ВАРИАНТЫ МЕТОДА СОПРЯЖЁННЫХ ГРАДИЕНТОВ*

В. Н. Малозёмов
v.malozemov@spbu.ru

29 октября 2015 г.

1°. Напомним описание основного варианта метода сопряжённых градиентов для минимизации на $\mathbb{R}^n$ квадратичной функции

$$Q(x) = \tfrac{1}{2}\langle Dx, x\rangle + \langle c, x\rangle$$

с симметричной положительно определённой матрицей D [1].

Нулевой шаг. Берём произвольное начальное приближение $x_0 \in \mathbb{R}^n$ и вычисляем градиент $g_0 = Q'(x_0) = Dx_0 + c$. Если $g_0 = \mathbb{O}$, то x_0 — точка минимума. Вычисления прекращаются. Иначе полагаем $s_1 = -g_0$.

k-й шаг. Пусть имеются x_{k-1}, $g_{k-1} \neq \mathbb{O}$, s_k. Последовательно вычисляем

$$t_k = \frac{\langle g_{k-1}, g_{k-1}\rangle}{\langle Ds_k, s_k\rangle}, \tag{1}$$

$$x_k = x_{k-1} + t_k s_k,$$

$$g_k = Q'(x_k) = g_{k-1} + t_k D s_k.$$

Если $g_k = \mathbb{O}$, то x_k — точка минимума. Вычисления прекращаются. В противном случае находим

$$b_k = \frac{\langle g_k, g_k\rangle}{\langle g_{k-1}, g_{k-1}\rangle}, \tag{2}$$

$$s_{k+1} = -g_k + b_k s_k. \tag{3}$$

Описание метода завершено.

Метод сопряжённых градиентов обладает следующими свойствами:

$$\langle Ds_k, s_j\rangle = 0 \quad \text{при} \quad k \neq j;$$

$$\langle g_k, s_j\rangle = 0 \quad \text{при} \quad j \in 1 : k; \tag{4}$$

$$\langle g_k, g_j\rangle = 0 \quad \text{при} \quad k \neq j. \tag{5}$$

---

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

Из последнего соотношения, в частности, следует, что $g_n = \mathbb{O}$ (если равенство $g_k = \mathbb{O}$ не встретится при $k < n$), то есть, по крайней мере, x_n будет точкой минимума.

2°. Для последовательности $x_0, x_1, \ldots$, построенной методом сопряжённых градиентов, можно получить другое представление. На рис. 1 введены векторы p_k и p_{k+1}, исходя из условий

$$p_{k+1} = -g_k + \lambda(p_k + g_k), \tag{6}$$

$$\langle p_{k+1}, p_k + g_k \rangle = 0. \tag{7}$$

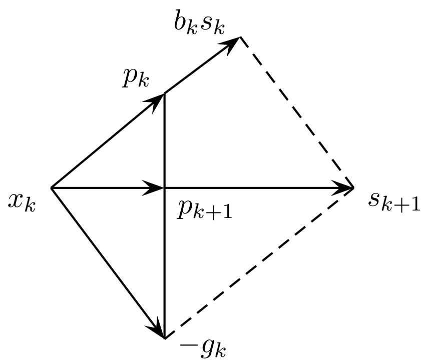

Рис. 1

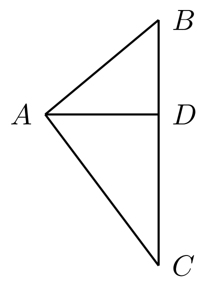

Рис. 2

На рис. 2 выделен треугольник ABC с $\overrightarrow{AB} = p_k$, $\overrightarrow{AC} = -g_k$ и $\overrightarrow{AD} = p_{k+1}$. Угол CAB прямой, поскольку, согласно (4), $\langle s_k, g_k \rangle = 0$. Вектор $p_{k+1} = \overrightarrow{AD}$ строится из условия, что отрезок AD перпендикулярен CB, то есть в прямоугольном треугольнике CAB из вершины A опускается перпендикуляр на сторону CB. При этом $\overrightarrow{AC} + \overrightarrow{CB} = \overrightarrow{AB}$, так что $\overrightarrow{CB} = p_k + g_k$.

Умножим равенство (6) скалярно на $p_k + g_k$. Учитывая (7) и ортогональность векторов p_k и g_k, получаем

$$0 = -\langle g_k, g_k \rangle + \lambda\big(\|p_k\|^2 + \|g_k\|^2\big).$$

Отсюда следует, что $\lambda = \frac{\|g_k\|^2}{\|p_k\|^2 + \|g_k\|^2}$ и

$$p_{k+1} = -g_k + \frac{\|g_k\|^2 (p_k + g_k)}{\|p_k\|^2 + \|g_k\|^2} = \frac{\|p_k\|^2(-g_k) + \|g_k\|^2 p_k}{\|p_k\|^2 + \|g_k\|^2}. \tag{8}$$

К этому добавим условие $p_1 = -g_0$.

ТЕОРЕМА. *Справедливо равенство*

$$p_k = \frac{\|p_k\|^2}{\|g_{k-1}\|^2} s_k, \quad k = 1, 2, \dots \tag{9}$$

Доказательство. Обозначим $\beta_k = \frac{\|p_k\|^2}{\|g_{k-1}\|^2}$, так что формулу (9) можно переписать в виде $p_k = \beta_k s_k$. При $k = 1$ равенство (9) выполняется. Сделаем индукционный переход от k к $k+1$.

Во-первых, отметим, что векторы p_k и g_k ортогональны, поскольку $\langle s_k, g_k \rangle = 0$. Далее, в силу (8), индукционного предположения (9) и (3) имеем

$$p_{k+1} = \frac{\|p_k\|^2}{\|p_k\|^2 + \|g_k\|^2}\Big(-g_k + \frac{\|g_k\|^2}{\|p_k\|^2} p_k\Big) =$$
$$= \frac{\|p_k\|^2}{\|p_k\|^2 + \|g_k\|^2}\Big(-g_k + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} s_k\Big) = \frac{\|p_k\|^2}{\|p_k\|^2 + \|g_k\|^2} s_{k+1}.$$

Остаётся учесть, что

$$\|p_{k+1}\|^2 = \frac{\|p_k\|^4}{\big(\|p_k\|^2 + \|g_k\|^2\big)^2}\Big(\|g_k\|^2 + \frac{\|g_k\|^4}{\|p_k\|^2}\Big) = \frac{\|p_k\|^2 \|g_k\|^2}{\|p_k\|^2 + \|g_k\|^2}.$$

Для коэффициента перед s_{k+1} получаем представление

$$\frac{\|p_k\|^2}{\|p_k\|^2 + \|g_k\|^2} = \frac{\|p_{k+1}\|^2}{\|g_k\|^2} = \beta_{k+1}.$$

Теорема доказана. □

Нетрудно проверить, что справедливо равенство

$$t_k s_k = \alpha_k p_k, \tag{10}$$

где

$$\alpha_k = \frac{\langle p_k, p_k \rangle}{\langle D p_k, p_k \rangle}.$$

Действительно, согласно (9), (1) и определению β_k

$$\alpha_k = \frac{t_k}{\beta_k} = \frac{\|g_{k-1}\|^2 \beta_k}{\langle D s_k, s_k \rangle \beta_k^2} = \frac{\|p_k\|^2}{\langle D p_k, p_k \rangle}.$$

3°. Воспользуемся векторами p_k для описания метода сопряжённых градиентов, имея в виду, что векторы p_k отличаются от s_k лишь положительным коэффициентом.

Нулевой шаг. Берём произвольное начальное приближение $x_0 \in \mathbb{R}^n$ и вычисляем градиент $g_0 = Q'(x_0) = Dx_0 + c$. Если $g_0 = \mathbb{O}$, то x_0 — точка минимума. Вычисления прекращаются. Иначе полагаем $p_1 = -g_0$.

k**-й шаг.** Пусть имеются x_{k-1}, $g_{k-1} \neq \mathbb{O}$ и p_k. Последовательно вычисляем

$$\alpha_k = \frac{\langle p_k, p_k\rangle}{\langle Dp_k, p_k\rangle}, \tag{11}$$
$$x_k = x_{k-1} + \alpha_k p_k,$$
$$g_k = Q'(x_k) = g_{k-1} + \alpha_k D p_k.$$

Если $g_k = \mathbb{O}$, то x_k — точка минимума. Вычисления прекращаются. В противном случае находим

$$p_{k+1} = \frac{\|p_k\|^2(-g_k) + \|g_k\|^2 p_k}{\|p_k\|^2 + \|g_k\|^2}.$$

Описание метода завершено.

Данный вариант метода сопряжённых градиентов назовём *геометрическим вариантом.*

4°. Для шага α_k в геометрическом варианте метода сопряжённых градиентов наряду с формулой (11) справедливо представление

$$\alpha_k = \operatorname*{argmin}_{\alpha>0} Q(x_{k-1} + \alpha p_k). \tag{12}$$

Проверим это.

Функция $\varphi(\alpha) = Q(x_{k-1} + \alpha p_k)$ является выпуклой и

$$\varphi'(\alpha) = \big\langle Q'(x_{k-1} + \alpha p_k), p_k\big\rangle = \langle g_{k-1} + \alpha D p_k, p_k\rangle =$$
$$= \langle g_{k-1}, p_k\rangle + \alpha\langle Dp_k, p_k\rangle.$$

При этом согласно (9), (3), (4) и определению β_k

$$\langle g_{k-1}, p_k\rangle = \langle g_{k-1}, \beta_k s_k\rangle = \big\langle g_{k-1}, \beta_k(-g_{k-1} + b_{k-1}s_{k-1})\big\rangle =$$
$$= -\beta_k\langle g_{k-1}, g_{k-1}\rangle = -\langle p_k, p_k\rangle,$$

так что

$$\varphi'(\alpha) = -\langle p_k, p_k\rangle + \alpha\langle Dp_k, p_k\rangle.$$

Отсюда следует, что $\varphi'(0) < 0$ и $\varphi'(\alpha) = 0$ только при $\alpha = \alpha_k$. Это гарантирует справедливость формулы (12).

В силу (10) шаг t_k вида (1) в основном варианте метода сопряжённых градиентов допускает аналогичное представление

$$t_k = \operatorname*{argmin}_{t>0} Q(x_{k-1} + ts_k).$$

Таким образом, в обоих вариантах неявно присутствует одномерная минимизация (линейный поиск). В следующих пунктах будет описан вариант метода сопряжённых градиентов без точного линейного поиска.

5°. Параллельно с последовательностью $\{x_k\}$, построенной с помощью основного варианта метода сопряжённых градиентов, рассмотрим ещё одну последовательность $\{\hat{x}_k\}$ вида

$$\hat{x}_k = \hat{x}_{k-1} + \hat{t}_k \hat{s}_k, \quad k = 1, 2, \ldots,$$

где $\hat{x}_0 = x_0$ и $\hat{s}_k = s_k$ при всех $k = 1, 2, \ldots$ В качестве шага $\hat{t}_k$ возьмём произвольное положительное число (например, обеспечивающее неравенство $Q(\hat{x}_{k-1} + \hat{t}_k \hat{s}_k) < Q(\hat{x}_{k-1})$). Выясним, как связаны последовательности $\{x_k\}$ и $\{\hat{x}_k\}$.

Введём обозначения:

$$\begin{gathered}
\theta_k = \hat{t}_k / t_k \quad (\text{так что } \hat{t}_k = \theta_k t_k),\\
\hat{z}_k = \hat{x}_k - \hat{x}_{k-1} = \hat{t}_k \hat{s}_k, \quad z_k = x_k - x_{k-1} = t_k s_k,\\
\hat{r}_k = \hat{g}_k - \hat{g}_{k-1} = D\hat{z}_k, \quad r_k = g_k - g_{k-1} = Dz_k,\\
u_k = \hat{x}_k - x_k, \qquad\qquad v_k = \hat{g}_k - g_k.
\end{gathered}$$

В силу равенства $\hat{x}_0 = x_0$ имеем $u_0 = \mathbb{O}$, $v_0 = \mathbb{O}$. Далее

$$\hat{z}_k = \hat{t}_k \hat{s}_k = \theta_k t_k s_k = \theta_k z_k,$$

$$\hat{r}_k = D\hat{z}_k = \theta_k D z_k = \theta_k r_k,$$

$$u_k = (\hat{x}_{k-1} + \hat{z}_k) - (x_{k-1} + z_k) = u_{k-1} + (1 - \tfrac{1}{\theta_k})\hat{z}_k, \tag{13}$$

$$\begin{aligned}
v_k = \hat{g}_k - g_k &= (\hat{g}_k - \hat{g}_{k-1}) + (\hat{g}_{k-1} - g_{k-1}) - (g_k - g_{k-1}) =\\
&= \hat{r}_k + v_{k-1} - r_k = v_{k-1} + (1 - \tfrac{1}{\theta_k})\hat{r}_k.
\end{aligned} \tag{14}$$

Из последней формулы и равенства $v_0 = \mathbb{O}$ следует, что

$$v_k = \sum_{j=1}^{k} (1 - \tfrac{1}{\theta_j})\hat{r}_j = \sum_{j=1}^{k} (\theta_j - 1) r_j. \tag{15}$$

Умножим обе части соотношения (15) скалярно на $\hat{z}_k = \theta_k z_k$. Получим

$$\langle \hat{g}_k - g_k, \hat{z}_k \rangle = \sum_{j=1}^{k} (\theta_j - 1) \langle r_j, \hat{z}_k \rangle. \tag{16}$$

Согласно (4)

$$\langle g_k, \hat{z}_k\rangle = \theta_k\langle g_k, z_k\rangle = \theta_k t_k\langle g_k, s_k\rangle = 0$$

и при $j < k$

$$\langle r_j, \hat{z}_k\rangle = \theta_k\langle Dz_j, z_k\rangle = \theta_k t_j t_k\langle Ds_j, s_k\rangle = 0.$$

Формула (16) принимает вид

$$\langle \hat{g}_k, \hat{z}_k\rangle = (\theta_k - 1)\langle r_k, \hat{z}_k\rangle = \tfrac{\theta_k - 1}{\theta_k}\langle \hat{r}_k, \hat{z}_k\rangle.$$

Учитывая, что $\langle \hat{r}_k, \hat{z}_k\rangle = \langle D\hat{z}_k, \hat{z}_k\rangle > 0$, получаем

$$\frac{\theta_k - 1}{\theta_k} = \frac{\langle \hat{g}_k, \hat{z}_k\rangle}{\langle \hat{r}_k, \hat{z}_k\rangle}.$$

Теперь рекуррентные соотношения (13) и (14) можно переписать так:

$$u_k = u_{k-1} + \frac{\langle \hat{g}_k, \hat{z}_k\rangle}{\langle \hat{r}_k, \hat{z}_k\rangle}\hat{z}_k,\ k = 1, 2, \ldots; \quad u_0 = \mathbb{O};$$
$$v_k = v_{k-1} + \frac{\langle \hat{g}_k, \hat{z}_k\rangle}{\langle \hat{r}_k, \hat{z}_k\rangle}\hat{r}_k,\ k = 1, 2, \ldots; \quad v_0 = \mathbb{O}.$$

При этом

$$x_k = \hat{x}_k - u_k, \quad g_k = \hat{g}_k - v_k.$$

6°. Переходим к описанию варианта метода сопряжённых градиентов без точного линейного поиска.

Нулевой шаг. Берём произвольное начальное приближение $\hat{x}_0 \in \mathbb{R}^n$. Вычисляем градиент $\hat{g}_0 = Q'(\hat{x}_0) = D\hat{x}_0 + c$. Если $\hat{g}_0 = \mathbb{O}$, то $\hat{x}_0$ — точка минимума. Вычисления прекращаются. Иначе полагаем

$$x_0 = \hat{x}_0,\ g_0 = \hat{g}_0,\ s_1 = -g_0,\ u_0 = \mathbb{O},\ v_0 = \mathbb{O}.$$

k-й шаг. Пусть имеются

$$\hat{x}_{k-1},\ x_{k-1}, \hat{g}_{k-1} \neq \mathbb{O},\ g_{k-1} \neq \mathbb{O},\ s_k,\ u_{k-1},\ v_{k-1}.$$

Произвольно выбираем шаг $\hat{t}_k > 0$. Находим $\hat{z}_k = \hat{t}_k s_k$,

$$\hat{x}_k = \hat{x}_{k-1} + \hat{z}_k,$$
$$\hat{g}_k = Q'(\hat{x}_k) = \hat{g}_{k-1} + D\hat{z}_k.$$

Если $\hat{g}_k = \mathbb{O}$, то $\hat{x}_k$ — точка минимума. Вычисления прекращаются.

Пусть $\hat{g}_k \neq \mathbb{O}$. Тогда вычисляем

$$\hat{r}_k = \hat{g}_k - \hat{g}_{k-1}, \quad \hat{c}_k = \frac{\langle \hat{g}_k, \hat{z}_k \rangle}{\langle \hat{r}_k, \hat{z}_k \rangle},$$
$$u_k = u_{k-1} + \hat{c}_k \hat{z}_k, \quad v_k = v_{k-1} + \hat{c}_k \hat{r}_k,$$
$$x_k = \hat{x}_k - u_k, \quad g_k = Q'(x_k) = \hat{g}_k - v_k.$$

Если $g_k = \mathbb{O}$, то x_k — точка минимума. Вычисления прекращаются. Иначе полагаем $s_{k+1} = -g_k + b_k s_k$, где b_k имеет вид (2).

Описание метода завершено.

Из описания видно, что в методе сопряжённых градиентов без точного линейного поиска наряду с последовательностью $\{\hat{x}_k\}$ строится стандартная последовательность $\{x_k\}$ метода сопряжённых градиентов. Это гарантирует сходимость метода не более чем за n шагов.

7°. Следуя Соренсену, отметим, что коэффициент b_k допускает другое представление:
$$b_k = \frac{\langle g_k, \hat{r}_k \rangle}{\langle s_k, \hat{r}_k \rangle}. \tag{17}$$
Действительно, согласно (4) и (5)

$$\begin{aligned}
\langle g_k, \hat{r}_k \rangle &= \theta_k \langle g_k, g_k - g_{k-1} \rangle = \theta_k \langle g_k, g_k \rangle, \\
\langle s_k, \hat{r}_k \rangle &= \theta_k \langle s_k, g_k - g_{k-1} \rangle = -\theta_k \langle s_k, g_{k-1} \rangle = \\
&= -\theta_k \langle -g_{k-1} + b_{k-1} s_{k-1}, g_{k-1} \rangle = \theta_k \langle g_{k-1}, g_{k-1} \rangle.
\end{aligned}$$

Отсюда и из (2) следует (17).

8°. История метода сопряжённых градиентов за период с 1948 по 1976 годы (с обширной аннотированной библиографией) представлена в обзорной статье [2]. Современному состоянию в этой области посвящена книга [3].

ЛИТЕРАТУРА

1. Малозёмов В. Н. *О методе сопряжённых градиентов* / Семинар «DHA & CAGD». Избранные доклады. 28 апреля 2012 г. (http://dha.spb.ru/reps12.shtml#0428)

2. Golub G. H., O'Leary D. P. *Some History of the Conjugate Gradient and Lanczos Algorithms: 1948–1976* // SIAM Rev. 1989. Vol. 31, No. 1, pp. 50–102.

3. Pytlak R. *Conjugate Gradient Algorithms in Nonconvex Optimization.* Berlin: Springer, 2009.