forall\, x \in \mathbb{R}^n, \quad \forall\, i \in I. \tag{12}$$

С каждой точкой $x \in \mathbb{R}^n$ свяжем оптимизационную задачу: найти

$$\min_{w \in \mathbb{R}^n} \max_{i \in I} \left\{ (f_i(x) - f(x))M + \langle f_i'(x), w\rangle + \frac{1}{2}\|w\|^2 \right\}. \tag{13}$$

Этот минимум достигается в единственной точке $w(x) \in \mathbb{R}^n$. Обозначим через

$$p(x) = \max_{i \in I} \left\{ (f_i(x) - f(x))M + \langle f_i'(x), w(x)\rangle + \frac{1}{2}\|w(x)\|^2 \right\}.$$

Из необходимых условий минимума задачи (13) следует существование коэффициентов

$$\lambda_i(x) \geqslant 0, \quad i \in I, \quad \sum_{i=1}^{s} \lambda_i(x) = 1,$$

таких, что

$$w(x) = -\sum_{i=1}^{s} \lambda_i(x) f_i'(x) = \sum_{i=1}^{s} \lambda_i(x)(A_i x + b_i),$$
$$p(x) = \sum_{i=1}^{s} \lambda_i(x)(f_i(x) - f(x))M - \frac{1}{2}\|w(x)\|^2 \leqslant 0.$$

Несложно доказать утверждения:

Лемма 2. 1) *Функция $p(x)$ тогда и только тогда равна нулю, когда точка x является точкой минимума функции f на $\mathbb{R}^n$.*

2) *Если вектор $w(x) = 0_n$, то $p(x) = 0$.*

Таким образом, если точка x^* — точка минимума функции f на $\mathbb{R}^n$, то $w(x^*) = 0_n$.

В дальнейшем будет показано, что если точка x не есть точка минимума функции f на $\mathbb{R}^n$, то направление $w(x)$ является направлением спуска функции f в точке x.

Опишем метод минимизации функции f на всем пространстве $\mathbb{R}^n$.

Выберем произвольную точку $x_0 \in \mathbb{R}^n$. Если точка x_0 есть точка минимума функции f на $\mathbb{R}^n$, то процесс останавливается.

Пусть уже найдена точка $x_k \in \mathbb{R}^n$. Если точка x_k является точкой минимума функции f на $\mathbb{R}^n$, то процесс закончен. Предположим, что точка x_k не является точкой минимума функции f на $\mathbb{R}^n$. Найдем направление спуска $w(x_k)$, т. е., решим оптимизационную задачу (13) при $x = x_k$. Тогда найдутся такие коэффициенты

$$\lambda_i(x_k) = \lambda_i^k \geqslant 0, \quad i \in I, \quad \sum_{i=1}^{s} \lambda_i^k = 1,$$

что справедливо

$$w(x_k) = -\sum_{i=1}^{s} \lambda_i^k f_i'(x_k), \quad w(x_k) \neq 0_n,$$
$$p(x_k) = \sum_{i=1}^{s} \lambda_i(x_k)(f_i(x_k) - f(x_k))M - \frac{1}{2}\|w(x_k)\|^2 < 0.$$

Положим $\hat{\alpha} = \dfrac{1}{M}$, $x_{k+1} = x_k + \hat{\alpha} w(x_k)$.

Лемма 3. *В каждой точке x_k выполняется неравенство*

$$f(x_{k+1}) - f(x_k) \leqslant \frac{1}{M} p(x_k). \tag{14}$$

$\lhd$ В силу свойств функций f_i справедливо

$$f_i(x + \alpha w) = f_i(x) + \alpha\langle f_i'(x), w\rangle + \frac{\alpha^2}{2}\langle A_i w, w\rangle \quad \forall\, i \in I,\ \forall\, x \in \mathbb{R}^n,\ w \in \mathbb{R}^n,\ \forall\, \alpha.$$

Тогда из (12) для любого $i \in I$ имеем

$$\begin{aligned} f_i(x_k + \hat{\alpha} w(x_k)) - f(x_k) &\leqslant f_i(x_k) - f(x_k) + \hat{\alpha}\langle f_i'(x_k), w(x_k)\rangle + \frac{\hat{\alpha}^2}{2} M \|w(x_k)\|^2 \\ &= f_i(x_k) - f(x_k) + \frac{1}{M}\langle f_i'(x_k), w(x_k)\rangle + \frac{1}{2M}\|w(x_k)\|^2. \end{aligned}$$

Отсюда вытекает неравенство

$$\max_{i \in I} f_i(x_k + \hat{\alpha} w(x_k)) - f(x_k) = f(x_{k+1}) - f(x_k)$$
$$\leqslant \max_{i \in I}\left\{ f_i(x_k) - f(x_k) + \frac{1}{M}\langle f_i'(x_k), w(x_k)\rangle + \frac{1}{2M}\|w(x_k)\|^2 \right\}$$
$$= \frac{1}{M}\max_{i \in I}\left\{ (f_i(x_k) - f(x_k))M + \langle f_i'(x_k), w(x_k)\rangle + \frac{1}{2}\|w(x_k)\|^2 \right\} = \frac{1}{M} p(x_k) < 0. \ \rhd$$

Неравенство (14) показывает, что процесс является релаксационным, следовательно, последовательность $\{f(x_k)\}$ является сходящейся.

Следствие 1. *Если последовательность* $\{x_k\}$ *бесконечна, то* $p(x_k) \to 0$ *при* $k \to +\infty$.

Так как для любой сильно выпуклой функции множество

$$\mathcal{L}(x_0) = \{x \in \mathbb{R}^n \bigm| f(x) \leqslant f(x_0)\}$$

компактно, то последовательность $\{x_k\}$, генерируемая этим методом принадлежит лебегову множеству $\mathcal{L}(x_0)$.

Лемма 4. *Для данного метода в каждой точке* x_k *справедливо неравенство*

$$p(x_k) \leqslant m\left(f(x^*) - f(x_k)\right). \tag{15}$$

$\lhd$ Для точек x_k, x^* и для каждого индекса $i \in I$ верно неравенство

$$f_i(x^*) - f_i(x_k) \geqslant \langle f_i'(x_k), x^* - x_k\rangle + \frac{m}{2}\|x^* - x_k\|^2, \quad i \in I.$$

Тогда, умножая это неравенство на m, получим

$$\langle f_i'(x_k), m(x^* - x_k)\rangle \leqslant m(f_i(x^*) - f_i(x_k)) - \frac{m^2}{2}\|x^* - x_k\|^2, \quad i \in I.$$

Положим в (13) $w = m(x^* - x_k)$. Тогда имеем

$$\begin{aligned} p(x_k) \leqslant \max_{i \in I}\Big\{&(f_i(x_k) - f(x_k))M + \langle f_i'(x_k), m(x^* - x_k)\rangle \\ &+ \frac{m^2}{2}\|x^* - x_k\|^2\Big\} \leqslant \max_{i\in I}\Big\{(f_i(x_k) - f(x_k))m + m(f_i(x^*) - f_i(x_k))\Big\} \\ &\leqslant m \max_{i \in I}\{f_i(x^*) - f(x_k)\} \leqslant m(f(x^*) - f(x_k)). \ \rhd \end{aligned}$$

Следствие 2. *Из формул* (14) *и* (15) *вытекает неравенство*

$$f(x_{k+1}) - f(x_k) \leqslant \frac{m}{M}(f(x^*) - f(x_k)). \tag{16}$$

Следовательно, последовательность $\{x_k\}$ *является сходящейся и ее предельной точкой является точка минимума функции* f *на* $\mathbb{R}^n$.

Теорема 2. *В данном методе при произвольной начальной точке* $x_0 \in \mathbb{R}^n$ *последовательность* $\{x_k\}$ *сходится к точке минимума* x^* *функции* f *со скоростью геометрической прогрессии:*

$$f(x_{k+1}) - f(x^*) \leqslant q(f(x_k) - f(x^*)),$$

где $q = 1 - \dfrac{m}{M}$, *и существует положительное число* Q, *что справедливо неравенство*

$$\|x_k - x^*\| \leqslant Q(\sqrt{q})^k.$$

$\lhd$ Из (16) получим

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leqslant f(x_k) - f(x^*) + \frac{m}{M}(f(x^*) - f(x_k)) \\ &= \left(1 - \frac{m}{M}\right)(f(x_k) - f(x^*)) = q(f(x_k) - f(x^*)). \end{aligned}$$

Следовательно, для каждого фиксированного $k > 0$ имеем

$$f(x_k) - f(x^*) \leqslant q^k(f(x_0) - f(x^*)).$$

Поскольку точка x^* есть точка минимума функции f на $\mathbb{R}^n$, то

$$f(x_k) - f(x^*) \geqslant \frac{m}{2}\|x_k - x^*\|^2.$$

Поэтому

$$\|x_k - x^*\|^2 \leqslant \frac{2}{m}(f(x_k) - f(x^*))q^k.$$

Обозначим через

$$Q = \sqrt{\frac{2}{m}(f(x_0) - f(x^*))},$$

тогда

$$\|x_k - x^*\| \leqslant \sqrt{\frac{2}{m}(f(x_k) - f(x^*))q^k} \leqslant Q(\sqrt{q})^k. \ \triangleright$$

Рассмотрим более подробно вычислительный аспект решения задачи (13). Для того чтобы найти направление $w(x_k)$, необходимо решить задачу квадратичного программирования

$$\min_{\lambda\in\Lambda} \left\{\langle G(x_k)\lambda, \lambda\rangle + M\langle q(x_k), \lambda\rangle\right\},$$

где

$$\Lambda = \{\lambda = (\lambda_1, \dots, \lambda_s) \in \mathbb{R}^s \mid \sum_{i=1}^{s} \lambda_i = 1,\ \lambda_i \geqslant 0,\ i \in I\},$$

$G(x_k)$ — матрица Грама векторов $f_i'(x_k)$, т. е.,

$$G(x_k) = \begin{pmatrix} \langle f_1'(x_k), f_1'(x_k)\rangle & \dots & \langle f_1'(x_k), f_s'(x_k)\rangle \\ & \cdots & \\ \langle f_s'(x_k), f_1'(x_k)\rangle & \dots & \langle f_s'(x_k), f_s'(x_k)\rangle \end{pmatrix},$$

и

$$q(x_k) = (q_1(x_k), \dots, q_s(x_k)) \in \mathbb{R}^s, \quad q_i(x_k) = f(x_k) - f_i(x_k),\ i \in I.$$

Таким образом, необходимо определить коэффициенты λ_i, $i \in I$, и выразить через них вектор $w(x_k)$ по формуле $w(x_k) = -\sum\limits_{i=1}^{s} \lambda_i^k f_i'(x_k)$.

Пример 2. В качестве тестовой функции была рассмотрена функция максимума [5]

$$f(x) = \max_{i\in I} f_i(x), \quad x \in \mathbb{R}^{10}, \quad I = 1, \dots, 5,$$

где

$$\begin{aligned} & f_i(x) = \langle A_i x, x\rangle + \langle b_i, x\rangle, \quad b_i \in \mathbb{R}^{10}, \quad i \in I, \\ & a_i(j,k) = e^{j/k}\cos(j\cdot k)\sin i, \quad j < k, \quad a_i(j,k) = a_i(k,j), \quad j > k, \\ & a_i(j,j) = 2\frac{|\sin i|\cdot i}{j} + \sum_{k\neq j} |a_i(j,k)|, \\ & b_i(j) = e^{j/i}\sin(i\cdot j), \quad j,k \in 1:10. \end{aligned}$$

Матрицы A_i, $i \in I$, являются положительно определенными и в таблице 1 приведены, соответственно, максимальное ($\lambda_{\max}$) и минимальное ($\lambda_{\min}$) собственные числа матриц

$2A_i$, значение каждой из функций f_i, $i \in I$, в точке минимума и норма градиента этих функций в точке минимума.

Для данной функции $M = 36.32$, $m = 0.67$. Точка $x^* = (-0.0546, -0.0241, -0.0057, 0.0231, 0.0558, -0.2434, 0.0685, 0.1321, 0.0772, 0.0336)^T \in \mathbb{R}^{10}$ — точка минимума функции $f(x)$ на $\mathbb{R}^n$, $f(x^*) = -0.72576$ — решение задачи, $\|w(x^*)\| = 0.000092$, $R(x^*) = \{2, 3, 4, 5\}$, где

$$R(x^*) = \{i \in I \mid f_i(x^*) = f(x^*)\}.$$

Таблица 1

i	1	2	3	4	5
$\lambda_{\max}$	53.947	59.585	9.470	55.017	72.613
$\lambda_{\min}$	14.301	16.661	2.754	19.337	25.919
$f_i(x^*)$	-311.19750	-0.725756	-0.725756	-0.725756	-0.725756
$\|f_i'(x^*)\|$	12805.888	155.679	37.765	14.479	5.934

Из таблицы видно, что первая функция не является активной в оптимальной точке, но поскольку ее градиент в x^* достаточно большой, то в точке $x^* + \alpha f_1'(x^*)$ при небольших α она становится активной.

Очевидно, что в большинстве случаях заранее не известно число M. Поэтому можно использовать данный алгоритм при различных M, начиная с заведомо большого. Для данного примера оптимальная точка x^* была получена при $M = 145.28$ за 85 шагов, при $M = 72.64$ за 38 шагов, при $M = 36.32$ за 19 шагов, при $M = 18.16$ за 11 шагов. При $M = 9$ и меньше алгоритм не сходился. Задача считалась решенной, если выполнялось условие $\|w(x_k)\| < 10^{-4}$.

Литература

1. *Демьянов В. Ф., Малоземов В. Н.* Введение в минимакс.—М.: Наука, 1972.—368 с.
2. *Пшеничный Б. Н.* Метод линеаризации.—М.: Наука, 1983.—136 с.
3. *Гантмахер Ф. Р.* Теория матриц.—М.: Наука, 1967.—552 с.
4. *Мишина А. П., Проскуряков И. В.* Высшая алгебра. СМБ.—М.: Физматгиз, 1962.—300 с.
5. *Lemarechal C.* An extension of Davidon methods to nondifferentiable problems // Mathematical programming.—1975.—Study 3.—P. 95–100.

Статья поступила 18 декабря 2006 г.

Полякова Людмила Николаевна
Санкт-Петербург, Санкт-Петербургский государственный университет
E-mail: `lyudmila.polyakova@pobox.spbu.ru`