Теорія оптимальних рішень

УДК 519.8

А.Ф. ГОДОНОГА, Б.М. ЧУМАКОВ

ВЕРОЯТНОСТНО-ГРАДИЕНТНЫЙ МЕТОД РЕШЕНИЯ НЕКОТОРЫХ ЗАДАЧ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

Рассматриваются два стохастических варианта градиентного метода с программным способом регулировки шага. Указаны определенные достаточные условия, при которых описанные алгоритмы сходятся к множеству оптимальных решений с вероятностью единица.

Задачи оптимизации занимают значительное место в теории и практики принятия обоснованных решений. Несмотря на нынешнее, поистине революционное, развитие вычислительных средств, интерес к вопросу, эффективности способов получения таких решений, остается актуальным. В данной работе рассматривается вероятностный подход к решению некоторых оптимизационных задач в сочетании с идеями градиентного метода, в котором программно реализуется выбор шагового множителя градиентного метода.

Рассматривается следующая задача:

$$\begin{cases} F(x) \to \min \\ x \in X, \end{cases} \tag{1}$$

где X – выпуклое компактное множество эвклидово пространство E^n. ε – окрестность множества X, определяется следующим образом: $V(X,\varepsilon) = \bigcup\limits_{x\in X} V(x,\varepsilon)$. $V(x,\varepsilon)$ сферическая окрестность радиуса $\varepsilon > 0$ точки $x \in E^n$, $V(x,\varepsilon) = \{y \in E^n : \|x-y\| < \varepsilon\}$. Пусть, при некотором $\varepsilon > 0$, функция $F(x)$ – выпуклая и непрерывно дифференцируемая [1, 2] (с непрерывным градиентом) на $V(X,\varepsilon)$. Следовательно, для $\forall x \in X$ определен следующий вектор:

$$GF(x) = \left(g_{F,1}(x), \ldots, g_{F,n}(x)\right) = \\ = \left(\frac{\partial F(x)}{\partial x_1}, \ldots, \frac{\partial F(x)}{\partial x_i}, \ldots, \frac{\partial F(x)}{\partial x_n}\right),$$

где $GF(x)$ – градиент функции $F(x)$ и норма $\|GF(x)\| = \sqrt{\sum_{i=1}^{n} g_{F,i}{}^2(x)}$ – непрерывная функция на X и, следовательно, существует константа $C < \infty$, при которой $\|GF(x)\| \le C$ для $\forall x \in X$. Тогда следует, что справедливы и неравенства вида

$$\|g_{F,i}(x)\| \le C, \forall i = \overline{1,n}, \ \forall x \in X.$$

Численный метод, рассматривающий для решения задачи (1), имеет итерационный характер. А именно, если предположить что находимся на k-ой итерации и точка x^k известна, то последовательно выполняются следующие действия:

1) в серии $m_k \ge 1$ независимых испытаний наблюдаются значения случайной величины ξ^k, заданной *apriori* известным законом распределения вероятностей

ξ^k	1	2	$\cdots$	n
P	P_1^k	P_2^k	$\cdots$	P_n^k

Таким образом, на каждой k-ой итерации генерируется, в соответствии с заданным законом распределения вероятностей, множество $I_k = \{i_1, i_2, \dots, i_{m_k}\}$, элементы которого являются независимыми наблюдениями над случайной величиной ξ^k, причем предполагается, что

$$P_i^k \ge \underline{P} > 0, \forall i = \overline{1,n}, \forall k = 0,1,\dots \tag{2}$$

В частности, можно ограничиться всего лишь одним наблюдением на каждой итерации, т. е. $m_k = 1$;

2) вычисляется направление движения $g^k{}_F(x^k)$ по правилу:

$$g_F{}^k(x^k) = \left(g_1^k, g_2^k, \dots, g_i^k, \dots, g_n^k\right), \ g_i^k = \begin{cases} g_i^{k-1}, & \text{если } i \notin I_k \\ \dfrac{\partial F(x^k)}{\partial x_i}, & \text{если } i \in I_k \end{cases}, \ \forall i = \overline{1,n}. \tag{3}$$

3) точка x^{k+1} определяется следующим образом:

$$x^{k+1} = X \prod \left(\tilde{x}^{k+1}\right), \quad \text{где} \quad \tilde{x}^{k+1} = x^k - \rho_k \eta^k, \tag{4}$$

где $\prod_X \left(\tilde{x}^{k+1}\right)$ – проекция на множество X элемента $\tilde{x}^{k+1} \in E^n$. Начальная точка x^0 может быть произвольным элементом из множества X.

Пусть числовая последовательность $\{\rho_k\}$ удовлетворяет классическим условиям

$$\rho_k \ge 0, \rho_k \to 0, \sum_{k=1}^{\infty} \rho_k = \infty \tag{5}$$

и, кроме того, допустим, что существует число $\bar{\varepsilon} > 0,$ такое, что для $\forall r \in (0, \bar{\varepsilon}]$ и $\forall \nu \in (0,1),$ сходится следующий ряд [3]:

$$\sum_{k=0}^{\infty} \nu^{L(k,r)} < \infty, \quad \text{где} \quad L(k,r) = \begin{cases} 0, \text{ если } \rho_k \ge r \quad \text{или } k = 0 \\ s_k, \text{ если } \sum\limits_{l=k-s_k}^{k} \rho_l < r \quad \text{и} \sum\limits_{l=k}^{k} \rho_l \ge r \end{cases}, \tag{6}$$

т. е. s_k – наибольшее из всех целых чисел $j \ge 0,$ при которых $\sum\limits_{l=k-j}^{k} \rho_l < r.$

В работе [3] доказано, что существует такая числовая последовательность $\{\rho_k\},$ удовлетворяющая условиям (5) – (6), в частности, если $\rho_k = \dfrac{R}{(k+1)^{\alpha}},$ $R > 0,$ $\alpha \in (0,1].$

Последовательность векторов $\{\eta^k\}$ строится следующим образом:

$$\eta^k = \begin{cases} \dfrac{g^k}{\|g^k\|}, \text{ если } g^k \ne \bar{0}, \ k = 1,2,\ldots \\ \bar{0}, \quad \text{при } g^k = \bar{0} \end{cases}. \tag{7}$$

Справедливо следующее утверждение.

Теорема 1. Пусть выполняются условия (2) – (7). Тогда, для произвольного фиксированного числа $\varepsilon > 0,$ все элементы случайной последовательности $\{x^k\}_{k \ge 0},$ почти наверное [4], принадлежат множеству $V(X^*, 2\varepsilon),$ за исключением, быть может, конечного числа ее членов.

Схема доказательства теоремы. 1. Если $X \subset V(X^*, 2\varepsilon),$ то утверждение очевидно. 2. При $X \setminus V(X^*, 2\varepsilon) \ne \varnothing$ рассматриваются два этапа. На первом этапе доказывается существование подпоследовательности $\{x^{k_l}\} \subset \{x^k\}_{k \ge 0},$ которая почти наверное содержится в $V(X^*, \varepsilon),$ т. е. $P\{\exists \{x^{k_l}\} \subset$ $\subset \{x^k\}_{k \ge 0} : x^{k_l} \in V(X^*, \varepsilon)\} = 1.$ Для этого, при некотором $q > 0$ и натуральном числе $K_q < \infty,$ определяется событие

$$A = \left\{\exists K_q : \forall k \ge K_q, \|x^k - x^*\| \ge \varepsilon, \ \text{ т. е. } x^k \notin V(X^*, \varepsilon), \forall x^* \in X^*\right\},$$

причем, предпологается что вероятность реализации данного события $P(A) \geq q$. Используя механизм построения векторов η_k, свойства целевой функции $F(x)$ и лемму Бореля – Кантелли [4] относительно последовательности событий $\{\bar{B}_F^k\}$, где событие $B_F^k = \{$по крайней мере один раз, на какой-то итерации вида $j = \overline{k - s_k, k}$ генерируется каждое возможное значение случайной величины $\{\ \xi^k\ \};\ k = 0,1,...\}$, получается противоречие состоящее в том, что $q \leq P(A) \leq 0$.

На втором этапе, используя уже полученный результат первого этапа, доказывается, что все, кроме, быть может, конечного числа членов последовательности $\{x^k\}$, содержатся в $V(X^*, 2\varepsilon)$ с вероятностью 1.

Далее рассматривается следующее обобщение задачи (1):

$$\begin{cases} F(x) \to \min, \\ \varphi(x) \leq 0, \\ x \in X. \end{cases} \tag{8}$$

Предположим, что задача (8) имеет решение. Пусть, при некотором $\varepsilon > 0$, функции $F(x)$ и $\varphi(x)$ – выпуклые и непрерывно дифференцируемые (с непрерывными градиентами) на $V(X, \varepsilon)$. Следовательно, для $\forall x \in X$ определены векторы

$$GF(x) = \left(g_{F,1}(x), \ldots, g_{F,n}(x)\right) = \left(\frac{\partial F(x)}{\partial x_1}, \ldots, \frac{\partial F(x)}{\partial x_i}, \ldots, \frac{\partial F(x)}{\partial x_n}\right),$$

$$G\varphi(x) = \left(g_{\varphi,1}(x), \ldots, g_{\varphi,n}(x)\right) = \left(\frac{\partial \varphi(x)}{\partial x_1}, \ldots, \frac{\partial \varphi(x)}{\partial x_i}, \ldots, \frac{\partial \varphi(x)}{\partial x_n}\right).$$

Тогда, очевидно, нормы $\|GF(x)\| = \sqrt{\sum_{i=1}^{n} g_{F,i}{}^2(x)}$, $\|G\varphi(x)\| = \sqrt{\sum_{i=1}^{n} g_{\varphi,i}{}^2(x)}$ – непрерывные функции на X и, следовательно, существует постоянная $C < \infty$, при которой $\|GF(x)\| \leq C$, $\|G\varphi(x)\| \leq C$ для $\forall x \in X$. Непосредственно следует, что справедливы и неравенства $\|g_{F,i}(x)\| \leq C, \|g_{\varphi,i}(x)\| \leq C,\ \forall i = \overline{1, n},\ \forall x \in X$.

Численный метод, который предлагается для решения задачи (8), имеет формально тот же вид, как алгоритм для решения предыдущей задачи. Только в данном случае на каждой итерации k рассматриваются две серии $m_k \geq 1$, $l_k \geq 1$ независимых испытаний в которых наблюдаются две случайные величины ξ^k, ψ^k, заданные дискретными законами распределения вероятностей:

ξ^k	1	2	$\cdots$	n
P	$P^k_{\xi,1}$	$P^k_{\xi,2}$	$\cdots$	$P^k_{\xi,n}$

ψ^k	1	2	$\cdots$	n
P	$P^k_{\psi,1}$	$P^k_{\psi,2}$	$\cdots$	$P^k_{\psi,n}$

Точнее, на каждой итерации k генерируются множества $I_k = \{i_1, i_2, \ldots, i_{m_k}\}$, $J_k = \{j_1, j_2, \ldots, j_{l_k}\}$, элементы которых представляют независимые реализации величин ξ^k, ψ^k, соответственно, с заданными распределениями вероятностей, причем

$$P^k_{\xi,i} \ge \underline{P}_{\xi} > 0,\ \forall i = \overline{1,n},\ \forall k = 0,1,\ldots$$

$$P^k_{\psi,i} \ge \underline{P}_{\psi} > 0,\ \forall i = \overline{1,n},\ \forall k = 0,1,\ldots \tag{9}$$

В частности, также можно брать $m_k = l_k = 1$, т. е. можно ограничиться лишь одним наблюдением ξ^k, ψ^k на каждой итерации. Итак:

1) векторы $g^k_F(x^k)$, $g^k_\varphi(x^k)$ определяются по правилу:

$$g^k_F(x^k) = \left(g^k_{F,1}, \ldots, g^k_{F,i}, \ldots, g^k_{F,n}\right),\ g^k_{F,i} = \left[\begin{array}{ll} g^{k-1}_{F,i}, & \text{если } i \notin I_k, \\ \dfrac{\partial F(x^k)}{\partial x_i}, & \text{если } i \in I_k, \end{array}\right.$$

$$g^k_\varphi(x^k) = \left(g^k_{\varphi,1}, \ldots, g^k_{\varphi,i}, \ldots, g^k_{\varphi,n}\right),\ g^k_{\varphi,i} = \left[\begin{array}{ll} g^{k-1}_{\varphi,i}, & \text{если } i \notin J_k \\ \dfrac{\partial \varphi(x^k)}{\partial x_i}, & \text{если } i \in J_k \end{array}\right. ; \tag{10}$$

2) аналогично определяется точка x^{k+1},

$$x^{k+1} = \prod\nolimits_X \left(\tilde{x}^{k+1}\right),\ \tilde{x}^{k+1} = x^k - \rho_k \eta^k; \tag{11}$$

3) последовательность векторов $\{\eta^k\}$ строится также как и в (7);

4) относительно направления движения g^k, последуем идеи работы [5], но определим

$$g^k = g^k(x^k) = \begin{cases} g^k_F(x^k), & \text{если } \varphi(x^k) \le \tau \\ g^k_\varphi(x^k), & \text{если } \varphi(x^k) > \tau \end{cases},\ \tau > 0. \tag{12}$$

Имеет место, следующее утверждение.

Теорема 2. Пусть последовательность случайных точек $\{x^k\}$ строится в соответствии с правилами (4) – (7), (9) – (12). Тогда, для произвольного числа $\varepsilon > 0$, существует такое число $\tau > 0$, при котором все элементы последовательности $\left\{x^k\right\}_{k\geq 0}$, почти наверное принадлежат множеству $V\left(X^*, 2\varepsilon\right)$, за исключением, быть может, конечного числа ее членов.

Следует отметить, что, в основном, схема доказательства данной теоремы примерно совпадает со схемой доказательства теоремы 1. Основное отличие состоит в следующем. На первом этапе доказательства, при достаточно малом положительным значении τ, для заданного числа $\varepsilon > 0$, можно найти общее число $\delta = \min\left\{\delta_F, \delta_\varphi\right\} > 0$, при котором

$$\left(GF(x^k), x^k - x^*\right) \geq 2\delta \left\|GF(x^k)\right\| \cdot \left\|x^k - x^*\right\|, \text{ если } \varphi(x^k) \leq \tau,$$
$$\left(G\varphi(x^k), x^k - x^*\right) \geq 2\delta \left\|G\varphi(x^k)\right\| \cdot \left\|x^k - x^*\right\|, \text{ если } \varphi(x^k) > \tau.$$

Тем самым, при достаточно больших значений k, для которых выполняются условия $x^k \notin V\left(X^*, \varepsilon\right)$ и $\varphi(x^k) \leq \tau$, получаются неравенства вида $\left({g^k}_F(x^k), x^k - x^*\right) \geq \delta \left\|{g_F}^k(x^k)\right\| \cdot \left\|x^k - x^*\right\|$, а при $x^k \notin V\left(X^*, \varepsilon\right)$ и $\varphi(x^k) > \tau$, неравенства – $\left({g^k}_\varphi(x^k), x^k - x^*\right) \geq \delta \left\|{g_\varphi}^k(x^k)\right\| \cdot \left\|x^k - x^*\right\|$. Эти неравенства являются, соответственно, следствиями следующих событий:

а) B_F^k={по крайней мере один раз, на какой-то итерации вида $j = \overline{k - s_k, k}$ генерируется каждое возможное значение случайной величины ξ^k };

б) B_φ^k={по крайней мере один раз, на какой-то итерации вида $j = \overline{k - s_k, k}$ генерируется каждое возможное значение случайной величины ψ^k }.

Далее определяются события вида $B_1^k = B_F^k \bigcap B_\varphi^k$, причем для $\overline{B}_1^k$, также как и для $\overline{B}_F^k$ из первой теоремы, можно получить неравенство $P\left(\overline{B_1^k}\right) \leq \nu^{s_k}$, в соответствии с которым, почти наверное, существует подпоследовательность последовательности $\left\{x^k\right\}$, которая полностью содержится в $V\left(X^*, \varepsilon\right)$. Второй этап доказательства аналогичен соответствующему этапу доказательства первой теоремы.

В ряде научных работ, в частности [6], доказано, что некоторые методы случайного поиска, при решении определенных классов оптимизационных задач, являются эффективней, нежели детерминированные алгоритмы. Поэтому, отказ от точного вычисления градиента, и тем более применение принципа случайного подбрасывания индексов частных производных, которых следовало бы «время от время» обновить, может не только существенно сократить объем вычислений, но и ускорить процесс достижения приемлемых решений в реальном масштабе времени. Некоторые детерминированные аспекты, рассмотренных здесь методов, анализируются в работе [7].

А.Ф. Годонога, Б.М. Чумаков

ЙМОВІРНІСНО-ГРАДІЄНТНИЙ МЕТОД РОЗВ'ЯЗАННЯ ДЕЯКИХ ЗАДАЧ
ОПУКЛОЇ ОПТИМІЗАЦІЇ

Розглядаються два стохастичних варіанта градієнтного методу з програмним засобом регулювання кроком. Зазначені певні умови, за яких описані алгоритми сходяться до множини оптимальних рішень з імовірністю одиниця.

A.F. Godonoga, B.M. Chumakov

PROBALISTIC-GRADIENT METHOD FOR SOLVING CERTAIN PROBLEMS
CONVEX OPTIMIZATION

The consider two variants of stochastic gradient method with the programmatically regulation of step. The shown are certain sufficient conditions under which the described algorithms converge to the set of optimal solutions with probability one.

1. *Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М.* Методы оптимизации. – М.: Наука, 1978. – 352 с.
2. *Шор Н.З.* Методы минимизации недифференцируемых функций и их приложения. – Киев: Наук. думка, 1979. – 199 с.
3. *Ермольев Ю.М., Гайворонский А.А.* Стохастический метод решения минимаксных задач // Кибернетика. – 1983. – № 4. – С. 92 – 97.
4. *Ширяев А.Н.* Вероятность. – М.: Наука, 1980. – 575 с.
5. *Поляк Б.Т.* Один общий метод решения экстремальных задач // Докл. АН СССР. – 1967. – **174**. – № 1. – С. 33 – 36.
6. *Шор Н.З., Щепакин М.Б.* Об оценке скорости сходимости метода случайного поиска // Кибернетика. – 1974. – № 4. – С. 55 – 58.
7. *Balan P., Godonoaga A.* Solutionarea unui model diferentiabil cu modificarea succesiva a componentelor gradientilor // Analele ATIC – Chisinau: Evrica, 2007. – P. 25 – 34.

Получено 03.04.2014