УДК 519.8

СУБГРАДИЕНТНЫЙ АЛГОРИТМ МИНИМИЗАЦИИ С ПРЕОБРАЗОВАНИЕМ ПРОСТРАНСТВА $\alpha(\varepsilon)$ – АЛГОРИТМ

Журбенко Н. Г.

Институт кибернетики им. В.М. Глушкова НАН Украины

E-mail: zhurbnick@yandex.ua

Аннотация. Приводится субградиентный алгоритм минимизации выпуклой функции в конечномерном евклидовом пространстве с преобразованием пространства. Алгоритм основан на процедуре одномерного спуска и является в некотором смысле монотонным. Алгоритм обеспечивает решение задачи с заданной точностью за конечное число итераций. Приводится оценка трудоемкости алгоритма при решении задачи $\varepsilon-$оптимизации.

Ключевые слова. Оптимизация, $\varepsilon-$субградиент, преобразование пространства, отсекающая плоскость, эллипсоид локализации, трудоемкость алгоритма.

Введение

Более 40 лет назад был разработан субградиентный алгоритм минимизации с растяжением пространства в направлении разности двух последовательных градиентов – $r-$алгоритм [1]. Практика использования $r-$алгоритма показывает, что до сих пор он является одним из наиболее эффективных алгоритмов негладкой оптимизации. Однако теоретическое исследование эффективности $r-$алгоритма далеко не закончено (даже для класса выпуклых функций).

В данной статье приводится описание $\alpha(\varepsilon)-$алгоритма – результаты автора по разработке субградиентного алгоритма сравнимой с $r-$алгоритмом эффективностью и с его основными характеристиками: использование процедуры одномерной минимизации и растяжение пространства с большими коэффициентами.

Задача $\bar{\varepsilon}$ –минимизации

Рассматривается задача минимизации ограниченной снизу выпуклой функции $f(x)$ в $n-$мерном евклидовом пространстве R^n:

$f^* = \min\{f(x) | x \in R^n\}$, X^* – множество точек минимума функции $f(x)$. Для заданного числа $\bar{\varepsilon} \ge 0$ введем $\bar{\varepsilon}$–оптимальное множество $X^*(\bar{\varepsilon})$,

$$X^*(\bar{\varepsilon}) = \{x \in R^n | f(x) \le f^* + \bar{\varepsilon}.$$

Произвольную точку $\tilde{x} \in X^*(\bar{\varepsilon})$ и значение функции $\tilde{f} = f(\tilde{x})$ будем называть решением задачи $\bar{\varepsilon}$–оптимизации ($\bar{\varepsilon}$–решением).

Пусть задан шар начальной локализации $\bar{\varepsilon}$–решения $D(z, R)$: $z \in R^n$ – центр шара, $R > 0$ – его радиус. Будем предполагать, что шар $D(z, R)$ содержит хотя бы одну точку минимума $x^* \in X^*$. Поэтому $D(z, R) \cap X^*(\bar{\varepsilon}) \ne \varnothing$. Далее характеристика множества начальной локализации будет уточнена (см. замечание к утверждению 10).

ε–субградиент

Мы будем использовать несколько отличное от классического определение ε–субградиента [2 – 4]. Вектор $g \in R^n$ называется $(\varepsilon, \tilde{f})$–субградиентом функции $f(x)$ в точке z, если для $\forall x$ выполняется неравенство:

$$f(x) \ge \tilde{f} + (g, x - z) - \varepsilon, \tag{1}$$

где $\tilde{f} \ge f^*; \varepsilon \in R^1$. Значение $\tilde{f}$ на k–ой итерации алгоритма решения задачи $\bar{\varepsilon}$ – оптимизации обычно равно полученному рекордному значению функции $f(x)$. Если значение f^* известно априори, то $\tilde{f} = f^*$. Обычное классическое определение ε – субградиента соответствует определению $(\varepsilon, \tilde{f})$–субградиента (1) при $\tilde{f} = f(z)$, $\varepsilon \ge 0$. Обобщенный градиент (субградиент) функции $f(x)$, в т. z в классическом смысле [5] является $(0, f(z))$–субградиентом: $\partial f(z) = G(0, f(z), z)$.

Через $G(\varepsilon, \tilde{f}, z)$, $\partial f(z)$ будем обозначать множество $(\varepsilon, \tilde{f})$–субградиентов и множество обобщенных градиентов функции $f(x)$ в т. z соответственно. В дальнейшем предполагается, что имеются алгоритмы вычисления $f(x)$ и $g(x) \in \partial f(x)$ в произвольной точке x. Пусть $g \in G(\varepsilon_1, \tilde{f}_1, z_1)$; $f^* \le f_2 \le f(z_2)$, тогда $g \in G(\varepsilon_2, \tilde{f}_2, z_2)$, где

$$\varepsilon_2 = \varepsilon_1 + \tilde{f}_2 - \tilde{f}_1 - (g, z_2 - z_1). \qquad (2)$$

Таким образом, $(\varepsilon, \tilde{f})$ – субградиент в некоторой точке z_1 является $(\varepsilon, \tilde{f})$ – субградиентом в любой другой точке z_2. Формула (2) определяет правило пересчета параметров $(\varepsilon, \tilde{f})$ – субградиента. $G(\varepsilon, \tilde{f}, z)$ по своим свойствам аналогично множеству $\partial f(z)$ (выпуклость, замкнутость и т.д., например, $\{0 \in G(\varepsilon, \tilde{f}, z); \varepsilon \le \bar{\varepsilon}\} \Leftrightarrow \{\tilde{f} \le f^* + \bar{\varepsilon}\}$).

Обозначим $\tilde{X}(\bar{\varepsilon}, \tilde{f}) = \{x \in R^n | f(x) \le \tilde{f} - \bar{\varepsilon}\}$.

Отметим два простых (но важных для дальнейшего) утверждения.

Утверждение 1. *Если* $\tilde{X}(\bar{\varepsilon}, \tilde{f}) = \varnothing$, *то* $\tilde{f}$ *является* $\bar{\varepsilon}$ *– решением.*

$$\left(\{\tilde{X}(\bar{\varepsilon}, \tilde{f}) = \varnothing\} \Rightarrow \{\forall x : f(x) > \tilde{f} - \bar{\varepsilon}\} \Rightarrow \{f^* > \tilde{f} - \bar{\varepsilon}\} \right).$$

Утверждение 2. *Если* $\tilde{f} \ge f^* + 2\bar{\varepsilon}$, *то* $X^*(\bar{\varepsilon}) \subset \tilde{X}(\bar{\varepsilon}, \tilde{f})$.

$$(\{x \in X^*(\bar{\varepsilon})\} \Rightarrow \{f(x) \le f^* + \bar{\varepsilon}\}. \{f(x) \le f^* + \bar{\varepsilon}; \tilde{f} \ge f^* + 2\bar{\varepsilon}\} \Rightarrow$$
$$\{x \in \tilde{X}(\bar{\varepsilon}, \tilde{f})\}.)$$

Агрегированный ε – субградиент

Пусть $P(z, \eta) = \{x \in R^n \,|(x - z, \eta) = 0\}$ – плоскость, проходящая через точку z с нормалью $\eta \in R^n, |\eta| > 0$. $P^+(z, \eta) = \{x \in R^n \,|(x - z, \eta) \ge 0\}$ – полупространство, определяемое (секущей) плоскостью $P(z, \eta)$. Следующее утверждение соответствует обычному построению отсекающих плоскостей "со сдвигом" для множества $\tilde{X}(\bar{\varepsilon}, \tilde{f})$. Пусть $g \in G(\varepsilon, \tilde{f}, z)$; $h = (\bar{\varepsilon} - \varepsilon)/|g|$; $\tilde{z} = z - hg/|g|$. Легко показать, что тогда $\tilde{X}(\bar{\varepsilon}, \tilde{f}) \subset P^+(x - \tilde{z}, -g)$.

Замечание. Если $\tilde{f} = f^*$ (задача с известным значением минимума), то субградиент $g \in \partial f(z)$ является (ε, f^*) – субградиент с $\varepsilon = -(f(z) - f^*)$.

Для случая обычной задачи минимизации ($\bar{\varepsilon}=0$) величина "сдвига" $h=(\bar{\varepsilon}-\varepsilon)/|g|=(f(z)-f^{*})/|g|$ в точности соответствует шаговому множителю метода Поляка [6].

Легко видеть, что $P^{+}(x-\tilde{z},-g)$ определяется неравенством: $(x-z,g)\le-(\bar{\varepsilon}-\varepsilon)$, а вектор «сдвига» $\tilde{z}-z$ отсекающей плоскости можно определить как результат решения следующей задачи $\min\{1/2\,|\,y\,|^{2}:(y,g)\le-\tilde{\varepsilon}\}$, где $\tilde{\varepsilon}=\bar{\varepsilon}-\varepsilon.$ Обобщим этот результат для множества субградиентов. Пусть $g_i\in G(\varepsilon_i,\tilde{f},z),i=1,\dots,m,\tilde{\varepsilon}_i=\bar{\varepsilon}-\varepsilon_i.$ Тогда $\tilde{X}(\bar{\varepsilon},\tilde{f})$ содержится в пересечении подпространств: $(x-z,g_i)\le-\tilde{\varepsilon}_i,$ (не исключается, что это пересечение пусто). Вектор «сдвига» y для множества ε–субградиентов $G(\varepsilon_i,\tilde{f},z)$ из точки z определим решением следующей задачи:

$$\min 1/2\,\|\,y\,\|^{2} \tag{3}$$

$$(y,g_i)\le-\tilde{\varepsilon}_i;i=1,\dots,m. \tag{4}$$

Если система (4) несовместна, то вектор y не определен.

Утверждение 3. *Пусть система ограничений* (4) *– несовместна. Тогда* $\tilde{f}$ *является решением задачи* $\bar{\varepsilon}$*–оптимизации и выпуклая оболочка множества* ε*–субградиентов* $G(\varepsilon_i,\tilde{f},z)$ *содержит 0.*

Утверждение 4. *Пусть:*

a) $\max\{\tilde{\varepsilon}_i;i=1,\dots,m\}>0$*;*

b) система (4) совместна;

c) y,λ *– оптимальные значения прямых и двойственных переменных задачи* (3)*,* (4)*;*

d) $g=\sum_{i=1,m}\lambda_i g_i/\sum_{i=1,m}\lambda_i;$

e) $\varepsilon=\sum_{i=1,m}\lambda_i\varepsilon_i/\sum_{i=1,m}\lambda_i.$ (5)

Тогда:

$$g \in G(\varepsilon, \widetilde{f}, z), \tag{6}$$

$$y = -\frac{\bar{\varepsilon} - \varepsilon}{|g|^2} g. \tag{7}$$

Если условие b) не выполнено, то положим $g = 0$.

Вектор g, определяемый в утверждении 4, будем называть *агрегированным* ε–субградиентом множества субградиентов $G(\varepsilon_i, \widetilde{f}, z)$.

Геометрический смысл агрегированного ε–субградиента: агрегированный ε–субградиент принадлежит выпуклой оболочке множества субградиентов $G(\varepsilon_i, \widetilde{f}, z)$ и обеспечивает максимальный сдвиг отсекающей плоскости.

Утверждение 5. *Пусть для всех субградиентов* $\varepsilon_i = \varepsilon < \bar{\varepsilon}, i = 1,\ldots,m$. *Тогда агрегированный субградиент является вектором наименьшей длины выпуклой оболочки множества субградиентов (* $g = Nr\{g_1,\ldots,g_m\}$ *).*

Таким образом, для множества ε–субградиентов $G(\varepsilon_i, \widetilde{f}, z)$ с одинаковыми значениями ε_i агрегированный ε–субградиент совпадает с кратчайшим вектором выпуклой оболочки $G(\varepsilon_i, \widetilde{f}, z)$. Именно этот вектор обычно используется при построении ε–субградиентных методов оптимизации. Введенное выше определение агрегированного ε–субградиента полезно в следующих смыслах. Во-первых, при решении задачи (3 – 4) может оказаться, что некоторое значение $\lambda_i > 0$ даже если соответствующее значение $\varepsilon_i > \bar{\varepsilon}$. Обычно такие ε–субградиенты при решении задачи $\bar{\varepsilon}$–оптимизации не используются. Во-вторых, при использовании $g = Nr\{g_1,\ldots,g_m\}$ все ε–субградиенты при выборе направления сдвига из точки z оказываются «равноправными» по отношению к значениям ε_i :

вектор $g = Nr\{g_1,\ldots,g_m\}$ не зависит от ε–параметров. Однако значения ε–параметров существенны для локализации $\bar{\varepsilon}$ – решения.

Использование введенного агрегированного ε–субградиента позволяет получать модификации ε–субградиентных методов минимизации по обычной схеме их построения.

На основе утверждения 5 разработан оригинальный алгоритм решения задачи проектирования на политоп (определения вектора $g = Nr\{g_1,\ldots,g_m\}$) [8].

Утверждение 6. *Для заданной точки* z, *направления (спуска)* η $(|\eta|=1)$ *и* $\bar{\varepsilon} > 0$ *можно гарантировать вычисление такого* ε – *субградиен – та* $g \in G(\varepsilon, \tilde{f}, z)$, *что* $\varepsilon \le \bar{\varepsilon}$ *и* $(g,\eta) \ge 0.$

Алгоритм генерации ε–субградиента утверждения 6 имеет много общего с известными алгоритмами пополнения множеств ε–субградиентов [3], [9].

Локализаторы $\bar{\varepsilon}$–решения

Следующие два утверждения касаются построения специальных эллипсоидов минимального объема, содержащих подмножества шара простой структуры (сегмент и сектор).

Обозначим $W = D(z,R) \cap P^+(\tilde{z},\eta)$ – часть шара (сегмент) $D(z,R)$, определяемая секущей плоскостью $P(\tilde{z},\eta_1)$, где $\tilde{z} = z + h\eta; 0 \le h < R, |\eta| = 1.$ W^* – эллипсоид минимального объема с центром в точке $\tilde{z}$, содержащий сегмент W $\tilde{D}^*(z,\tilde{R})$ – шар минимального объема, содержащий эллипсоид W^*, $V(K)$ – объем тела $K.$ $R_\alpha(\eta) = (\alpha - 1)\eta^T\eta + I$ – оператор растяжения пространства [5] по направлению η с коэффициентом $\alpha.$

Утверждение 7. $\tilde{R} = \sqrt{R^2 - h^2}$; $\tilde{D}^*(z,\tilde{R}) = \Omega W^* = R_{\alpha^*}(\eta) W^*$, *где*

$$\alpha^* = \sqrt{(R+h)/(R-h)};$$

$$q = V(W^*)/V(D(z,\widetilde{R})) = \sqrt{(R-h)(R+h)}\,((R-h)/(R+h))^{n/2}.$$

Обозначим $W = D(z,R) \bigcap P^+(z,\eta_1) \bigcap P^+(z,\eta_2)$ – часть шара $D(z,R)$ (сектор), определяемая секущими плоскостями $P^+(z,\eta_1), P^+(z,\eta_2)$; W^* – эллипсоид минимального объема с центром в т. z, содержащий сектор W, $\widetilde{D}^*(z,\widetilde{R})$ – шар минимального объема, содержащий W^*; $\eta^* = (\eta_2 - \eta_1)/|\eta_2 - \eta_1|$, $\xi^* = (\eta_2 + \eta_1)/|\eta_2 + \eta_1|$, $\{\widetilde{\eta}_i, i = 1,\ldots,n-2\}$ – ортонормированный базис подпространства ортогонального векторам η^*, ξ^*.

Пусть $(\eta_1,\eta_2) = \cos(\varphi) \le 0$, $\psi = \pi - \varphi$ (ψ – угол между плоскостями $P(z,\eta_1), P(z,\eta_2)$).

Утверждение 8.

$$\widetilde{R} = \sqrt{2}\cos(\psi/2)R, \tag{8}$$

$$\widetilde{D}^*(z,\widetilde{R}) = \Omega W^* = \{R_{\alpha^*}(\eta^*)\textstyle\prod_{i=1}^{n-2} R_{\widetilde{\alpha}}(\widetilde{\eta}_i)\}\ W^*, \tag{9}$$

где

$$\alpha^* = ctg(\psi/2),\ \widetilde{\alpha} = \sqrt{2}\cos(\psi/2), \tag{10}$$

$$q = V(W^*)/V(D(z,R)) = \sin(\psi). \tag{11}$$

Заметим, что построение эллипсоида W^* приведено в работе [7]. Приведенный оператор преобразования эллипсоида W^* в шар Ω отличен от используемого в [7] «одноранговового эллипсоидального оператора». Существенное отличие состоит в том, что все собственные числа самосопряженного оператора Ω не меньше 1.

Оператор преобразования Ω формулы (9) можно представить в следующем более компактном виде: $\Omega = \widetilde{\alpha} R_{\alpha^*/\widetilde{\alpha}}(\eta^*) R_{1/\widetilde{\alpha}}(\xi^*)$, $(\Omega = R_{\alpha^*}(\eta^*)\prod_{i=1}^{n-2} R_{\widetilde{\alpha}}(\widetilde{\eta}_i) = R_{\alpha^*/\widetilde{\alpha}}(\eta^*) R_{1/\widetilde{\alpha}}(\xi^*)(R_{\widetilde{\alpha}^*}(\eta^*) R_{\widetilde{\alpha}}(\xi^*)\prod_{i=1}^{n-2} R_{\widetilde{\alpha}}(\widetilde{\eta}_i)) = \widetilde{\alpha} R_{\alpha^*/\widetilde{\alpha}}(\eta^*) R_{1/\widetilde{\alpha}}(\xi^*).)$

Из утверждения 8 следует, что объем эллипсоида W^* меньше объема шара $\widetilde{D}^*(z,\widetilde{R})$ только для $\psi < \pi/2$ $(\varphi > \pi/2)$. Поэтому для использования такого эллипсоида в качестве локализации ε–решения необходим алгоритм генерации двух ε–субградиентов, угол между которыми тупой. Описание одного из таких алгоритмов приведено ниже.

Генерация сектора (алгоритм AlgSectorLocation)

Приведем итеративный алгоритм (AlgSectorLocation) генерации (по крайней мере) двух отсекающих плоскостей $P(z,\eta_1)$, $P(z,\eta_2)$ таких, что $(\eta_1,\eta_2) = \cos(\varphi) \le -1+\delta$, для $\forall \delta > 0$.

Обозначим $E = \{e_1, e_2, \dots, e_m\}$ – множество векторов $e_i \in R^n, i = 1,\dots m$, $Nr\{E\}$ – вектор выпуклой оболочки множества E минимальной длины.

Описание алгоритма AlgSectorLocation

1 – ая итерация.

Вычисляем $g_1 \in \partial f(z)$. Если $g_1 = 0$, то останов (задача минимизации $f(x)$ решена), иначе $e_1 = -g_1/|g_1|$.

Используя процедуру одномерной минимизации по направлению e_1, вычисляем ε–субградиент $g_2 \in G(\varepsilon, \widetilde{f}, z)$, для которого $\varepsilon \le \bar{\varepsilon}$, $(g_2, e_1) \ge 0$ (см. утверждение 6). Если $g_2 = 0$, то останов (задача $\bar{\varepsilon}$–минимизации $f(x)$ решена).

Полагаем:

$$G_1 = \{g_1, g_2\},\ e_2 = -g_2/|g_2|,\ E_1 = \{e_1, e_2\},$$

$$p_1 = Nr\{E_1\} = 1/2(e_1 + e_2),\ \eta_1 = e_1,\ \eta_2 = e_2,\ \cos(\varphi_1) = (\eta_1, \eta_2) \le 0.$$

Если $p_1 = 0$, то останов (задача $\bar{\varepsilon}$ – минимизации $f(x)$ решена). Если $\cos(\varphi_1) \le -1+\delta$, то останов, иначе переход на следующую итерацию.

Пусть на итерации $k\,(k=1,2,\ldots)$ получены множество E_k, $(m_k = |E_k|; 2 \le m_k \le n)$ и вектор $p_k = Nr\{E_k\} \ne 0.$

Итерация $(k+1)$ $(k=1,2\ldots).$

Используя процедуру одномерной минимизации по направлению p_k, вычисляем ε–субградиент $g_{m+1} \in G(\varepsilon, \tilde{f}, z)$, для которого $\varepsilon \le \bar{\varepsilon}$, $(g_{m+1}, p_k) \ge 0$ (см. утверждение 6). Если $g_{m+1} = 0$, то останов (задача $\bar{\varepsilon}$–минимизации $f(x)$ решена). Заметим, что в результате одномерной минимизации значение $\tilde{f}$ может уменьшиться. В этом случае производится пересчет $(\varepsilon, \tilde{f})$ параметров векторов множества G_k согласно (2).

Полагаем:

$$\tilde{G}_{k+1} = G_k \bigcup g_{m+1},\ e_{m+1} = -g_{m+1} / |g_{m+1}|,$$

$$\tilde{E}_{k+1} = E_k \bigcup e_{m+1},\ p_{k+1} = Nr\{\tilde{E}_{k+1}\} = \sum_{i \in I} \lambda_i e_i,$$

где I – множество индексов i, для которых $\lambda_i > 0.$

Если $p_{k+1} = 0$, то останов (задача минимизации $f(x)$ решена). Заметим, что если $p_{k+1} \ne 0$, то $n \ge |I| \ge 2$ и $\lambda_{m+1} > 0$ (следствие $(g, \eta) \ge 0$).

Полагаем $G_{k+1} = \{g_i \in \tilde{G}_{k+1} \mid \lambda_i > 0\}$, $E_{k+1} = \{e_i \in \tilde{E}_{k+1} \mid \lambda_i > 0\}$ (E_{k+1} – множество «активных» векторов из $\tilde{E}_{k+1}$ относительно операции $Nr\{.\}$, $n \ge |E_{k+1}| \ge 2$). Перенумеруем вектора множества E_{k+1}, таким образом, чтобы $E_{k+1} = \{e_i, i = 1,2\ldots m;\}$, $\lambda_1 \ge \lambda_2 \ge \ldots \lambda_m > 0$ (вектора G_{k+1} перенумеровываются соответственно E_{k+1}).

Полагаем $\eta_1 = e_1$, $\eta_2 = \xi / |\xi|$, где $\xi = \sum\limits_{i=2}^{m} \dfrac{\lambda_i}{1-\lambda_1} e_i$; $\cos(\varphi_{k+1}) = (\eta_1, \eta_2)$.

Если $\cos(\varphi_{k+1}) \le -1 + \delta$,то останов, иначе переход на следующую итерацию.

Утверждение 9. *Алгоритм AlgSectorLocation конечен. При останове алгоритма выполняется одно из следующих условий: а) задача* $\bar{\varepsilon}$ *– минимиза-*

ции $f(x)$ *решена; b) получены такие отсекающие плоскости* $P(z,\eta_1),P(z,\eta_2)$*, что* $(\eta_1,\eta_2)=\cos(\varphi_{k+1})\le -1+\delta$*. Справедливы следующие оценки значений* $|p_{k+1}|$ *и* $\sin(\psi_{k+1})$ $\quad(\psi_{k+1}=\pi-\varphi_{k+1})$

$$|p_{k+1}|\le 1/\sqrt{2+k}; \tag{12}$$

$$\cos(\varphi_{k+1})=(\eta_1,\eta_2)=\frac{p_{k+1}^2-\lambda_1}{\sqrt{p_{k+1}^2(1-2\lambda_1)+\lambda_1^2}}\equiv F(p_{k+1}^2,\lambda_1). \qquad (13$$

$$\sin^2(\psi_{k+1})\le\frac{p_{k+1}^2(1-p_{k+1}^2)}{p_{k+1}^2(1-2/n)+(1/n)^2}\equiv\Phi(p_{k+1}^2). \tag{14}$$

(*Здесь и далее* $p_{k+1}^2=|p_{k+1}|^2$.)

Приведем некоторые свойства формулы (13). Нетрудно видеть, что $\dfrac{\partial}{\partial\lambda_1}F(p_{k+1}^2,\lambda_1)=\dfrac{p_{k+1}^4-p_{k+1}^2}{(p_{k+1}^2(1-2\lambda_1)+\lambda_1^2)^{3/2}}<0$ для $p_{k+1}^2<1$. Так как $\lambda_1\ge 1/n$, то отсюда следует оценка:

$$\cos(\varphi_{k+1})\le\frac{p_{k+1}^2-1/n}{\sqrt{p_{k+1}^2(1-2/n)+(1/n)^2}}. \tag{15}$$

Пусть $\psi_{k+1}=\pi-\varphi_{k+1}$ (ψ_{k+1} – угол между плоскостями $P(z,\eta_1),P(z,\eta_2)$). Тогда из (15) следует

$$\sin^2(\psi_{k+1})\le\frac{p_{k+1}^2(1-p_{k+1}^2)}{p_{k+1}^2(1-2/n)+(1/n)^2}\equiv\Phi(p_{k+1}^2). \tag{16}$$

Можно показать, что функция $\Phi(p_{k+1}^2)$ правой части (14) монотонно возрастает для $p_{k+1}^2\in[0,1/n]$, причем $\Phi(0)=0$, $\Phi(1/n)=1$. Поэтому оценка (14) содержательна лишь для $p_{k+1}^2<1/n$. Например, $\Phi(1/n^2)=\dfrac{1}{2}(1+1/n)$. Поэтому, если $p_{k+1}^2\le 1/n^2$, то $\sin(\psi_{k+1})\le\dfrac{\sqrt{2}}{2}(1+1/n)$ $(\approx\dfrac{\sqrt{2}}{2}$ для $n>>1)$. Отметим, что при этом, в соответствии с (10), (11), $q=\sin(\psi_{k+1})\approx 0.7$

(любобытно заметить, что коэффициент уменьшения объма локлизации в методе центров тяжести для $n >> 1$ $q = 1 - 1/n \approx 0.632$); $\alpha^* = ctg(\psi_{k+1}/2) \approx 1.4$; $\widetilde{\alpha} = \sqrt{2}\cos(\psi_{k+1}/2) \approx 1.3$.

Из приведенных оценок (12), (14), следует, что алгоритм AlgSector-Location гарантирует коэффициент $q \approx 0.7$ не более чем за n^2 итераций. Автор, конечно, понимает, что эта оценка удручающе плоха. Однако это *оценка* трудоемкости и можно надеятся, что при реработе алгоритма она практически не будет реализовываться (пример такого алгоритма всем известен – симплекс алгоритм).

$\alpha(\varepsilon)$ – Алгоритм

В основе предлагаемого ε–субградиентного алгоритма минимизации лежит следующая простая идея. Задано число (параметр алгоритма) q, $0 < q < 1$.

Пусть на итерации k получены: точка x_k, A_k – невырожденное линейное преобразование ($B_k = A_k^{-1}$), $\widetilde{f}_k$ – рекордное значение функции $f(x)$, R_k– параметр эллипсоида локализации E_k множества $\widetilde{X}(\bar{\varepsilon}, \widetilde{f}_k)$: $\widetilde{X}(\bar{\varepsilon}, \widetilde{f}_k) \subset E_k \equiv \{| x \in R^n \,||\, A_k(x - x_k) | \leq R_k\}$.

Итерация $k+1$алгоритма производится в преобразованном пространстве $Y_k = A_k X$, где X – исходное пространство переменных. Образом эллипсоида E_k в пространстве Y_k является шар $D(y_k, R_k)$, где $y_k = A_k x_k$. На итерации $k+1$ устанавливается факт решения задачи $2\bar{\varepsilon}$ – оптимизации или строится эллипсоид локализации множества $\widetilde{Y}(\bar{\varepsilon}, \widetilde{f}_{k+1}) \subset \widetilde{E}_k \equiv$ $\{y \in Y_k \,||\, (\Omega_k(y - y_{k+1}) | \leq R_{k+1}\}$. При этом $V(\widetilde{E}_{k+1}) \leq qV(D(y_k, R_k))$. Построение этого эллипсоида и параметра $\widetilde{f}_{k+1}$ производится на основании использования процедур одномерной минимизации в соответствии с алгоритмом AlgSectorLocation. При этом на каждой итерации этого алгоритма вычисляются значения коэффициентов уменьшения объемов эллипсоидов локализации множества $\widetilde{X}(\bar{\varepsilon},\widetilde{f}_k)$ q_1 и q_2 согласно утверждениям 7 и 8 соответственно. Пусть $q_{k+1}=\min\{q_1,q_2\}$. Если $q_{k+1}\le q$, то алгоритм AlgSectorLocation прекращает работу.

Возможны два случая.

1. Если $q_{k+1}=q_1$, то оператор Ω_{k+1} определяется оператором Ω утверждения (7). Параметры утверждения (7) определяются следующим образом: $\eta=\widetilde{g}_k/|\widetilde{g}_k|$, $h=(\bar{\varepsilon}-\varepsilon)/|\widetilde{g}_k)|$, где $\widetilde{g}_k$ – агрегированный субградиент текущего множества субградиентов алгоритма AlgSectorLocation. Если $h>R_k$, то останов – задача решена. Иначе переход в новую точку (соответствует сдвигу отсекающей плоскости агрегированного субградиента в преобразованном пространстве): $x_{k+1}=x_k-hB_k\widetilde{g}_k/|\widetilde{g}_k|$.
2. Если $q_{k+1}=q_2>q_1$, то оператор Ω_{k+1} определяется оператором Ω утверждения (8). Параметры утверждения (8) определяются следующим образом: $\eta_1=e_1$, $\eta_2=\xi/|\xi|$, где e_1,ξ – текущие вектора алгоритма AlgSectorLocation. Перехода в новую точку не происходит: $x_{k+1}=x_k$.

Преобразование пространства на итерации $k+1$ определяется оператором Ω_{k+1}: $A_{k+1}=\Omega_{k+1}A_k$.

На основании [11] можно доказать следующее утверждение об эффективности приведенного алгоритма.

Утверждение 10. *Для числа итераций k алгоритма, за которые он обеспечивает решение 2ε–оптимизации, справедлива следующая оценка:*

$$k\le \left]n\frac{\ln(1/\gamma)}{\ln(1/q)}\right[,$$

где γ – относительная точность решения задачи ($\gamma=\bar{\varepsilon}/(RC)$, где C оценка сверху норм субградиентов в исходном шаре локализации $D(z,R)$).

Замечания

1. При доказательстве утверждения 10 предполагается, что для шара начальной локализации $\bar{\varepsilon}$–решения $D(z,R)$ выполняется следующие условия. Пусть C–оценка сверху норм субградиентов в точках шара ограничена: $|\,g(x)\,|\le C\mid g(x)\in\partial f(x); x\in D(z,R)$. Шар $D(z,R)$ содержит такой шар $d(x^*,r_{\bar{\varepsilon}})=\{x:|\,x-x^*\,|\le r_{\bar{\varepsilon}}\}$ (окрестность некоторой точки минимума $x^*\in X^*$), что все его точки являются $\bar{\varepsilon}$–решениями: $d(x^*,r_{\bar{\varepsilon}})\subset X^*(\bar{\varepsilon})$. Учитывая ограниченность норм субградиентов, можно считать, что $r_{\bar{\varepsilon}}\ge\bar{\varepsilon}/C$.

2. Поясним смысл используемой относительной точности γ. Из ограниченности норм субградиентов, имеем следующую начальную оценку: $f(z)-f^*\le RC$. Для точки $\bar{\varepsilon}$–решения $\tilde{x}\in X^*(\bar{\varepsilon})$ $f(\tilde{x})-f^*\le\bar{\varepsilon}=\gamma RC$, начальная оценка улучшена в γ раз.

Численная эффективность $\alpha(\varepsilon)$–алгоритма.

Тестовые задачи:

$$f_1(x)=\sum_{i=1}^{n}\rho_n^{i-1}x_i^2,\ \ f_2(x)=\sum_{i=1}^{n}\rho_n^{i-1}\,|\,x_i\,|,$$

где $\rho_n=10^{6/(n-1)}$. Степень вытянутости линий уровня («овражности») функций не зависит от размерности и равна 10^6. Начальная точка $x_i=1.0, i=1,\ldots n$. Параметр точности решения по функционалу $\bar{\varepsilon}=10^{-6}$.

Обозначения колонок таблиц: nVarbl – число переменных; qVolum – значение параметра уменьшения объема локализации $\bar{\varepsilon}$ –решения на итерации (параметр алгоритма); nIter – число итераций; nLStep – среднее число применения алгоритма одномерной минимизации на одной итерации; Alpha – среднее значение коэффициента растяжения пространства; r – алгоритм – число итераций r–алгоритма.

Таблица 1

Минимизация функции $f_1(x)$

nVarbl	qVolum	nIter	nLStep	Alpha	r – агоритм
10	0.99	107	1.364	1.661	72
10	0.7	56	3.214	3.035	
20	0.99	195	1.297	1.621	103
20	0.7	86	3.686	2.724	
40	0.99	360	1.236	1.611	157
40	0.7	134	4.127	2.761	
50	0.99	435	1.205	1.6	190
50	0.7	153	4.255	2.759	
100	0.99	711	1.136	1.592	577
100	0.7	243	4.407	2.744	

Таблица 2

Минимизация функции $f_2(x)$

nVarbl	qVolum	nIter	nLStep	Alpha	r – агоритм
10	0.99	413	1.165	1.855	141
10	0.7	133	3.015	3.38	
20	0.99	1274	1.095	1.552	204
20	0.7	289	4.173	2.842	
40	0.99	1930	1.09	1.508	349
40	0.7	374	6.035	2.607	
50	0.99	2594	1.076	1.465	435
50	0.7	455	6.868	2.601	
100	0.9	4062	2.597	1.755	888
100	0.7	1559	9.201	2.547	

Приведенные численные эксперименты демонстрируют достаточно высокую эффективность $\alpha(\varepsilon)-$алгоритма, сравнимую с эффективностью $r-$алгоритма [5] (по числу итераций). Результаты выявляют следующие интересные его особенности.

Среднее число применения алгоритма одномерной минимизации на одной итерации оказывается малым в сравнении с гарантированной его оценкой (n^2).

Даже если параметр гарантируемого уменьшения объема локализации $\varepsilon-$решения на каждой итерации задается малым (0.99), тем не менее, среднее значение коэффициента растяжения пространства оказывается существенно больше единицы.

Заключение

Качественная интерпретация $\alpha(\varepsilon)$–алгоритма состоит в следующем. Алгоритм относится к классу методов с преобразованием пространства [5]. На каждой итерации алгоритма преобразование состоит в применении операторов растяжения пространства по ортогональным направлениям. Параметры преобразования определяются построением эллипсоидов локализации ε–решения. Эллипсоиды локализации строятся на основе информации, получаемой в результате применения процедуры одномерной минимизации. На каждой итерации обеспечивается уменьшение объема локализации ε–решения в не менее чем заданное число q раз (параметр агоритма). Если в результате процедур одномерной минимизации происходит существенное улучшение рекордного значения функции, то преобразование пространства определяется оператором растяжения по направлению агрегированного субградиента. В противном случае, за конечное число одномерных процедур минимизации гарантируется генерации (по крайней мере) двух $\varepsilon-$субградиентов с настолько тупым углом между ними, что он обеспечивает построение эллипсоида локализации $\varepsilon-$решения с требуемым уменьшением объема локализации. При этом преобразование пространства определяется операторами растяжения по $n-1$ ортогональным направлениям. Причем с максимальным коэффициентом производится растяжение по направлению разности двух ортонормированных субградиентов из выпуклой оболочки построенных субградиентов. Таким образом, образно выражаясь, $\alpha(\varepsilon)$–алгоритм объединяет алгоритмы Н.З. Шора с растяжением пространства по субградиенту и по разности двух последовательных субградиентов.

Получена оценка трудоемкости $\alpha(\varepsilon)$–алгоритма для решения задачи $\bar{\varepsilon}$–оптимизации. Численные эксперименты показывают, что (по числу итераций) эффективность $\alpha(\varepsilon)$–алгоритма сравнима с эффективностью $r-$

алгоритма Н.З. Шора. Однако трудоемкость одной итерации $\alpha(\varepsilon)$– алгоритма существенно больше.

Список литературы

[1] Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – Киев: Наук. думка, 1971. – №3. – С. 51 – 59.

[2] Brondsted A. and Rockafellar R.T. On the subdifferentiability of convex functions // Proc. Amer. Math. Soc. 16, 1965. – Pp. 605 – 611.

[3] Lemarechal C., Mifflin K. Nonsmooth Optimization. – Oxford: Pergamon Press, 1978. – 180 p.

[4] Демьянов В.Ф., Васильев Л.В. Недифференцируемая оптимизация. – М.: Наука, 1981. – 384 с.

[5] Шор Н.З. Методы минимизации недифференцируемых функций и их применение. – Киев: Наук.думка, 1979. – 200 с.

[6] Поляк Б.Т. Минимизация негладких функционалов // Журн. вычислит. математики и матем. физики, 1969. – Т.9, № 3. – С. 507 – 521.

[7] Стецюк П.И. Ортогонализующие линейные операторы в выпуклом программировании (Часть I) // Кибернетика и системный анализ, 1997. – № 3. – С. 97 – 119.

[8] Журбенко Н.Г. Алгоритм проектирования на политоп // Теорія оптимальних рішень, 2008. – № 7. – С. 125 – 132.

[9] Ржевский С.В. Монотонные методы выпуклого программирования. – Киев: Наук. думка, 1993. – 319 с.

[10] Журбенко Н.Г. Об одном классе методов минимизации с преобразованием пространства // Методы решения экстремальных задач, 1996. – С. 68 – 80.

[11] Журбенко Н.Г. Оценка эффективности одного класса ε–субградиентных методов минимизации с преобразованием пространства // Оптимизация и ее приложения, 1997. – С. 49 – 54.