ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 1998, том 38, № 10, с. 1626–1635

УДК 519.658:519.85

ИСПОЛЬЗОВАНИЕ ДИСКРЕТНОГО ГРАДИЕНТА ПРИ МИНИМИЗАЦИИ ЛИПШИЦЕВЫХ ФУНКЦИЙ

© 1998 г. А. М. Багиров

(370148 Баку, ул. З. Халилова, 23, Бакинский гос. ун-т, фак-т прикл. матем. и кибернетики, Азербайджан)

Поступила в редакцию 13.06.96 г.
Переработанный вариант 13.03.98 г.

Для минимизации липшицевых функций предлагается метод, в котором используется понятие дискретного градиента. Исследуется сходимость метода. Приводятся результаты вычислительного эксперимента.

1. ВВЕДЕНИЕ

Работа посвящена построению численного метода для решения следующей задачи безусловной минимизации:

$$f(u) \longrightarrow \inf, \quad u \in \mathbb{E}_n, \tag{1.1}$$

где $\mathbb{E}_n$ есть n-мерное евклидово пространство векторов $u = (u_1, \ldots, u_n)$, функция f в общем случае удовлетворяет локальному условию Липшица.

Задача (1.1), когда функция f является выпуклой, изучалась многими авторами [1]–[4]. Но случай, когда функция f лишь локально-липшицева, исследован менее детально. В этом случае для построения численного метода минимизации используются либо обобщенные градиенты Кларка [5], либо другие обобщения градиента, которые получаются из субдифференциала Кларка тем или иным путем [6]–[8]. В [9] доказана сходимость метода обобщенных градиентов Кларка со специальным выбором направления спуска в задачах минимизации липшицевых функций. Достаточно интересный класс методов изучался в работах [10], [11].

В настоящей статье для построения численного метода минимизации функции f вместо субдифференциала Кларка используется его непрерывная аппроксимация. Такой подход позволяет использовать некоторые идеи гладкой оптимизации при построении и обосновании численных методов минимизации липшицевых функций.

Для построения непрерывной аппроксимации субдифференциала Кларка используется понятие дискретного градиента функции f (см. [12], [13]). В методе дискретного градиента используются лишь значения минимизируемой функции. Такой метод может быть эффективным, например, в случае, когда субдифференциал локально-липшицевой функции имеет сложный вид, а также в случае, когда аналитический вид таких функций неизвестен и их значения получаются с помощью некоторого дополнительного процесса.

В статье приняты следующие обозначения:

$$\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$$

есть скалярное произведение в $\mathbb{E}_n$, $\|u\| = \langle u, u\rangle^{1/2}$ – евклидова норма, $f'(u, g)$ – производная функции f в точке u по направлению $g \in \mathbb{E}_n$, $\partial f(u)$ – субдифференциал Кларка функции f в точке $u \in \mathbb{E}_n$.

2. ДИСКРЕТНЫЙ ГРАДИЕНТ И НЕПРЕРЫВНАЯ АППРОКСИМАЦИЯ СУБДИФФЕРЕНЦИАЛА КЛАРКА

Понятие дискретного градиента введено в [12], [13]. Напомним его определение. Пусть

$$S_1 = \{g \in \mathbb{E}_n \mid \|g\| = 1\}, \quad V = \{e \in \mathbb{E}_n \mid |e_j| = 1,\ j = 1, 2, \ldots, n\},$$

$$P = \{z(\lambda) \mid z(\lambda) \in \mathbb{E}_1,\ z(\lambda) > 0,\ \lambda > 0,\ \lambda^{-1}z(\lambda) \longrightarrow 0,\ \lambda \longrightarrow 0\},$$

$$I(g, \alpha) = \{i \in \{1, 2, \ldots, n\} \mid |g_i| \geq \alpha\},$$

где $\alpha \in (0, n^{-1/2}]$ – некоторое фиксированное число. Множество $I(g, \alpha)$ непусто для любого $g \in S_1$.

Определим операторы $H_i^j : \mathbb{E}_n \longrightarrow \mathbb{E}_n$ для $i, j = 1, 2, \ldots, n$ следующим образом:

$$H_i^j g = \begin{cases} (g_1, \ldots, g_j, 0, \ldots, 0), & \text{если} \quad j < i, \\ (g_1, \ldots, g_{i-1}, 0, g_{i+1}, \ldots, g_j, 0, \ldots, 0), & \text{если} \quad j \geq i. \end{cases}$$

Пусть $e(\beta) = (\beta e_1, \beta^2 e_2, \ldots, \beta^n e_n)$, где $\beta \in (0, 1]$. Рассмотрим следующие векторы:

$$u_i^j(g, e, z, \lambda, \beta) = u + \lambda g - z(\lambda) H_i^j e(\beta),$$

где $g \in S_1$, $e \in V$, $i \in I(g, \alpha)$, $z \in P$, $\lambda > 0$, $j = 0, 1, \ldots, n$, $j \neq i$.

Ясно, что $H_i^0 = 0 \in \mathbb{E}_n$ для любого $i = 1, 2, \ldots, n$. Поэтому

$$u_i^0(g, e, z, \lambda, \beta) = u + \lambda g.$$

Кроме того, из определения оператора H_i^j следует, что $H_i^i g = H_i^{i-1} g$ $(g \in S_1)$. Следовательно,

$$u_i^i(g, e, z, \lambda, \beta) = u_i^{i-1}(g, e, z, \lambda, \beta).$$

Если $i = n$, то $u_n^n(g, e, z, \lambda, \beta) = u_n^{n-1}(g, e, z, \lambda, \beta)$.

В дальнейшем, если задано $\lambda_0 > 0$, выражения $z_1 \leq z_2$ и $z_k \longrightarrow 0$ $(z_k \in P, k \longrightarrow +\infty)$ будут означать, что $z_1(\lambda) \leq z_2(\lambda)$ для любых $\lambda \in [0, \lambda_0]$ и $\sup\{z_k(\lambda), \lambda \in [0, \lambda_0]\} \longrightarrow 0$ при $k \longrightarrow +\infty$.

Определение 1. Дискретным градиентом функции f в точке $u \in \mathbb{E}^n$ будем называть вектор $\Gamma^i(u, g, e, z, \lambda, \beta) = (\Gamma_1^i, \ldots, \Gamma_n^i) \in \mathbb{E}_n$ $(g \in S_1, i \in I(g, \alpha))$, координаты которого имеют следующий вид:

$$\Gamma_j^i = [z(\lambda) e_j(\beta)]^{-1} [f(u_i^{j-1}(g, e, z, \lambda, \beta)) - f(u_i^j(g, e, z, \lambda, \beta))], \quad j = 1, 2, \ldots, n, \quad j \neq i,$$

$$\Gamma_i^i = (\lambda g_i)^{-1} \left[f(u_i^n(g, e, z, \lambda, \beta)) - f(u) - \sum_{j=1, j \neq i}^{n} \Gamma_j^i (\lambda g_j - z(\lambda) e_j(\beta)) \right].$$

Для фиксированных $z \in P$, $\lambda > 0$ и $\beta \in (0, 1]$ рассмотрим отображение

$$D_0(u, z, \lambda, \beta) = \operatorname{clco}\{v \in \mathbb{E}_n \mid \exists (g \in S_1, e \in V, i \in I(g, \alpha)) \colon v = \Gamma^i(u, g, e, z, \lambda, \beta)\}.$$

Положим

$$B(u) = \bigcap_{z_0 \in P, \lambda_0 > 0, \beta_0 > 0} \operatorname{cl} \operatorname{co} \bigcup \{D_0(u, z, \lambda, \beta) \colon z \in P, z < z_0, \lambda < \lambda_0, \beta < \beta_0\}.$$

Из непрерывности (локальной липшицевости) функции f следует непрерывность (локальная липшицевость) отображения $u \longrightarrow D_0(u, z, \lambda, \beta)$ при фиксированных $z \in P$, $\lambda > 0$ и $\beta \in (0, 1]$ в метрике Хаусдорфа [13]. Если функция f локально-липшицева, то многозначное отображение $u \longrightarrow D_0(u, z, \lambda, \beta)$ при фиксированных $z \in P$, $\lambda > 0$ и $\beta \in (0, 1]$ компактнозначно. В [12], [13] доказаны следующие две леммы.

Лемма 1. *Для любых* $g \in S_1$, $e \in V$, $i \in I(g, \alpha)$, $z \in P$, $\lambda > 0$, $\beta > 0$ *имеем*

$$f(u + \lambda g) - f(u) = \lambda \langle \Gamma^i(u, g, e, z, \lambda, \beta), g \rangle. \tag{2.1}$$

Лемма 2. *Если функция f дифференцируема по направлению $g \in S_1$, то существует $v \in B(u)$ такое, что*

$$f'(u, g) = \langle v, g \rangle.$$

Лемма 3. *Пусть функция f удовлетворяет условию Липшица на множестве $S_\delta(u) = \{y \in \mathbb{E}_n \mid \|y - u\| < \delta\}$, $u \in \mathbb{E}_n$, $\delta > 0$. Тогда существуют такие $z_0 \in P$, $\lambda_0 > 0$ и $R > 0$, что*

$$\sup\{\|v\| \colon v \in D_0(u, z, \lambda, \beta), z \in P, z \leq z_0, \lambda \in (0, \lambda_0], \beta \in (0, 1]\} \leq R.$$

Доказательство. Для $\delta > 0$ найдутся такие $z_0 \in P$, $\lambda_0 > 0$, что $u_i^j(g, e, z, \lambda, \beta) \in S_\delta(u)$ для любых $z \leq z_0$, $\lambda \in (0, \lambda_0]$, $\beta \in (0, 1]$, $g \in S_1$, $e \in V$, $i \in I(g, \alpha)$, $j = 0, 1, \ldots, n$, $j \neq i$, Пусть L – константа Липшица функции f на множестве $S_\delta(u)$. Тогда из определения дискретного градиента сразу следует, что $|\Gamma^i_j| \le L$ для любых $j = 1, 2, \ldots, n, j \ne i$.

Для $j = i$ из (2.1) с учетом неравенства $|g_i| \ge \alpha$ получаем, что

$$\left|\Gamma^i_i\right| \le (\lambda|g_i|)^{-1}\left[|f(u+\lambda g) - f(u)| + \lambda \sum_{j=1,\, j\ne i}^{n} \left|\Gamma^i_j g_j\right|\right] \le nL/\alpha.$$

Таким образом,

$$\|\Gamma^i(u, g, e, z, \lambda, \beta)\| \le (L/\alpha)(\alpha^2 n + n^2 - \alpha^2)^{1/2}$$
$$\forall g \in S_1,\ e \in V,\ i \in I(g, \alpha),\ z \in P,\ z \le z_0,\ \lambda \in (0, \lambda_0],\ \beta \in (0, 1].$$

Обозначим $R = (L/\alpha)(\alpha^2 n + n^2 - \alpha^2)^{1/2}$. Тогда получим, что $\|v\| \le R$ для любых $v \in D_0(u, z, \lambda, \beta)$ и $z \in P$, $z \le z_0$, $\lambda \in (0, \lambda_0]$, $\beta \in (0, 1]$. Лемма доказана.

Используя отображение $D_0(u, z, \lambda, \beta)$, можно аппроксимировать субдифференциал Кларка достаточно широкого класса негладких функций. В [14] доказана следующая

Теорема 1. *Пусть f – полугладкая функция и ее производная по направлению $f'(u, g)$ полунепрерывна сверху по u при фиксированном $g \in \mathbb{E}_n$. Тогда*

$$\partial f(u) = B(u).$$

Следствие 1. Пусть выполняются все условия теоремы 1. Тогда для любого $\varepsilon > 0$ существуют $z_0 \in P$, $\lambda_0 > 0$ и $\beta_0 > 0$ такие, что для любых $z \in P$, $z < z_0$, $\lambda \in (0, \lambda_0)$, $\beta \in (0, \beta_0)$

$$H(\partial f(u), D_0(u, z, \lambda, \beta)) < \varepsilon, \tag{2.2}$$

где

$$H(A, B) = \max\{\max_{u \in A} \min_{v \in B} \|u - v\|, \max_{v \in B} \min_{u \in A} \|u - v\|\},$$

$A, B \subset \mathbb{E}_n$ – метрика Хаусдорфа.

Обозначим

$$\partial f(u + S_\delta) = \bigcup\{\partial f(y), y \in S_\delta(u)\},$$

где $S_\delta = S_\delta(0)$.

Пусть $u \in \mathbb{E}_n$ – любая точка и в этой точке функция f удовлетворяет следующему условию: для любых $\varepsilon > 0$, $\mu > 0$ найдутся такие $z_0 \in P$, $\lambda_0 > 0$, $\beta_0 > 0$, $\tau > 0$, что

$$D_0(y, z, \lambda, \beta) \subset \partial f(u + S_\varepsilon) + S_\mu \tag{2.3}$$

при любых $y \in S_\tau(u)$, $z \in P$, $z < z_0$, $\lambda \in (0, \lambda_0)$, $\beta \in (0, \beta_0)$.

В дальнейшем рассматриваются только те функция, которые удовлетворяют локальному условию Липшица и для которых верно соотношение (2.3). Соотношение (2.2) показывает, что класс таких функций достаточно широк.

3. ПОСТРОЕНИЕ НАПРАВЛЕНИЯ СПУСКА МИНИМИЗИРУЕМОЙ ФУНКЦИИ

Обратимся к построению направления спуска минимизируемой функции. Сначала докажем следующую лемму.

Лемма 4. *Пусть в точке $u \in \mathbb{E}_n$ для заданных $z \in P$, $\lambda > 0$, $\beta > 0$ и $\delta > 0$ выполняется неравенство*

$$\min\{\|v\|: v \in D_0(u, z, \lambda, \beta)\} \ge \delta.$$

Тогда существует $g^0 \in S_1$ такое, что

$$f(u + \lambda g^0) - f(u) \le -\lambda\delta.$$

Доказательство. Пусть $v^0 \in D_0(u, z, \lambda, \beta)$ таково, что

$$\|v^0\| = \min\{\|v\|: v \in D_0(u, z, \lambda, \beta)\} \ge \delta.$$

Обозначим $g^0 = -\|v^0\|^{-1}v^0$. Так как $D_0(u, z, \lambda, \beta)$ – компактное и выпуклое множество, то

$$\max\{\langle v, g^0\rangle, v \in D_0(u, z, \lambda, \beta)\} = -\|v_0\| \le -\delta.$$

Отсюда и из (2.1) имеем

$$f(u+\lambda g^0)-f(u) = \lambda\langle\Gamma^i(u, g^0, e, z, \lambda, \beta), g^0\rangle \le \lambda\max\{\langle v, g^0\rangle, v\in D_0(u, z, \lambda, \beta)\} = -\lambda\|v^0\| \le -\lambda\delta.$$

Лемма доказана.

Лемма 4 указывает способ вычисления направления убывания функции. Для этого необходимо решить следующую задачу минимизации:

$$\|v\| \longrightarrow \min, \quad v\in D_0(u, z, \lambda, \beta). \tag{3.1}$$

В общем случае вычисление множества $D_0(u, z, \lambda, \beta)$ достаточно трудно. Поэтому решить задачу (3.1) не всегда представляется возможным. Учитывая это, задачу (3.1) можно заменить более простой:

$$\|v\| \longrightarrow \min, \quad v\in \bar{D}, \tag{3.2}$$

где множество $\bar{D}$ получается путем "последовательного улучшения" аппроксимации множества $D_0(u, z, \lambda, \beta)$ и при этом $\bar{D} \subset D_0(u, z, \lambda, \beta)$. Множество $\bar{D}$ является выпуклой оболочкой конечного числа точек. Поэтому (3.2) является задачей квадратичного программирования. Для решения задач типа (3.2) разработаны специальные алгоритмы (см. [15], [16]).

Для построения направления убывания функции f предлагается следующий

Алгоритм

Пусть заданы $z\in P$, $\lambda>0$, $\beta\in(0, 1]$, некоторое число $c\in(0, 1)$ и достаточно малое число $\delta>0$.

Шаг 1. Взять любые $g^1\in S_1$, $e\in V$, $i\in I(g^1, \alpha)$ и вычислить дискретный градиент $v^1=\Gamma^i(u, g^1, e, z, \lambda, \beta)$. Положить $\bar{D}_1(u)=\{v_1\}$. Пусть уже построено множество $\bar{D}_k(u)$.

Шаг 2. Вычислить вектор $\|w^k\|=\min\{\|w\|: w\in\bar{D}_k(u)\}$. Если

$$\|w^k\|\le\delta, \tag{3.3}$$

то останов.

Шаг 3. Вычислить направление $g^{k+1}=-\|w^k\|^{-1}w^k$.

Шаг 4. Если

$$f(u+\lambda g^{k+1})-f(u)\le-\lambda c\|w^k\|, \tag{3.4}$$

то останов.

Шаг 5. Вычислить дискретный градиент $v^{k+1}=\Gamma^i(u, g^{k+1}, e, z, \lambda, \beta)$, $i\in I(g^{k+1}, \alpha)$, построить множество $\bar{D}_{k+1}(u)=\mathrm{co}\{\bar{D}_k(u)\cup\{v^{k+1}\}\}$, положить $k:=k+1$ и перейти к шагу 2.

Сначала покажем, что если не выполняются оба условия для прекращения работы алгоритма, то новый дискретный градиент $v^{k+1}\notin\bar{D}_k(u)$, т.е. алгоритм позволяет улучшить аппроксимацию множества $D_0(u, z, \lambda, \beta)$. Действительно, в этом случае $\|w^k\|>\delta$ и $f(u+\lambda g^{k+1})-f(u)>-\lambda c\|w^k\|$ и поэтому из (2.1) получаем, что

$$f(u+\lambda g^{k+1})-f(u) = \lambda\langle\Gamma^i(u, g^{k+1}, e, z, \lambda, \beta), g^{k+1}\rangle = \lambda\langle v^{k+1}, g^{k+1}\rangle > -\lambda c\|w^k\|.$$

Отсюда имеем

$$\langle v^{k+1}, w^k\rangle < c\|w^k\|^2. \tag{3.5}$$

С другой стороны, так как $w^k=\operatorname{argmin}\{\|w\|: w\in\bar{D}_k(u)\}$, то из необходимого условия для минимума следует, что для любого $w\in\bar{D}_k(u)$

$$\langle w^k, w-w^k\rangle\ge 0,$$

или

$$\langle w^k, w\rangle\ge\|w^k\|^2.$$

Тогда отсюда и из (3.5) получаем, что $v^{k+1}\notin\bar{D}_k(u)$.

Теперь покажем, что описанный алгоритм является конечным. Для этого достаточно получить верхнюю оценку числа m вычисленных дискретных градиентов, после которых выполнится неравенство

$$\|w^m\| \le \delta. \tag{3.6}$$

Ясно, что при произвольном $t \in [0, 1]$

$$\|w^{k+1}\|^2 \le \|tv^{k+1} + (1-t)w^k\|^2,$$

или

$$\|w^{k+1}\|^2 \le \|w^k\|^2 + 2t\langle w^k, v^{k+1} - w^k\rangle + t^2\|v^{k+1} - w^k\|^2.$$

Из леммы 3 следует, что существует $R > 0$ такое, что $\|v^{k+1} - w^k\| \le 2R$. Тогда, учитывая неравенство (3.5), имеем

$$\|w^{k+1}\|^2 \le \|w^k\|^2 - 2t(1-c)\|w^k\|^2 + 4t^2R^2.$$

При $t = (1-c)(2R)^{-2}\|w^k\|^2 \in (0, 1)$ получаем

$$\|w^{k+1}\|^2 \le \{1 - [(1-c)(2R)^{-1}\|w^k\|]^2\}\|w^k\|^2. \tag{3.7}$$

Для заданного $\delta \in (0, R)$ подсчитаем верхнюю оценку числа m вычисленных дискретных градиентов, после которых выполнится неравенство (3.6).

Из (3.7) и условия $\|w^k\| > \delta$ $(k = 1, 2, \dots, m-1)$ получаем неравенство

$$\|w^{k+1}\|^2 \le \{1 - [(1-c)(2R)^{-1}\delta]^2\}\|w^k\|^2.$$

Обозначим $r = 1 - [(1-c)(2R)^{-1}\delta]^2$. Ясно, что $r \in (0, 1)$. Следовательно,

$$\|w^m\|^2 \le r\|w^{m-1}\|^2 \le \dots \le r^{m-1}\|w^1\|^2 \le r^{m-1}R^2.$$

Если $r^{m-1}R^2 \le \delta^2$, то выполняется также неравенство (3.6) и поэтому

$$m \le 2\log_2(\delta/R)/\log_2 r + 1.$$

Таким образом, доказана

Теорема 2. *Пусть функция f, определенная на $\mathbb{E}_n$, удовлетворяет локальному условию Липшица, число $R > 0$ таково, что в точке $u \in \mathbb{E}_n$ для заданных $z \in P$, $\lambda > 0$, $\beta \in (0, 1]$*

$$\max\{\|v\|: v \in D_0(u, z, \lambda, \beta)\} \le R$$

и $c \in (0, 1)$, $\delta \in (0, R)$ – заданные числа. Тогда максимум за m шагов в алгоритме для вычисления направления спуска выполнится одно из условий (3.3), (3.4), где

$$m \le 2\log_2(\delta/R)/\log_2 r + 1, \quad r = 1 - [(1-c)(2R)^{-1}\delta]^2.$$

4. ОПИСАНИЕ АЛГОРИТМА И ИССЛЕДОВАНИЕ ЕГО СХОДИМОСТИ

Для решения задачи (1.1) предлагается следующий метод с использованием отображения $D_0(u, z, \lambda, \beta)$.

Пусть заданы последовательности $\{\delta_k\}$, $\{z_k\}$, $\{\lambda_k\}$, $\{\beta_k\}$ ($\delta_k > 0$, $z_k \in P$, $\lambda_k > 0$, $\beta_k > 0$, $\delta_k \longrightarrow 0$, $z_k \longrightarrow 0$, $\lambda_k \longrightarrow 0$, $\beta_k \longrightarrow 0$, $k \longrightarrow +\infty$), число $c \in (0, 1)$ и любая начальная точка $u^0 \in \mathbb{E}_n$. Обозначим

$$M(u^0) = \{u \in \mathbb{E}_n | f(u) \le f(u^0)\}.$$

Пусть уже найдено k-е приближение u^k, Тогда $(k+1)$-е приближение u^{k+1} строится с помощью следующего алгоритма.

Шаг 1. Положить $s = 0$ и $u_s^k = u^k$.

Шаг 2. Применить алгоритм для построения направления спуска при $u = u_s^k$, $\delta = \delta_k$, $z = z_k$, $\lambda = \lambda_k$, $\beta = \beta_k$ и $c \in (0, 1)$. После останова этого алгоритма для некоторого $m > 0$ либо определяются элемент $\|v_s^k\| = \min\{\|v\|: v \in \overline{D}_m(u_s^k)\}$ и направление $g_s^k = -\|v_s^k\|^{-1}v_s^k$ такие, что

$$f(u_s^k + \lambda_k g_s^k) - f(u_s^k) \le -c\lambda_k\|v_s^k\|, \tag{4.1}$$

либо выполняется неравенство $\| v_s^k \| \le \delta_k$.

Шаг 3. Если

$$\left\| v_s^k \right\| \le \delta_k, \tag{4.2}$$

то положить $u^{k+1} = u_s^k$, $k := k + 1$ и перейти к шагу 1.

Шаг 4. Построить итерацию

$$u_{s+1}^k = u_s^k + \sigma_s g_s^k, \tag{4.3}$$

где σ_s определяется из условия

$$f(u_s^k + \sigma_s g_s^k) = \inf\{ f(u_s^k + \sigma g_s^k), \sigma \ge 0\}. \tag{4.4}$$

Шаг 5. Положить $s = s + 1$ и перейти к шагу 2.

Заметим, что $(k+1)$-е приближение u^{k+1} строится в результате выполнения некоторого этапа. В самом начале этапа выбираются параметры $z_k \in P$, $\lambda_k > 0$, $\beta_k > 0$, $\delta_k > 0$, и они остаются постоянными до конца этапа. Этап заканчивается при выполнении условия (4.2).

Перейдем к исследованию сходимости предложенного метода. Верна

Теорема 3. *Пусть функция f локально-липшицева, удовлетворяет условию* (2.3) *и $M(u^0)$ – ограниченное множество для любого $u^0 \in \mathbb{E}_n$. Тогда любая предельная точка последовательности $\{u^k\}$, полученная методом* (4.3), (4.4), *принадлежит множеству $U^0 = \{u \in \mathbb{E}_n | 0 \in \partial f(u)\}$.*

Доказательство. Так как $M(u^0)$ – ограниченное множество, а функция f локально-липшицева, то $f_* = \inf\{f(u), u \in \mathbb{E}_n\} > -\infty$. С другой стороны, если выполняется условие (4.1), то

$$f(u_{s+1}^k) - f(u_s^k) \le f(u_s^k + \lambda_k g_s^k) - f(u_s^k) \le -c\lambda_k \left\| v_s^k \right\| < 0. \tag{4.5}$$

Поэтому $f(u_{s+1}^k) < f(u_s^k)$ $(k, s = 0, 1, 2, \ldots)$ и, следовательно, $u_s^k \in M(u^0)$ для любых k, $s = 0, 1, \ldots$. Таким образом, $u^k \in M(u^0)$ для любого k. Из ограниченности снизу функции f получаем, что предел

$$\lim_{s \to \infty} f(u_s^k) \ge f_*$$

существует для любого фиксированного k. Из (4.5) следует

$$\left\| v_s^k \right\| \le (c\lambda_k)^{-1} [f(u_s^k) - f(u_{s+1}^k)].$$

Так как λ_k постоянно для фиксированного k, то существует такое конечное $s(k) > 0$, что $\| v_{s(k)}^k \| \le \delta_k$, где для некоторого $m > 0$

$$\left\| v_{s(k)}^k \right\| = \min\{ \|v\| : v \in \bar{D}_m(u_{s(k)}^k)\}.$$

Значит, условие (4.2) на каждом этапе выполняется за конечное число шагов. При этом $u^{k+1} = u_{s(k)}^k$. Тогда из включения $\bar{D}_m(u^{k+1}) \subset D_0(u^{k+1}, z_k, \lambda_k, \beta_k)$ получаем, что

$$\min\{ \|w\| : w \in D_0(u^{k+1}, z_k, \lambda_k, \beta_k)\} \le \left\| v_{s(k)}^k \right\| \le \delta_k \tag{4.6}$$

для любого $k = 1, 2, \ldots$.

Из ограниченности множества $M(u^0)$ следует ограниченность последовательности $\{u^k\}$. Тогда из теоремы Больцано–Вейерштрасса получаем, что последовательность $\{u^k\}$ имеет хотя бы одну предельную точку. Пусть u^* – любая предельная точка этой последовательности. Не умаляя общности, будем считать, что $u^k \longrightarrow u^*$ при $k \longrightarrow +\infty$. Допустим противное. Пусть $u^* \notin U^0$. Тогда существует $\delta > 0$ такое, что

$$\min\{\|v\| : v \in \partial f(u^*)\} \ge 4\delta.$$

Из полунепрерывности сверху отображения $\partial f(u)$ получаем, что для $\delta > 0$ найдется такое $\tau > 0$, что для любых $y \in S_\tau(u^*)$

$$\partial f(y) \subset \partial f(u^*) + S_{2\delta}$$

и, значит,

$$\partial f(u^* + S_\tau) \subset \partial f(u^*) + S_{2\delta}.$$

Отсюда следует, что

$$\inf\{\|v\|: v \in \partial f(u^* + S_\tau)\} \geq 2\delta.$$

Так как функция f удовлетворяет условию (2.3), то в точке u^* для $\tau > 0$ и $\mu = \delta$ найдутся такие $\varepsilon > 0$, $z_0 \in P$, $\lambda_0 > 0$, $\beta_0 > 0$, что

$$D_0(y, z, \lambda, \beta) \subset \partial f(u^* + S_\tau) + S_\mu$$

для любых $y \in S_\varepsilon(u^*)$, $z \in P$, $z < z_0$, $\lambda \in (0, \lambda_0)$, $\beta \in (0, \beta_0)$. Поэтому для любых $y \in S_\varepsilon(u^*)$, $z \in P$, $z < z_0$, $\lambda \in (0, \lambda_0)$, $\beta \in (0, \beta_0)$

$$\min\{\|v\|: v \in D_0(y, z, \lambda, \beta)\} \geq \delta. \tag{4.7}$$

Так как $u^k \longrightarrow u^*$, $z_k \longrightarrow 0$, $\lambda_k \longrightarrow +0$, $\beta_k \longrightarrow +0$, $\delta_k \longrightarrow +0$ при $k \longrightarrow +\infty$, то найдется такое $k_0 > 0$, что $u^k \in S_\varepsilon(u^*)$, $z_k < z_0$, $\lambda_k \in (0, \lambda_0)$, $\beta_k \in (0, \beta_0)$, $\delta_k < \delta$ для любых $k \geq k_0$. Тогда из (4.7) получаем, что

$$\min\{\|v\|: v \in D_0(u^{k+1}, z_k, \lambda_k, \beta_k)\} \geq \delta > \delta_k.$$

Это противоречит соотношению (4.6). Теорема доказана.

Теперь рассмотрим случай, когда f – выпуклая функция.

Теорема 4. *Пусть относительно функции f выполняются следующие условия*:

1) *f – собственная выпуклая функция*;

2) $\mathrm{dom} f \equiv \{u \in \mathbb{E}_n | f(u) < +\infty\} = \mathbb{E}_n$.

3) *функция f удовлетворяет условию* (2.3).

Пусть, кроме того, $M(u^0)$ – ограниченное множество для любого $u^0 \in \mathbb{E}_n$. Тогда любая предельная точка последовательности $\{u^k\}$, полученная методом (4.3), (4.4), *принадлежит множеству $U^* = \{u \in \mathbb{E}_n | f(u) = f_*\}$.*

Доказательство. При выполнении условий 1) и 2) функция f удовлетворяет локальному условию Липшица, является полугладкой и существует ее производная по любому направлению. Производная по направлению $f'(u, g)$ полунепрерывна сверху по u при фиксированном $g \in \mathbb{E}_n$. Субдифференциал Кларка функции f совпадает с субдифференциалом этой функции в смысле выпуклого анализа. Следовательно, выполняются все условия теоремы 1, и поэтому для функции f верно соотношение (2.2). Тем самым получаем, что все условия теоремы 3 выполняются. Так как f – выпуклая функция, то в этом случае множества U^0 и U^* совпадают. Таким образом, доказательство теоремы 4 следует из теоремы 3.

5. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Приведем результаты вычислительного эксперимента. Для отыскания решения задач липшицевой оптимизации в постановке (1.1) была разработана программа, реализующая метод дискретного градиента.

Отметим следующие особенности условия проведения вычислительного эксперимента. Линейный поиск производился методом дихотомии. Проектирование начала координат $\mathbb{E}_n$ на выпуклую оболочку системы векторов производилось по первому методу из [15, с. 338]. Вычисления проводились на ПЭВМ IBM PC/AT 386. При тестировании алгоритма какая-либо его "подстройка" к задаче не допускалась.

Во всех примерах число $c = 0.2$. Параметры $z_k(\lambda) = \lambda^{1.4}$, $\beta_k = 1$ для любых k и для всех примеров, а параметр λ выбирался следующим образом: $\lambda_k = \tau\lambda_{k-1}$, где $\tau = 0.75$ для всех примеров.

При описании результатов работы алгоритма используются следующие обозначения: $f = f(u)$ – минимизируемая функция; n – число неизвестных задачи; u^0 – начальная точка, u^* – решение задачи; f_* – значение задачи; $\delta_k = f(u^k) - f_*$ – точность, соответствующая точке u^k;

$$l = \sum_{\delta(u^k) \geq \delta} s(k)$$

есть общее число итераций для достижения точности $\delta > 0$, где $s(k)$ – число итераций на k-м этапе; m – число вычислений дискретных градиентов функции f для достижения точности $\delta > 0$.

Для проведения вычислительного эксперимента были использованы следующие тестовые задачи (см. [17], [18]).

Задача 1: $f(u) = \max\{u_1^2 + u_2^4, (2-u_1)^2 + (2-u_2)^2, 2e^{-u_1+u_2}\}$, $u \in \mathbb{E}_2$, $u^0 = (1, -0.1)$, $u^* = (1.1390, 0.8996)$, $f_* = 1.952$.

Задача 2: $f(u) = \max\{u_1^4 + u_2^2, (2-u_1)^2 + (2-u_2)^2, 2e^{-u_1+u_2}\}$, $u \in \mathbb{E}_2$, $u^0 = (2, 2)$, $u^* = (1, 1)$, $f_* = 2$.

Задача 3: $f(u) = \max\{5u_1 + u_2, -5u_1 + u_2, u_1^2 + u_2^2 + 4u_2\}$, $u \in \mathbb{E}_2$, $u^0 = (1, 1)$, $u^* = (0, -3)$, $f_* = -3$.

Задача 4: $f(u) = \max\{f_i(u), i = 1, 2, 3\}$, $f_1(u) = u_1^2 + u_2^2$, $f_2(u) = u_1^2 + u_2^2 + 10(-4u_1 - u_2 + 4)$, $f_3(u) = u_1^2 + u_2^2 + 10(-u_1 - 2u_2 + 6)$, $u \in \mathbb{E}_2$, $u^0 = (-1, 5)$, $u^* = (1.2, 2.4)$, $f_* = 7.2$.

Задача 5: $f(u) = \max\{-u_1 - u_2, -u_1 - u_2 + u_1^2 + u_2^2 - 1\}$, $u \in \mathbb{E}_2$, $u^0 = (0.5, 0.5)$, $u^* = (2^{-1/2}, 2^{-1/2})$, $f_* = -2^{1/2}$.

Задача 6: $f(u) = -u_1 + 20\max\{u_1^2 + u_2^2 - 1, 0\}$, $u \in \mathbb{E}_2$, $u^0 = (0.8, 0.6)$, $u^* = (1, 0)$, $f_* = -1$.

Задача 7: $f(u) = -u_1 + 2(u_1^2 + u_2^2 - 1) + 1.75|u_1^2 + u_2^2 - 1|$, $u \in \mathbb{E}_2$, $u^0 = (-1, 1)$, $u^* = (1, 0)$, $f_* = -1$.

Задача 8: $f(u) = \max\{f_1(u), f_1(u) + 10f_2(u), f_1(u) + 10f_3(u), f_1(u) + 10f_4(u)\}$, $f_1(u) = u_1^2 + u_2^2 + 2u_3^2 + u_4^2 - 5u_1 - 5u_2 - 21u_3 + 7u_4$; $f_2(u) = u_1^2 + u_2^2 + u_3^2 + u_4^2 + u_1 - u_2 + u_3 - u_4 - 8$; $f_3(u) = u_1^2 + 2u_2^2 + u_3^2 + 2u_4^2 - u_1 - u_4 - 10$; $f_4(u) = u_1^2 + u_2^2 + u_3^2 + 2u_1 - u_2 - u_4 - 5$; $u \in \mathbb{E}_4$, $u^0 = (0, 0, 0, 0)$, $u^* = (0, 1, 2, -1)$, $f_* = -44$.

Задача 9: $f(u) = \max\{b_i\|u - a^i\|^2, i = 1, \ldots, 10\}$, $u \in \mathbb{E}_5$, $b = (1, 5, 10, 2, 4, 3, 1.7, 2.5, 6, 3.5)$, $a^1 = (0, 0, 0, 0, 0)$, $a^2 = (2, 1, 1, 1, 3)$, $a^3 = (1, 2, 1, 1, 2)$, $a^4 = (1, 4, 1, 2, 2)$, $a^5 = (3, 2, 1, 0, 1)$, $a^6 = (0, 2, 1, 0, 1)$, $a^7 = (1, 1, 1, 1, 1)$, $a^8 = (1, 0, 1, 2, 1)$, $a^9 = (0, 0, 2, 1, 0)$, $a^{10} = (1, 1, 2, 0, 0)$, $u^0 = (0, 0, 0, 0, 1)$, $u^* = (1.12434, 0.97945, 1.47770, 0.92023, 1.12429)$, $f_* = 22.60016$.

Задача 10: $f(u) = \max\{u_i^2, i = 1, 2, \ldots, n\}$, $u \in \mathbb{E}_n$, $u^0 = (i, i = 1, 2, \ldots,]n/2[, -i, i =]n/2[+1, \ldots, n)$, $u^* = (0, \ldots, 0)$, $f_* = 0$.

Задача 11: $f(u) = \max\{|u_i|, i = 1, 2, \ldots, n\}$, $u \in \mathbb{E}_n$, $u^0 = (i, i = 1, 2, \ldots,]n/2[, -i, i =]n/2[+1, \ldots, n)$, $u^* = (0, \ldots, 0)$, $f_* = 0$.

Задача 12: $f(u) = \sum_{i=1}^n i^3 |u_i|^i$, $u \in \mathbb{E}_n$, $u^0 = (\gamma, \gamma/2, \ldots, \gamma/n)$, $\gamma = 10$, $u^* = (0, \ldots, 0)$, $f_* = 0$.

Задача 13: $f(u) = 4|u_1 - u_2| + |u_1 + u_2| + |u_2 - u_3| + |u_2 + u_3| + |u_3 - u_4| + |u_3 + u_4|$, $u \in \mathbb{E}_4$, $u^0 = (1, 2, 1, 1)$, $u^* = (0, 0, 0, 0)$, $f_* = 0$.

Задача 14: $f(u) = \sum_{j=1}^{100} |\sum_{i=1}^n (u_i - u_i^*) t_j^{i-1}|$, $u \in \mathbb{E}_n$, $t_j = 0.01j$, $j = 1, \ldots, 100$, $u^0 = (0, \ldots, 0)$, $u^* = (1/n, \ldots, 1/n)$, $f_* = 0$.

Задача 15: $f(u) = |u_1 - 1| + 100|u_2 - |u_1||$, $u \in \mathbb{E}_2$, $u^0 = (-1.2, 1)$, $u^* = (1, 1)$, $f_* = 0$.

Задача 16: $f(u) = |u_1 - 1| + 100|u_2 - |u_1|| + 90|u_4 - |u_3|| + |u_3 - 1| + 10.1(|u_2 - 1| + |u_4 - 1|) + 4.95(|u_2 + u_4 - 2| - |u_2 - u_4|)$, $u \in \mathbb{E}_4$, $u^0 = (0, 0, 0, 0)$, $u^* = (1, 1, 1, 1)$, $f_* = 0$.

Задача 17: $f(u) = \sum_{j=1}^{100} |\sum_{i=1}^n (u_i - u_i^*) t_j^{i-1}| - \max\{|\sum_{i=1}^n (u_i - u_i^*) t_j^{i-1}|, j = 1, \ldots, 100\}$, $u \in \mathbb{E}_n$, $t_j = 0.01j$, $j = 1, \ldots, 100$, $u^* = (1/n, \ldots, 1/n)$, $f_* = 0$.

В табл. 1 приведены данные, характеризующие работу алгоритма при отыскании решений задач 1–17.

6. НЕКОТОРЫЕ ВЫВОДЫ

По результатам численных экспериментов можно сделать следующие выводы.

Таблица 1

N	n	$\delta = 1.0E\text{–}02$		$\delta = 1.0E\text{–}03$		$\delta = 1.0E\text{–}04$	
		l	m	l	m	l	m
1	2	5	8	6	10	11	23
2	2	3	5	3	5	3	5
3	2	5	8	5	8	17	44
4	2	7	11	8	13	8	13
5	2	1	1	1	1	2	3
6	2	6	12	8	16	15	36
7	2	4	5	5	7	9	17
8	4	17	43	31	100	41	143
9	5	18	44	23	72	30	113
10	5	27	47	33	66	37	80
	10	102	193	110	212	115	239
	15	214	411	219	424	230	455
11	5	27	47	37	74	48	103
	10	116	216	133	266	146	302
	15	250	485	267	526	288	594
12	5	91	92	91	92	91	92
	10	89	89	90	92	93	114
	15	94	94	95	96	101	152
13	4	17	36	27	75	35	106
14	5	57	140	82	244	95	299
	10	66	154	140	596	208	1025
	20	75	211	174	865	340	1982
15	2	8	17	8	17	18	45
16	4	36	121	45	165	60	240
17	5	52	113	108	346	129	439
	10	57	127	115	514	189	993
	15	57	105	83	373	312	2162

Предложенный алгоритм находит приближенное решение с заданной точностью, используя не такое уж большое количество вычислений дискретных градиентов. Здесь исключение составляют только задачи 14 и 17 с $n = 15$, 20 и с точностью $\delta = 0.0001$. Но следует отметить, что эти задачи очень трудоемки и функции в них имеют много точек недифференцируемости.

По мере приближения к точному решению среднее число вычисленных в одной итерации дискретных градиентов возрастает. Это связано с необходимостью более точной аппроксимации субдифференциала по мере приближения к точному решению. Поэтому по мере приближения к точному решению время выполнения одной итерации увеличивается.

Результаты численных экспериментов зависят от выбора параметров $z \in P$, $\lambda > 0$, $\beta > 0$ и $c \in \in (0, 1)$. При более удачном выборе этих параметров для каждой конкретной задачи можно получить более хорошие результаты.

Предложенный алгоритм позволил решить все задачи с заданной точностью. Можно считать, что метод дискретного градиента – надежный и эффективный метод решения задач негладкой оптимизации с размерностью по крайней мере $n \le 20$.

Результаты численных экспериментов показывают, что метод дискретного градиента можно использовать для минимизации как выпуклых, так и невыпуклых полугладких функций.

Автор благодарит рецензентов за полезные замечания, способствовавшие улучшению статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Шор Н.З.* Методы минимизации недифференцируемых функций и их приложения. Киев: Наук. думка, 1979.
2. *Демьянов В.Ф., Васильев Л.В.* Недифференцируемая оптимизация. М.: Наука, 1981.
3. *Нурминский Е.А.* Численные методы выпуклой оптимизации. М.: Наука, 1990.
4. *Ржевский С.В.* Монотонные методы выпуклого программирования. Киев: Наук. думка, 1993.
5. *Кларк Ф.* Оптимизация и негладкий анализ. М.: Наука, 1988.
6. *Михалевич В.С., Гупал А.М., Норкин В.И.* Методы невыпуклой оптимизации. М.: Наука, 1987.
7. *Дорофеев П.А.* О некоторых свойствах метода обобщенного градиента // Ж. вычисл. матем. и матем. физ. 1985. Т. 25. № 2. С. 181–189.
8. *Завриев С.К., Перевозчиков А.Г.* Метод стохастического обобщенного градиента для решения минимаксных задач со связанными переменными // Ж. вычисл. матем. и матем. физ. 1990. Т. 30. № 4. С. 491–500.
9. *Перевозчиков А.Г.* О сходимости метода обобщенных градиентов Кларка в задачах минимизации липшицевых функций // Ж. вычисл. матем. и матем. физ. 1992. Т. 32. № 2. С. 208–216.
10. *Kiwiel K.C.* Methods of descent for nondifferentiable optimization. Berlin: Springer, 1985.
11. *Zowe J.* Nondifferentiable optimization – a motivation and a short introduction into the subgradient and the bundle concept // ASI Proc. Comput. Math. Program. 1985. Vol. F15.
12. *Багиров А.М.* Об одном методе аппроксимации субдифференциала // Ж. вычисл. матем. и матем. физ. 1992. Т. 32. № 4. С. 652–658.
13. *Багиров А.М.* О непрерывной аппроксимации субдифференциала функции максимума // Кибернетика и системный анализ. 1993. № 4. С. 180–184.
14. *Багиров А.М., Гасанов А.А.* О методе аппроксимации квазидифференциала // Ж. вычисл. матем. и матем. физ. 1995. Т. 35. № 4. С. 511–519.
15. *Демьянов В.Ф., Малоземов В.Н.* Введение в минимакс. М.: Наука, 1972.
16. *Wolfe P.* Finding the nearest point in a polytope // Math. Program. 1976. V. 11. № 2. P. 128–149.
17. *Lemarechal C.* Numerical experiments in nonsmooth optimization // Progress in Nondifferentiable Optimization. Laxenburg, Austria, 1982. P. 61–84.
18. *Нестеров Ю.Е., Пурмаль Е.И.* Анализ эффективности методов негладкой оптимизации. М.: ЦЭМИ АН СССР, 1984.