О МИНИМИЗАЦИИ ФУНКЦИИ МАКСИМУМА СИЛЬНО ВЫПУКЛЫХ ФУНКЦИЙ*

Л.Н. Полякова

lnpol07@mail.ru

11 декабря 2014 г.

Аннотация. Рассматривается задача безусловной минимизации функции максимума сильно выпуклых функций с постоянным шагом. Конструируется оптимизационный алгоритм, в котором для нахождения направления спуска приходится решать задачу квадратичного программирования. Доказывается, что генерируемая этим алгоритмом последовательность сходится к точке минимума целевой функции с линейной скоростью.

1°. Введение. Пусть

$$f(x) = \max_{i \in I} f_i(x),$$

где $f_i(x)$, $i \in I = \{1, \ldots, s\}$, — дважды непрерывно дифференцируемые сильно выпуклые на $\mathbb{R}^n$ функции. Обозначим через $m_i > 0$, $i \in I$, их константы сильной выпуклости. То есть,

$$f_i(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f_i(x_1) + (1-\lambda) f_i(x_2) - m_i \lambda (1-\lambda) ||x_1 - x_2||^2,$$

$$\forall\, \lambda \in [0,1], \quad \forall\, x_1, x_2 \in \mathbb{R}^n, \quad i \in I.$$

Тогда функция f также является сильно выпуклой с константой сильной выпуклости равной $m = \min\limits_{i \in I} m_i > 0$.

Рассмотрим оптимизационную задачу:

$$f(x) \to \min_{x \in \mathbb{R}^n}. \tag{1}$$

Решение задачи (1) существует и точка минимума функции f на $\mathbb{R}^n$ единственна.

*Семинар по конструктивному негладкому анализу и недифференцируемой оптимизации «CNSA & NDO»: `http://www.apmath.spbu.ru/cnsa/`

Выпишем критерии оптимальности для задачи (1):

для того чтобы точка $x^* \in \mathbb{R}^n$ была точка минимума функции f на $\mathbb{R}^n$, необходимо и достаточно, чтобы

$$0_n \in \partial f(x^*) = co \left\{ \bigcup_{i \in R(x^*)} f'(x^*) \right\}, \tag{2}$$

где

$$R(x^*) = \{i \in I \mid f_i(x^*) = f(x^*)\}.$$

Заметим, что условие (2) можно переформулировать таким образом:

для того чтобы точка $x^* \in \mathbb{R}^n$ была точка минимума функции f на $\mathbb{R}^n$, необходимо и достаточно, чтобы нашлись такие числа $\lambda_i(x^*)$,

$$\lambda_i(x^*) \geqslant 0 \quad \forall\, i \in I, \quad \sum_{i=1}^{s} \lambda_i(x^*) = 1,$$

что

$$0_n = \sum_{i=1}^{s} \lambda_i(x^*) f_i'(x^*)$$

и

$$\lambda_i(x^*)(f_i(x^*) - f(x^*)) = 0 \quad \forall\, i \in I.$$

Предположим, что существует такая константа $M \geqslant \max\{m, 1\}$, что выполнено неравенство

$$\langle f_i''(x)w, w\rangle \leqslant M||w||^2 \quad \forall\, x \in \mathbb{R}^n, \quad \forall\, w \in \mathbb{R}^n, \quad \forall\, i \in I, \tag{3}$$

где $f_i''(x)$, $i \in I$, — матрицы вторых производных функций f_i в точке x.

С каждой точкой $x \in \mathbb{R}^n$ свяжем оптимизационную задачу:

$$\max_{i \in I} \left\{ (f_i(x) - f(x))M + \langle f_i'(x), w\rangle + \frac{1}{2}||w||^2 \right\} \to \min_{w \in \mathbb{R}^n}. \tag{4}$$

Решение этой задачи $w(x)$ также единственно.

Обозначим

$$\nu(x) = \max_{i \in I} \left\{ (f_i(x) - f(x))M + \langle f_i'(x), w(x)\rangle + \frac{1}{2}||w(x)||^2 \right\}.$$

Из необходимых и достаточных условий минимума задачи (4) следует существование таких коэффициентов

$$\lambda_i(x) \geqslant 0, \quad i \in I, \quad \sum_{i=1}^{s} \lambda_i(x) = 1,$$

что

$$w(x) = -\sum_{i=1}^{s} \lambda_i(x) f_i'(x),$$

$$\nu(x) = \sum_{i=1}^{s} \lambda_i(x)(f_i(x) - f(x))M - \frac{1}{2}||w(x)||^2 \leqslant 0.$$

ЛЕММА 1. *Справедливы следующие утверждения:*

1) *функция $\nu(x)$ тогда и только тогда равна нулю, когда точка $x \in \mathbb{R}^n$ является точкой минимума функции f на $\mathbb{R}^n$;*

2) *если $w(x) = 0_n$, то $\nu(x) = 0$.*

Доказательство. Следует из необходимых и достаточных условий минимума задачи (4). □

Таким образом, если точка $x^* \in \mathbb{R}^n$ — точка минимума функции f на $\mathbb{R}^n$, то $w(x^*) = 0_n$.

В дальнейшем будет показано, что если точка $x \in \mathbb{R}^n$ не есть точка минимума функции f на $\mathbb{R}^n$, то направление $w(x)$ является направлением спуска функции f в точке x.

2°. Алгоритм минимизации. Опишем метод минимизации функции f на всем пространстве.

Выберем произвольную точку $x_0 \in \mathbb{R}^n$. Если $\nu(x_0) = 0_n$, то x_0 есть точка минимума функции f на $\mathbb{R}^n$, и процесс останавливается.

Пусть уже найдена точка $x_k \in \mathbb{R}^n$. Если $\nu(x_k) = 0_n$, то x_k — точка минимума функции f на $\mathbb{R}^n$, и процесс закончен.

Пусть $\nu(x_k) \neq 0_n$, тогда x_k не является точкой минимума функции f на $\mathbb{R}^n$. Определим направление спуска $w(x_k)$, то есть, решим оптимизационную задачу (4) при $x = x_k$. Тогда найдутся такие коэффициенты

$$\lambda_i(x_k) \geqslant 0, \quad i \in I, \quad \sum_{i=1}^{s} \lambda_i(x_k) = 1,$$

что справедливо

$$w(x_k) = -\sum_{i=1}^{s} \lambda_i(x_k) f_i'(x_k) \neq 0_n,$$

$$\nu(x_k) = \sum_{i=1}^{s} \lambda_i(x_k)(f_i(x_k) - f(x_k))M - \frac{1}{2}||w(x_k)||^2 < 0.$$

Положим

$$\bar{\alpha} = \frac{1}{M} \quad \text{и} \quad x_{k+1} = x_k + \bar{\alpha} w(x_k).$$

ЛЕММА 2. *В каждой точке $x_k \in \mathbb{R}^n$ выполняется неравенство*

$$f(x_{k+1}) - f(x_k) \leqslant \frac{1}{M}\nu(x_k) \leqslant 0. \qquad (5)$$

Доказательство. В силу свойств функций f_i справедливо равенство

$$f_i(x + \alpha w) = f_i(x) + \alpha\langle f_i'(x), w\rangle + \frac{\alpha^2}{2}\langle f_i''(x + \eta_i w)w, w\rangle,$$

$$\eta_i \in (0,1) \quad \forall i \in I, \quad \forall x \in \mathbb{R}^n, \quad w \in \mathbb{R}^n, \quad \forall \alpha \in [0,1].$$

Тогда из (3) для любого $i \in I$ имеем

$$f_i(x_k + \bar{\alpha}w(x_k)) - f(x_k) \leqslant f_i(x_k) - f(x_k) + \bar{\alpha}\langle f_i'(x_k), w(x_k)\rangle + \frac{\bar{\alpha}^2}{2}M||w(x_k)||^2 =$$

$$= f_i(x_k) - f(x_k) + \frac{1}{M}\langle f_i'(x_k), w(x_k)\rangle + \frac{1}{2M}||w(x_k)||^2.$$

Отсюда имеем

$$\max_{i \in I} f_i(x_k + \bar{\alpha}w(x_k)) - f(x_k) \leqslant$$

$$\leqslant \max_{i \in I}\left\{f_i(x_k) - f(x_k) + \frac{1}{M}\langle f_i'(x_k), w(x_k)\rangle + \frac{1}{2M}||w(x_k)||^2\right\} =$$

$$= \frac{1}{M}\max_{i \in I}\left\{(f_i(x_k) - f(x_k))M + \langle f_i'(x_k), w(x_k)\rangle + \frac{1}{2}||w(x_k)||^2\right\} = \frac{1}{M}\nu(x_k) < 0.$$

Таким образом,

$$f(x_{k+1}) - f(x_k) \leqslant \frac{1}{M}\nu(x_k).$$

Лемма доказана. □

Неравенство (5) показывает, что процесс является релаксационным.

СЛЕДСТВИЕ 1. *Если последовательность $\{x_k\}$ бесконечна, то справедливо утверждение*

$$f(x_{k+1}) - f(x_k) \to 0.$$

Доказательство. Предположим, что $f(x_{k+1}) - f(x_k) \not\to 0$. Тогда найдутся такие положительные числа $a > 0$, $K > 0$ и подпоследовательность индексов $\{k_s\}$, что справедливо неравенство

$$f(x_{k_s+1}) - f(x_{k_s}) < -a \quad \forall\, k_s > K.$$

Отсюда в силу монотонного убывания последовательности $\{f(x_k)\}$ вытекает, что функция f неограничена снизу. Это противоречит ограниченности снизу функции f на $\mathbb{R}^n$. □

СЛЕДСТВИЕ 2. *Если последовательность* $\{x_k\}$ *бесконечна, то*

$$\nu(x_k) \to 0 \quad \textit{при } k \to +\infty.$$

Так как для любой сильно выпуклой функции множество

$$\mathcal{L}(x_0) = \{x \in \mathbb{R}^n \mid f(x) \leqslant f(x_0)\}$$

компактно, то последовательность $\{x_k\}$, генерируемая этим методом, содержится в множестве $\mathcal{L}(x_0)$.

ЛЕММА 3. *Для данного метода в каждой точке* $x_k \in \mathbb{R}^n$ *справедливо неравенство:*

$$\nu(x_k) \leqslant m\left(f(x^*) - f(x_k)\right), \tag{6}$$

где x^* *— точка минимума функции* f *на* $\mathbb{R}^n$.

Доказательство. Для точек x_k, x^* и для каждого индекса $i \in I$ верно неравенство

$$f_i(x^*) - f_i(x_k) \geqslant \langle f_i'(x_k), x^* - x_k\rangle + \frac{m}{2}||x^* - x_k||^2, \quad i \in I.$$

Умножив его на m, получим

$$\langle f_i'(x_k), m(x^* - x_k)\rangle \leqslant m(f_i(x^*) - f_i(x_k)) - \frac{m^2}{2}||x^* - x_k||^2, \quad i \in I.$$

Положим в (4) $w = m(x^* - x_k)$. Тогда

$$\nu(x_k) \leqslant \max_{i\in I}\left\{(f_i(x_k) - f(x_k))M + \langle f_i'(x_k), m(x^* - x_k)\rangle + \frac{m^2}{2}||x^* - x_k||^2\right\} \leqslant$$

$$\leqslant \max_{i\in I}\left\{(f_i(x_k) - f(x_k))m + m(f_i(x^*) - f_i(x_k))\right\} \leqslant$$

$$\leqslant m \max_{i\in I}\{f_i(x^*) - f(x_k)\} \leqslant m(f(x^*) - f(x_k)).$$

Лемма доказана. □

СЛЕДСТВИЕ 3. *Из формул (5) и (6) вытекает неравенство*

$$f(x_{k+1}) - f(x_k) \leqslant \frac{m}{M}(f(x^*) - f(x_k)). \tag{7}$$

Таким образом, последовательность $\{x_k\}$ является сходящейся и ее предельной точкой является точка x^*.

ТЕОРЕМА 1. *В данном методе при произвольной начальной точке $x_0 \in \mathbb{R}^n$ последовательность $\{x_k\}$ сходится к точке минимума x^* функции f со скоростью геометрической прогрессии:*

$$f(x_{k+1}) - f(x^*) \leqslant q(f(x_k) - f(x^*)),$$

где $q = 1 - \dfrac{m}{M}$, и существует положительное число Q, что справедливо неравенство

$$||x_k - x^*|| \leqslant Q(\sqrt{q})^k.$$

Доказательство. Из (7) получим

$$f(x_{k+1}) - f(x^*) \leqslant f(x_k) - f(x^*) + \frac{m}{M}(f(x^*) - f(x_k)) =$$

$$= \left(1 - \frac{m}{M}\right)(f(x_k) - f(x^*)) = q(f(x_k) - f(x^*)).$$

Следовательно, для каждого фиксированного $k > 0$ имеем

$$f(x_k) - f(x^*) \leqslant q^k(f(x_0) - f(x^*)).$$

Поскольку x^* есть точка минимума функции f на $\mathbb{R}^n$, то

$$f(x_k) - f(x^*) \geqslant \frac{m}{2}||x_k - x^*||^2.$$

Поэтому

$$||x_k - x^*||^2 \leqslant \frac{2}{m}(f(x_k) - f(x^*)) \leqslant \frac{2}{m}(f(x_0) - f(x^*))q^k.$$

Обозначим

$$Q = \sqrt{\frac{2}{m}(f(x_0) - f(x^*))},$$

тогда

$$||x_k - x^*|| \leqslant \sqrt{\frac{2}{m}(f(x_k) - f(x^*))q^k} = Q(\sqrt{q})^k.$$

Теорема доказана. □

3°. Вычислительный аспект решения вспомогательной задачи. Основная вычислительная трудность в алгоритме — это определение направления спуска $w(x_k)$. Покажем, что задачу (4) можно свести к задаче квадратичного программирования.

Обозначим

$$z = \max_{i \in I} \left\{ (f_i(x) - f(x))M + \langle f_i'(x), w(x) \rangle \right\}.$$

Тогда задача (4) эквивалентна следующей задаче квадратичного программирования в пространстве $\mathbb{R}^n \times \mathbb{R}$: минимизировать функцию

$$z + \frac{1}{2}\langle w, w\rangle$$

на множестве

$$T = \{[w, z] \in \mathbb{R}^n \times \mathbb{R} \mid \langle f_i'(x), w(x)\rangle - z \leqslant (f(x) - f_i(x))M,\ i \in I\}.$$

З а м е ч а н и е. Пшеничным Б.Н. [2] и Полаком Е. [4] для нахождения направления спуска $w(x_k)$ решалась оптимизационная задача:

$$w(x_k) = \arg \min_{w\in\mathbb{R}^n} \max_{i\in I} \left\{ f_i(x_k) - f(x_k) + \langle f_i'(x_k), w\rangle + \frac{1}{2}||w||^2 \right\},$$

а шаг выбирался из условия одномерной минимизации или из правила Армихо. Была доказана теорема, аналогичная теореме 1.

ЛИТЕРАТУРА

1. Демьянов В.Ф., Малоземов В.Н. *Введение в минимакс.* М., Наука, 1972. 368 с.

2. Пшеничный Б.Н. *Метод линеаризации.* М. Наука. 1983. 136 с.

3. Пшеничный Б.Н., Данилин Ю.М. *Численные методы в экстремальных задачах.* М. Наука. 1975. 320 с.

4. Polak E. *Basics of minimax algorithms.* // In Nonsmooth Optimization and Related Topics. Ed. Clarke F., Demyanov V.F., Giannessi F. N.Y.: Plenum Press, 1989. P. 343–369.