П. А. Дорофеев, Одна схема итеративных методов минимизации,
Ж. вычисл. матем. и матем. физ., 1986, том 26, номер 4, 536–544

ЖУРНАЛ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

Том 26, 1986 № 4

УДК 519.615.7

ОДНА СХЕМА ИТЕРАТИВНЫХ МЕТОДОВ МИНИМИЗАЦИИ

ДОРОФЕЕВ П. А.

(Ленинград)

Предлагается схема, позволяющая строить сходящиеся итеративные методы минимизации широкого класса функций. В качестве примера рассматривается задача минимизации произвольной локально липшицевой функции, заданной с ошибками.

§ 1. Определения. Вспомогательные утверждения

1. Пусть функция F: $E_n \to E_1$ удовлетворяет на E_n локальному условию Липшица и обладает в любой точке $x \in E_n$ производной по всякому направлению e, причем

$$(1.1)\qquad F'(x,e)=\lim_{\lambda\to0+}\frac{F(x+\lambda e)-F(x)}{\lambda}=\max_{g\in\partial F(x)}(g,e),$$

где $\partial F(x)$ — субдифференциал Кларка (см., например, [1]) функции F в точке x.

Следуя терминологии, принятой в [2], будем называть такие функции субдифференцируемыми локально липшицевыми (с.д.л.). В частности, выпуклые и непрерывно дифференцируемые функции принадлежат выделенному классу.

Рассмотрим задачу

$$(1.2)\qquad \min\{f(x)\,|\,x\in D\},$$

где

$$(1.3)\qquad D=\{x\,|\,p(x)\leqslant 0\}\neq\varnothing,$$

считая f и p с.д.л.-функциями. Предположим также, что

$$(1.4)\qquad 0\notin\partial p(x)\qquad \forall x\in \operatorname{fr} D=\{y\,|\,p(y)=0\},$$

т. е. выполняется аналог условия Слейтера.

Вектор $g\in E_n$ назовем возможным в широком смысле направлением произвольного множества $Q\subset E_n$ в точке $x\in Q$, если существуют число $\gamma\geqslant0$ и последовательность $\{x_k\}$ такие, что [2]

$$x_k\in Q,\quad x_k\neq x,\quad x_k\to x,\qquad (x_k-x)\|x_k-x\|^{-1}\to g_0,\quad g=\gamma g_0.$$

Очевидно, множество таких направлений $\Gamma_Q(x)$ является конусом с вершиной в начале координат. При этом [2]

$$(1.5)\qquad \Gamma_D(x)=\begin{cases} E_n, & p(x)<0,\\ \{e\,|\,p'(x,e)\leqslant0\}, & p(x)=0.\end{cases}$$

2. Как будет показано в дальнейшем, задачу (1.2) можно в определенном смысле решить, если удастся построить семейство множеств $B(x)$, $x\in E_n$, обладающих следующими свойствами.

Условие 1. $B(x)$ являются непустыми выпуклыми компактами при всех $x\in E_n$.

Условие 2. Точечно-множественное отображение $x \to B(x)$ ограничено в окрестности любой точки и полунепрерывно сверху при всяком $x \in E_n$.

Положим

$$\|h_B(x)\| = \min_{g \in B(x)} \|g\|, \qquad h_B(x) \in B(x).$$

Условие 3. *Если* $x \in D$, $e \in \Gamma_D$ и $(h_B(x), e) > 0$, то $f'(x, e) > 0$; если же $x \notin D$, $e \in E_n$ и $(h_B(x), e) > 0$, то $p'(x, e) > 0$.

Замечание 1. Согласно свойству 3,

$$h_B(x) = \gamma_1(x) \nabla f(x), \qquad \gamma_1(x) \geqslant 0,$$

в любой точке дифференцируемости функции f, лежащей внутри D. Аналогично,

$$h_B(x) = \gamma_2(x) \nabla p(x), \qquad \gamma_2(x) \geqslant 0,$$

в любой точке дифференцируемости функции p, лежащей вне D.

Можно доказать, что $0 \in B(x)$, если в точке $x \in D$ функция f имеет локальный минимум на множестве D. (Этот факт в явном виде далее не используется.) Убедимся, что семейство множеств

$$B_0(x) = \begin{cases} \partial f(x), & p(x) < 0, \\ \mathrm{co}\{\partial f(x) \cup \partial p(x)\}, & p(x) = 0, \\ \partial p(x), & p(x) > 0, \end{cases}$$

удовлетворяет условиям 1—3.

Некоторую сложность может представлять только проверка условия 3 при $x \in \mathrm{fr}\, D$. В этом случае $h_{B_0}(x)$ является субградиентом с.д.л.-функции $F(y) = \max(f(y) - f(x), p(y))$ в точке x. Поэтому, в силу (1.1), $\max(f'(x, e), p'(x, e)) > 0$, если $(h_{B_0}(x), e) > 0$. Теперь из (1.5) получаем требуемое.

Определим γ-окрестность, $\gamma \geqslant 0$, множества $Q \subset E_n$ равенством

$$[Q]_\gamma = \{x \mid \inf_{y \in Q} \|y - x\| \leqslant \gamma\}.$$

Ясно, что множество $[Q]_\gamma$ выпукло и замкнуто, если таково Q.

По данному семейству $B(x)$, $x \in E_n$, построим новое $B_*(x) = [B(x)]_{\varepsilon(x)}$, где функция $\varepsilon(x)$ неотрицательна и полунепрерывна сверху на E_n:

$$(1.6) \qquad \limsup \varepsilon(x_k) \leqslant \varepsilon(x),$$

если $x_k \to x$.

Лемма 1. *Если семейство $B(x)$, $x \in E_n$, удовлетворяет условиям 1—3, то этим же свойством обладает и семейство $B_*(x)$.*

Доказательство. Справедливость условия 2 непосредственно следует из (1.6). Так как минимальный по норме вектор в $B(x)$ сонаправлен соответствующему вектору в $B_*(x)$, то условие 3 также выполняется.

Для произвольной с.д.л.-функции F положим

$$\|H(F, x)\| = \max_{g \in \partial F(x)} \|g\|.$$

Зафиксируем $\Delta_k \geqslant 0$, $\alpha_k \geqslant 0$, $k=1, 2$, и введем функции

$$\varepsilon_1(x)=\begin{cases} \Delta_1, & p(x)<0, \\ \max(\Delta_1, \Delta_2), & p(x)=0, \\ \Delta_2, & p(x)>0, \end{cases}$$

$$\varepsilon_2(x)=\begin{cases} \alpha_1\|H(f,x)\|, & p(x)<0, \\ \max(\alpha_1\|H(f,x)\|,\ \alpha_2\|H(p,x)\|), & p(x)=0, \\ \alpha_2\|H(p,x)\|, & p(x)>0. \end{cases}$$

Семейства множеств $B_k(x)=[B_0(x)]_{\varepsilon_k(x)}$, $k=1, 2$, естественно рассматривать в ситуациях, когда обобщенные градиенты функций f и p вычисляются приближенно с заданными абсолютными или относительными погрешностями. Поскольку, как нетрудно убедиться, функции ε_1, ε_2 полунепрерывны сверху, то условия 1—3 здесь выполняются.

3. Определим множества

$$V_k=\{x\in D\,|\,0\in B_k(x)\},\qquad k=0, 1, 2.$$

Очевидно, $V_0\subset V_1\cap V_2$. При этом для того, чтобы с.д.л.-функция f достигала в точке $x\in D$ минимума на D, необходимо (а для выпуклых функций f, p и достаточно), чтобы $x\in V_0$ (см. [1], [2]).

Оценим степень близости V_2 к V_0. Имеем

$$\|h_{B_0}(x)\|\leqslant\varepsilon_2(x)\quad \forall x\in V_2. \tag{1.7}$$

Пусть x' — точка дифференцируемости f. Так как

$$\|h_{B_0}(x')\|=\|H(f, x')\|=\|\nabla f(x')\|\qquad \forall x'\in \operatorname{int} D,$$

то, согласно (1.7), при $\alpha_1<1$

$$\nabla f(x')=\mathbf{0}\qquad \forall x'\in V_2\cap\operatorname{int} D. \tag{1.8}$$

Обозначим через $L_F(Q)$ константу Липшица локально липшицевой функции F на множестве Q ($L_F(Q)<\infty$, если Q ограничено). В силу (1.7),

$$\|h_{B_0}(x)\|\leqslant\begin{cases} \alpha_1 L_f(Q), & x\in V_2\cap\operatorname{int} D, \\ \max(\alpha_1 L_f(Q),\ \alpha_2 L_p(Q)), & x\in V_2\cap\operatorname{fr} D, \end{cases} \tag{1.9}$$

для любого множества Q такого, что $\operatorname{int} Q\supset D$.

Все введенные выше понятия с тривиальными изменениями переносятся и на случай, когда $D=E_n$.

4. Рассмотрим задачу минимизации функции F на множестве Q (для произвольных F и Q) и некоторый итеративный метод ее решения, генерирующий бесконечную ограниченную последовательность $\{x_k\}\subset E_n$. Выделим множество $V\subset Q$, которое будем считать множеством обобщенных решений указанной задачи. Таким образом, решается задача (F, Q, V).

Обозначим через T множество предельных точек последовательности $\{x_k\}$. Положим

$$F_*=\lim\inf F(x_k),\qquad F^*=\lim\sup F(x_k),$$

$$T_*=\{x\in T\,|\,F(x)=F_*\},\qquad T^*=\{x\in T\,|\,F(x)=F^*\}.$$

Для наших целей удобно следующее определение (см. [3]).

Будем говорить, что данный метод решает задачу (F, Q, V), если

$$[F_*, F^*]=F(T\cap V)=\{a\in E_1\,|\,\exists x\in T\cap V: a=F(x)\},\ T_*\cup T^*\subset V. \tag{1.10}$$

Согласно этому определению, метод, решающий задачу (F, Q, V), всегда позволяет выделить предельную точку x последовательности $\{x_k\}$, для которой $F(x)$ совпадает со значением функции F в какой-то точке множества $T \cap V$. Если же значения $F(x_k)$ могут быть вычислены точно, то из равенства $\liminf F(x_k) = F(x)$, $x \in T$, следует, что $x \in V$. Кроме того, соотношения (1.10) дают информацию о множестве T в целом. Так,

$$T \subset \{x \in Q \mid F(x) \leqslant \sup_{y \in V} F(y)\}.$$

§ 2. Основные результаты

В этом параграфе устанавливаются теоремы о сходимости достаточно широкого класса аналогов метода обобщенного градиента [4]—[6]. Настоящая работа построена таким образом, что доказательства указанных теорем только в деталях отличаются от доказательства соответствующего утверждения в [3] (см. также [7]) и поэтому, как правило, не приводятся.

Допустим, что для задачи (1.2) выделено какое-то семейство множеств $B(x)$, $x \in E_n$, удовлетворяющее условиям 1—3.

Зафиксируем $x_0 \in E_n$ и последовательность $\{\lambda_k\}$ такую, что

$$(2.1) \qquad \lambda_k \to 0+, \qquad \sum_{k=0}^{\infty} \lambda_k = \infty.$$

Пусть $V_B = \{x \in D \mid 0 \in B(x)\}$.

1. Положим при $k \geqslant 0$

$$(2.2) \qquad x_{k+1} = x_k - \lambda_k(\gamma_k g_k + a_k).$$

Здесь $g_k \in B(x_k)$, a_k — помеха, в приложениях имеющая стохастический характер (см. § 3), $\gamma_k > 0$ — нормирующие множители, обладающие свойствами

$$(2.3) \qquad \lambda_k \gamma_k \|g_k\| \to 0;$$

из предположений $x_k \to x$, $0 \notin B(x)$ следует, что

$$(2.4) \qquad \liminf \gamma_k > 0.$$

(Смысл условия (2.4) разъясняется ниже.) В частности, при $a \leqslant -1$

$$\gamma_k = \begin{cases} 1, & \|g_k\| \leqslant 1, \\ \|g_k\|^a, & \|g_k\| > 1, \end{cases}$$

заведомо удовлетворяют (2.3), (2.4).

Теорема 1. *Если последовательность $\{x_k\}$ ограничена, ряд $\sum_{k=0}^{\infty} \lambda_k a_k$ сходится и*

$$(2.5) \qquad 0 \notin B(x) \qquad \forall x \notin D,$$

то метод (2.2) *решает задачу* (f, D, V_B).

Доказательство лишь кратко наметим. Пусть T — множество предельных точек $\{x_k\}$ и $x \in T \setminus B_B$. Рассмотрим подпоследовательность $\{x_{k_m}\} \to x$, $\delta > 0$, индекс

$$\tau_m(\delta) = \max_{i < k_m} \{i \mid \|x_i - x_{k_m}\| > \delta\},$$

а также

$$\|H_B(y)\| = \max_{g\in B(y)} \|g\|, \qquad \gamma_\delta(x) = \max_{y\in[x]_\delta} \|H_B(y)\|.$$

Зафиксируем $\alpha \in]0, 1[$. Тогда с учетом (2.5) для любой предельной точки x_δ подпоследовательности $\{x_{\tau_m(\delta)}\}$

$$(2.6) \qquad \gamma_\delta(x)\,(h_B(x),\, x_\delta - x) \geqslant \alpha\|h_B(x)\|^2\|x_\delta - x\| > 0$$

при всех достаточно малых δ.

Если $p^* = \limsup p(x_k) > 0$, то в силу (2.6) и условия 3 можно построить точку $x' \in T$, для которой $p(x') > p^*$. Это противоречие означает, что $T \subset D$.

Теперь если $\delta \to 0+$, $(x_\delta - x)\|x_\delta - x\|^{-1} \to e$, то $e \in \Gamma_D(x)$. Отсюда вытекает заключение теоремы.

Установить ограниченность последовательности $\{x_k\}$ бывает непросто. Поэтому полезно модифицировать метод (2.2).

Предположим, что значения функции $f(x)$ вычисляются приближенно с абсолютной погрешностью $\Delta \geqslant 0$, а значения $p(x)$ могут быть вычислены точно (иначе неясен способ построения множеств $B(x)$).

Зафиксируем $t > 2\Delta$. Положим

$$\Phi(x) = \max(f(x) - f(x_0),\ p(x)),$$

$$(2.7) \qquad x_{k+1} = \begin{cases} x_k - \lambda_k(\gamma_k g_k + a_k), & \Phi(x_k) \leqslant t, \\ x_0, & \Phi(x_k) > t. \end{cases}$$

По отношению к методу (2.7) теорема 1 остается справедливой. Теперь для ограниченности $\{x_k\}$ достаточна ограниченность множества $Q_t = \{x \mid \Phi(x) \leqslant t\}$ (при этом можно взять $\gamma_k = 1$), а (2.5) приобретает вид

$$(2.8) \qquad \mathbf{0} \notin B(x) \qquad \forall x \in Q_t \setminus D.$$

Обсудим полученные результаты в случае, когда $B(x) = B_1(x)$ или $B(x) = B_2(x)$.

Обозначим через $d(Q)$ диаметр множества $Q \subset E_n$, считая $d(Q) = \infty$ для неограниченных множеств.

Условия (2.5), (2.8) эффективно проверяемы в предположении о выпуклости функции p. Если известна точка x' такая, что $p(x') < 0$ (ее существование следует из (1.4)), то, например, для $B_1(x)$ условие (2.8) выполняется при $\Delta_2 d(Q_t) < -p(x')$.

Для гладкой выпуклой функции p, $\alpha_2 < 1$ и $B(x) = B_2(x)$ условия (2.5), (2.8) выполняются на основании (1.4).

Любопытно поведение описанных методов в ситуации, когда $B(x) = B_2(x)$, $\alpha_1 < 1$, $D = E_n$ и функция f гладкая. Тогда из ограниченности множества $Q_t = \{x \mid f(x) \leqslant f(x_0) + t\}$ с учетом (1.8) следует, что метод (2.7) решает задачу $(f,\ E_n,\ V_0(t))$, где $V_0(t) = \{x \in Q_t \mid \nabla f(x) = \mathbf{0}\}$. Таким образом, в этом случае в известном смысле безразлично, ищутся ли градиенты $\nabla f(x_k)$ точно или с заданной относительной погрешностью.

Вообще, как видно из (1.7)—(1.9), желательно по возможности уменьшать относительную погрешность вычисления обобщенных градиентов именно в точках, близких к $\operatorname{fr} D$.

2. Допустим, что множество D имеет достаточно простой вид, так что операция проектирования $\pi(x)$ точки x на D может быть выполнена точно. Эта операция для невыпуклых D определена, вообще говоря, неод-

нозначно. Условимся в качестве $\pi(x)$ при каждом x брать произвольное фиксированное решение указанной задачи.

Под $B(x)$, $x\in D$, будем в этом пункте подразумевать $B(x)=[\partial f(x)]_{\varepsilon(x)}$, предполагая, что функция $\varepsilon(x)$ неотрицательна и полунепрерывна сверху на D.

Введем множество

$$\tilde{V}_B=\{x\in\operatorname{int} D\,|\,0\in B(x)\}\cup\{x\in\operatorname{fr} D\,|\,B(x)\cap -K(\partial p(x))\neq\varnothing\},$$

где $K(\partial p(x))$ — коническая оболочка $\partial p(x)$. Отметим, что $\tilde{V}_B$ (в отличие, скажем, от V_1) не меняется при замене функции p, определяющей множество D посредством (1.3), на $\tilde{p}=\gamma p$, $\gamma>0$, и $\tilde{V}_B\supset V_0$.

Следующее утверждение является аналогом соответствующей леммы из [3] и доказывается так же.

Лемма 2. *Если $x\in\operatorname{fr} D\backslash\tilde{V}_B$, то найдется $\gamma>0$, для которого*

$$0\notin[\operatorname{co}\{\partial f(x)\cup\partial\gamma p(x)\}]_{\varepsilon(x)}.$$

Положим

$$x_{k+1}=\pi[x_k-\lambda_k(\gamma_k g_k+a_k)], \tag{2.9}$$

считая, как и ранее, что $g_k\in B(x_k)$, $\lambda_k\|g_k\|\gamma_k\to 0$, а (2.4) выполняется при $x\in D$.

Теорема 2. *Если последовательности $\{x_k\}$ и $\{\gamma_k^{-1}a_k\}$ ограничены, ряд $\sum_{k=0}^\infty\lambda_k a_k$ сходится, то метод (2.9) решает задачу $(f, D, \tilde{V}_B)$.*

При $\varepsilon(x)=\mathrm{const}$, $a_k=0$ эта теорема доказана в [3].

Следствие. Допустим, что функции f, p выпуклы, причем f достигает своего минимума на D, и $x^*\in\tilde{V}_B$ — предельная точка $\{x_k\}$, для которой $\limsup f(x_k)=f(x^*)$. Тогда

$$f(x^*)-\min_{x\in D} f(x)\leqslant\varepsilon(x^*)\min_{x\in V_0}\|x-x^*\|. \tag{2.10}$$

Если множество D не ограничено, то обеспечить ограниченность последовательности $\{x_k\}$ в (2.9) удастся, возможно, за счет возвратов к начальному приближению (см. (2.7)).

Замечание 2. Подобный же анализ решения «составной» задачи $\min\{f(x)\,|\,p(x)\leqslant 0,\ x\in Q\}$ при надлежащих предположениях относительно f, p и Q не представляет принципиальных сложностей.

3. Укажем некоторые способы ослабления наложенных выше условий. Ограничения, о которых пойдет речь, могут быть убраны независимо друг от друга.

А. Откажемся от предположения $\sum_{k=0}^\infty\lambda_k=\infty$. Тогда либо теорема 1 для методов (2.2), (2.7) остается справедливой, либо

$$x_k\to x\notin V_B. \tag{2.11}$$

Аналогичное утверждение верно и для метода (2.9) с заменой V_B на $\tilde{V}_B$.

Эта альтернатива делает ясным смысл условия (2.4): ряд $\sum_{k=0}^\infty\lambda_k\gamma_k$ должен расходиться именно в предположении (2.11).

Б. Теоремы 1, 2 справедливы для методов типа (2.2), (2.7), (2.9) с усреднением фиксированного числа направлений спуска (см., например, [7]).

В. Обратимся к задаче (1.2), считая f и p произвольными локально липшицевыми функциями. Допустим, что для этой задачи можно выделить семейство множеств $B(x)$, $x\in E_n$, обладающих свойствами 1, 2, а также следующим, более слабым, чем 3, свойством: если $x\in D$, $e\in\Gamma_D(x)$ и (см. (2.6))

(2.12) $$\|H_B(x)\|(h_B(x), e)\geqslant\|e\|\|h_B(x)\|^2>0,$$

то

$$\liminf_{\lambda\to0+}\frac{f(x+\lambda e)-f(x)}{\lambda}>0,$$

если же $x\notin D$, $e\in E_n$, то из (2.12) следует, что

$$\liminf_{\lambda\to0+}\frac{p(x+\lambda e)-p(x)}{\lambda}>0.$$

Отметим, что, в силу (2.12), векторы $h_B(x)$ и e сонаправлены для одноточечных множеств $B(x)$.

Тогда теорема 1 справедлива для методов (2.2), (2.7), построенных по семейству $B(x)$ (т. е. с $g_k\in B(x_k)$).

При $n=1$ множества $B_k(x)$, $k=0, 1, 2$, как нетрудно убедиться, обладают нужными свойствами.

Неясно, однако, как строить такие семейства множеств в общем случае. Эту трудность удается в определенном смысле преодолеть, используя подход, предложенный в [8].

Замечание 3. Полученные результаты можно применить, в частности, для изучения методов переменной метрики (см., например, [9]) минимизации непрерывно дифференцируемых функций на E_n. В этом случае $B(x)=\{A(x)\nabla f(x)\}$, где $A(x)$ – положительно-определенная матрица. Из (2.12) вытекает неравенство $f'(x, e)>0$. Следовательно, если функция $B(x)$ непрерывна, то теорема 1 справедлива.

§ 3. О минимизации произвольных функций, удовлетворяющих локальному условию Липшица

1. Пусть f — некоторая локально липшицевая на E_n функция и Q — множество в E_n. Следуя [8], зафиксируем $\gamma>0$ и свяжем с f непрерывно дифференцируемую функцию

$$f_\gamma(x)=(2\gamma)^{-n}\int_{\Pi_\gamma} f(x+y)\,dy,$$

где $\Pi_\gamma=\{y=(y^1,\dots,y^n)\,|\,|y^i|\leqslant\gamma,\ i=1, 2,\dots,n\}$. Тогда [8]

(3.1) $$|f(x)-f_\gamma(x)|\leqslant\gamma n^{1/2}L_f([Q]_{\gamma n^{1/2}})\qquad\forall x\in Q,$$

$$\nabla f_\gamma(x)=(2\gamma)^{-n}\int_{\Pi_\gamma}\nabla f(x+y)\,dy.$$

Последний интеграл является интегралом Лебега от определенной и непрерывной почти всюду на E_n по мере Лебега функции $\nabla f(x+y)$.

Допустим, что вместо значений $f(x)$ известны приближения $\tilde f(x)$ такие, что

$$|f(x)-\tilde f(x)|\leqslant\Delta\qquad\forall x\in E_n.$$

Рассмотрим случайный вектор $\theta(\omega)=(\theta^1(\omega),\dots,\theta^n(\omega))$, $\omega\in\Omega$, равномерно распределенный в Π_γ. Положим, опуская в дальнейшем зависимость от ω,

$$\eta(x,\theta)=(2\gamma)^{-1}\sum_{i=1}^{n}[\tilde{f}(x+\theta+(\gamma-\theta^i)e_i)-\tilde{f}(x+\theta-(\gamma+\theta^i)e_i)]e_i,$$

где e_i, $i=1, 2, \ldots, n$,— орты пространства E_n. Очевидно,

$$\|\eta(x,\theta)\|\leqslant n^{1/2}\left\{L_f([Q]_{\gamma n^{1/2}})+\frac{\Delta}{\gamma}\right\}\quad \forall x\in Q,\quad \forall\omega\in\Omega. \tag{3.2}$$

Обозначив через $\mathbf{M}\eta(x, \theta)$ математическое ожидание случайного вектора $\eta(x, \theta)$, имеем

$$\|\mathbf{M}\eta(x,\theta)-\nabla f_\gamma(x)\|\leqslant\frac{\Delta}{\gamma}n^{1/2}. \tag{3.3}$$

2. Пусть D — ограниченное множество, удовлетворяющее условиям, оговоренным в п. 2 из § 2. Сохраним также использованные в этом пункте обозначения.

Вместо задачи минимизации на D локально липшицевой функции f будем решать задачу

$$\min\{f_\gamma(x)\,|\,x\in D\}$$

с гладкой целевой функцией f_γ, близкой к f при малых γ.

Возьмем точку $x_0\in E_n$ и детерминированную последовательность $\{\lambda_k\}$, для которой, помимо (2.1),

$$\sum_{k=0}^{\infty}\lambda_k^2<\infty. \tag{3.4}$$

Используем также последовательность независимых случайных векторов $\{\theta_k\}$, равномерно распределенных в Π_γ. Положим

$$x_{k+1}=\pi[x_k-\lambda_k\eta(x_k,\ \theta_k)]. \tag{3.5}$$

Как отмечалось, $\{x_k\}$ в (3.5) являются случайными векторами, т. е. зависят от $\omega\in\Omega$.

Замечание 4. Согласно сказанному выше, в качестве $\pi(x)$ при каждом x берется произвольное фиксированное решение задачи проектирования. Если операция проектирования определена неоднозначно, то это может привести к затруднениям, связанным с возможной неизмеримостью $\{x_k\}$ из (3.5). Указанную трудность можно преодолеть, выделяя какую-нибудь непрерывную ветвь функции $\pi(x)$.

Теорема 3. *Для почти всех* $\omega\in\Omega$ *метод* (3.5) *решает задачу* $(f_\gamma, D, \tilde{V}_B)$, *где* $B(x)=[\nabla f_\gamma(x)]_{(\Delta/\gamma)n^{1/2}}$.

Доказательство. Пусть $\mathfrak{A}_k$ есть σ-алгебра, порожденная случайными векторами $\{\theta_i\}_{i=0}^{k-1}$ и

$$a_k=\eta(x_k,\theta_k)-\mathbf{M}\{\eta(x_k,\theta_k)\,|\,\mathfrak{A}_k\}.$$

На основании (3.2), (3.4) и теоремы о сходимости мартингалов [10], ряд $\sum_{k=0}^{\infty}\lambda_k a_k$ сходится почти наверное. Теперь из (3.3) и теоремы 2 следует требуемое.

Если $x\in\tilde{V}_B\cap\operatorname{int}D$, то

$$(2\gamma)^{-n}\left\|\int_{\Pi_\gamma}\nabla f(x+y)\,dy\right\|\leqslant\frac{\Delta}{\gamma}n^{1/2},$$

т. е. норма среднего значения субградиента $\nabla f(x+y)$ в «кубе» Π_γ не превосходит $(\Delta/\gamma)n^{1/2}$.

Отсюда видно, что в нетривиальном случае $\Delta>0$ естественно брать $\gamma=\Delta^{1/2}$. Будем далее считать, что это именно так.

Следствие. Если функции f и p выпуклы, то, с учетом (2.10), (3.1),

$$\limsup \tilde{f}(x_k)-\min_{x\in D} f(x)\leqslant (n\Delta)^{1/2}\{d(D)+2L_f([D]_{(n\Delta)^{1/2}})\}+\Delta$$

для почти всех $\omega\in\Omega$.

Подобный подход может быть использован и в сочетании с методом (2.7).

3. Реализация метода (3.5) при сделанных относительно D предположениях не представляет сложности. На практике разумно использовать в дополнение к (3.5) какую-либо процедуру одномерного поиска в направлении $-\eta(x_k, \theta_k)$, не выводящую из множества D. Если эта процедура приводит к уменьшению рекордного значения

$$t_k=\min_{i=0,1,\ldots,k} \tilde{f}(x_i)$$

на некоторую фиксированную до начала всего процесса величину $q>2\Delta$, то полученная точка берется в качестве нового приближения. В противном случае делается шаг согласно (3.5).

Такая модификация позволяет снизить ограничительность требования о детерминированном характере последовательности $\{\lambda_k\}$ без изменения асимптотических свойств метода.

Одним из способов повышения эффективности рассмотренного алгоритма является также использование операции усреднения направлений спуска (см., например, [7], [11]).

Литература

1. *Пшеничный Б. Н.* Необходимые условия экстремума. М.: Наука, 1982.
2. *Демьянов В. Ф., Васильев Л. В.* Недифференцируемая оптимизация. М.: Наука, 1981.
3. *Дорофеев П. А.* О некоторых свойствах метода обобщенного градиента. – Ж. вычисл. матем. и матем. физ., 1985, т. 25, № 2, с. 181–189.
4. *Шор Н. З.* Методы минимизации недифференцируемых функций и их приложения. Киев: Наук. думка, 1979.
5. *Поляк Б. Т.* Один общий метод решения экстремальных задач. – Докл. АН СССР, 1967, т. 174, № 1, с. 33–36.
6. *Ермольев Ю. М.* Методы стохастического программирования. М.: Наука, 1976.
7. *Дорофеев П. А.* Многошаговый стохастический ε-субградиентный метод минимизации выпуклой функции. – Ж. вычисл. матем. и матем. физ., 1984, т. 24, № 4, с. 608–611.
8. *Гупал А. М.* Стохастические методы решения негладких экстремальных задач. Киев: Наук. думка, 1979.
9. *Поляк Б. Т.* Введение в оптимизацию. М.: Наука, 1983.
10. *Ширяев А. Н.* Вероятность. М.: Наука, 1980.
11. *Поляк Б. Т.* Сравнение скорости сходимости одношаговых и многошаговых алгоритмов оптимизации при наличии помех. – Изв. АН СССР. Техн. кибернетика, 1977, № 1, с. 9–12.

Поступила в редакцию 2.X.1984
Переработанный вариант 9.I.1985