КВАЗИНЬЮТОНОВСКИЕ МЕТОДЫ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ*

В. Н. Малозёмов — v.malozemov@spbu.ru

Е. К. Чернэуцану — katerinache@yandex.ru

16 апреля 2015 г.

Аннотация. В докладе анализируется общая схема построения квазиньютоновских методов безусловной минимизации, предложенная Ю. М. Данилиным [1]. Особенность квазиньютоновских методов состоит в том, что они позволяют найти точку минимума выпуклой квадратичной функции от n переменных не более чем за n шагов.

1°. Начнём с минимизации квадратичной функции:

$$f(x) := \tfrac{1}{2}\langle Dx, x\rangle + \langle c, x\rangle \to \min_{x\in\mathbb{R}^n}.$$

Здесь D — симметричная положительно определённая матрица.

Точка минимума x_* функции $f(x)$ на $\mathbb{R}^n$ существует, единственна и допускает представление

$$x_* = -D^{-1}c.$$

Вычисление x_* по этой формуле можно заменить итерационной схемой, в которой вместо атрибутов квадратичной функции (матрицы D и вектора c) используются значения функции f и её градиента.

2°. Возьмём некоторую линейно независимую систему векторов $y_0, y_1, \ldots, y_{n-1}$ и произвольные векторы $x_0, x_1, \ldots, x_{n-1}$. Положим

$$r_i = f'(x_i + y_i) - f'(x_i) = Dy_i, \quad i \in 0 : n-1. \tag{1}$$

Очевидно, что векторы $r_0, r_1, \ldots, r_{n-1}$ линейно независимы. Запишем

$$f'(x_i) - f'(x_*) = D(x_i - x_*).$$

*Семинар по конструктивному негладкому анализу и недифференцируемой оптимизации «CNSA & NDO»: `http://www.apmath.spbu.ru/cnsa/`

Умножим последнее равенство скалярно на y_i. Учитывая, что $f'(x_*) = \mathbb{O}$, получаем

$$\bigl\langle D(x_i - x_*), y_i \bigr\rangle = \bigl\langle f'(x_i), y_i \bigr\rangle$$

или (в силу симметричности матрицы D и (1))

$$\langle r_i, x_* \rangle = \langle r_i, x_i \rangle - \bigl\langle f'(x_i), y_i \bigr\rangle.$$

Обозначив $d_i = \langle r_i, x_i \rangle - \bigl\langle f'(x_i), y_i \bigr\rangle$, придём к системе линейных уравнений относительно x_*:

$$\langle r_i, x_* \rangle = d_i, \quad i \in 0 : n-1. \tag{2}$$

Таким образом, задача минимизации квадратичной функции сводится к решению системы линейных уравнений (2), матрица и правая часть которой зависят только от градиента целевой функции.

3°. Выбор точек x_i находится в нашем распоряжении. В качестве x_0 возьмём произвольную точку из $\mathbb{R}^n$, а выбор остальных точек x_k подчиним условию

$$\langle r_i, x_k \rangle = d_i, \quad i \in 0 : k-1;\ k = 1, \ldots, n. \tag{3}$$

Подробнее: вычислив r_0 и d_0, точку x_1 строим так, чтобы

$$\langle r_0, x_1 \rangle = d_0; \tag{4}$$

точка x_2 должна удовлетворять двум уравнениям

$$\langle r_0, x_2 \rangle = d_0, \quad \langle r_1, x_2 \rangle = d_1$$

и так далее. Очевидно, что $x_n = x_*$.

Положим для определённости

$$x_{k+1} = x_k + \alpha_k p_k, \quad k \in 0 : n-1. \tag{5}$$

Уравнение (4) примет вид

$$\langle r_0, x_0 + \alpha_0 p_0 \rangle = d_0$$

или (с учётом определения d_0)

$$\alpha_0 \langle r_0, p_0 \rangle = -\langle f'(x_0), y_0 \rangle.$$

В качестве p_0 можно взять произвольный ненулевой вектор.

Допустим, что вектор x_k удовлетворяет системе уравнений (3) и потребуем, чтобы $\langle r_i, x_{k+1} \rangle = d_i$ при $i \in 0 : k$. Согласно (5) это условие можно переписать в виде

$$\langle r_i, x_k + \alpha_k p_k \rangle = d_i, \quad i \in 0 : k. \tag{6}$$

При $i = k$ получаем
$$\alpha_k \langle r_k, p_k \rangle = -\langle f'(x_k), y_k \rangle. \tag{7}$$
Для выполнения условия (6) при $i \in 0 : k-1$ в силу (3) достаточно, чтобы
$$\langle r_i, p_k \rangle = 0, \quad i \in 0 : k-1; \; k \in 1 : n-1. \tag{8}$$
Условие (7) используется для определения α_k.

4°. Обратимся к соотношению (8). Будем искать p_k в виде
$$p_k = H_k w_k,$$
где H_k — квадратная матрица порядка n. В качестве w_k обычно берут антиградиент $-f'(x_k)$. Имеем
$$\langle H_k^T r_i, w_k \rangle = 0, \quad i \in 0 : k-1. \tag{9}$$
Чтобы получить условие, аналогичное (3), потребуем, чтобы вектор $H_k^T r_i$ не зависел от k:
$$H_k^T r_i = z_i, \quad i \in 0 : k-1. \tag{10}$$
Тогда условие (9) примет вид
$$\langle z_i, w_k \rangle = 0, \quad i \in 0 : k-1; \; k \in 1 : n-1. \tag{11}$$
В частности, при $z_i = y_i$ матрица H_k должна удовлетворять условию
$$H_k^T D y_i = y_i, \quad i \in 0 : k-1. \tag{12}$$

Обозначим через Y матрицу со столбцами $y_0, y_1, \ldots, y_{n-1}$. Для H_n получим матричное уравнение
$$H_n^T D Y = Y.$$
Отсюда следует, что
$$H_n^T = Y (DY)^{-1} = D^{-1}.$$
В связи с этим условие (12) называют *квазиньютоновским*, а условие (10) — *обобщённым квазиньютоновским*. Отметим, что в вычислениях матрица H_n не участвует (последней используется матрица H_{n-1}).

5°. Матрицы H_k будем вводить последовательно:

$$H_{k+1} = H_k + \Delta H_k, \quad k \in 0, 1, \dots, n-2.$$

В качестве H_0 берётся произвольная матрица (обычно $H_0 = I$, где I — единичная матрица порядка n).

Соотношение (10) при $k = 1$ примет вид

$$(H_0 + \Delta H_0)^T r_0 = z_0,$$

откуда следует, что

$$(\Delta H_0)^T r_0 = z_0 - H_0^T r_0.$$

Допустим, что H_k удовлетворяет условию (10). Потребуем, чтобы $H_{k+1}^T r_i = z_i$ при $i \in 0 : k$, то есть чтобы

$$(H_k + \Delta H_k)^T r_i = z_i, \quad i \in 0 : k. \tag{13}$$

При $i = k$ получаем

$$(\Delta H_k)^T r_k = z_k - H_k^T r_k. \tag{14}$$

Для выполнения условия (13) при $i \in 0 : k-1$ достаточно, чтобы

$$(\Delta H_k)^T r_i = \mathbb{O}, \quad i \in 0 : k-1. \tag{15}$$

Таким образом, обобщённое квазиньютоновское условие (10) выполняется для последовательности матриц $H_{k+1} = H_k + \Delta H_k$, где H_0 — произвольная матрица, а ΔH_k удовлетворяет условию (14) при $k \in 0 : n-2$ и условию (15) при $k \in 1 : n-2$.

6°. Соотношениям (14) и (15) можно удовлетворить, например, следующим образом (*схема Хуанга*):

$$(\Delta H_k)^T = \frac{z_k u_k^T}{\langle u_k, r_k \rangle} - \frac{H_k^T r_k v_k^T}{\langle v_k, r_k \rangle}, \tag{16}$$

если

$$\begin{gathered} \langle u_k, r_k \rangle \neq 0, \ \langle v_k, r_k \rangle \neq 0, \quad k \in 0 : n-2; \\ \langle u_k, r_i \rangle = 0, \ \langle v_k, r_i \rangle = 0, \quad i \in 0 : k-1, \ k \in 1 : n-2. \end{gathered} \tag{17}$$

Для того чтобы схема была полной, к (16) нужно добавить условия (7) и (11).

Обычно полагают

$$\begin{aligned} u_k &= t_{k1} z_k + t_{k2} H_k^T r_k, \\ v_k &= t_{k3} z_k + t_{k4} H_k^T r_k, \end{aligned}$$

где $t_{k1}, \dots, t_{k4}$ — произвольные числа, обеспечивающие выполнение условий (17). Тем самым, конкретная реализация схемы Хуанга определяется матрицей второго порядка

$$T_k = \begin{pmatrix} t_{k1} & t_{k2} \\ t_{k3} & t_{k4} \end{pmatrix}.$$

7°. Положим в схеме Хуанга

$$T_k = \begin{pmatrix} \rho_k & -1 \\ 1 & 0 \end{pmatrix},$$

так что

$$\begin{aligned} u_k &= \rho_k z_k - H_k^T r_k, \\ v_k &= z_k. \end{aligned}$$

Параметр ρ_k выберем из условия равенства знаменателей в формуле (16), то есть из условия $\langle u_k, r_k \rangle = \langle v_k, r_k \rangle$. Имеем

$$\langle \rho_k z_k - H_k^T r_k, r_k \rangle = \langle z_k, r_k \rangle,$$

откуда следует, что

$$\rho_k = 1 + \frac{\langle H_k^T r_k, r_k \rangle}{\langle z_k, r_k \rangle}.$$

Формула (16) принимает вид

$$\begin{aligned} (\Delta H_k)^T = \frac{z_k (\rho_k z_k - H_k^T r_k)^T}{\langle z_k, r_k \rangle} - \frac{H_k^T r_k z_k^T}{\langle z_k, r_k \rangle} = \\ = \left(1 + \frac{\langle H_k^T r_k, r_k \rangle}{\langle z_k, r_k \rangle}\right) \frac{z_k z_k^T}{\langle z_k, r_k \rangle} - \frac{H_k^T r_k z_k^T + z_k r_k^T H_k}{\langle z_k, r_k \rangle}. \end{aligned}$$

Отметим, что в правой части этого равенства стоит симметричная матрица.

Положим $H_0 = I$. Тогда матрицы H_k будут симметричными. При этом

$$\begin{aligned} H_{k+1} = H_k + \left(1 + \frac{\langle H_k^T r_k, r_k \rangle}{\langle z_k, r_k \rangle}\right) \frac{z_k z_k^T}{\langle z_k, r_k \rangle} - \frac{H_k^T r_k z_k^T + z_k r_k^T H_k}{\langle z_k, r_k \rangle} = \\ = \left(I - \frac{z_k r_k^T}{\langle z_k, r_k \rangle}\right) H_k \left(I - \frac{r_k z_k^T}{\langle z_k, r_k \rangle}\right) + \frac{z_k z_k^T}{\langle z_k, r_k \rangle} = \\ = Q_k^T H_k Q_k + \frac{z_k z_k^T}{\langle z_k, r_k \rangle}. \end{aligned} \qquad (18)$$

Здесь

$$Q_k = I - \frac{r_k z_k^T}{\langle z_k, r_k \rangle}.$$

Отметим, что $Q_k r_k = \mathbb{O}$.

Дальнейший выбор параметров осуществим так:

$$w_k = -f'(x_k), \quad p_k = H_k w_k, \quad z_k = y_k = \alpha_k p_k,$$

где α_k удовлетворяет условию (7). Учитывая, что

$$r_k = f'(x_k + \alpha_k p_k) - f'(x_k) = f'(x_{k+1}) - f'(x_k) = \alpha_k D p_k = D z_k,$$

перепишем условие (7) в виде

$$\alpha_k^2 \langle D p_k, p_k \rangle = -\alpha_k \big\langle f'(x_k), p_k \big\rangle.$$

В качестве α_k возьмём величину

$$\alpha_k = -\frac{\big\langle f'(x_k), p_k \big\rangle}{\langle D p_k, p_k \rangle}. \tag{19}$$

Отметим, что α_k является точкой минимума функции

$$\varphi_k(\alpha) = f(x_k + \alpha p_k).$$

Действительно,

$$\begin{aligned} \varphi_k'(\alpha) &= \big\langle f'(x_k + \alpha p_k), p_k \big\rangle = \big\langle D(x_k + \alpha p_k) + c, p_k \big\rangle = \\ &= \big\langle f'(x_k), p_k \big\rangle + \alpha \langle D p_k, p_k \rangle, \end{aligned}$$

поэтому уравнение $\varphi_k'(\alpha) = 0$ имеет единственное решение $\alpha = \alpha_k$.

Условие $\varphi_k'(\alpha_k) = 0$ можно записать в виде

$$\big\langle f'(x_{k+1}), p_k \big\rangle = 0. \tag{20}$$

8°. В дальнейшем будем использовать обозначение $f_k' = f'(x_k)$.

ТЕОРЕМА. *Если градиенты $f_0', f_1', \ldots, f_{k-1}'$ отличны от нуля, то*

1) $z_{k-1} \neq \mathbb{O}$;

2) $\langle f_k', z_i \rangle = 0$, $i \in 0 : k-1$;

3) H_k — *положительно определённая матрица;*

4) $H_k D z_i = z_i$, $i \in 0 : k-1$;

5) $\langle D z_i, z_j \rangle = 0$ *при* $i \neq j$; $i,\ j \in 0 : k-1$.

Д о к а з а т е л ь с т в о проведём индукцией по k. При $k = 1$ имеем $p_0 = -H_0 f_0' = -f_0' \neq \mathbb{O}$,

$$\alpha_0 = -\frac{\langle f_0', -f_0' \rangle}{\langle D p_0, p_0 \rangle} > 0,$$

так что $z_0 = \alpha_0 p_0 \neq \mathbb{O}$. Согласно (20), $\langle f_1', z_0 \rangle = \alpha_0 \langle f_1', p_0 \rangle = 0$.

Покажем, что матрица H_1 положительно определена. Приняв во внимание формулу (18) и тот факт, что $\langle z_0, r_0 \rangle = \langle z_0, D z_0 \rangle > 0$, запишем

$$\langle H_1 x, x \rangle = \| Q_0 x \|^2 + \frac{\langle z_0, x \rangle^2}{\langle z_0, r_0 \rangle} \geqslant 0.$$

Допустим, что $\langle H_1 x, x \rangle = 0$. Тогда $\langle z_0, x \rangle = 0$ и $Q_0 x = \mathbb{O}$. Из последнего равенства следует, что

$$\mathbb{O} = x - \frac{r_0 \langle z_0, x \rangle}{\langle z_0, r_0 \rangle} = x,$$

то есть $x = \mathbb{O}$. Положительная определённость матрицы H_1 установлена.

Далее, $H_1 D z_0 = H_1 r_0 = z_0$, поскольку $Q_0 r_0 = \mathbb{O}$. Условие 5) при $k = 1$ бессодержательно. Чтобы и по нему создать базу индукции, рассмотрим случай $k = 2$.

По условию теоремы $f_1' \neq \mathbb{O}$ и по доказанному матрица H_1 положительно определена. Значит, $p_1 = -H_1 f_1' \neq \mathbb{O}$. Согласно (19),

$$\alpha_1 = -\frac{\langle f_1', -H_1 f_1' \rangle}{\langle D p_1, p_1 \rangle} > 0,$$

так что $z_1 = \alpha_1 p_1 \neq \mathbb{O}$. Теперь имеем

$$\begin{aligned} \langle z_1, D z_0 \rangle &= -\alpha_1 \langle H_1 f_1', D z_0 \rangle = -\alpha_1 \langle f_1', H_1 D z_0 \rangle = \\ &= -\alpha_1 \langle f_1', z_0 \rangle = 0 \end{aligned}$$

и

$$\langle f_2', z_0 \rangle = \langle f_1' + D z_1, z_0 \rangle = \langle f_1', z_0 \rangle = 0.$$

Равенство $\langle f_2', z_1 \rangle = 0$ справедливо в силу (20).

Проверим, что матрица H_2 положительно определена. Запишем

$$\langle H_2 x, x \rangle = \langle H_1 Q_1 x, Q_1 x \rangle + \frac{\langle z_1, x \rangle^2}{\langle z_1, r_1 \rangle} \geqslant 0.$$

Мы приняли во внимание, что $\langle z_1, r_1 \rangle = \langle z_1, D z_1 \rangle > 0$. Если $\langle H_2 x, x \rangle = 0$, то $Q_1 x = \mathbb{O}$ и $\langle z_1, x \rangle = 0$. Отсюда следует, что

$$\mathbb{O} = Q_1 x = x - \frac{r_1 \langle z_1, x \rangle}{\langle z_1, r_1 \rangle} = x,$$

то есть $x = \mathbb{O}$. Положительная определённость матрицы H_2 установлена.

На основании равенства $Q_1 r_1 = \mathbb{O}$ получаем

$$H_2 D z_1 = H_2 r_1 = z_1.$$

Далее

$$Q_1 D z_0 = \left(I - \frac{r_1 z_1^T}{\langle z_1, r_1\rangle}\right) D z_0 = D z_0,$$

$$Q_1^T D z_0 = \left(I - \frac{z_1 r_1^T}{\langle z_1, r_1\rangle}\right) z_0 = z_0 - \frac{\langle D z_1, z_0\rangle}{\langle z_1, r_1\rangle} = z_0.$$

поэтому

$$H_2 D z_0 = Q_1^T H_1 Q_1 D z_0 = Q_1^T H_1 D z_0 = Q_1^T z_0 = z_0.$$

При $k = 2$ теорема доказана.

Сделаем индукционный переход от k к $k+1$. Так как $f_k' \neq \mathbb{O}$ и H_k — положительно определённая матрица, то $p_k = -H_k f_k' \neq \mathbb{O}$, $\alpha_k > 0$, $z_k \neq \mathbb{O}$. Матрица H_{k+1} положительно определена, что следует из положительной определённости матрицы H_k и условия $z_k \neq \mathbb{O}$.

При $i \in 0 : k-1$ в силу условий 4) и 2) имеем

$$\langle D z_i, z_k\rangle = -\alpha_k \langle H_k f_k', D z_i\rangle = -\alpha_k \langle f_k', z_i\rangle = 0.$$

Согласно (20), $\langle f_{k+1}', z_k\rangle = 0$. При $i \in 0 : k-1$

$$\langle f_{k+1}', z_i\rangle = \langle f_k' + D z_k, z_i\rangle = \langle f_k', z_i\rangle = 0.$$

Наконец, так же, как при $k = 2$, проверяется, что $H_{k+1} D z_k = z_k$ и $H_{k+1} D z_i = Q_k^T z_i = z_i$ при $i \in 0 : k-1$.

Теорема доказана. □

9°. Теперь можно описать метод минимизации квадратичной функции $f(x)$ с положительно определённой матрицей D, в котором используются только значения функции $f(x)$ и её градиента $f'(x)$.

Возьмём произвольное начальное приближение x_0. Если $f'(x_0) = \mathbb{O}$, то x_0 — точка минимума. Процесс заканчивается. Иначе полагаем $H_0 = I$.

Общая $(k+1)$-я итерация, перед началом которой имеются x_k, $f'(x_k) \neq \mathbb{O}$ и H_k, состоит из следующих шагов:

- находим направление спуска $p_k = -H_k f'(x_k)$;
- вычисляем α_k как точку минимума функции $\varphi_k(\alpha) = f(x_k + \alpha p_k)$;
- определяем $z_k = \alpha_k p_k$ и $x_{k+1} = x_k + z_k$;

- если $f'(x_{k+1}) = \mathbb{O}$, то процесс заканчивается. Иначе находим $r_k = f'(x_{k+1}) - f'(x_k)$;

- осуществляем подготовку к следующему шагу

$$H_{k+1} = \left(I - \frac{z_k r_k^T}{\langle z_k, r_k \rangle}\right) H_k \left(I - \frac{r_k z_k^T}{\langle z_k, r_k \rangle}\right) + \frac{z_k z_k^T}{\langle z_k, r_k \rangle}.$$

Таким образом, построены x_{k+1}, $f'(x_{k+1}) \neq \mathbb{O}$ и H_{k+1}. Описание метода завершено.

Подчеркнём, что в формировании матрицы H_{k+1} наряду с матрицей H_k участвуют векторы $z_k = x_{k+1} - x_k$ и $r_k = f'(x_{k+1}) - f'(x_k)$.

Последовательность $z_0, z_1, \ldots$ по свойству 5) из теоремы состоит из D-ортогональных векторов, которые, в частности, являются линейно независимыми. Их количество не может превысить n. Значит, по крайней мере, $f'(x_n) = \mathbb{O}$. Другими словами, точка минимума будет найдена не более, чем за n итераций.

10°. Описанный метод называется BFGS-методом по именам его авторов Бройдена [2], Флетчера [3], Гольдфарба [4] и Шанно [5]. Он может применяться для минимизации не только квадратичных функций, но и для минимизации произвольных гладких и даже негладких функций. По этому поводу см. книгу [6].

ЛИТЕРАТУРА

1. Пшеничный Б. Н., Данилин Ю. М. *Численные методы в экстремальных задачах* М.: Наука, 1975.

2. Broyden C. G. *The convergence of a class of double-rank minimization algorithms* // Jornal of the Institute of Mathematics and Its Applications **6** (1970): 76–90.

3. Fletcher R. *A New Approach to Variable Metric Algorithms* // Computer Journal **13** (1970): 317–322.

4. Goldfarb D. *A Family of Variable Metric Updates Derived by Variational Means* // Mathematics of Computation **24** (1970): 23–26.

5. Shanno D. F. *Conditioning of quasi-Newton methods for function minimization* // Math. Comput. **24** (1970): 647–656.

6. Nocedal J., Wright S. J. *Numerical Optimization.* 2nd edition. USA: Springer, 2006.