

Н. А. Соколов (Москва, ЦЭМИ РАН). **О новых модификациях обобщенного седлового варианта метода уровней.**

В докладе предложены новые модификации обобщенного седлового варианта метода уровней [1], [2] для отыскания седловой точки выпукло-вогнутой функции $f(x, y)$, эффективное множество $\bar{G}$ которой содержится в декартовом произведении многогранников $G = G_x \times G_y$ и, вообще говоря, не совпадает с G , т. е. функция f определена конечными значениями не во всех точках G . «Классический» седловой вариант метода уровней ($G \equiv \bar{G}$) описан в [3], [4]. Рассмотренные модификации метода близки к описанным в [5]; установлена оценка сходимости этих модификаций.

1. Постановка задачи, описание оракула. Пусть $\bar{G}_x$ и $\bar{G}_y$ — выпуклые компакты с непустыми внутренностями, расположенные в конечномерных евклидовых пространствах E_x и E_y , и пусть $G_x \subset E_x$ и $G_y \subset E_y$ — выпуклые многогранники, содержащие, соответственно, $\bar{G}_x$ и $\bar{G}_y$. Обозначим $G = G_x \times G_y$, $\bar{G} = \bar{G}_x \times \bar{G}_y$, $E = E_x \times E_y$.

Рассмотрим функцию $f(z) = f(x, y)$, $z = (x, y)$, определенную конечными значениями на множестве $\bar{G}$. Относительно f примем следующие допущения:

А) функция f выпукло-вогнута на $\bar{G}$, т. е. при любом фиксированном $x^0 \in \bar{G}_x$ функция $f(x^0, y)$ вогнута по $y \in \bar{G}_y$ и при любом фиксированном $y^0 \in \bar{G}_y$ функция $f(x, y^0)$ выпукла по $x \in \bar{G}_x$;

В) функция f удовлетворяет условию Липшица с постоянной L , т. е. при любых $(x, y), (x', y'), (x'', y'') \in \bar{G}$ имеют место неравенства

$$|f(x, y') - f(x, y'')| \leq L\|y' - y''\|, \quad |f(x', y) - f(x'', y)| \leq L\|x' - x''\|. \quad (1)$$

Сделанные предположения относительно $\bar{G}$ и $f(z)$ обеспечивают непустоту множества $Z^* = X^* \times Y^*$ седловых точек функции $f(z)$ на $\bar{G}$, т. е. для любых точек $z = (x, y) \in \bar{G}$, $z^* = (x^*, y^*) \in Z^*$ выполнены неравенства $f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)$.

Кроме того, из А), В) вытекает субдифференцируемость функции f по x и супердифференцируемость функции f по y в каждой точке $(x, y) \in \bar{G}$. Для произвольной точки $y \in \bar{G}_y$, если $x \in \text{int } \bar{G}_x$ и l_x — произвольный элемент множества $\partial f_x(x, y)$, то $\|l_x\| \leq L$; для точек $x \in \partial G_x$ субдифференциал $\partial f_x(x, y)$ неограничен, но существует субградиент $l_x \in \partial f_x(x, y)$, для которого $\|l_x\| \leq L$. Аналогично, для произвольной точки $x \in \bar{G}_x$, если $y \in \text{int } \bar{G}_y$ и l_y — произвольный элемент множества $\partial f_y(x, y)$, то $\|l_y\| \leq L$; для граничных точек $(y \in \partial G_y)$ супердифференциал $\partial f_y(x, y)$ неограничен, но существует суперградиент $l_y \in \partial f_y(x, y)$, для которого $\|l_y\| \leq L$.

Далее предполагается, что многогранники G_x и G_y заданы двумя системами линейных неравенств. Что же касается выпуклого компакта $\bar{G} = \bar{G}_x \times \bar{G}_y \subset G$ и функции f , определенной на нем, то они задаются не аналитически, а алгоритмически, т. е. с помощью некоторого алгоритма, называемого *оракулом*. Ответы оракула подчиняются следующим правилам.

1) Если $z^0 \in \text{int } \bar{G}$, то оракул вычисляет $f(z^0)$ и находит некоторые субградиент $l_x(z^0) \in \partial_x f(z^0)$ и суперградиент $l_y(z^0) \in \partial_y f(z^0)$, т. е. ответом оракула (типа 1) является набор

$$f(z^0), l_x(z^0), l_y(z^0). \quad (2)$$

2) Если $z^0 = (x^0, y^0) \notin \bar{G}$, т. е. если $x^0 \in G_x \setminus \bar{G}_x$ и/или $y^0 \in \bar{G}_y \setminus \bar{G}_y$, то оракул находит гиперплоскость (одну или несколько), разделяющую точку z^0 и множество $\bar{G}$, т. е. вычисляет такие вектор $a_z(z^0) \in E$ единичной длины и скаляр $\alpha(z^0) \geq 0$ (отклик типа 2), что для всех $z \in \bar{G}$

$$\langle a_z(z^0), z - z^0 \rangle + \alpha_z(z^0) \leq 0. \quad (3)$$

3) Остался случай, когда точка z^0 принадлежит границе множества $\bar{G}$. В [1] для всех точек из $\partial \bar{G}$ оракул выдает информацию типа (2). В [2] оракул обрабатывает

граничные точки $\partial \bar{G}$ так же, как точки множества $G \setminus \bar{G}$, т. е. выдает одну или две гиперплоскости вида (3). В общем случае оракул в граничной точке множества $\bar{G}$ может использовать любой из предложенных способов (либо оба сразу).

2. Описание новых модификаций обобщенного седлового варианта метода уровней. В предложенных ранее седловых вариантах метода уровней [1], [2] отдельная итерация алгоритма, связанная с вычислением очередной точки $z_{k+1} \in G$, заключается в решении одной задачи линейного программирования и одной задачи квадратичного программирования; получены оценки скорости сходимости этих методов (порядка $O(k^{-1/2})$) при условии, что множества $\bar{G}_x$ и $\bar{G}_y$ замкнуты и имеют непустые внутренности.

Рассмотрим новые модификации обобщенного варианта метода уровней для нахождения на выпуклом компакте $\bar{G}$ седловых точек выпукло-вогнутой функции $f(z)$. Обобщение производится в нескольких направлениях: новый метод является «гибридом» седловых методов [1], [2]; можно произвольным образом нормировать векторы $l(z_i) = (l_x(z_i), -l_y(z_i))$ и $a_z(z_i) = (a_x(z_i), a_y(z_i))$; вместо одного, постоянного для всех итераций алгоритма числа λ используется набор, вообще говоря, различных чисел λ_i , $i = 1, 2, \dots$; разрешено добавлять на каждой итерации несколько линейных ограничений любого типа; вместо единичной матрицы в задаче квадратичного программирования используется произвольная симметричная положительно определенная матрица; вспомогательные задачи линейного и квадратичного программирования решаются приближенно.

Зададим множества $S_0 = \emptyset$, $\Lambda_0 = \emptyset$ и число $n_0 = 0$, также выберем числа $\lambda_0 \in (0, 1)$ и $F_0 > 0$ (например, $\lambda_0 = 1/2$, $F_0 = 1$) и произвольную точку $z_1 \in G$. Введем положительно определенную матрицу симметричную H (с собственными числами $0 < h_{\min} \leq \dots \leq h_{\max}$ и числом обусловленности $\mu = h_{\max}/h_{\min}$).

Пусть уже построено k ($k \geq 1$) точек $z_i \in G$, $i = 1, 2, \dots, k$, определены множества Λ_{k-1} , S_{k-1} , $|S_{k-1}| = n_{k-1}$. С точкой z_k свяжем m_k ($m_k \geq 1$) откликов оракула $J_k = \{n_{k-1} + 1, \dots, n_k\}$, $n_k = n_{k-1} + m_k$ (типы откликов оракула описаны ранее). Отклики типа 1 обозначим J'_k , они порождают три подтипа неравенств. К подтипу J'_{0k} , $|J'_{0k}| = m'_{0k}$, отнесем неравенства вида $F_j(z_k) \|l_j(z_k)\|^{-1} g[\langle l_{xj}(z_k), x - x_k \rangle - \langle l_{yj}(z_k), y - y_k \rangle] + t_y - t_x \leq 0$, к подтипам J'_{xk} и J'_{yk} , $|J'_{xk}| = |J'_{yk}| = m'_{1k}$, отнесем, соответственно, неравенства вида $F_0[\langle l_{xj}(z_k), x - x_k \rangle + f(z_k)] - t_x \leq 0$, $-F_0[\langle l_{yj}(z_k), y - y_k \rangle + f(z_k)] + t_y \leq 0$. Здесь $F_j(z_k)$ ($j \in J'_{0k}$) — положительные числа (задаваемые пользователем), $f(z_k)$, $l_j(z_k) = (l_{xj}(z_k), -l_{yj}(z_k)) \neq (0, \dots, 0)$ — информация, выдаваемая оракулом (см. (2)). Итак, множество J'_k разбито на три взаимно непересекающиеся подмножества J'_{0k} , J'_{xk} и J'_{yk} , $|J'_k| = m'_k = m'_{0k} + 2m'_{1k}$.

Оставшиеся отклики оракула (типа 2) обозначим $J''_k \equiv J_k \setminus J'_k$, $|J''_k| = m''_k$. Они определяют неравенства вида $A_j(z_k) [\langle a_{xj}(x_k), x - x_k \rangle + \langle a_{yj}(y_k), y - y_j \rangle + \alpha_j(z_k)] + t_y - t_x \leq 0$, где числа $A_j(z_k) > 0$, $j \in J''_k$, задаются пользователем, $a_j(z_k) = (a_{xj}(x_k), a_{yj}(y_k))$, $\|a_j(z_k)\| = 1$, $j \in J''_k$. В итоге имеем $m_k = m'_k + m''_k$. Положим $S_k = S_{k-1} \cup J_k$, $S'_k = S'_{k-1} \cup J'_k$, $S''_k = S''_{k-1} \cup J''_k$.

Поиск точки $z_{k+1} \in G$ начинается с решения задачи $\mathcal{A}_k$ и задачи $\tilde{\mathcal{A}}_k$, ей двойственной. Задача линейного программирования $\mathcal{A}_k$ имеет вид:

$$\mathcal{A}_k : \begin{cases} t_y - t_x \rightarrow \max, & (x, y) \in G, \\ F_0[\langle l_{xj}(z_i), x - x_i \rangle + f(z_i)] - t_x \leq 0, & j \in J'_{xi}, \\ -F_0[\langle l_{yj}(z_i), y - y_i \rangle + f(z_i)] + t_y \leq 0, & j \in J'_{yi}, \quad 1 \leq i \leq k, \\ F_j(z_i) \|l_j(z_i)\|^{-1} \langle l_j(z_i), z - z_i \rangle + t_y - t_x \leq 0, & j \in J'_{0i}, \\ A_j(z_i) [\langle a_j(z_i), z - z_i \rangle + \alpha_j(z_i)] + t_y - t_x \leq 0, & j \in J''_i, \end{cases} \quad (4)$$

переменными которой являются точка $z = (x, y)$ и два числа t_x и t_y . Обозначим Δ_k — максимальное значение линейной формы задачи $\mathcal{A}_k$. По построению $\Delta_k \geq 0$ для всех $k \geq 1$ и последовательность $\{\Delta_k\}$ монотонно невозрастает.

Если множество $J'_k \neq \emptyset$, то вычисляется k -е приближение z_k^* к множеству Z^* :

$$\begin{aligned} x_k^* &= \sum_{i=1}^k \left[\sum_{j \in J'_{yi}} v_j + \sum_{j \in J'_{0i}} w_j \right] x_i / \sum_{i=1}^k \left[\sum_{j \in J'_{yi}} v_j + \sum_{j \in J'_{0i}} w_j \right], \\ y_k^* &= \sum_{i=1}^k \left[\sum_{j \in J'_{xi}} u_j + \sum_{j \in J'_{0i}} w_j \right] y_i / \sum_{i=1}^k \left[\sum_{j \in J'_{xi}} u_j + \sum_{j \in J'_{0i}} w_j \right], \end{aligned} \quad (5)$$

где $u_j \geq 0$, $j \in J'_{xi}$, $v_j \geq 0$, $j \in J'_{yi}$, $w_j \geq 0$, $j \in J'_{0i}$, $1 \leq i \leq k$ — компоненты оптимального плана задачи $\tilde{\mathcal{A}}_k$, соответствующие первым трем первым группам неравенств в задаче $\mathcal{A}_k$. Можно показать, что, начиная с некоторого номера $\bar{k}$, при $k \geq \bar{k}$ множество $J'_k \neq \emptyset$ и оба знаменателя положительны. Итерационный процесс оканчивается, если точка z_k^* «близка» к множеству седловых точек Z^* , в противном случае доопределяется множество $\Lambda_k = \Lambda_{k-1} \cup \{\lambda_j: \lambda_j \in (0, 1), j \in J'_{0k}, \lambda_j = \lambda_0, j \in J'_k \setminus J'_{0k}\}$, вводятся масштабные множители $F_j(z_k) > 0$, $j \in J'_k \setminus J'_{0k}$, $A_j(z_k) > 0$, $j \in J''_k$, и для нахождения точки z_{k+1} решается задача квадратичного программирования $\mathcal{B}_k$ вида

$$\mathcal{B}_k: \quad \langle H(z - z_k), z - z_k \rangle \rightarrow \min, \quad z \in G_k(\Delta_k; \Lambda_k), \quad (6)$$

$$\begin{aligned} G_k(t; \Lambda_k) &= \{z = (x, y) \in G: \quad t_y - t_x \geq \lambda_0 t, \\ &\quad F_0[\langle l_{xj}(z_i), x - x_i \rangle + f(z_i)] - t_x \leq 0, \quad j \in J'_{xi}, \quad 1 \leq i \leq k, \\ &\quad -F_0[\langle l_{yj}(z_i), y - y_i \rangle + f(z_i)] + t_y \leq 0, \quad j \in J'_{yi}, \quad 1 \leq i \leq k, \\ &\quad F_j(z_i) \|l_j(z_i)\|^{-1} \langle l_j(z_i), z - z_i \rangle + \lambda_j t \leq 0, \quad j \in J'_{0i}, \quad 1 \leq i \leq k, \\ &\quad A_j(z_i) [\langle a_j(z_i), z - z_i \rangle + \alpha_j t] + \lambda_j t \leq 0, \quad j \in J''_i, \quad 1 \leq i \leq k\}; \end{aligned} \quad (7)$$

непустой при фиксированном $t \in [0, \Delta_k]$ выпуклый многогранник $G_k(t; \Lambda_k) \subset E$ является проекцией на E многогранника, определяемого при помощи условий задачи $\mathcal{A}_k$ и дополнительного условия $t_y - t_x \geq \lambda_0 t$. Далее итерационный процесс повторяется для новой точки z_{k+1} .

3. Обоснование сходимости седловых вариантов обобщенного метода уровней. В качестве меры близости точки $z = (x, y)$ к множеству Z^* естественно (см. [1]–[4]) взять сумму уклонов точек x и y от множеств X^* и Y^* , т. е. $\Delta(z) = (\varphi(x) - \varphi^*) + (\psi^* - \psi(y)) = \varphi(x) - \psi(y)$, где $\varphi^* = \psi^* = f(x^*, y^*)$ — оптимальное значение пары задач $\varphi(x) \equiv \max_{y \in \bar{G}_y} f(x, y) \rightarrow \min_{x \in \bar{G}_x} \psi(y) \equiv \min_{x \in \bar{G}_x} f(x, y) \rightarrow \max_{y \in \bar{G}_y}$. Функция $\Delta(z)$ неотрицательна для всех точек $z \in G$ и обращается в нуль лишь при $z \in Z^*$. Будем использовать метод до тех пор, пока не получим неравенство

$$\Delta(z_k^*) \leq \varepsilon, \quad (8)$$

где $\varepsilon > 0$ — наперед заданное число. Точку z_k^* , вычисленную по формулам (5) и удовлетворяющую (8), назовем ε -седловой точкой функции $f(z)$ на $\bar{G}$.

Положим

$$\begin{aligned} \hat{F}_0 &= \begin{cases} 1/F_0, & S'_k \setminus S'_{0k} \neq \emptyset, \\ 0, & S'_k \setminus S'_{0k} = \emptyset, \end{cases} \quad \hat{F}_k = \begin{cases} \max_{1 \leq i \leq k} \max_{j \in J'_{0i} \neq \emptyset} \|l_j(z_i)\|/F_j(z_i), & S'_{0k} \neq \emptyset, \\ 0, & S'_{0k} = \emptyset, \end{cases} \\ \tilde{F}_k &= \max\{\hat{F}_0, \hat{F}_k\}, \quad k = 1, 2, \dots, \quad \tilde{F} = \sup_{k \geq 1} \{\tilde{F}_k\}. \end{aligned}$$

Теорема. Пусть f — выпукло-вогнутая функция, определенная конечными значениями на выпуклом компакте $\bar{G} = \bar{G}_x \times \bar{G}_y$ и удовлетворяющая на нем условию Липшица (1), а $G = G_x \times G_y$ — многогранник, содержащий $\bar{G}$, каждый из

многогранников G_x и G_y имеет диаметр d , каждый из компактов $\overline{G}_x$ и $\overline{G}_y$ содержит шар радиуса $r > 0$. Пусть седловой вариант метода уровней (4)–(7) задается с помощью положительно определенной симметричной матрицы H с числом обусловленности μ и трех последовательностей чисел:

$$\begin{aligned}\Lambda_k &= \{\lambda_j : 0 < \underline{\lambda} \leq \lambda_j \leq \overline{\lambda} < 1, j \in J'_{0i}, \lambda_j = \lambda_0 \in (\underline{\lambda}, \overline{\lambda}), j \in J'_i \setminus J'_{0i}, 1 \leq i \leq k\}, \\ F_k &= \{F_0\} \cup \{F_j(z_i) : 0 < F_j(z_i) \leq \overline{F} < \infty, \frac{\|l_j(z_i)\|}{F_j(z_i)} \leq \tilde{F} < \infty, j \in J'_{0i}, 1 \leq i \leq k\}, \\ A_k &= \{A_j(z_i) : 0 < \underline{A} \leq A_j(z_i) \leq \overline{A} < \infty, j \in J''_i, 1 \leq i \leq k\},\end{aligned}$$

где $1/\tilde{F} \leq F_0 \leq \overline{F}/(L\sqrt{2})$, если $S'_{0k} \neq S'_k \neq \emptyset$, и пусть $\overline{B} = \max\{\overline{A}, \overline{F}\}$.

Тогда седловой вариант метода уровней вырабатывает при $k > K(\underline{A}r)$ точку z_k^* , мера близости которой к седловому множеству Z^* допускает оценку

$$\Delta(z_k^*) \leq CQk^{-1/2}, \quad (9)$$

где $C \equiv [2L(d-r)/(\underline{A}r) + \tilde{F}]$, $Q = \overline{B}d\sqrt{2\mu}\underline{\lambda}^{-1}(1 - \overline{\lambda}^2)^{-1/2}$, $K(z) \equiv (Q/z)^2$.

В частности, для любого $\varepsilon > 0$ можно найти ε -седловую точку $f(z)$ на $\overline{G}$ за k_ε итераций, причем $k_\varepsilon \leq 1 + \max\{K(\varepsilon/C), K(\underline{A}r)\}$.

В неравенстве (9) обычно константы L , r , a , возможно, и $\underline{A}$, $\tilde{F}$, $\overline{B}$, $\underline{\lambda}$, $\overline{\lambda}$, определяющие C и Q , неизвестны. Поэтому для получения оценки (9) одновременно с задачей A_k вида (4) будем решать, например, некоторую (другую) задачу линейного программирования A'_k , отличающуюся от задачи A_k отсутствием переменных t_x и t_y в последней группе ограничений, а также отсутствием всех нормирующих множителей A_j , $j \in S''_k$, F_0 , F_j , $j \in S'_{0k}$ (точнее, все эти числа равны 1). Пусть Δ'_k — максимальное значение линейной формы задачи A'_k . Тогда легко проверяется, что $\Delta'_k \geq \Delta_k/\tilde{F}$. Кроме того, для $k > K(\underline{A}r)$ множество $S'_k \neq \emptyset$ и можно определить точку $z_k'^* = (x_k'^*, y_k'^*) \in \overline{G}$ с помощью формул, аналогичных формулам (5). Можно показать, что последовательность $\{\Delta'_k\}$ стремится к нулю с той же скоростью, что и $\{\Delta_k\}$ (точнее, если $\Delta_k < \underline{A}r$, то $\Delta(z_k'^*) \leq \Delta'_k \leq C\Delta_k$).

Для поиска ε -седловой точки в предлагаемом варианте метода уровней можно использовать правило остановки: $\min\{\Delta'_k, \min_{j \in J'_i \neq \emptyset: 1 \leq i \leq k} \{\|l_j(z_i)\|/(d\sqrt{2})\}\} \leq \varepsilon$. Конечной точкой можно считать либо точку $z_k'^*$, если $\Delta'_k \leq \varepsilon$, либо точку z_k^* , если наименьшая норма вектора $l_j(z_k^*)$, $1 \leq k^* \leq k$, достаточно мала. Итак, ε -седловая точка функции $f(z)$ на множестве G находится для любого $\varepsilon > 0$ за конечное число шагов, равное $O(\varepsilon^{-2})$. Для уменьшения трудоемкости метода уровней рекомендуется решать задачу A'_k не на каждой итерации, а, например, через 10 итераций.

В докладе рассматриваются также варианты обобщенного седлового варианта метода уровней для случая, когда на каждой итерации одна и/или обе вспомогательные задачи (A_k, B_k) решаются приближенно; приводится оценка скорости сходимости и для этих вариантов.

Работа выполнена при финансовой поддержке РФФИ, проект № 05-01-00491.

СПИСОК ЛИТЕРАТУРЫ

1. Бэр К., Гольштейн Е. Г., Соколов Н. А. Метод отыскания седловой функции, область определения которой содержится в многограннике. — Экономика и матем. методы, 2001, т. 37, № 3, с. 97–105.
2. Гольштейн Е. Г. Обобщенный седловой вариант метода уровней. — Журнал вычисл. матем. и матем. физики, 2001, т. 41, № 8, с. 1139–1147.
3. Lemaréchal C., Nemirovskii A., Nesterov Yu. New Variants of Bundle Methods. — Math. Progr., 1995, v. 69, p. 111–147.

4. Гольштейн Е. Г., Немировский А. С., Нестеров Ю. Е. Метод уровней, его обобщения и приложения. — Экономика и матем. методы, 1995, т. 31, в. 3, с. 164–180.
5. Соколов Н. А. Новые варианты метода обобщенного метода уровней для минимизации выпуклой недифференцируемой функции, не все значения которой конечны. — Обозрение прикл. и промышл. матем., 2006, т. 13, в. 5, с. 927–930.