Е. А. Нурминский, Об одном классе методов выпуклого программирования, *Ж. вычисл. матем. и матем. физ.*, 1986, том 26, номер 8, 1150–1159

ЖУРНАЛ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

Том 26, 1986 № 8

УДК 519.853.3

ОБ ОДНОМ КЛАССЕ МЕТОДОВ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

НУРМИНСКИЙ Е. А.

(Киев)

Задача выпуклой оптимизации сводится к нахождению наименьшего скалярного корня ε-субградиентного точечно-множественного отображения. Это сведе́ние служит основой для разработки класса методов отделяющих плоскостей, сходимость которых доказывается в общем виде. Для явного учета ограниченности памяти ЭВМ доказывается сходимость метода отделяющих плоскостей с отбрасыванием несущественной информации.

Введение

Целью настоящей работы является развитие подхода, объединяющего ряд методов недифференцируемой (выпуклой) оптимизации. Этот подход не только дает возможность однообразно рассмотреть ряд известных алгоритмов и прояснить механизм их сходимости, достоинства и недостатки, но и предложить новые методы.

В основе предлагаемого подхода лежит использование аппарата ε-субградиентных отображений, уже приобретших определенную популярность в выпуклом программировании [1]–[6]. Последнее связано как с практическими преимуществами использования ε-субградиентов, так и с их более регулярными аналитическими свойствами [7]–[12], что позволяет развить более содержательную теорию.

В качестве основного будем рассматривать конечномерное евклидово пространство E со скалярным произведением xy, $x\in E$, $y\in E$. Множество вещественных чисел будем обозначать через R, множество неотрицательных чисел — через R_+.

Норма вектора x определяется и обозначается обычным образом: $\|x\|=(xx)^{1/2}$; U — единичный шар в E. Введем также норму множества X:

$$\|X\|=\sup_{x\in X}\|x\|.$$

Опорную функцию множества $A\subset E$ обозначим

$$(A)_x=\sup_{a\in A}(ax).$$

Ниже исследуются выпуклые функции на E такие, что $f(x)>-\infty$ для всех x. Область определения f обозначим через $\operatorname{dom} f=\{x:f(x)<\infty\}$.

Пусть $\varepsilon\in R_+$. Множество $\partial_\varepsilon f(x)$ векторов g таких, что $f(y)-f(x)\geqslant \geqslant g(y-x)-\varepsilon$ для всех $y\in E$, называется ε-субдифференциалом f в точке x. При $x\in\operatorname{int}\operatorname{dom} f$ множество $\partial_\varepsilon f(x)$ выпукло, замкнуто и ограничено.

В дальнейшем будем предполагать, что $0\in\operatorname{int}\operatorname{dom} f$ и $f(0)=0$. Между f и $\partial_\varepsilon f$ существует [12] соотношение

$$\inf_x f(x)=-\inf_{0\in\partial_\varepsilon f(0)}\varepsilon,$$

сводящее задачу выпуклой многомерной оптимизации к задаче нахождения наименьшего (скалярного) корня многозначного отображения. На указанном соотношении основаны дальнейшие результаты.

§ 1. Общая схема алгоритма

Обозначим

$$D=\{(\varepsilon, g): g\in\partial_\varepsilon f(0), \varepsilon\in R_+\}\subset R_+\times E,$$

и пусть $D(\varepsilon)$ — сечение этого множества, соответствующее фиксированному ε, $D(\varepsilon)=\partial_\varepsilon f(0)$. Легко показать, что D выпукло и $D(\varepsilon')\subset D(\varepsilon'')$ при $\varepsilon'\leqslant\varepsilon''$. В дальнейшем множество D будет мажорироваться и минорироваться (по включению) последовательностями множеств $\{D_k^I\}$, $\{D_k^o\}$, $D_k^I\subset D\subset D_k^o$, $k=1, 2, \ldots$. Сечения D_k^I и D_k^o при фиксированном ε будем обозначать $D_k^I(\varepsilon)$ и $D_k^o(\varepsilon)$ соответственно. Общая схема алгоритма на структурном псевдоязыке программирования выглядит следующим образом.

Алгоритм отделяющих плоскостей:
положить $D_0^o=R_+\times E$, $D_0^I=\{(\varepsilon, 0)\ \varepsilon\geqslant E\}$, где

$$0\leqslant-\inf_x f(x)\leqslant E<\infty;$$

положить $k=0$.

Если не выполнен критерий оптимальности, переходим к основному циклу алгоритма:
найти

$$\inf_{0\in D_k^o(\varepsilon)}\varepsilon=\varepsilon_k;$$

положить $A_k=\{(\varepsilon_k, 0)\}\subset R_+\times E$;

построить гиперплоскость $z^k\in R\times E$, отделяющую A_k и D_k^I, т. е. найти z^k такое, что

$$(A_k)_{z^k}+(D_k^I)_{-z^k}\leqslant 0; \tag{1}$$

если $z^k=(\theta_k, x^k)$, $\theta_k\in R_+$, $x^k\in E$, то найти

$$(D)_{-z^k}=-\tilde g^k x^k-\tilde\varepsilon_k\theta_k=-v_k, \tag{2}$$

модифицировать D_k^I:

$$D_{k+1}^I=\mathrm{co}\,\{D_k^I, (\tilde\varepsilon_k, \tilde g^k)\},$$

модифицировать D_k^o:

$$D_{k+1}^o=D_k^o\cap\{\bar g:\bar g z^k\geqslant v_k,\ \bar g\in R_+\times E\},$$

увеличить счетчик итераций $k:=k+1$.

Конец основного цикла алгоритма.

Конец алгоритма отделяющих плоскостей.

Приведенная схема требует некоторых пояснений. Ключевыми моментами в работе алгоритма являются задачи (1), (2). Для обеспечения сходимости алгоритма на решение задачи (1) необходимо наложить определенные требования, которые здесь названы принципом равномерной отделимости.

Будем считать, что правило построения гиперплоскостей x^k, $k=1, 2, \ldots$, отделяющих от нуля члены последовательности множеств X_k,

$k=1, 2, \ldots$, удовлетворяет принципу равномерной отделимости, если из существования $\lambda>0$ такого, что $X_k+\lambda U$, $k=1, 2, \ldots$, отделимы от нуля, следует существование $\lambda'>0$ такого, что каждое x^k отделяет от нуля соответствующее множество $X_k+\lambda' U$, т. е.

$$(X_k+\lambda' U)_{-x^k}\leqslant 0, \quad \text{или} \quad (X_k)_{-x^k}\leqslant -\lambda'\|x^k\|.$$

Далее будет предполагаться, что решение задачи (1) удовлетворяет принципу равномерной отделимости, точнее — его очевидной модификации, связанной с тем, что в данном случае происходит отделение множеств $D_k{}^I$ не от начала координат, а от некоторых точек в $R_+\times E$.

Решение (2) заключается в определении $\tilde{g}$, $\tilde{\varepsilon}$ таких, что

$$\sup_{g\in\partial_\varepsilon f(0),\ \varepsilon\geqslant 0}\{-xy-\theta\varepsilon\}=-x\tilde{g}-\theta\tilde{\varepsilon}. \tag{3}$$

Для простоты индекс k итерации алгоритма опускается. Задачу (3) можно переписать в виде

$$\sup_{g\in\partial_\varepsilon f(0),\ \varepsilon\geqslant 0}\{-xg-\theta\varepsilon\}=\sup_{\varepsilon\geqslant 0}\{\sup_{g\in\partial_\varepsilon f(0)}\{-xg\}-\theta\varepsilon\}=$$

$$=\sup_{\varepsilon\geqslant 0}\{(\partial_\varepsilon f(0))_{-x}-\theta\varepsilon\}=\theta f\left(-\frac{x}{\theta}\right)$$

(см. [12, теорема 1]). Пара $\tilde{\varepsilon}$, $\tilde{g}$ определяется так:

$$\tilde{g}\in\partial f(\tilde{x}), \quad \tilde{x}=-x/\theta, \qquad \tilde{\varepsilon}=\tilde{x}\tilde{g}-f(\tilde{x})\geqslant 0.$$

Естественно, неявно предполагается, что $\tilde{x}\in\operatorname{dom} f$. Последнее можно гарантировать, например ограничившись рассмотрением конечных выпуклых функций, где $\operatorname{dom} f=E$.

Сводя вместе условия применимости алгоритма, получаем следующую теорему сходимости.

Теорема 1. *Пусть $f(x)$ — конечная выпуклая функция, $f(0)=0$,*

$$f_*=\inf_x f(x)\geqslant -E>-\infty.$$

Тогда

$$\lim_{k\to\infty}\varepsilon_k=-f_*.$$

Доказательство. Обозначим

$$\lim_{k\to\infty}D_k{}^o=D_\circ{}^o, \qquad \operatorname{cl}(\lim_{k\to\infty}D_k{}^I)=D_\circ{}^I,$$

где в последнем выражении имеется в виду замыкание предельного множества. Указанные пределы существуют в силу монотонного характера соответствующих последовательностей.

Аналогично предыдущему через $D_\cdot{}^o(\varepsilon)$ и $D_\cdot{}^I(\varepsilon)$ будем обозначать сечения по ε соответствующих множеств.

Если теорема неверна, то

$$\lim_{k\to\infty}\varepsilon_k=\varepsilon_\circ<-f_*=\inf_{0\in\partial_\varepsilon f(0)}\leqslant\inf_{0\in D_k{}^I(\varepsilon)}\varepsilon.$$

Переходя к пределу, получаем

$$\varepsilon_\circ<\inf_{0\in D_\cdot{}^I(\varepsilon)}\varepsilon.$$

Это означает, что существует $\lambda>0$ такое, что $D_\cdot{}^I+\lambda U$ отделимо от точки $(\varepsilon_\cdot, 0)$. Легко видеть, что при этом и $D_k{}^I+\lambda U$ отделимы от $(\varepsilon_k, 0)$ для

$k=1, 2, \ldots$. (Сохранено U как обозначение единичного шара, но на этот раз в $R\times E$.)

Действительно, тогда существует $\bar{z}=(\bar{\theta}, \bar{x})$ такое, что $\bar{\theta}\varepsilon.+ +(D.^I+\lambda U)_{-\bar{z}}\leqslant 0$. Можно убедиться, что $\bar{\theta}\geqslant 0$, откуда в силу монотонности последовательностей $\{\varepsilon_k\}$ и $\{D_k^I\}$ имеем

$$\bar{\theta}\varepsilon_k+(D_k^I+\lambda U)_{-\bar{z}}\leqslant\bar{\theta}\varepsilon.+(\bar{D}.+\lambda U)_{-\bar{z}}\leqslant 0.$$

Тогда в силу принципа равномерной отделимости существует $\lambda'>0$ такое, что $(\theta_k, x^k)=z^k$ отделяет $D_k^I+\lambda' U$ от точки $(\varepsilon_k, 0)$:

$$\theta_k\varepsilon_k+(D_k^I+\lambda' U)_{-z^k}\leqslant 0, \text{ или } \theta_k\varepsilon_k+(D_k^I)_{-z^k}\leqslant-\lambda'\|z^k\|.$$

Нормируя на $\|z^k\|$ и переходя к пределу при $k\to\infty$, $z^k/\|z^k\|\to z^\cdot$, $\theta_k/\|z^k\|\to\theta.$, получаем $\theta.\varepsilon.+(D.^I)_{-z^\cdot}\leqslant-\lambda.'$. Поскольку

$$\varepsilon_k=\inf_{0\in D_k^O(\varepsilon)}\varepsilon=\inf_{0\in D_{k-1}^O(\varepsilon)\cap\{\bar{g}z^k\geqslant v_k\}}\varepsilon\geqslant\max\{\inf_{0\in D_{k-1}^O(\varepsilon)}\varepsilon, \inf_{\varepsilon\theta_k\geqslant v_k}\varepsilon\}=$$

$$=\max\{\varepsilon_{k-1}, \inf_{\varepsilon\theta_k\geqslant v_k}\varepsilon\},$$

то

$$\varepsilon_k\theta_k\geqslant\max\{\theta_k\varepsilon_{k-1}, v_k\}\geqslant v_k=-(D)_{-z^k}=-(D_{k+1}^I)_{-z^k}.$$

Опять нормируя и переходя к пределу, получаем $\varepsilon.\theta.\geqslant-(D.^I)_{-z^\cdot}$. При этом $0=-(D.^I)_{-z^\cdot}+(D.^I)_{-z^\cdot}\leqslant\theta.\varepsilon.+(D.^I)_{-z^\cdot}\leqslant-\lambda'<0$, т. е. противоречие, которое доказывает теорему.

§ 2. Построение отделяющей гиперплоскости

Основным элементом описанного выше алгоритма является решение вспомогательной задачи (1) — нахождения гиперплоскости, отделяющей множество D_k^I от точки $(\varepsilon_k, 0)$. Для обеспечения сходимости алгоритма решение этой задачи должно удовлетворять принципу равномерной отделимости, сформулированному выше. В остальном это правило может быть произвольно, что дает возможность предложить целое семейство методов выпуклой оптимизации. В данном параграфе будет описан один общий способ построения отделяющих плоскостей, доказано, что он удовлетворяет условию равномерной отделимости, и приведены примеры его использования в алгоритмах выпуклой оптимизации.

С общей точки зрения в алгоритмах отделяющих плоскостей задача (1) может быть сведена к задаче построения гиперплоскости, отделяющей начало координат от множества $X\subset R_+\times E$, заданного своими вершинами:

$$X=\mathrm{co}\{x^i, i=1, 2, \ldots, I\}=\left\{\sum_{i=1}^I\lambda_i x^i, \lambda_i\geqslant 0, \sum_{i=1}^I\lambda_i=1\right\}.$$

Будем считать, что в $R\times E$ задана положительно-однородная выпуклая норма M: для $x\in R\times E$

$$\|x\|_M\geqslant 0, \qquad \|x\|_M=0 \text{ влечет } x=0,$$

$$\|x+y\|_M\leqslant\|x\|_M+\|y\|_M, \qquad \|\lambda x\|_M=\lambda\|x\|_M, \qquad \lambda\geqslant 0.$$

Такая норма может быть представлена как опорная функция некоторого выпуклого замкнутого ограниченного множества χ_M:

$$\|x\|_M=(\chi_M)_x, \qquad 0\in\operatorname{int}\chi_M.$$

Единичный шар в норме (метрике) M обозначим через U_M. Обычную евклидову норму по-прежнему будем обозначать через $\|x\|$ и единичный евклидов шар — через U. Множество U_M ограничено в евклидовой метрике.

Для построения отделяющей гиперплоскости рассмотрим следующую вспомогательную задачу:

$$\min_{x\in X}\|x\|_M,$$

или

$$\min \|x\|_M,$$

$$x=\sum_{i=1}^{I}\lambda_i x^i, \tag{4}$$

$$\sum_{i=1}^{I}\lambda_i=1, \qquad \lambda_i\geqslant 0, \quad i=1,2,\ldots,I.$$

Для наших целей представляют интерес оптимальные значения не столько прямых переменных, сколько двойственных.

Теорема 2. *Вектор u^* оптимальных двойственных переменных, соответствующих ограничению (4), задает гиперплоскость, отделяющую начало координат от множества X и удовлетворяющую принципу равномерной отделимости.*

Доказательство. Для проверки принципа равномерной отделимости необходимо рассмотреть случай $0\notin X$, или

$$\min_{x\in X}\|x\|=\sup_{\rho U\cap X=\varnothing}\rho=R>0.$$

В норме M

$$\min_{x\in X}\|x\|_M=\sup_{\rho U_M\cap X=\varnothing}\rho=R_M.$$

Пусть

$$\theta_M=\inf_{U_M\subset\theta U}\theta.$$

В силу ограниченности U_M будет $\theta_M<\infty$. Тогда

$$R_M\geqslant\sup_{\rho\theta U\cap X=\varnothing}\rho=\frac{1}{\theta}\sup_{\rho U\cap X=\varnothing}\rho=\frac{R}{\theta} \tag{5}$$

для всех $\theta>\theta_M$. Беря супремум правой части (5) по $\theta>\theta_M$, получаем

$$\min_{x\in X}\|x\|_M\geqslant\frac{R}{\theta_M}. \tag{6}$$

Задача, двойственная к (4), имеет вид

$$\max_{u\in\chi_M}\ \min_{i=1,2,\ldots,I} ux^i, \tag{7}$$

откуда для оптимального решения u^* задачи (7) получаем с учетом (6) неравенства

$$u^*x^i\geqslant R/\theta_M, \qquad i=1,2,\ldots,I.$$

Умножая их на $\lambda_i \geqslant 0$, $\Sigma\lambda_i = 1$, и суммируя, получаем $u^* x \geqslant R/\theta_M$ для любого $x \in X$, что дает оценку

$$(X)_{-u^*} \leqslant -R/\theta_M. \tag{8}$$

Поскольку $\|u^*\| \leqslant \|\chi_M\|$, то правую часть (8) можно усилить:

$$(X)_{-u^*} \leqslant -\frac{R}{\theta_M \|\chi_M\|} \|u^*\| = -\frac{R}{\theta_M \|\chi_M\|} (U)_{u^*},$$

откуда следует, что u^* отделяет начало координат от множества $X + \lambda' U$, где $\lambda' = R/(\theta_M \|\chi_M\|) > 0$, что и требовалось доказать.

Покажем сейчас, что ряд известных алгоритмов выпуклого программирования может быть получен из изложенной общей схемы выбором специфической нормы $\|\cdot\|_M$. Для краткости рассматриваются лишь принципиальные черты алгоритмов.

Метод секущих плоскостей [13]. Определим в пространстве $R \times E$ с элементами $\bar{x} = (\theta, x)$ норму

$$\|\bar{x}\|_C = |\theta| + M \sum_{i=1}^{n} |x_i|,$$

где x_i, $i = 1, 2, \ldots, n$, — координаты вектора $x \in E$. Константа $M > 0$ — некоторая достаточно большая величина, критерий выбора которой будет изложен ниже.

Подзадача (1) построения на k-й итерации отделяющей гиперплоскости имеет в этом случае вид

$$\min_{\bar{x} \in D_k{}^I} \|\bar{x}\|_C = \min_{g \in D_k{}^I(\varepsilon),\ \varepsilon \geqslant 0} \left\{ |\varepsilon - \varepsilon_k| + M \sum_{i=1}^{n} |g_i| \right\}. \tag{9}$$

Для достаточно больших M решение задачи (9) имеет $g = 0$. Тогда (9) может быть переписано в виде

$$\min_{0 \in D_k{}^I(\varepsilon),\ \varepsilon \geqslant 0} |\varepsilon - \varepsilon_k|.$$

Поскольку $\varepsilon_k < \varepsilon$: $0 \in D_k{}^I(\varepsilon)$, то эта задача сводится к следующей:

$$\min_{0 \in D_k{}^I(\varepsilon),\ \varepsilon \geqslant 0} \varepsilon = \min \varepsilon \text{ при } \varepsilon = \sum_{m=1}^{k} \lambda_m \tilde{\varepsilon}_m,$$

$$0 = \sum_{m=1}^{k} \lambda_m \tilde{g}^m, \qquad 1 = \sum_{m=1}^{k} \lambda_m, \qquad \lambda_m \geqslant 0.$$

В силу теоремы 2, отделяющая гиперплоскость имеет вид $(1, x)$, где x — двойственная переменная, соответствующая ограничению $0 = \Sigma\lambda_m \tilde{g}^m$. Сама двойственная задача такова:

$$\max_{|x_i| \leqslant M,\ i=1,2,\ldots,n} \ \min_{\lambda_m + \ldots + \lambda_k = 1,\ \lambda_m \geqslant 0,\ m=1,2,\ldots,k} \sum_{m=1}^{k} \lambda_m (\tilde{\varepsilon}_m - x\tilde{g}^m) = \tag{10}$$

$$= \max_{|x_i| \leqslant M,\ i=1,2,\ldots,n} \ \min_{m=1,2,\ldots,k} \{\tilde{\varepsilon}_m - x\tilde{g}^m\}.$$

Учитывая, что $\tilde{\varepsilon}_m = x^m \tilde{g}^m - f(x^m)$, задачу (10) можно переписать таким образом:

$$-\min_{|x_i|\leq M,\ i=1,2,\dots,M}\ \max_{m=1,2,\dots,k}\ \{f(x^m)+\tilde{g}^m(x^m-x^m)\},$$

что и является методом секущих плоскостей.

Метод спуска [14]. Сохраняя обозначения предыдущего примера, вводим в $R\times E$ норму $\|\bar{x}\|_L=M|\theta|+\|x\|$. Задача (1) требует решения

$$\min_{x\in D_k{}^I-A_k}\ \|\bar{x}\|_L=\min_{g\in D_k{}^I(\varepsilon)}\ \{M|\varepsilon-\varepsilon_k|+\|g\|\}.$$

При достаточно больших M в этом решении $\varepsilon=\varepsilon_k$, что приводит к задаче

$$\min_{g\in D_k{}^I(\varepsilon_k)}\ \|g\|=\min_{g=\Sigma\lambda_m\tilde{g}^m,\ 1=\Sigma\lambda_m,\ \lambda_m\geq 0}\ \|g\|.$$

Из практических соображений удобнее решать эквивалентную задачу

$$(11)\qquad \min\|g\|^2 \text{ при } g=\sum_{m=1}^{k}\lambda_m\tilde{g}^m,\quad 1=\sum_{m=1}^{k}\lambda_m,\quad \lambda_m\geqslant 0.$$

Двойственная переменная x^k, соответствующая (11), с точностью до знака совпадает с решением прямой задачи. Поскольку в этом методе происходит потеря компоненты отделяющей гиперплоскости, соответствующей ε-координате, вектор x^k используется далее в поиске минимума по направлению.

§ 3. Реализуемые методы

Приведенные выше результаты представляют собой теоретическую основу для разработки практически реализуемых алгоритмов. В первую очередь при реализации методов следует учесть конечность памяти ЭВМ, что не позволяет неограниченно накапливать информацию об аппроксимациях $D_k{}^o$, $D_k{}^I$. Поскольку внешняя аппроксимация $D_k{}^o$ используется в алгоритме лишь для определения ε_k, что можно сделать рекуррентно, $D_k{}^o$ может быть учтено неявным образом, без хранения в памяти ЭВМ всех отделяющих плоскостей. Критическим является накопление информации, связанной с уточнением внутренней аппроксимации $D_k{}^I$. Теоретическая схема, изложенная выше, предполагает сохранение всех точек, выпуклой оболочкой которых является $D_k{}^I$. В действительности, конечно. ограниченность памяти ЭВМ заставляет вводить какой-либо процесс отбрасывания излишней информации без потери сходимости алгоритма и, желательно, с минимальным ухудшением его скорости сходимости. В действительности возможно даже ускорение реальной сходимости за счет упрощения вспомогательной задачи (1). Ниже излагается подобный алгоритм с достаточно произвольным правилом упрощения аппроксимации $D_k{}^I$. Условия, которым должно удовлетворять правило сокращения запоминаемой информации, допускают значительную свободу в способах реализации. В минимальных случаях объем запоминаемой информации может лишь линейно зависеть от размерности задачи.

Сохраняя обозначения, введенные выше, опишем

Основной цикл реализуемого алгоритма: найти

$$\inf_{0\in D_k{}^o(\varepsilon)}\ \varepsilon=\varepsilon_k;$$

положить

$$A_k=\{\xi^k\}=\{(\varepsilon_k, 0)\}\subset R_+\times E;$$

построить гиперплоскость $z^k=(\theta_k, x^k)$, отделяющую множества A_k и D_k^I:

$$\| z^k \| = \min_{g\in D_k^I} \| g-\xi^k \|, \qquad z^k+\xi^k \in D_k^I;$$

заменить множество D_k^I на замкнутое множество $\tilde{D}_k^I$, выполнив при этом следующие условия:

$$\tilde{D}_k^I\subset D_k^I, \qquad \pi=(E, 0)\in\tilde{D}_k^I,$$

$$\min_{g\in\tilde{D}_k^I} \| g-\xi^k \| = \min_{g\in D_k^I} \| g-\xi^k \| = \| z^k \|;$$

(Заметим сразу, что последнее условие влечет за собой $z^k+\xi^k\in\tilde{D}_k^I$.)
найти $(D)_{-z^k}=-\tilde{g}^k x^k-\tilde{\varepsilon}_k\theta_k=-v_k$;
модифицировать D_k^I:

$$D_{k+1}^I=\mathrm{co}\,\{\tilde{D}_k^I, (\tilde{\varepsilon}_k, \tilde{g}^k)\};$$

модифицировать D_k^o:

$$D_{k+1}^o=D_k^o \cap \{g: gz^k \geqslant v_k\};$$

увеличить счетчик итераций: $k:=k+1$.

Конец основного цикла реализуемого алгоритма.

Описанный алгоритм отличается от общей схемы, во-первых, наличием дополнительной операции упрощения внутренней аппроксимации D_k^I и, во-вторых, однозначным определением способа построения отделяющей гиперплоскости z^k. Сходимость алгоритма утверждает

Теорема 3. *Пусть $f(x)$ — конечная выпуклая функция, $f(0)=0$, $\varepsilon_*=-f_*=-\inf_x f(x)\geqslant -E>-\infty$. Тогда*

$$\lim_{k\to\infty} \varepsilon_k=\varepsilon_*.$$

Доказательство. Поскольку $\xi^k\notin D$, а $z^k+\xi^k\in D_k^I\subset D$, то достаточно показать, что

$$\lim_{k\to\infty} \| z^k \| = 0.$$

Покажем прежде всего, что указанный предел существует. По построению, $z^k+\xi^k\in\tilde{D}_k^I$ и, следовательно, $\zeta_\lambda^{k+1}=\lambda(z^k+\xi^k)+(1-\lambda)\pi\in\tilde{D}_k^I\subset D_{k+1}^I$ для $\lambda\in[0, 1]$. Тогда $\tilde{z}_\lambda^{k+1}=\zeta_\lambda^{k+1}-\xi^{k+1}$ для всех $\lambda\in[0, 1]$ удовлетворяет соотношению $\tilde{z}_\lambda^{k+1}+\xi^{k+1}\in D_{k+1}^I$. При $\lambda=(E-\varepsilon_{k+1})/(E-\varepsilon_k)$ будет $\tilde{z}_\lambda^{k+1}=\lambda z^k$ и, следовательно,

$$\| z^{k+1} \| = \min_{z+\xi^{k+1}\in D_{k+1}^I} \| z \| \leqslant \min_{z+\xi^{k+1}\in\tilde{D}_k^I} \| z \| \leqslant \| \tilde{z}_\lambda^{k+1} \| \leqslant \lambda \| z^k \| \leqslant \| z^k \|.$$

Из монотонности $\{\|z^k\|\}$ следует существование предела

$$\lim_{k\to\infty} \| z^k \| = \sigma \geqslant 0.$$

Обозначим через $K=\{n_k, k=1, 2,\ldots\}$ подпоследовательность индексов таких, что

$$\lim_{k\to\infty,\, k\in K} z^k = z^*$$

существует. По доказанному выше, $\|z^*\|=\sigma$. Докажем, что

$$\lim_{k\to\infty,\, k\in K} \inf (g^k-\xi^k) z^k \geqslant \sigma^2. \tag{12}$$

Напомним, что $z^k=(\theta_k, x^k)$ и $g^k=(\tilde{\varepsilon}_k, \tilde{g}^k)$, где $\tilde{g}^k\in\partial f(\tilde{x}^k)$, $\tilde{x}^k=-x^k/\theta_k$.

Поскольку для достаточно больших k имеем $\|z^k\|\leqslant 2\sigma$ и $\theta_k\geqslant\gamma>0$, то $\tilde{x}^k$ равномерно ограничены. Это означает, что $\tilde{g}^k$, $\tilde{\varepsilon}_k$ ограничены тоже: $\|g^k\|\leqslant G$, причем $G>\sigma$. Предположим теперь, что, в отличие от (12),

$$\lim_{k\to\infty,\, k\in K} \inf (g^k-\xi^k) z^k \leqslant \sigma^2-\delta^2$$

для некоторого $\delta>0$. Тогда

$$\begin{aligned} \|z^{k+1}\|^2 &\leqslant \inf_{\lambda\in[0,1]} \|\xi^{k+1}-\xi^k+\lambda z^k+(1-\lambda)(g^k-\xi^k)\|^2 \leqslant \\ &\leqslant \|\xi^{k+1}-\xi^k\|^2+2(G+\sigma)\|\xi^{k+1}-\xi^k\|+ \\ &+\inf_{\lambda\in[0,1]} \|\lambda z^k+(1-\lambda)(g^k-\xi^k)\|^2. \end{aligned} \tag{13}$$

Первые два слагаемых стремятся к нулю при $k\to\infty$, поэтому сосредоточимся на последнем:

$$\begin{aligned} &\inf_{\lambda\in[0,1]} \|\lambda z^k+(1-\lambda)(g^k-\xi^k)\|^2\leqslant \\ &\leqslant \inf_{\lambda\in[0,1]} \{\sigma^2(1-\lambda^2)-2\lambda(1-\lambda)\delta^2+\lambda^2G^2\}\leqslant \\ &\leqslant \sigma^2+\inf_{\lambda\in[0,1]} \{\lambda^2(G^2+2\delta^2)-2\lambda\delta^2\}. \end{aligned}$$

Вычисление инфимума дает оценку

$$\inf_{\lambda\in[0,1]} \|\lambda z^k+(1-\lambda)g^k\|^2\leqslant \sigma^2-\frac{\delta^4}{G^2+2\delta^2}.$$

Переходя в (13) к пределу, получаем

$$\lim_{k\to\infty,\, k\in K} \|z^{k+1}\|^2\leqslant\sigma^2-\frac{\delta^4}{G^2+2\delta^2}<\sigma^2,$$

что невозможно, и, следовательно, (12) доказано.

Далее заметим, что, по построению, $\xi^{k+1}z^k\geqslant g^kz^k$. Используя (12), получаем $\varepsilon_{k+1}\theta_k\geqslant\sigma^2+\xi^kz^k-\gamma=\sigma^2+\varepsilon_k\theta_k-\gamma$ для произвольно малых $\gamma>0$, лишь k достаточно велико. Тогда $\varepsilon_{k+1}\geqslant\varepsilon_k+\sigma^2/\theta_k-\gamma/\theta_k\geqslant\varepsilon_k+\sigma-\gamma/\theta_k$. Поскольку $\theta_k\geqslant\geqslant\theta>0$ для некоторого θ, переход к пределу дает $\varepsilon_*\geqslant\varepsilon_*+\sigma-\gamma/\theta$ для произвольного $\gamma>0$, что равносильно $\varepsilon_*\geqslant\varepsilon_*+\sigma>\varepsilon$.

Полученное противоречие завершает доказательство.

Литература

1. *Rockafellar T. R.* The multiplier method of Hestenes and Powell applied to convex programming.– J. Optimizat. Theory and Appl., 1973, v. 12(6), p. 555–562.
2. *Нурминский Е. А., Желиховский А. А.* E-Субградиентный метод решения негладких экстремальных задач.– Кибернетика, 1977, № 1, с. 109–113.
3. *Шор Н. З.* Новые направления в развитии методов негладкой оптимизации.– Кибернетика, 1977, № 6, с. 87–91.
4. *Ржевский С. В.* E-Субградиентный метод решения задач выпуклого программирования.– Ж. вычисл. матем. и матем. физ., 1981, т. 21, № 5, с. 1126–1132.
5. *Васильев Л. В., Демьянов В. Ф.* Метод обобщенных (e, m, t)-градиентов при наличии ограничений.– Вестн. ЛГУ. Матем., механ., астрон., 1979, № 19, с. 19–23.
6. *Демьянов В. Ф., Васильев Л. В., Лизина С. А.* Минимизация выпуклых функций при помощи e-субградиентов.– В кн.: Управление динамич. системами. Л.: Изд-во ЛГУ, 1978, с. 3–22.

7. *Нурминский Е. А.* О непрерывности e-субградиентных отображений.— Кибернетика, 1977, № 5, с. 148–149.
8. *Lemarechal C., Nurminski E.* Sur la differentiabilite′ de la fonction d'appui du sous-differential approche.— C. r. Acad. sci. Ser. A, 1980, t. 290, p. 855–858.
9. *Auslender A.* Sur la differentiabilite′ de la fonction d'appui du sous-differential a e-pres.— C. r. Acad. sci., 1981, t. 292, p. 221–224.
10. *Hirriart-Urruty J.-B.* Lipschitz r-continuity of the approximate subdifferential of a convex function.— Math. scand., 1980, v. 47, p. 123–134.
11. *Демьянов В. Ф., Лупиков И. М.* Функции экстремума по E-субдифференциальному отображению.— Вестн. ЛГУ. Матем., механ., астрон., 1983, № 1, с. 27–32.
12. *Нурминский Е. А.* Глобальные свойства e-субградиентных отображений.— Кибернетика, 1986, № 1, с. 120–122.
13. *Kelley J. E.* The cutting plane method for solving convex programs.— J. Soc. Industr. and Appl. Math., 1960, v. 8(4), p. 703–712.
14. *Lemarechal C.* An algorithm for minimizing convex functions.— In: Proc. IFIP Congress. Stockholm: North-Holland Publ. Co., 1974, p. 553–556.

Поступила в редакцию 9.I.1985