Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

СХОДИМОСТЬ В МЕТОДЕ СОПРЯЖЕННЫХ ГРАДИЕНТОВ

РАДОСТИН П. ИВАНОВ, ЦВЕТАНА Х. НЕДЕВА

Изучаются класс строго выпуклых функций и однопараметрическое семейство градиентных методов для нахождения экстремума функций. Частными случаями этих методов являются метод наискорейшего спуска и различные варианты метода сопряженных градиентов.

Введение. Идея применения сопряженных градиентов, по-видимому, принадлежит Хестенесу и Стифелю [1], исследования которых относятся к решению линейных систем уравнений. Далее, в 60-х годах, метод сопряженных градиентов широко применялся для решения экстремальных задач. Так, в статье Флетчера и Пауэля [2] предлагается конкретный вариант метода сопряженных градиентов для нахождения экстремума квадратичной функции и доказывается его сходимость за конечное число итераций. В работе [3] предлагается один вариант метода сопряженных градиентов для решения общей задачи на безусловный экстремум. Для случая сильно выпуклых, дважды или трижды дифференцируемых функций, многие авторы исследовали вопросы сходимости и нахождения оценок для скорости сходимости этих методов. Так, для одного из вариантов в статье [4] доказана сходимость, а в работе [5] получены оценки для скорости сходимости без предположения о существовании второй производной. В работе [6] дано необходимое и достаточное условие сходимости метода. Отметим, что некоторые варианты метода исследованы при предположении только дифференцируемости и выпуклости минимизируемой функции. Достаточно полную библиографию по этим вопросам можно найти, например, в работах [7, 8, 9].

Многие из авторов рассматривают постановку экстремальной задачи в эвклидовом пространстве, хотя, как правило, их результаты без труда переносятся на случай гильбертова пространства. Мы иногда будем ссылаться на такие работы, чтобы избежать формальное повторение доказательств.

В данной статье изучается класс строго выпуклых функций в банаховых пространствах. Получена шкала оценок для скорости сходимости метода, которые лучше, чем оценки, приведенные в работе [10] для выпуклых функций. Для постановки задачи в гильбертовом пространстве предложен класс однотипных вариантов метода сопряженных градиентов, который включает в себя метод, описанный в статье [3] и доказана сходимость этих вариантов. Для метода, описанного в работе [5], доказана сходимость при более слабых предположениях о минимизируемой функции.

1. Основные теоремы. Определение. *Функцию $\varphi(x)$, определенную на выпуклом множестве Q банахова пространства E, будем называть ε-выпуклой, если существуют такие константы $\varepsilon\geqq 0$ и $\lambda>0$, что*

$$\varphi(\alpha x'+(1-\alpha)x'')\leqq\alpha\varphi(x')+(1-\alpha)\varphi(x'')-\lambda\alpha(1-\alpha)\|x'-x''\|^{2+\varepsilon} \tag{1}$$

для любых α, x', x'', *для которых* $0 \leq \alpha \leq 1$, $x', x'' \in E$, $\| x' - x'' \| \leq 1$.

Заметим, что при $\varepsilon = 0$ функция $\varphi(x)$ является сильно выпуклой.

Приведем простейший пример: функция $\varphi(x) = |x|^q$ при $q > 2$ является $(q-2)$-выпуклой на числовой оси, а при $1 < q \leq 2$ будет сильно выпуклой, т. е. 0-выпуклой.

Ясно, что любая ε-выпуклая функция является строго выпуклой, однако обратное утверждение неверно, что видно из следующего примера:

$$\varphi(x) = \begin{cases} \exp(1 - 1/x^2), & x \neq 0, \ |x| \leq \sqrt{2/3} \\ 0 & , x = 0. \end{cases}$$

Легко показать, что эта функция в окрестности нуля не является ε-выпуклой ни для какого $\varepsilon \geqq 0$ и $\lambda > 0$.

Для класса строго выпуклых функций, более широкого чем класс ε-выпуклых функций, докажем следующую теорему:

Теорема 1.1. *Пусть функция* $\varphi(x)$ *определена в рефлексивном банаховом пространстве* E, *непрерывна и* $\varphi((x'+x'')/2) \leqq (\varphi(x') + \varphi(x''))/2 - \psi(\|x' - x''\|)$, *где* $\psi(t)$ *неотрицательная непрерывная функция, для которой* $\psi(t) = 0$, *только если* $t = 0$, а $\| x' - x'' \| \leqq 1$. *Тогда множество* $M_0 = \{x : x \in E, \varphi(x) \leq \varphi(x^0)\}$ *слабо компактно.*

Доказательство. Пусть x^0 любая точка пространства E и h — любое, $1/4 \leq \| h \| \leq 1/2$. Исследуем функцию $\varphi(x)$ по направлению h. Имеем

$$\varphi(x^0 + (k-1)h) = \varphi([(x_0 + kh) + (x^0 + (k-2)h]/2)$$
$$\leq (\varphi(x^0 + kh) + \varphi(x^0 + (k-2)h))/2 - \psi(2\|h\|)$$

или

$$\varphi(x^0 + kx) \geqq 2\varphi(x^0 + (k-1)h) - \varphi(x^0 + (k-2)h) + 2C, \tag{2}$$

где $C = \psi(2\|h\|)$. Покажем, что для любого $i \geqq 1$ имеем

$$\varphi(x^0 + kh) \geqq (i+1)\varphi(x^0 + (k-i)h) - i\varphi(x^0 + (k-i-1)h) + i(i+1)C. \tag{3}$$

Докажем этот факт индуктивно. Действительно, для $i = 1$ неравенства (2) и (3) совпадают. Теперь допустим, что для некоторого $i \geqq 1$ неравенство (3) имеет место. Учитывая неравенство (2) из (3), получаем

$$\varphi(x^0 + kh) \geqq (i+1)[2\varphi(x^0 + (k-i-1)h) - \varphi(x^0 + (k-i-2)h) + 2C]$$
$$- i\varphi(x^0 + (k-i-1)h) + i(i+1)C$$

или $\varphi(x^0 + kh) \geqq (i+2)\varphi(x^0 + (k-i-1)h) - (i+1)\varphi(x^0 + (k-i-2)h) + (i+1)(i+2)C$. Этим показано, что для $i+1$ неравенство (3) может иметь место.

Отсюда при $i = k-1$ получаем

$$\varphi(x^0 + kh) \geqq k\varphi(x^0 + h) - (k-1)\varphi(x^0) + k(k-1)C. \tag{4}$$

Так как любая сильно непрерывная выпуклая функция $\varphi(x)$, определенная в банаховом пространстве E, является слабо полунепрерывной снизу и так как ограниченные, выпуклые и замкнутые множества рефлексивного пространства E являются слабо компактными, то функция $\varphi(x)$ достигает своего минимума на шаре с центром в x^0 и радиусом $1/2$. Обозначим этот минимум через m. Далее, так как $\psi(t)$ непрерывная функция, то она дости-.

гает минимума на отрезке $[1/2, 1]$. Этот минимум, ввиду условия теоремы, отличен от нуля. Обозначим его через δ. Теперь, для любых h, $1/4 \leqq \| h \| \leqq 1/2$, и $k \geqq 1$ из неравенства (4) получаем $\varphi(x^0 + kh) \geqq km - (k-1)\varphi(x^0) + k(k-1)\delta$. Ясно, что для всех достаточно больших k правая сторона этого неравенства будет больше, чем $\varphi(x^0)$. Кроме того, любая точка $x \in E$, которая достаточно далеко от точки x^0, представима в виде $x = x^0 + kh$, $1/4 \leqq \| h \| \leqq 1/2$. Отсюда следует, что множество M_0 ограничено. Теперь из выпуклости и замкнутости множества M_0 и из рефлексивности пространства E следует утверждение теоремы.

Из теоремы 1.1. легко следует важная теорема существования и единственности минимума для функции $\varphi(x)$:

Теорема 1.2. *Пусть выполнены предположения теоремы* 1.1. *Тогда существует единственная точка* x^* *такая, что* $\varphi(x^*) \leqq \varphi(x)$ *для любого* $x \in E$.

Дальше мы ограничимся рассмотрением только ε-выпуклых функций.

Лемма 1.1. *Пусть* $\varphi(x)$ *непрерывная функция, определенная в банаховом пространстве* E *и имеющая производную Фреше. Тогда условие*

$$\langle \nabla\varphi(x') - \nabla\varphi(x''), x' - x'' \rangle \geqq \lambda \| x' - x'' \|^{2+\varepsilon}, \ \lambda > 0, \ \| x' - x'' \| \leqq 1 \tag{5}$$

является необходимым и достаточным условием для ε-*выпуклости функции* $\varphi(x)$.

Здесь символом $\langle \nabla\varphi, h \rangle$ обозначено значение линейного функционала $\nabla\varphi \in E'$ на элементе $h \in E$.

Остановимся кратко на требовании о существовании производной Фреше для функции $\varphi(x)$. В этом параграфе такое сильное требование не является необходимым, так как мы будем получать неравенства, в которых участвуют два элемента пространства E. По сути дела мы будем рассматривать непрерывную выпуклую функцию на прямой, а, как известно, такие функции ведут себя довольно хорошо. Однако мы несколько ограничим общность некоторых неравенств, поскольку в конечном итоге они будут применяться к функции, имеющей производную Фреше.

Доказательство леммы 1.1. Необходимость. Из выпуклости функции $\varphi(x)$ следует, что $0 \leqq (\varphi(x') - \varphi(\alpha x' + (1-\alpha)x''))/(1-\alpha) + (\varphi(x'') - \varphi(\alpha x' + (1-\alpha)x''))/\alpha$, отсюда, после предельного перехода при $\alpha \to +0$, получаем

$$\varphi(x') - \varphi(x'') \leqq \langle \nabla\varphi(x'), x' - x'' \rangle. \tag{6}$$

Теперь из неравенства (1), полагая $\alpha = 1/2$, и из неравенства (6) получаем

$$\lambda \| x' - x'' \|^{2+\varepsilon}/4 \leqq (\varphi(x') - \varphi((x' + x'')/2))/2 + (\varphi(x'') - \varphi((x' + x'')/2))/2$$
$$\leqq \langle \nabla\varphi(x') - \nabla\varphi(x''), \ x' - x'' \rangle/4,$$

т. е. $\langle \nabla\varphi(x') - \nabla\varphi(x''), x' - x'' \rangle \geqq \lambda \| x' - x'' \|^{2+\varepsilon}$.

Достаточность. Пусть $\| x - y \| \leqq 1$. Тогда стандартной техникой [7] можем получить

$$\varphi(\alpha x + (1-\alpha)y) - \alpha\varphi(x) - (1-\alpha)\varphi(y) = -\int_0^1 \alpha(1-\alpha)\langle \nabla\varphi(x - \tau(1-\alpha)(x-y))$$
$$- \nabla\varphi(y + \tau\alpha(x-y)), x - y \rangle d\tau \leqq -\alpha(1-\alpha)\lambda \| x - y \|^{2+\varepsilon}/(2+\varepsilon),$$

а это и есть ε-выпуклость функции $\varphi(x)$. Лемма доказана.

Лемма 1.2. *Пусть $\varphi(x)$ ε-выпуклая и дифференцируемая по Фреше функция. Тогда для любых $x, y \in E$, $\|x-y\|>1$ имеем*

$$\langle \nabla\varphi(x)-\nabla\varphi(y),\ x-y\rangle \geqq \lambda\|x-y\|^2/2^\varepsilon. \tag{7}$$

Доказательство. Пусть k такое целое число, что $k-1<\|x-y\|\leqq k$ и $h=(y-x)/k$. Так как $\|h\|\leqq 1$, то имеет место неравенство (5), которое мы напишем следующим образом:

$$\langle \nabla\varphi(x+(i-1)h)-\nabla\varphi(x+ih),\ -h\rangle \geqq \lambda\|h\|^{2+\varepsilon},\ i=1, 2, \dots, k.$$

Суммируя по i, получим

$$\langle \nabla\varphi(x)-\nabla\varphi(y), x-y\rangle \geqq \lambda\|x-y\|^{2+\varepsilon}/k^\varepsilon. \tag{8}$$

Так как $\|x-y\|>1$, то $k\geqq 2$ и, следовательно, $\|x-y\|/k>(k-1)/k\geqq 1/2$ Окончательно из неравенства (8) получаем неравенство (7).

Лемма 1.3. *Пусть $\varphi(x)$ ε-выпуклая и дифференцируемая по Фреше функция. Тогда для любых $x, y\in E$, $\|x-y\|>1$ выполнено*

$$\varphi(\alpha x+(1-\alpha)y)\leqq \alpha\varphi(x)+(1-\alpha)\varphi(y)-\lambda\alpha(1-\alpha)\|x-y\|^2/2^{1+\varepsilon}. \tag{9}$$

Доказательство. Так же как и в лемме 1.1, обозначая $1/\|x-y\|=r$, имеем

$$\begin{aligned}
&\varphi(\alpha x+(1-\alpha)y)-\alpha\varphi(x)-(1-\alpha)\varphi(y)\\
&=-\alpha(1-\alpha)\int_0^{1-r}\langle \nabla\varphi(x-\tau(1-\alpha)(x-y))-\nabla\varphi(y+\tau\alpha(x-y)),\ x-y\rangle d\tau\\
&\quad-\alpha(1-\alpha)\int_{1-r}^{1}\langle \nabla\varphi(x-\tau(1-\alpha)(x-y))-\nabla\varphi(y+\tau\alpha(x-y)),\ x-y\rangle d\tau\\
&\leqq-\alpha(1-\alpha)\int_0^{1-r}\lambda(1-\tau)\|x-y\|^2/2^\varepsilon d\tau-\alpha(1-\alpha)\int_{1-r}^{1}\lambda\|x-y\|^{2+\varepsilon}(1-\tau)^{1+\varepsilon}d\tau\\
&=-\alpha(1-\alpha)\lambda\|x-y\|^2(1-1/\|x-y\|^2)/2^{1+\varepsilon}-\alpha(1-\alpha)\lambda/(2+\varepsilon)\\
&\leqq-\alpha(1-\alpha)\lambda\|x-y\|^2/2^{1+\varepsilon}.
\end{aligned}$$

Лемма доказана.

Лемма 1.4. *Если $\varphi(x)$ ε-выпуклая функция и $x^*\in E$ точка минимума, то для любого $x\in E$, $\|x-x^*\|\leqq 1$ выполнено*

$$\|x-x^*\|^{2+\varepsilon}\leqq 2(\varphi(x)-\varphi(x^*))/\lambda. \tag{10}$$

Если, кроме того, функция $\varphi(x)$ дифференцируема по Фреше, то имеют место неравенства

$$\|x-x^*\|\leqq\lambda^{-1/(1+\varepsilon)}\|\nabla\varphi(x)\|^{1/(1+\varepsilon)}, \tag{11}$$

$$0\leqq\varphi(x)-\varphi(x^*)\leqq\lambda^{-1/(1+\varepsilon)}\|\nabla\varphi(x)\|^{(2+\varepsilon)/(1+\varepsilon)}. \tag{12}$$

Доказательство. Из неравенства (1) при $\alpha=1/2$ имеем

$$\varphi(x^*)\leqq\varphi((x+x^*)/2)\leqq(\varphi(x)+\varphi(x^*))/2-\lambda\|x-x^*\|^{2+\varepsilon}/4,$$

отсюда следует неравенство (10). Хорошо известно, что $\langle\nabla\varphi(x^*), x-x^*\rangle=0$. Теперь из неравенства (5) получаем

$$\lambda \| x - x^* \|^{2+\varepsilon} \leqq \langle \nabla \varphi(x), x - x^* \rangle \leqq \| \nabla \varphi(x) \| \| x - x^* \|.$$

Отсюда следует неравенство (11). Неравенство (12) следует из неравенств:

$$0 \leqq \varphi(x) - \varphi(x^*) \leqq \langle \nabla \varphi(x), x - x^* \rangle \leqq \| \nabla \varphi(x) \| \| x - x^* \| \leqq \lambda^{-1/(1+\varepsilon)} \| \nabla \varphi(x) \|^{(2+\varepsilon)/(1+\varepsilon)}.$$

Лемма доказана.

Во втором пункте нам будет необходима следующая лемма [11, с. 94]:

Лемма 1.5. *Если числовая последовательность такова, что* $\mu_k - \mu_{k-1} \geqq \tau_k \mu_k^p$, $\tau_k > 0$, $k = 0, 1, \ldots$, *где* $p > 1$, *то для всех* $m = 1, 2, \ldots$ *имеем*

$$\mu_m \leqq \Big(\sum_{k=0}^{m-1} \tau_k \Big)^{-1/(p-1)}. \tag{13}$$

Доказательство аналогично доказательству приведенной в [11] лемме.

2. Методы, сходимость, оценки. Определение. *Последовательность элементов* x^k, $k = 0, 1, \ldots$ *из банахова пространства* E *будем называть релаксационной относительно функции* $\varphi(x)$, *если для всех* $k = 0, 1, 2, \ldots$ *выполнено* $\varphi(x^{k+1}) \leqq \varphi(x^k)$.

Теорема 2.1. *Пусть* $\varphi(x)$ ε-*выпуклая и дифференцируемая в* E *функция, а* $\{x^k\}_0^\infty$, *релаксационная последовательность, для которой* $\varphi(x^{k+1}) < \varphi(x^k)$ *и* $\| x^k - x^* \| \leqq 1$, $k = 0, 1, \ldots$, *где* x^* *точка минимума функции* $\varphi(x)$ *на* E. *Тогда имеют место следующие оценки*:

$$\varphi(x^m) - \varphi(x^*) \leqq \Big[\lambda^{2/(2+\varepsilon)} \sum_{k=0}^{m-1} (\varphi(x^k) - \varphi(x^{k+1})) / \| \nabla \varphi(x^k) \|^2\Big]^{-(2+\varepsilon)/\varepsilon}, \tag{14}$$

$$\| x^m - x^* \| \leqq \Big[\lambda 2^{-\varepsilon/(2+\varepsilon)} \sum_{k=0}^{m-1} (\varphi(x^k) - \varphi(x^{k+1})) / \| \nabla \varphi(x^k) \|^2\Big]^{-1/\varepsilon}. \tag{15}$$

Доказательство. Положив $\mu_k = \varphi(x^k) - \varphi(x^*)$, из леммы 1.4. имеем

$$\| \nabla \varphi(x^k) \|^2 \geqq \lambda^{2/(2+\varepsilon)} \mu_k^{(2+2\varepsilon)/(2+\varepsilon)},$$

$$\mu_k - \mu_{k+1} = (\varphi(x^k) - \varphi(x^{k+1})) \| \nabla \varphi(x^k) \|^2 / \| \nabla \varphi(x^k) \|^2,$$

$$\mu_k - \mu_{k+1} \geqq \lambda^{2/(2+\varepsilon)} (\varphi(x^k) - \varphi(x^{k+1})) \mu_k^{(2+2\varepsilon)/(2+\varepsilon)} / \| \nabla \varphi(x^k) \|^2.$$

Теперь, если положим $\tau_k = \lambda^{(2+\varepsilon)/\varepsilon} (\varphi(x^k) - \varphi(x^{k+1})) / \| \nabla \varphi(x^k) \|^2$ и $p = (1+2\varepsilon)/(2+\varepsilon)$, то из леммы 1.5 непосредственно получаем неравенство (14). Отсюда и из неравенства (10) получаем неравенство (15). Теорема доказана.

Заметим, что неравенства (14) и (15) дают шкалу оценок в зависимости от значения ε. Известные оценки [10]

$$\varphi(x^m) - \varphi(x^*) \leqq \delta^2 \Big[\sum_{k=0}^{m-1} (\varphi(x^k) - \varphi(x^{k+1})) / \| \nabla \varphi(x^k) \|^2\Big]^{-1}, \tag{16}$$

где $\delta = \operatorname{diam} \{x : x \in E, \ \varphi(x) \leqq \varphi(x^0)\}$ и

$$\varphi(x^m) - \varphi(x^*) \leqq (\varphi(x^0) - \varphi(x^*)) \exp \Big[- \sum_{k=0}^{m-1} \lambda (\varphi(x^k) - \varphi(x^{k+1})) / \| \nabla \varphi(x^k) \|^2\Big] \tag{17}$$

для выпуклых и, соответственно, для сильно выпуклых функций, можно рассматривать как предельные для этой шкалы оценки.

Из неравенства (14) следуют известные оценки [10].

Теорема 2.2. *Пусть $\varphi(x)$ выпуклая функция, имеющая производную. Тогда, для того чтобы процесс, определенный релаксационной последовательностью x^k, $k=0, 1, 2, \ldots$ сходился, достаточно, чтобы ряд*

$$(18)\qquad \sum_{k=0}^{\infty}(\varphi(x^k)-\varphi(x^{k+1}))/\|\nabla\varphi(x^k)\|^2$$

расходился.

Опишем хорошо известный алгоритм для минимизации функции $\varphi(x)$ на пространстве E. Пусть начальное приближение x^0 любая точка из E, а $-s^0$ любое направление убывания функции $\varphi(x)$ в точке x^0. Определяем $x^1=x^0-\beta_0 s^0$, где β_0 определяется из условия $\varphi(x^0-\beta_0 s^0)=\min\{\varphi(x^0-\beta s^0) \mid \beta\geqq 0\}$.

Опишем k-ую итерацию:

1. Пусть известны k-ое приближение x^k и k-ое направление s^k.
2. Минимизируем функцию $\varphi(x^k-\beta s^k)$ по $\beta\geqq 0$ и определяем β_k так, что $\varphi(x^k-\beta_k s^k)=\min\{\varphi(x^k-\beta s^k) \mid \beta\geqq 0\}$.
3. Определяем $x^{k+1}=x^k-\beta_k s^k$.
4. Находим такое s^{k+1}, чтобы $\langle\nabla\varphi(x^{k+1}), s^{k+1}\rangle>0$.
5. Если такого направления не существует, то $x^{k+1}=x^*$, иначе переходим к пункту 1.

В предположении, что функция $\nabla\varphi(x)$ удовлетворяет условию Липшица с константой L, т. е.

$$(19)\qquad \|\nabla\varphi(x)-\nabla\varphi(y)\|\leqq L\,|x-y|,$$

оценим скорость сходимости описанного итерационного процесса.

Теорема 2.3. *Пусть $\varphi(x)$ ε-выпуклая функция с производной, удовлетворяющей условию (19), а последовательность x^k, $k=0, 1, \ldots$ получена при помощи описанного алгоритма, т. е. $x^{k+1}=x^k-\beta_k s^k$, $k=0, 1, \ldots$, $\varphi(x^k-\beta_k s^k)=\min\{\varphi(x^k-\beta s^k) \mid \beta\leqq 0\}$, $\|x^k-x^*\|\leqq 1$, $k=0, 1, \ldots$, где x^0 любая точка пространства E, а $-s^0$ направление убывания функции $\varphi(x)$ в точке x^0. Пусть далее*

$$(20)\qquad \alpha_k=\langle\nabla\varphi(x^k), s^k\rangle/\|\nabla\varphi(x^k)\|\,\|s^k\|>0.$$

Тогда имеют место следующие оценки:

$$(21)\qquad \varphi(x^m)-\varphi(x^*)\leqq[\sum_{k=0}^{m-1}\alpha_k^2/2L]^{-(2+\varepsilon)/\varepsilon}\lambda^{\varepsilon/2}$$

$$(22)\qquad \|x^m-x^*\|\leqq[(\lambda 2^{(2+\varepsilon)/(2+2\varepsilon)}/L)\sum_{k=0}^{m-1}\alpha_k^2]^{-1/\varepsilon}.$$

Доказательство. Нам будет необходимо известное неравенство [7] $\varphi(x)-\varphi(y)\geqq\langle\nabla\varphi(x), x-y\rangle-(L/2)\|x-y\|^2$. Имеем $\varphi(x^k)-\varphi(x^{k+1})\geqq\varphi(x^k)-\varphi(x^k-\beta s^k)\geqq\langle\nabla\varphi(x^k), \beta s^k\rangle-(L/2)\|x^k-x^k+\beta s^k\|^2=\beta\langle\nabla\varphi(x^k), s^k\rangle-\beta^2 L\,\|s^k\|^2/2$. Так как эти неравенства имеют место для всех $\beta\geqq 0$, то при $\beta=\alpha_k\|\nabla\varphi(x^k)\|/L\|s^k\|$ получим $\varphi(x^k)-\varphi(x^{k+1})\geqq\alpha_k^2\|\nabla\varphi(x^k)\|^2/2L$. Теперь утверждение теоремы следует из неравенств (14) и (15).

До сих пор мы рассматривали общую схему некоторых градиентных методов для минимизации функции $\varphi(x)$. Ниже рассмотрим однопараметрическое семейство методов в предположении, что пространство E гильбертово, включающее в себя метод наискорейшего спуска и известные варианты метода сопряженных градиентов. Этот класс методов вкладывается в общую схему и мы его получим, детализируя определения направления s^k, а именно $s^k = \nabla\varphi(x^k) - \xi_k s^{k-1}$, $k=0, 1, 2, \ldots$ Можно показать [7], что $\langle \nabla\varphi(x^{k+1}), s^k\rangle = 0$, $k=0, 1, 2, \ldots$, $\langle \nabla\varphi(x^k), s^k\rangle = \|\nabla\varphi(x^k)\|^2$, $\|s^k\|^2 = \|\nabla\varphi(x^k)\|^2 + \xi_k^2\|s^{k-1}\|^2$, $k=0, 1, \ldots$ Дальше имеем:

$$\alpha_k^2 = \langle \nabla\varphi(x^k), S^k\rangle / \|\nabla\varphi(x^k)\|\,\|S^k\| = \|\nabla\varphi(x^k)\|^2/\|s^k\|^2 \tag{23}$$
$$= \|\nabla\varphi(x^k)\|^2/(\|\nabla\varphi(x^k)\|^2 + \xi_k^2\|s^{k-1}\|).$$

Теперь мы готовы рассмотреть некоторые конкретные методы в предположении, что минимизируемая функция является ε-выпуклой.

Следствие 1. *Если ξ_k таково, что $|\xi_k| \leqq c\|\nabla\varphi(x^k)\|/\|s^{k-1}\|$, $c=\mathrm{const}>0$, $k=0, 1, \ldots$, то итерационный процесс сходится со скоростью:*

$$\varphi(x^m) - \varphi(x^*) \leqq (c_1/m)^{(2+\varepsilon)/\varepsilon},\ c_1 = \mathrm{const}>0,\ \|x^m - x^*\| \leqq (c_2/m)^{1/\varepsilon},\ c_2 = \mathrm{const}>0$$

или $\varphi(x^m) - \varphi(x^*) \leqq O(1/m^{(2+\varepsilon)/\varepsilon})$, $\|x^m - x^*\| \leqq O(1/m^{1/\varepsilon})$.

Доказательство следует из неравенств (21), (22) и (23).

Заметим, что полученная скорость сходимости процесса лучше, как естественно надо ожидать, известной оценки [10] для выпуклой функции. Частным случаем следствия 1 является

Следствие 2. *Если $\xi_k = 0$, $k=0, 1, \ldots$, т. е. для метода наискорейшего спуска, получаем, что скорость сходимости процесса равна* $O(1/m^{(2+\varepsilon)/\varepsilon})$.

Следствие 3. *Если* $\xi_k = \|\nabla\varphi(x^k)\|^{\eta/2}/\|\nabla\varphi(x^{k-1})\|^{\eta/2}$, $\eta \geqq 2$, *то процесс сходится.*

Доказательство. Из равенства $\|s^k\|^2 = \|\nabla\varphi(x^k)\|^2 + \xi_k^2\|s^{k-1}\|^2$ имеем:

$$\|s^k\|^2 = \|\nabla\varphi(x^k)\|^2 + \sum_{i=1}^{k}\|\nabla\varphi(x^{k-i})\|^2 \prod_{j=k-i+1}^{k}\xi_j^2$$

$$= \|\nabla\varphi(x^k)\|^2 + \sum_{i=1}^{k}\|\nabla\varphi(x^{k-i})\|^{2-\eta}\|\nabla\varphi(x^k)\|^{\eta} = \|\nabla\varphi(x^k)\|^{\eta}\sum_{i=0}^{k}\|\nabla\varphi(x^{k-i})\|^{2-\eta}.$$

Следовательно, для α_k получим $\alpha_k^2 = \|\nabla\varphi(x^k)\|^{2-\eta}/\Sigma_{i=0}^k\|\nabla\varphi(x^{k-i})\|^{2-\eta}$. Если функция $\varphi(x)$ имеет ограниченную производную для всех k, т. е. $\|\nabla\varphi(x^k)\| \leqq \gamma < \infty$, $k=0, 1, \ldots$, в частности, для ε-выпуклых функций с липшицовой производной этот факт имеет место, то $\|\nabla\varphi(x^k)\|^{2-\eta} \geqq \gamma^{2-\eta} > 0$, так как $\eta \geqq 2$. Следовательно, последовательность $\|\nabla\varphi(x^k)\|^{2-\eta}$, $k=0, 1, \ldots$ не стремится к нулю и по критерию Абеля [12] ряд $\Sigma_{k=0}^{\infty}\alpha_k^2$ расходится. Однако по теореме 2.2 это обеспечивает сходимость процесса. Следствие доказано.

Заметим, что при $\eta=2$ можно вычислить $\alpha_k^2 = 1/(k-1)$ и теперь нетрудно получить оценку

$$\varphi(x^m) - \varphi(x^*) \leqq O\{[\sum_{k=0}^{m-1}(k+1)^{-1}]^{-(2+\varepsilon)/\varepsilon}\} = O((\ln m)^{-(2+\varepsilon)/\varepsilon}).$$

Такая скорость сходимости неудовлетворительна. Некоторые численные эксперименты дают основание выдвинуть гипотезу, что оптимальные значения для η находятся в отрезке [4, 6]. В конце заметим, что при $\eta=4$ получаем хорошо известный из [3] метод сопряженных градиентов.

Рассмотрим другой вариант метода сопряженных градиентов [5, 6, 7, 9, 13].

Следствие 4. *Если* $\xi_k = \langle \nabla\varphi(x^k), \nabla\varphi(x^k) - \nabla\varphi(x^{k-1})\rangle / \| \nabla\varphi(x^{k-1}) \|^2$, *то процесс сходится.*

Доказательство. Оценим ξ_k сверху, а α_k снизу:

$$\xi_k \leqq \| \nabla\varphi(x^k) \| \, \| \nabla\varphi(x^k) - \nabla\varphi(x^{k-1}) \| / \| \nabla\varphi(x_k^{k-1}) \|^2$$

$$\leqq L\beta_{k-1} \| \nabla\varphi(x^k) \| \, \| s^{k-1} \| / \| \nabla\varphi(x^{k-1}) \|^2,$$

$$\alpha_k^2 \geqq \| \nabla\varphi(x^k) \|^2 / (\| \nabla\varphi(x^k) \|^2 + L^2\beta_{k-1}^2 \| \nabla\varphi(x^k) \|^2 \, \| s^{k-1} \|^2 / \| \nabla\varphi(x^{k-1}) \|^4)$$

$$= 1/(1 + \beta_{k-1}^2 \| s^{k-1} \|^2 L^2 / \| \nabla\varphi(x^{k-1}) \|^4).$$

Покажем, что $\beta_k \| s^k \| \to 0$, $k \to \infty$. Действительно,

$$\varphi(x^k) - \varphi(x^{k+1}) = \int_0^1 \langle \nabla\varphi(x^{k+1} + \tau\beta_k s^k), \beta_k s^k \rangle d\tau$$

$$- \langle \nabla\varphi(x^{k+1}), \beta_k s^k \rangle \geqq \lambda\beta_k^{2+\varepsilon} \| s^k \|^{2+\varepsilon} / (2+\varepsilon),$$

и так как $\lim_{k\to\infty} (\varphi(x^k) - \varphi(x^{k+1})) = 0$, то $\lim_{k\to\infty} \beta_k \| s^k \| = 0$. Теперь, если допустим, что процесс не сходится, будем иметь $\liminf_{k\to\infty} \| \nabla\varphi(x^k) \| > 0$ и, следовательно, с некоторого номера k_0 будет выполнено $\| \nabla\varphi(x^k) \|^4 \geqq a > 0$. Отсюда получаем $\alpha_k^2 \geqq 1/(1 + \beta_{k-1}^2 \| s^{k-1} \|^2 L^2 / a) > 1/2$ для всех достаточно больших k. Следовательно, ряд $\Sigma_{k=0}^{\infty} \alpha_k^2$ расходится, и по теореме 2.2 процесс сходится, что противоречит сделанному допущению. Следствие доказано.

ЛИТЕРАТУРА

1. M. Hestenes, E. Stiefel. Methods of conjugate gradients for solving linear system *J. Res. Nat. Bur. Stand.*, **49**, 1952, 409.
2. R. Fletcher, M. Powel. A rapidly convergent descent method for minimization. *Comput. J.*, **6**, 1963, No. 2, 163—168.
3. R. Fletcher, C. Reeves. Function minimization by conjugate gradients. *Comput. J.*, **7**, 1964, No. 2, 149—153.
4. J. Daniel. The conjugate gradient method for linear and nonlinear operator equations. *SIAM J. Numer. Anal.*, **4**, 1967, 10.
5. Б. Т. Поляк. Метод сопряженных градиентов в задачах на экстремум. *Ж. вычисл. мат. и мат. физ.*, **9**, 1969, № 4, 807—821.
6. Melanie L. Lenard. Practical convergence conditions for unconstrained optimization. *Math. programming*, **4**, 1973, No. 3, 309—323.
7. Ф. П. Васильев. Лекции по методам решения экстремальных задач. Москва, 1974.
8. У. И. Зангвилл. Нелинейное программирование. Москва, 1973.
9. Э. Полак. Численные методы оптимизации. Москва, 1974.
10. В. Г. Карманов. Математическое программирование. Москва, 1975.
11. В. Ф. Демьянов, А. М. Рубинов. Приближенные методы решения экстремальных задач. Ленинград, 1968.
12. Я. А. Тагамлицки. Диференциално смятане. София, 1971.
13. А. Фиакко, Г. Мак-Кормик. Нелинейное программирование. Методы последовательной безусловной минимизации. Москва, 1972.

Единый центр науки и подготовки кадров по математике и механике
1000 София П. Я. 373

Поступила 28. 3. 1975
в переработанном виде 20. 6. 1976