МЕТОДЫ ОПТИМИЗАЦИИ

ОБ ОДНОМ ПОДХОДЕ К ПОСТРОЕНИЮ ОПТИМАЛЬНЫХ МЕТОДОВ МИНИМИЗАЦИИ ГЛАДКИХ ВЫПУКЛЫХ ФУНКЦИЙ

Нестеров Ю. Е.

(Москва)

1. ВВЕДЕНИЕ

В статье рассматриваются итеративные методы решения экстремальной задачи

$$\min\{f(x)\,|\,x\in H\},\tag{1}$$

где H — гильбертово пространство; $f(x)$ — выпуклая функция из класса $C^{1,1}(H)$, т. е. такая, что для любых x, $y\in H$, $x\neq y$ отношение $\|f'(x)-f'(y)\|/\|x-y\|$ ограничено сверху некоторой константой $L=L(f)>0$. Будем говорить, что функция $f(x)$ является сильно выпуклой с константой $m(f)\geqslant 0$, если для любых x, $y\in H$ справедливо неравенство

$$f(x)\geqslant f(y)+\langle f'(y), x-y\rangle+\frac{m(f)}{2}\|x-y\|^2.\tag{2}$$

В частности, сильная выпуклость при $m(f)=0$ означает обычную выпуклость функции $f(\cdot)$. В дальнейшем $\mathscr{F}_m{}^L(H)$ — множество функций $f(\cdot)$, которые принадлежат классу $C^{1,1}(H)$ с константами $L(f)\leqslant L$ и являются сильно выпуклыми с константами $m(f)\geqslant m$.

Важной характеристикой любого итеративного метода решения задачи (1) является глобальная оценка его скорости сходимости на классе $\mathscr{F}_m{}^L(H)$. Под глобальной оценкой скорости сходимости метода $\mathfrak{m}$ на этом классе мы понимаем некоторую функцию $\xi_{\mathfrak{m}}(s, m, L, k)$ такую, что для любой функции $f(\cdot)\in\mathscr{F}_m{}^L(H)$ при любом $k\geqslant 0$ выполняется неравенство

$$f(x_k)-f(x^*)\leqslant C_1\xi_{\mathfrak{m}}(x_0-x^*, m, L, k),$$

где x_k — результат k-й итерации метода $\mathfrak{m}$, стартовавшего из точки x_0; x^* — решение задачи (1); C_1 — положительная константа. Чем выше скорость стремления к нулю функции $\xi_{\mathfrak{m}}(s, m, L, k)$ при $k\to\infty$, тем меньше максимальное число итераций, необходимых методу $\mathfrak{m}$ для получения ε — приближения к оптимальному значению f^* произвольной функции $f(\cdot)$ из класса $\mathscr{F}_m{}^L(H)$.

Однако порядок убывания функции $\xi_{\mathfrak{m}}(s, m, L, k)$ не может быть сколь угодно большим. Точную нижнюю оценку скорости уменьшения этой функции дает следующая теорема (приводим ее в удобной для дальнейшего форме).

Теорема 1. [1]. *Пусть метод $\mathfrak{M}$ строит минимизирующую последовательность точек $\{x_k\}_{k=0}^{\infty}$. Причем на k-й итерации метода значение функции $f(\cdot)$ и ее градиент вычисляются в N_k вспомогательных точках, $0\leqslant N_k\leqslant\infty$, выпуклая оболочка которых имеет афинную размерность, не превосходящую l. Тогда независимо от способа учета накопленной информации и правил формирования точек x_k, применяемых в методе $\mathfrak{M}$, при всех $k\geqslant 0$ справедлива оценка*

$$\xi_{\mathfrak{m}}(s, m, L, k)\geqslant C_2\|s\|^2\min\{((l+1)k)^{-2}, e^{-\sqrt{\frac{m}{L}}C_3k(l-1)}\},\tag{3}$$

где $C_2=C_2(m, L)$; C_3—абсолютная константа.

Будем говорить, что метод $\mathfrak{M}$ обладает оптимальной оценкой скорости сходимости на классе $\mathcal{F}_m^L(H)$, если функция $\xi_{\mathfrak{m}}(s, m, L, k)$ допускает оценку сверху вида правой части неравенства (3) с измененными значениями констант C_2 и C_3. Если, кроме того, в методе $\mathfrak{m}$ числа N_k, $0 \leqslant k < \infty$, равномерно ограничены, то этот метод назовем оптимальным на классе $\mathcal{F}_m^L(H)$.

Следует отметить, что свойство оптимальности выдвигает довольно серьезные требования к скоростным характеристикам методов минимизации. Так, до сих пор ни для одного из традиционных методов решения задачи (1) это свойство еще не доказано. Более того, в [1] удалось установить, что некоторые распространенные методы (метод градиентного спуска, сопряженных градиентов Флетчера — Ривса, сопряженных градиентов Полака — Рибьерра, проективный метод Зойтендейка и переменной метрики Давидона — Флетчера — Пауэлла) оптимальной оценкой скорости сходимости заведомо не обладают.

Первые методы, обладающие на классе $\mathcal{F}_m^L(H)$ оптимальной оценкой скорости сходимости, были получены в [1, 2]. Однако эти методы не были оптимальными на $\mathcal{F}_m^L(H)$ в вышеуказанном смысле, так как на каждой итерации в них требовалось решать вспомогательную задачу минимизации: двумерную в [1] и одномерную в [2]. Первый метод, оптимальный на классе $\mathcal{F}_m^L(H)$, был предложен в [3]. На базе метода [3] в [4] был получен достаточно представительный класс оптимальных на $\mathcal{F}_m^L(H)$ алгоритмов. Отметим, однако, что обоснование методов [3, 4] было чисто алгебраическим, лишенным какой-либо геометрической интерпретации. Дальнейший анализ этих методов позволил выявить их геометрическую структуру, выработать общую методику построения и обоснования оптимальных на $\mathcal{F}_m^L(H)$ алгоритмов. Изложению этого подхода и посвящена настоящая статья.

Остановимся вкратце на содержании работы. В разд. 2 рассмотрены общие принципы построения оптимальных алгоритмов исходя из предложенного способа оценки их скорости сходимости. Подробно исследуется одна из возможных схем таких методов. В разд. 3 проанализирована работа одной из версий методов сопряженных градиентов на квадратичных задачах. Показано, как можно построить модификацию этого метода, обладающую на классе $\mathcal{F}_m^L(H)$ оптимальной оценкой скорости сходимости.

2. ПРИНЦИП ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ МЕТОДОВ

Прежде всего остановимся на способе получения оценок скорости сходимости, который будет использоваться в данной работе.

Пусть $\{x_k\}_{k=0}^{\infty}$ — последовательность точек, вырабатываемая методом минимизации $\mathfrak{M}$: $\psi_k(x)$, $x \in H$ — последовательность функций таких, что для любого x из H $\lim\limits_{k\to\infty} \psi_k(x) = 0$. Предположим, что при любом $k \geqslant 0$ и $x \in H$ справедливо неравенство

$$f(x_k) \leqslant f(x) + \psi_k(x). \tag{4}$$

Тогда, очевидно,

$$f(x_k) - f(x^*) \leqslant \psi_k(x^*) \to 0 \text{ при } k \to \infty. \tag{5}$$

Таким образом, если метод $\mathfrak{M}$ обеспечивает выполнение неравенства (4), то в силу (5) оценка стремления к нулю числовой последовательности $\{\psi_k(x^*)\}$ дает глобальную оценку скорости сходимости метода $\mathfrak{M}$.

Способы (4), (5) при соответствующем выборе функций $\psi_k(\cdot)$ позволяют получать оценки скорости сходимости некоторых известных методов, например градиентного спуска. Однако более важным является тот факт, что попытка построить метод минимизации, непосредственно учитывающий структуру неравенства (4), приводит к оптимальным на $\mathcal{F}_m^L(H)$ алгоритмам.

Для упрощения изложения всюду далее будем считать, что параметры m и L класса $\mathcal{F}_m^L(H)$ известны.

Рассмотрим последовательность функций $\{\varphi_k(x)\}_{k=0}^{\infty}$ $x\in H$, задаваемую рекуррентными соотношениями

$$\varphi_0(x)=f(x_0)+\frac{A_0}{2}\|x-x_0\|^2,$$

$$\varphi_{k+1}(x)=(1-\alpha_k)\varphi_k(x)+\alpha_k\Big(f(y_k)+\langle f'(y_k),x-y_k\rangle+$$
$$+\frac{m}{2}\|x-y_k\|^2\Big),\quad k=0,1,\ldots, \tag{6}$$

где y_k — произвольные точки из H; α_k — произвольные числа из отрезка $[0, 1]$; A_0 — положительная константа. Положим

$$\psi_0(x)=\varphi_0(x)-f(x)=f(x_0)-f(x)+\frac{A_0}{2}\|x-x_0\|^2,$$

$$\psi_{k+1}(x)=(1-\alpha_k)\psi_k(x),\quad k=0,1,\ldots.$$

Тогда $\varphi_0(x)=f(x)+\psi_0(x)$. Далее, если $f(\cdot)\in\mathscr{F}_m{}^L(H)$, то по индукции получаем

$$\varphi_{k+1}(x)=(1-\alpha_k)\varphi_k(x)+\alpha_k\left(f(y_k)+\langle f'(y_k),x-y_k\rangle+\frac{m}{2}\|x-y_k\|^2\right)\leqslant$$
$$\leqslant(1-\alpha_k)(f(x)+\psi_k(x))+\alpha_k f(x)=f(x)+(1-\alpha_k)\psi_k(x)=f(x)+\psi_{k+1}(x).$$

Таким образом, при любом $k\geqslant 0$ и произвольном x из H справедливо неравенство

$$\varphi_k(x)\leqslant f(x)+\psi_k(x)=f(x)+\lambda_k\psi_0(x), \tag{7}$$

где $\lambda_0=1$; $\lambda_k=\prod\limits_{j=0}^{k-1}(1-\alpha_j)$, $k=0,1,\ldots$. Следовательно, для того чтобы воспользоваться способом оценок (4), (5), достаточно выбрать точки x_k, y_k и числа $\alpha_k\in[0, 1]$ так, чтобы при любом $k\geqslant 0$ и $x\in H$ выполнялось неравенство

$$f(x_k)\leqslant\varphi_k(x). \tag{8}$$

В этом случае в силу (7) будет справедлива оценка

$$f(x_k)-f^*\leqslant\lambda_k\psi_0(x^*)=\lambda_k\left[f(x_0)-f^*+\frac{A_0}{2}\|x_0-x^*\|^2\right]. \tag{9}$$

Покажем, как можно добиться выполнения неравенства (8). Заметим, что в силу формул пересчета (6) функции $\varphi_k(x)$ допускают следующее каноническое представление

$$\varphi_k(x)=\varphi_k^*+\frac{A_k}{2}\|x-v_k\|^2,$$

где φ_k^* — минимальное значение функции $\varphi_k(x)$ на H; v_k — ее точка минимума; числа A_k пересчитываются по формуле $A_{k+1}=(1-\alpha_k)A_k+\alpha_k m$, $k=0, 1,\ldots$.

Лемма 1. *Если при некотором $k\geqslant 0$ $f(x_k)\leqslant\varphi_k^*$, то при любом $y_k\in H$ справедливо неравенство*

$$f(y_k)+\langle f'(y_k),\ \alpha_k v_k+(1-\alpha_k)x_k-y_k\rangle-\frac{\alpha_k^2}{2(1-\alpha_k)A_k}\|f'(y_k)\|^2\leqslant\varphi_{k+1}^*.$$

Доказательство. Для упрощения выкладки опустим индексы у всех символов, относящихся к k-му шагу и пометим черточкой сверху все символы, относящиеся к $(k+1)$-му шагу.

Используя предположение леммы и выпуклость функции, получаем

$$\bar\varphi^*=(1-\alpha)\varphi(\bar v)+\alpha\left(f(y)+\langle f'(y),\bar v-y\rangle+\frac{m}{2}\|\bar v-y\|^2\right)\geqslant$$
$$\geqslant(1-\alpha)\varphi^*+\frac{(1-\alpha)A}{2}\|\bar v-v\|^2+\alpha(f(y)+\langle f'(y),\bar v-y\rangle)\geqslant$$
$$\geqslant(1-\alpha)f(x)+\frac{(1-\alpha)A}{2}\|\bar v-v\|^2+\alpha(f(y)+\langle f'(y),\bar v-y\rangle)\geqslant$$

$$\geqslant f(x) + \langle f'(y), \alpha v + (1-\alpha) x - y\rangle + \frac{(1-\alpha) A}{2} \| \bar{v} - v \|^2 +$$

$$+ \alpha \langle f'(y), \bar{v} - v\rangle \geqslant f(y) + \langle f'(y), \alpha v + (1-\alpha) x - y\rangle - \frac{\alpha^2}{2(1-\alpha) A} \| f'(y) \|^2.$$

Лемма доказана.

Как известно, для любой функции $f(\cdot)$ из класса $\mathscr{F}_m{}^L(H)$ справедливо неравенство

$$f(x) \leqslant f(y) + \langle f'(y), x - y\rangle + \frac{L}{2} \| x - y \|^2,$$

где x, y — произвольные точки из H. В частности, для $\bar{x} = y - f'(y)/L$ имеем

$$f(\bar{x}) \leqslant f(y) - \frac{1}{2L} \| f'(y) \|^2.$$

Сопоставив это неравенство с утверждением леммы 1, нетрудно заметить, что для выполнения неравенства $f(x_{k+1}) \leqslant \varphi_{k+1}$ достаточно выбрать точки x_{k+1}, y_k и число α_k из условий

$$L\alpha_k{}^2 = (1-\alpha_k) A_k,$$

$$\langle f'(y_k),\ \alpha_k v_k + (1-\alpha_k) x_k - y_k\rangle \geqslant 0, \tag{10}$$

$$f(x_{k+1}) \leqslant f(y_k) - \frac{1}{2L} \| f'(y_k) \|^2. \tag{11}$$

В частности, можно взять

$$y_k = (1-\alpha_k) x_k + \alpha_k v_k, \tag{12}$$

$$x_{k+1} = y_k - \frac{1}{L} f'(y_k). \tag{13}$$

В этом случае приходим к методу $\mathfrak{M}(m, L, x_0, A)$. Суть метода состоит в следующем.

1. Полагаем $A_0 = A$, где $A > 0$, $A \geqslant m$.

$$\varphi_0(x) = f(x_0) + \frac{A_0}{2} \| x - x_0 \|^2.$$

2. k-я итерация метода, $k \geqslant 0$, имеет вид:

а) вычисляем $\alpha_k > 0$ из уравнения

$$\alpha_k{}^2 L = (1-\alpha_k) A_k;$$

б) полагаем

$$v_k = \operatorname{argmin} \{\varphi_k(x) \mid x \in H\},$$

$$y_k = \alpha_k v_k + (1-\alpha_k) x_k,$$

$$A_{k+1} = (1-\alpha_k) A_k + \alpha_k m,$$

$$x_{k+1} = y_k - \frac{1}{L} f'(y_k),$$

$$\varphi_{k+1}(x) = (1-\alpha_k) \varphi_k(x) + \alpha_k \left(f(y_k) + \langle f'(y_k), x - y_k\rangle + \frac{m}{2} \| x - y_k \|^2 \right).$$

Оценим скорость сходимости метода.

Теорема 2. *Если метод $\mathfrak{M}(m, L, x_0, A)$ с $A > 0$, $A \geqslant m$, применяется для минимизации функции $f(\cdot)$ из класса $\mathscr{F}_m{}^L(H)$, то при любом $k \geqslant 0$ в случае $m > 0$ справедлива оценка*

$$f(x_k) - f^* \leqslant \frac{4m (f(x_0) - f^* + 0{,}5 A \| x_0 - x^* \|^2)}{(\sqrt{A} + \sqrt{m})^2 Q^{2k} + (\sqrt{A} - \sqrt{m})^2 Q^{-2k} - 2(A - m)}, \tag{14}$$

где $Q = 1 + 0{,}5\sqrt{q}$, $q = mL^{-1}$.

В случае $m = 0$ при условии существования точки x^* оценка (14) *переходит в следующую*

$$f(x_k) - f^* \leqslant \frac{4L (f(x_0) - f^* + 0{,}5 A \| x_0 - x^* \|^2)}{(2\sqrt{L} + k\sqrt{A})^2}. \tag{15}$$

Доказательство. В силу построения метода $\mathfrak{M}(m, L, x_0, A)$ для получения оценки его скорости сходимости можно использовать неравенство (9). А это значит, что достаточно оценить скорость стремления к нулю скалярных величин λ_k, где $\lambda_0=1, \lambda_k = \prod_{j=0}^{k-1}(1-\alpha_j)$, $k=1, 2, \ldots$; числа α_j определяются из уравнения $L\alpha_j^2=(1-\alpha_j)A_j$, а величины A_j пересчитываются по формуле

$$A_{j+1}=(1-\alpha_j)A_j+\alpha_j m.$$

Нетрудно убедиться в справедливости формулы

$$A_k=m+\lambda_k(A_0-m).$$

Поэтому, учитывая, что $\lambda_{k+1}=(1-\alpha_k)\lambda_k$, получаем уравнение для величин λ_k

$$L\left(\frac{\lambda_k-\lambda_{k+1}}{\lambda_k}\right)^2=\frac{\lambda_{k+1}}{\lambda_k}(m+\lambda_k(A-m)). \tag{16}$$

Положим $\mu_k=\lambda_k^{-1/2}$. Тогда уравнение (16) примет вид

$$1-\frac{\mu_k^2}{\mu_{k+1}^2}=\frac{1}{\mu_{k+1}}\left(q\mu_k^2+\frac{A-m}{L}\right)^{1/2},$$

т. е. $(\mu_{k+1}+\mu_k)(\mu_{k+1}-\mu_k)=\mu_{k+1}(q\mu_k^2+L^{-1}(A-m))^{1/2}$.

Отсюда в силу возрастания величин μ_k следует, что

$$\mu_{k+1}\geqslant\mu_k+\frac{1}{2}(q\mu_k^2+L^{-1}(A-m))^{1/2}. \tag{17}$$

Воспользовавшись неравенством (17) и начальным условием $\mu_0=1$, по индукции нетрудно показать, что при всех $k\geqslant 0$ (в случае $m>0$) справедлива оценка

$$\mu_k\geqslant\frac{1}{2}m^{-1/2}\left((\sqrt{A}+\sqrt{m})\left(1+\frac{1}{2}\sqrt{q}\right)^k-(\sqrt{A}-\sqrt{m})\left(1+\frac{1}{2}\sqrt{q}\right)^{-k}\right). \tag{18}$$

В случае $m=0$ справедлив предельный (при $m\to 0$) вариант оценки (18)

$$\mu_k\geqslant 1+\frac{1}{2}k\left(\frac{A}{L}\right)^{1/2}. \tag{19}$$

Осталось заметить, что оценки (18), (19) в силу соотношения $\lambda_k=\mu_k^{-2}$ переходят в оценки скорости стремления к нулю величин λ_k. Подставляя эти оценки в (9), получаем (14), (15). Теорема доказана.

Следствие. *При всех $m\geqslant 0$ метод $\mathfrak{m}(m, L, x_0, L)$ является оптимальным* на классе функций $\mathscr{F}_m^L(H)$.

Доказательство. Из неравенства (18) легко получить, что

$$\mu_k\geqslant 0{,}5\left(\exp\left(\frac{k}{3}\sqrt{q}\right)+\frac{k}{2}\sqrt{\frac{A}{L}}\right)\geqslant$$
$$\geqslant 0{,}5\max\left\{\exp\left(\frac{k}{3}\sqrt{q}\right),\ 1+\frac{k}{2}\sqrt{\frac{A}{L}}\right\}.$$

Следовательно, для величин λ_k справедливо неравенство

$$\lambda_k=\mu_k^{-2}\leqslant 4\min\left\{e^{-\frac{2}{3}\sqrt{q}k},\ \frac{4}{(k+2)^2}\right\}.$$

Подставляя это неравенство в (9) и учитывая, что $f(x_0)-f^*\leqslant 0{,}5L\|x_0-x^*\|^2$, получаем

$$f(x_k)-f^*\leqslant 4L\|x_0-x^*\|^2\min\left\{e^{-\frac{2}{3}\sqrt{q}k},\ \frac{4}{(k+2)^2}\right\}.$$

Таким образом, метод $\mathfrak{m}(m, L, x_0, L)$ обладает оптимальной оценкой скорости сходимости. Осталось заметить, что на каждой итерации этот метод лишь один раз вычисляет значение функции и градиент.

Приведем алгоритмическую форму метода $\mathfrak{m}(m, L, x_0, A)$, в которой учитывается возможность канонического представления функций $\varphi_k(x)$.

1. Полагаем $v_0=x_0$, $A_0=A$, где $A>0$, $A\geqslant m$.
2. k-я итерация метода $k\geqslant 0$ имеет вид:

а) вычисляем $\alpha_k>0$ из уравнения

$$\alpha_k^2 L=(1-\alpha_k)A_k;$$

б) полагаем

$$y_k=(1-\alpha_k)x_k+\alpha_k v_k,$$

$$x_{k+1}=y_k-\frac{1}{L}f'(y_k),$$

$$A_{k+1}=(1-\alpha_k)A_k+\alpha_k m,$$

$$v_{k+1}=(1-\alpha_k)\frac{A_k}{A_{k+1}}v_k+\alpha_k\frac{m}{A_{k+1}}y_k-\frac{\alpha_k}{A_{k+1}}f'(y_k).$$

При выводе метода использованы наиболее простые способы (12), (13), обеспечивающие выполнение условий (10), (11). Однако существует много других способов выбора точек x_k, y_k, удовлетворяющих соотношениям (10), (11). В частности, можно взять

$$y_k=\operatorname{argmin}\{f(x)\,|\,x=x_k+\lambda(v_k-x_k),\quad \lambda\in R\},$$

$$x_{k+1}=\operatorname{argmin}\{f(x)\,|\,x=y_k-\lambda f'(y_k),\quad \lambda\geqslant 0\}.$$

В этом случае приходим к методу $\mathfrak{m}_1(m, L, x_0, A)$, определяемому следующим образом.

1. Полагаем $v_0=x_0$, $A_0=A$, где $A>0$, $A\geqslant m$.
2. k-я итерация метода $k\geqslant 0$ имеет вид:

а) вычисляем $\alpha_k>0$ из уравнения

$$\alpha_k^2 L=(1-\alpha_k)A_k;$$

б) полагаем

$$y_k=\operatorname{argmin}\{f(x)\,|\,x=x_k+\lambda(v_k-x_k),\quad \lambda\in R\},$$
$$x_{k+1}=\operatorname{argmin}\{f(x)\,|\,x=y_k-\lambda f'(y_k),\quad \lambda\geqslant 0\},$$
$$A_{k+1}=(1-\alpha_k)A_k+\alpha_k m,$$

$$v_{k+1}=(1-\alpha_k)\frac{A_k}{A_{k+1}}v_k+\alpha_k\frac{m}{A_{k+1}}y_k-\frac{\alpha_k}{A_{k+1}}f'(y_k).$$

Не будем выписывать схемы других методов, соответствующих различным комбинациям стратегий удовлетворения условий (10), (11). Отметим лишь, что для любого метода такого типа удается получить оптимальную оценку скорости сходимости. Причем обоснование этой оценки сводится, как и у метода $\mathfrak{m}$, к анализу скорости убывания величин λ_k.

3. АНАЛИЗ ПОВЕДЕНИЯ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ

Методы сопряженных градиентов относятся к числу наиболее распространенных методов безусловной минимизации. В настоящее время известно более ста методов такого типа. Однако ни для одного из них пока не удалось получить оптимальные оценки скорости сходимости при решении задачи (1). Более того, в [1] построены специальные примеры, на которых отдельные версии методов сопряженных градиентов работают не лучше, чем простейший метод градиентного спуска. В то же время, как известно, методы сопряженных градиентов являются оптимальными методами для решения задачи минимизации квадратичной функции. В связи с этим складывается впечатление, что при переносе схем методов сопряженных градиентов с квадратичного случая на общий нелинейный учитываются не все особенности такого переноса.

В этом разделе проанализируем поведение одной версии метода сопряженных градиентов при решении квадратичной задачи с помощью подхода, описанного в предыдущем разделе. Будет показано, как можно осуществить перенос этого метода на общий нелинейный случай с сохранением оптимальной оценки скорости сходимости.

Рассмотрим следующий метод $\mathfrak{m}_2(x_0)$.

1. Полагаем $y_{-2}=y_{-1}=x_0$.

2. k-я итерация метода, $k\geqslant 0$, имеет следующий вид.

Полагаем

$$y_k=\operatorname{argmin}\{f(x)\,|\,x=x_k+\lambda(y_{k-2}-x_k),\quad \lambda\in R\}, \tag{20}$$

$$x_{k+1}=\operatorname{argmin}\{f(x)\,|\,x=y_k-\lambda f'(y_k),\quad \lambda\geqslant 0\}. \tag{21}$$

Применим метод $\mathfrak{m}_2(x_0)$ для решения задачи

$$\min\left\{f(x)=\frac{1}{2}\langle B(x-x^*),x-x^*\rangle\,|\,x\in H\right\}, \tag{22}$$

где B — линейный неотрицательно определенный самосопряженный оператор. Обозначим

$$\mathfrak{L}_k=\left\{x\,|\,x=x_0+\sum_{j=0}^{k-1}\lambda_j f'(y_j),\ \lambda_j\in R\right\}.$$

Теорема 3. *Пусть последовательности точек $\{x_k\}$, $\{y_k\}$ сформированы методом $\mathfrak{m}(x_0)$, примененным к задаче* (22). *Тогда при любом $k\geqslant 0$* **имеет** *место равенство*

$$y_k=\operatorname{argmin}\{f(x)\,|\,x\in\mathfrak{L}_k\}. \tag{23}$$

Если же $f(x)\in\mathcal{F}_m{}^L(H)$, то при любом $k\geqslant 0$ справедливы оценки

$$f(x_k)-f^*\leqslant\frac{4m\left(f(x_0)-f^*+\frac{A}{2}\|x_0-x^*\|^2\right)}{(\sqrt{A}+\sqrt{m})\,Q^{2k}+(\sqrt{A}-\sqrt{m})\,Q^{-2k}-2(A-m)} \tag{24}$$

в случае $m>0$ и

$$f(x_k)-f^*\leqslant\frac{4L\left(f(x_0)-f^*+\frac{A}{2}\|x_0-x^*\|^2\right)}{(2\sqrt{L}+k\sqrt{A})^2}. \tag{25}$$

в случае $m=0$, где $Q=1+0{,}5\sqrt{q}$; $q=m/L$; A — любое положительное число, $A\geqslant m$.

Доказательство. Соотношение (23) докажем по индукции. Из схемы метода $\mathfrak{m}_2(x_0)$ следует, что

$$y_0\equiv x_0\equiv\operatorname{argmin}\{f(x)\,|\,x\in\mathfrak{L}_0\equiv\{x_0\}\},$$

$$y_1\equiv x_1=\operatorname{argmin}\{f(x)\,|\,x\in\mathfrak{L}_1\}.$$

Предположим, что (23) выполнено при всех k, $0\leqslant k\leqslant m$, где $m\geqslant 1$. Из правил (20), (21) выбора точек y_k, x_k вытекает

$$\langle f'(x_{m+1}),\,f'(y_m)\rangle=0, \tag{26}$$

$$\langle f'(y_{m+1}),\,x_{m+1}-y_{m-1}\rangle=0. \tag{27}$$

Кроме того, из (20) и из квадратичности функции $f(x)$ следует

$$f'(y_{m+1})=(1+\lambda')f'(x_{m+1})-\lambda' f'(y_{m-1}). \tag{28}$$

Сопоставляя (26)—(28) и пользуясь индуктивным предположением, получаем

$$\langle f'(y_{m+1}),f'(y_m)\rangle=-\lambda'\langle f'(y_{m-1}),f'(y_m)\rangle=0,$$

$$\langle f'(y_{m+1}),\,y_m-y_{m-1}\rangle=\langle f'(y_{m+1}),\,x_{m+1}+\lambda'' f'(y_m)-y_{m-1}\rangle=0.$$

Следовательно,

$$y_{m+1}=\operatorname{argmin}\{f(x)\,|\,x=y_m+\lambda' f'(y_m)+\lambda''(y_m-y_{m-1}),\ \lambda'\lambda''\in R\}. \tag{29}$$

В свою очередь, пользуясь индуктивным предположением и соотношением (29), можно стандартным способом доказать справедливость равенства (23) для точки y_{m+1} (см. [5, с. 70–71]). Таким образом, первое утверждение теоремы доказано.

Для доказательства второго утверждения введем последовательность вспомогательных функций $\{\varphi_k(x)\}$, которые определяются рекуррентными соотношениями

$$\varphi_0(x)=f(x_0)+0{,}5A_0\|x-x_0\|^2, \tag{30}$$

$$\varphi_{k+1}(x)=(1-\alpha_k)\varphi_k(x)+\alpha_k\left(f(y_k)+\langle f'(y_k), x-y_k\rangle+\frac{m}{2}\|x-y_k\|^2\right). \tag{31}$$

В формулах (30), (31) $A_0=A$ — любое положительное число, не меньшее, чем m, числа $\alpha_k>0$ находится из уравнения

$$\alpha_k^2L=(1-\alpha_k)A_k,$$

а числа A_k пересчитываются по формуле

$$A_{k+1}=(1-\alpha_k)A_k+\alpha_k m.$$

Сразу же отметим, что при любом $k\geqslant 0$ точки x_k и точки $v_k=\operatorname{argmin}\{\varphi_k\times\times(x)\,|\,x\in H\}$ лежат в многообразии $\mathfrak{L}_k$. С помощью рассуждений, приведенных в начале разд. 2, нетрудно установить справедливость неравенства

$$\varphi_k=\min\{\varphi_k(x)\,|\,x\in H\}\leqslant f^*+\lambda_k\left(f(x_0)-f^*+\frac{A}{2}\|x_0-x^*\|^2\right),$$

где $\lambda_0=1$,

$$\lambda_k=\prod_{j_0}^{k-1}(1-\alpha_j).$$

Покажем, что в методе $\mathfrak{m}_2 f(x_k)\leqslant\varphi_k$. Действительно, $f(x_0)=\varphi_0$. Далее по лемме 1, если $f(x_k)\leqslant\varphi_k$, то

$$\varphi_{k+1}\geqslant f(y_k)+\langle f'(y_k),\quad \alpha_k v_k+(1-\alpha_k)x_k-y_k\rangle-$$
$$-\frac{\alpha_k^2}{2(1-\alpha_k)A_k}\|f'(y_k)\|^2=f(y_k)-\frac{1}{2L}\|f'(y_k)\|^2+\langle f'(y_k),\alpha_k v_k+$$
$$+(1-\alpha_k)x_k-y_k\rangle\geqslant f(x_{k+1})+\langle f'(y_k),\alpha_k v_k+(1-\alpha_k)x_k-y_k\rangle.$$

Но $y_k=\operatorname{argmin}\{f(x)\,|\,x\in\mathfrak{L}_k\}$. Поэтому $\langle f'(y_k),\ z-y_k\rangle=0$ для любого $z\in\mathfrak{L}_k$, в частности для $z=\alpha_k v_k+(1-\alpha_k)x_k$.

Таким образом, установлено, что для рассматриваемого метода при всех $k\geqslant 0$ справедливо неравенство

$$f(x_k)\leqslant\varphi_k\leqslant f^*+\lambda_k(f(x_0)-f^*+0{,}5A\|x_0-x^*\|^2).$$

Осталось воспользоваться оценками величин λ_k, которые были получены при доказательстве теоремы 2. Теорема доказана.

При выводе оценки скорости сходимости метода $\mathfrak{m}_2(x_0)$ характеристическое свойство (23) метода сопряженных градиентов использовалось лишь для доказательства неравенства

$$\langle f'(y_k),\ \alpha_k v_k+(1-\alpha_k)x_k-y_k\rangle\geqslant 0. \tag{32}$$

Однако если метод $\mathfrak{m}_2(x_0)$ применять для минимизации неквадратичных функций, то гарантировать выполнение равенства (23) заведомо не удастся. В этом случае соотношение (31) может оказаться нарушенным, а значит, может ухудшиться и оценка скорости сходимости метода. Поэтому при переносе метода $\mathfrak{m}_2(x_0)$ на общий случай необходимо принять специальные меры, обеспечивающие выполнение неравенства (31). Например, можно модифицировать этот метод, соединив его с методом $\mathfrak{m}_1$. Одна из возможных комбинаций методов $\mathfrak{m}_1$ и $\mathfrak{m}_2$ дает метод $\mathfrak{m}_3(m, L, x_0, A)$, определяемый следующим образом.

1. Полагаем $y_{-2}=y_{-1}=v_0=x_0$, $A_0=A$, где $A>0$, $A\geqslant m$.
2. k-я итерация, $k\geqslant 0$, метода имеет вид:
а) полагаем

$$\bar{y}_k=\operatorname{argmin}\{f(x)\,|\,x=x_k+\lambda(y_{k-2}-x_k),\quad \lambda\in R\},$$

$$y_k=\operatorname{argmin}\{f(x)\,|\,x=\bar{y}_k+\lambda(v_k-\bar{y}_k),\quad \lambda\in R\};$$

б) вычисляем $\alpha_k>0$ из уравнения

$$\alpha_k^2L=(1-\alpha_k)A_k;$$

в) полагаем

$$x_{k+1}=\operatorname{argmin}\{f(x)\,|\,x=y_k-\lambda f'(y_k),\quad \lambda\geqslant 0\},$$

$$A_{k+1}=(1-\alpha_k)A_k+\alpha_k m,$$

$$v_{k+1}=(1-\alpha_k)\frac{A_k}{A_{k+1}}v_k+\alpha_k\frac{m}{A_{k+1}}y_k-\frac{\alpha_k}{A_{k+1}}f'(y_k).$$

Нетрудно показать, что для квадратичных функций метод $\mathfrak{m}_3$ генерирует последовательность точек, удовлетворяющую соотношению (23). При минимизации неквадратичных функций этим методом справедливы оценки (14), (15). В то же время трудоемкость одной итерации метода $\mathfrak{m}_3$ лишь в 1,5 раза превышает трудоемкость методов $\mathfrak{m}_1$ и $\mathfrak{m}_2$.

ЛИТЕРАТУРА

1. *Немировский А. С., Юдин Д. Б.* Сложность задач и эффективность методов оптимизации. М.: Наука, 1979.
2. *Немировский А. С., Юдин Д. Б.* Информационная сложность задач математического программирования // Изв. АН СССР. Техн. кибернетика. 1983. № 1.
3. *Нестеров Ю. Е.* Метод решения задачи выпуклого программирования со скоростью сходимости $O(1/k^2)$ // Докл. АН СССР. 1983. Т. 269. № 3.
4. *Нестеров Ю. Е.* Об одном классе методов безусловной минимизации выпуклой функции, обладающих высокой скоростью сходимости // Журн. вычисл. математики и мат. физики. 1984. № 7.
5. *Поляк Б. Т.* Введение в оптимизацию. М.: Наука, 1983.

Поступила в редакцию
21 XII 1982