МЕТОД НЕСТЕРОВА МИНИМИЗАЦИИ ВЫПУКЛЫХ ФУНКЦИЙ*

М. В. Долгополик

maxim.dolgopolik@gmail.com

7 апреля 2016 г.

Аннотация. В докладе обсуждается субградиентный метод минимизации негладкой выпуклой функции на выпуклом множестве, предложенный Ю.Е. Нестеровым (см. видео-лекцию [1]).

1°. Субградиентные методы минимизации выпуклых функций. Рассмотрим задачу выпуклой оптимизации вида

$$f(x) \to \min_{x \in Q}, \qquad (1)$$

где $f \colon \mathbb{R}^n \to \mathbb{R}$ — заданная выпуклая функция, а $Q \subset \mathbb{R}^n$ — замкнутое выпуклое множество. Требуется построить численный метод решения данной задачи. При этом предполагается, что в каждой точке $x \in Q$ мы можем вычислять один из субградиентов функции f в этой точке, который мы обозначим $v(x)$.

Для того чтобы понять идею, лежащую в основе метода Нестерова (также называемого субградиентным методом с двойным усреднением), необходимо проследить историю развития субградиентных методов минимизации негладких выпуклых функций. Первым субградиентным методом минимизации негладких выпуклых функций был *метод Шора* [2]. Данный метод определяется следующим соотношением:

$$x_{k+1} = \mathrm{Pr}_Q(x_k - \alpha_k v(x_k)), \quad k \in \{0\} \cup \mathbb{N},$$

где Pr_Q — оператор проектирования на множество Q, $x_0 \in Q$ — начальное приближение, а α_k — заданные положительные числа, определяющие величину шага на каждой итерации.

Обозначим через x^* *произвольный* оптимальный план задачи (1) (мы предполагаем, что он существует). Следующее утверждение, содержит важную

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

оценку, которая позволяет установить условия сходимости метода Шора. Отметим, что схожие оценки будут появляться при исследовании всех остальных субградиентных методов, обсуждаемых в данном докладе.

ТЕОРЕМА 1. *Пусть последовательность $\{x_k\}$ построена по методу Шора и*

$$A_k = \sum_{i=0}^{k} \alpha_i.$$

Тогда для любого $k \in \mathbb{N}$ справедливо неравенство

$$\frac{1}{A_k}\sum_{i=0}^{k} \alpha_i f(x_i) - f(x^*) \leqslant \frac{1}{A_k}\left(\frac{1}{2}\|x_0 - x^*\|^2 + \frac{1}{2}\sum_{i=0}^{k}\alpha_i^2\|v(x_i)\|^2\right), \tag{2}$$

где $\|\cdot\|$ — евклидова норма. Отсюда, в частности, следует, что если последовательность $\{v(x_k)\}$ ограничена, $\alpha_k \to 0$ при $k \to \infty$ и

$$\sum_{k=0}^{\infty} \alpha_k = +\infty,$$

то

$$\lim_{k\to\infty} \frac{1}{A_k}\sum_{i=0}^{k}\alpha_i f(x_i) = f(x^*), \quad \liminf_{k\to\infty} f(x_k) = f(x^*). \tag{3}$$

Доказательство. Для любого $i \in \{0\} \cup \mathbb{N}$ обозначим $r_i = \|x_i - x^*\|^2$. По определению последовательности $\{x_k\}$ будет

$$r_{i+1} = \|\mathrm{Pr}_Q(x_i - \alpha_k v(x_i)) - x^*\|^2.$$

Заметим, что для любых $y \in \mathbb{R}^n$ и $x \in Q$ справедливо неравенство

$$\|\mathrm{Pr}_Q(y) - x\|^2 \leqslant \|y - x\|^2.$$

Поэтому

$$\begin{aligned} r_{i+1} \leqslant & \ \|x_i - \alpha_i v(x_i) - x^*\|^2 = \\ & = r_i - 2\alpha_i\langle v(x_i), x_i - x^*\rangle + \alpha_i^2\|v(x_i)\|^2 \quad \forall i \in \{0\} \cup \mathbb{N}. \end{aligned}$$

Данное неравенство зависит от i. Просуммируем его по всем i от 0 до k, где $k \in \{0\} \cup \mathbb{N}$ фиксировано. Получим

$$r_{k+1} - r_0 \leqslant -2\sum_{i=0}^{k}\alpha_i\langle v(x_i), x_i - x^*\rangle + \sum_{i=0}^{k}\alpha_i^2\|v(x_i)\|^2.$$

Так как $v(x_i)$ — субградиент функции f в точке x_i, то

$$f(x) - f(x_i) \geqslant \langle v(x_i), x - x_i \rangle \quad \forall x \in \mathbb{R}^n,$$

Поэтому

$$\sum_{i=0}^{k} \alpha_i (f(x_i) - f(x^*)) \leqslant \sum_{i=0}^{k} \alpha_i \langle v(x_i), x_i - x^* \rangle \leqslant$$
$$\leqslant \frac{1}{2} r_0 - \frac{1}{2} r_{k+1} + \frac{1}{2} \sum_{i=0}^{k} \alpha_i^2 \|v(x_i)\|^2 \leqslant \frac{1}{2} r_0 + \frac{1}{2} \sum_{i=0}^{k} \alpha_i^2 \|v(x_i)\|^2,$$

откуда следует, что

$$\frac{1}{A_k} \sum_{i=0}^{k} \alpha_i f(x_i) - f(x^*) \leqslant \frac{1}{A_k} \left(\frac{1}{2} \|x_0 - x^*\|^2 + \frac{1}{2} \sum_{i=0}^{k} \alpha_i^2 \|v(x_i)\|^2 \right).$$

Таким образом, неравенство (2) доказано.

Предположим теперь, что последовательность $\{v(x_k)\}$ ограничена, $\alpha_k \to 0$ при $k \to \infty$ и

$$\sum_{k=0}^{\infty} \alpha_k = +\infty.$$

Покажем, что в этом случае

$$\lim_{k \to \infty} \frac{1}{A_k} \left(\frac{1}{2} \|x_0 - x^*\|^2 + \frac{1}{2} \sum_{i=0}^{k} \alpha_i^2 \|v(x_i)\|^2 \right) = 0.$$

Отсюда и из неравенства (2) следуют соотношения (3).

Достаточно проверить, что

$$\lim_{k \to \infty} \frac{1}{A_k} \sum_{i=0}^{k} \alpha_i^2 = 0.$$

Для этого зафиксируем произвольное $\varepsilon > 0$. По предположению $\alpha_k \to 0$ при $k \to \infty$, поэтому найдётся $k_0 \in \mathbb{N}$ такое, что

$$\alpha_k < \frac{\varepsilon}{2} \quad \forall k \geqslant k_0.$$

Следовательно,

$$\alpha_k^2 = \alpha_k \alpha_k < \frac{\varepsilon}{2} \alpha_k \quad \forall k \geqslant k_0.$$

Отсюда имеем

$$\frac{\sum\limits_{i=k_0}^{k} \alpha_i^2}{\sum\limits_{i=k_0}^{k} \alpha_i} < \frac{\varepsilon}{2} \quad \forall k \geqslant k_0.$$

Значит для всех $k \geqslant k_0$

$$\frac{1}{A_k} \sum_{i=k_0}^{k} \alpha_i^2 \leqslant \frac{\sum\limits_{i=k_0}^{k} \alpha_i^2}{\sum\limits_{i=k_0}^{k} \alpha_i} < \frac{\varepsilon}{2}. \tag{4}$$

По условию $A_k \to +\infty$ при $k \to \infty$, поэтому найдётся $k_1 \in \mathbb{N}$ такое, что

$$\frac{1}{A_k} \sum_{i=0}^{k_0-1} \alpha_i^2 < \frac{\varepsilon}{2} \quad \forall k \geqslant k_1. \tag{5}$$

Объединив неравенств (4) и (5), получим, что для всех $k \geqslant \max\{k_0, k_1\}$

$$\frac{1}{A_k} \sum_{i=0}^{k} \alpha_i^2 = \frac{1}{A_k} \sum_{i=0}^{k_0-1} \alpha_i^2 + \frac{1}{A_k} \sum_{i=k_0}^{k} \alpha_i^2 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Так как $\varepsilon > 0$ было выбрано произвольно, то

$$\lim_{k \to \infty} \frac{1}{A_k} \sum_{i=0}^{k} \alpha_i^2 = 0,$$

что и требовалось доказать. $\square$

Замечание 1. Заметим, что в предыдущей теореме можно избавиться от условия ограниченности последовательности $v(x_k)$, если в методе Шора нормировать субградиент $v(x_k)$, т.е. положить

$$x_{k+1} = \mathrm{Pr}_Q \left(x_k - \alpha_k \frac{v(x_k)}{\|v(x_k)\|} \right) \quad \forall k \in \{0\} \cup \mathbb{N}.$$

Следующим шагом в развитии субградиентных методов, после метода Шора, стал *метод зеркального спуска*, предложенный Немировским и Юдиными [3]. Метод зеркального спуска определяется следующим соотношением

$$x_{k+1} = \arg\min_{x \in Q} \left\{ \left\langle \sum_{i=0}^{k} \alpha_i v(x_i), x \right\rangle + \frac{1}{2} \|x - x_0\|^2 \right\}, \quad k \in \{0\} \cup \mathbb{N},$$

где $x_0 \in Q$ — начальное приближение, а α_k — положительные числа. Как и в случае метода Шора, можно показать, что для метода зеркального спуска справедливо неравенство

$$\frac{1}{A_k}\sum_{i=0}^{k}\alpha_i f(x_i) - f(x^*) \leqslant \frac{1}{A_k}\left(\frac{1}{2}\|x_0 - x^*\|^2 + \frac{1}{2}\sum_{i=0}^{k}\alpha_i^2\|v(x_i)\|^2\right), \tag{6}$$

где $A_k = \sum\limits_{i=0}^{k}\alpha_k$. Следовательно, для того чтобы гарантировать сходимости метода зеркального спуска необходимо потребовать, чтобы $\alpha_k \to 0$ при $k \to \infty$ и

$$\sum_{k=0}^{\infty}\alpha_k = +\infty.$$

Заметим, что из условия $\alpha_k \to 0$ следует, что субградиенты $v(x_i)$ входят в сумму

$$\Big\langle \sum_{i=0}^{k}\alpha_i v(x_i), x\Big\rangle + \frac{1}{2}\|x - x_0\|^2,$$

на основе которой определяется каждая итерация метода зеркального спуска, с уменьшающимися весами. Однако, метод зеркального спуска можно модифицировать так, чтобы субградиенты $v(x_i)$ входили с одинаковыми весами. Такая модификация метода зеркального спуска, предложенная Ю.Е. Нестеровым [4], называется *субградиентным методом с двойственным усреднением.*

Субградиентный метод с двойственным усреднением определяется следующим соотношением

$$x_{k+1} = \arg\min_{x\in Q}\left\{\frac{1}{A_k}\Big\langle \sum_{i=0}^{k}\alpha_i v(x_i), x\Big\rangle + \frac{\gamma_k}{2A_k}\|x - x_0\|^2\right\} \quad k \in \{0\}\cup\mathbb{N},$$

где $x_0 \in Q$ — начальное приближение, α_k — положительные числа, а $\{\gamma_k\}$ — неубывающая последовательность положительных шкалирующих коэффициентов. Заметим, что при $\gamma_k \equiv 1$ субградиентный метод с двойственным усреднением преобразуется в метод зеркального спуска.

Следующая теорема содержит достаточные условия сходимости субградиентного метода с двойственным усреднением.

ТЕОРЕМА 2 (Нестеров, [4])**.** *Пусть последовательность $\{x_k\}$ построена по субградиентному методу с двойственным усреднением. Тогда для любого $k \in \mathbb{N}$ справедливо неравенство*

$$\frac{1}{A_k}\sum_{k=1}^{n}\alpha_k f(x_k) - f(x^*) \leqslant \frac{1}{A_k}\left(\frac{\gamma_k}{2}\|x_0 - x^*\|^2 + \sum_{i=0}^{k}\frac{\alpha_i^2}{2\gamma_i}\|v(x_i)\|^2\right). \tag{7}$$

Отсюда, в частности, следует, что если последовательность $v(x_k)$ ограничена и

$$\lim_{k\to\infty} \frac{\gamma_k}{A_k} = 0, \qquad \lim_{k\to\infty} \frac{1}{A_k} \sum_{i=0}^{k} \frac{\alpha_i^2}{\gamma_i} = 0, \tag{8}$$

то

$$\lim_{k\to\infty} \frac{1}{A_k} \sum_{k=1}^{n} \alpha_k f(x_k) = f(x^*), \quad \liminf_{k\to\infty} f(x_k) = f(x^*).$$

Замечание 2. Заметим, что соотношения (8) будут выполнены, если положить $\alpha_k \equiv 1$ и $\gamma_k = \sqrt{k+1}$. В этом случае сумма

$$\frac{1}{A_k}\Big\langle \sum_{i=0}^{k} \alpha_i v(x_i), x \Big\rangle + \frac{\gamma_k}{2A_k}\|x - x_0\|^2,$$

на основе которой определяются итерации субградиентного метода с двойственным усреднением, принимает вид

$$\frac{1}{k+1}\Big\langle \sum_{i=0}^{k} v(x_i), x \Big\rangle + \frac{1}{2\sqrt{k+1}}\|x - x_0\|^2.$$

Таким образом, данная сумма содержит *усреднение* (среднее арифметическое) всех субградиентов вычисленных за предыдущие итерации.

2°. Субградиентный метод с двойным усреднением. В основных оценках (см. (2), (6) и (7)), с помощью которых доказывается сходимость субградиентных методов, участвует не значение целевой функции f в точке x_k, а лишь выпуклая комбинация

$$\frac{1}{A_k} \sum_{i=0}^{k} \alpha_i f(x_i).$$

Вследствие этого для субградиентных методов не удаётся установить сходимость последовательности $\{f(x_k)\}$ к оптимальному значению $f(x^*)$. Удаётся лишь доказать более слабое утверждение: $\liminf\limits_{k\to\infty} f(x_k) = f(x^*)$. Таким образом, субградиентные методы (по крайней мере с теоретической точки зрения) не являются сходящимися.

Заметим, что поскольку функция f выпукла, то

$$f(y_k) \leqslant \frac{1}{A_k} \sum_{i=0}^{k} \alpha_i f(x_i), \quad y_k = \frac{1}{A_k} \sum_{i=0}^{k} \alpha_i x_i.$$

Поэтому, если выполнены достаточные условия сходимости, указанные выше, то $f(y_k) \to f(x^*)$ при $k \to \infty$. Однако, сходящаяся (по функции) последовательность выпуклых комбинаций $\{y_k\}$ не используется при построении исходной последовательности $\{x_k\}$ ни в одном из субградиентные методов, описанных в предыдущем разделе.

Так как последовательность $\{y_k\}$ является сходящейся по функции, то возникает желание каким-то образом использовать эту последовательность при построении субградиентного метода. Самый простой и естественный способ использования последовательности $\{y_k\}$, заключается в том, чтобы в качестве результата очередной итерации субградиентного метода брать не точку x_{k+1}, а точку y_{k+1}, то есть полагать в качестве нового приближения вектор

$$\widehat{x}_{k+1} = \frac{1}{A_{k+1}} \sum_{i=0}^{k+1} \alpha_i x_i.$$

Заметим, что в этом случае

$$\widehat{x}_{k+1} = \frac{A_k}{A_{k+1}} \Big(\frac{1}{A_k} \sum_{i=0}^{k} \alpha_i x_i \Big) + \frac{\alpha_{k+1}}{A_{k+1}} x_{k+1} = \frac{A_k}{A_{k+1}} \widehat{x}_k + \frac{\alpha_{k+1}}{A_{k+1}} x_{k+1}.$$

Если обозначить $\tau_k = \alpha_{k+1}/A_{k+1}$, то получим

$$\widehat{x}_{k+1} = (1-\tau_k)\widehat{x}_k + \tau_k x_{k+1}.$$

Таким образом, мы приходим к следующей схеме субградиентного метода:

1) вычислим

$$x_k^+ = \arg\min_{x \in Q} \left\{ \frac{1}{A_k} \Big\langle \sum_{i=0}^{k} \alpha_i v(x_i), x \Big\rangle + \frac{\gamma_k}{2A_k} \|x - x_0\|^2 \right\},$$

2) положим $\tau_k = \alpha_{k+1}/A_{k+1}$ и $x_{k+1} = (1-\tau_k)x_k + \tau_k x_k^+$.

Здесь $\{\alpha_k\}$ — заданная последовательность положительных чисел, а $\{\gamma_k\}$ — неубывающая последовательность положительных шкалирующих коэффициентов.

Описанный выше метод называется *субградиентным методом с двойным усреднением* [5], т.к. в нём применяется как усреднение субградиентов, так и усреднение исходной последовательности (при $\alpha_k \equiv 1$). Исследуем сходимость данного метода.

Получим сначала несколько вспомогательных оценок последовательности $\{f(x_k)\}$.

ЛЕММА 1. *Пусть последовательность $\{x_k\}$ построена по субградиентному методу с двойным усреднением. Тогда для любого $k \in \{0\} \cup \mathbb{N}$ справедливо неравенство*

$$A_k f(x_k) \leqslant \sum_{i=0}^{k} \alpha_i \big[f(x_i) + \langle v(x_i), x - x_i\rangle\big] + \frac{\gamma_k}{2}\|x - x_0\|^2 + B_k \quad \forall x \in Q, \tag{9}$$

где

$$B_k = \frac{1}{2}\sum_{i=0}^{k} \frac{\alpha_i^2}{\gamma_{i-1}}\|v(x_i)\|^2$$

и $\gamma_{-1} = \gamma_0$.

Доказательство. Обозначим

$$\ell_k(x) = \sum_{i=0}^{k} \alpha_i \big[f(x_i) + \langle v(x_i), x - x_i\rangle\big]$$

и

$$\psi_k^* = \min_{x \in Q}\left(\ell_k(x) + \frac{\gamma_k}{2}\|x - x_0\|^2\right).$$

С учётом данных обозначений неравенство (9) принимает вид

$$A_k f(x_k) \leqslant \psi_k^* + B_k \quad \forall k \in \{0\} \cup \mathbb{N}. \tag{10}$$

Для того чтобы доказать это неравенство воспользуемся методом математической индукции.

База индукции. Пусть $k = 0$. По определению

$$\psi_0^* = \min_{x \in Q}\left(\alpha_0\big[f(x_0) + \langle v(x_0), x - x_0\rangle\big] + \frac{\gamma_0}{2}\|x - x_0\|^2\right).$$

Легко видеть, что функция $h_0(x) := \alpha_0\langle v(x_0), x - x_0\rangle + (\gamma_0/2)\|x - x_0\|^2$ достигает глобального минимума на $\mathbb{R}^n$ в точке

$$\overline{x} = x_0 - \frac{a_0}{\gamma_0}v(x_0),$$

причём

$$h_0(\overline{x}) = -\frac{a_0^2}{2\gamma_0}\|v(x_0)\|^2.$$

Отсюда, учитывая, что $A_0 = a_0$ и $\gamma_{-1} = \gamma_0$, имеем

$$\psi_0^* = \min_{x \in Q}\big(a_0 f(x_0) + h_0(x)\big) \geqslant a_0 f(x_0) + h_0(\overline{x}) = A_0 f(x_0) - \frac{a_0^2}{2\gamma_{-1}}\|v(x_0)\|^2$$

или, что эквивалентно

$$A_0 f(x_0) \leqslant \psi_0^* + \frac{a_0^2}{2\gamma_{-1}} \|v(x_0)\|^2 =: \psi_0^* + B_0.$$

Таким образом база индукции доказана.

Индукционный переход. Предположим, что неравенство (10) выполнено для некоторого k. Покажем, что тогда это неравенство выполнено и для $k+1$. По определению

$$\psi_{k+1}^* = \min_{x \in Q} \left(\ell_{k+1}(x) + \frac{\gamma_{k+1}}{2} \|x - x_0\|^2 \right) = \\ = \min_{x \in Q} \left(\ell_k(x) + \frac{\gamma_{k+1}}{2} \|x - x_0\|^2 + \alpha_{k+1} \big[f(x_{k+1}) + \langle v(x_{k+1}), x - x_{k+1} \rangle \big] \right).$$

Напомним, что последовательность шкалирующих коэффициентов $\{\gamma_k\}$ не убывает. Поэтому

$$\psi_{k+1}^* \geqslant \min_{x \in Q} \left(\ell_k(x) + \frac{\gamma_k}{2} \|x - x_0\|^2 + \alpha_{k+1} \big[f(x_{k+1}) + \langle v(x_{k+1}), x - x_{k+1} \rangle \big] \right). \quad (11)$$

Обозначим

$$h_k(x) = \ell_k(x) + \frac{\gamma_k}{2} \|x - x_0\|^2 := \\ := \left(\sum_{i=0}^{k} \alpha_i \langle v(x_i), x \rangle + \frac{\gamma_k}{2} \|x - x_0\|^2 \right) + \sum_{i=0}^{k} \alpha_i \big[f(x_i) - \langle v(x_i), x_i \rangle \big].$$

Заметим, что функция $h_k(x)$ является сильно выпуклой с константой сильной выпуклости γ_k, так как $h_k''(x) = \gamma_k I_n$, где I_n — единичная матрица размерности n. Поэтому, как хорошо известно, для любых $x, y \in \mathbb{R}^n$ выполняется неравенство

$$h_k(x) \geqslant h_k(y) + \langle \nabla h_k(y), x - y \rangle + \frac{\gamma_k}{2} \|x - y\|^2. \quad (12)$$

Из определения субградиентного метода с двойным усреднением следует, что функция $h_k(x)$ достигает глобального минимума на множестве Q в точке x_k^+, причём $h_k(x_k^+) = \psi_k^*$. По необходимому и достаточному условию минимума гладкой выпуклой функции на выпуклом множестве будет

$$\langle \nabla h_k(x_k^+), x - x_k^+ \rangle \geqslant 0 \quad \forall x \in Q.$$

Отсюда, с учётом (12), имеем

$$h_k(x) \geqslant \psi_k^* + \frac{\gamma_k}{2} \|x - x_k^+\|^2.$$

Воспользовавшись данным неравенством в (11), получим

$$\psi_{k+1}^* \geqslant \min_{x\in Q}\left(\psi_k^* + \frac{\gamma_k}{2}\|x - x_k^+\|^2 + \alpha_{k+1}\big[f(x_{k+1}) + \langle v(x_{k+1}), x - x_{k+1}\rangle\Big).\right. \quad (13)$$

Поскольку мы предполагаем, что неравенство (10) выполнено для k, то

$$\psi_k^* \geqslant A_k f(x_k) - B_k.$$

По определению субградиента выпуклой функции будет

$$f(x_k) - f(x_{k+1}) \geqslant \langle v(x_{k+1}), x_k - x_{k+1}\rangle.$$

Поэтому

$$\psi_k^* \geqslant A_k\big[f(x_{k+1}) + v(x_{k+1}), x_k - x_{k+1}\rangle\big] - B_k.$$

Следовательно, учитывая (13), имеем

$$\psi_{k+1}^* \geqslant \min_{x\in Q}\Big(A_k\big[f(x_{k+1}) + v(x_{k+1}), x_k - x_{k+1}\rangle\big] - B_k + \frac{\gamma_k}{2}\|x - x_k^+\|^2 +$$
$$+ \alpha_{k+1}\big[f(x_{k+1}) + \langle v(x_{k+1}), x - x_{k+1}\rangle\Big). \quad (14)$$

Из определения субградиентного метода с двойным усреднением следует, что

$$(A_k + \alpha_{k+1})x_{k+1} := A_{k+1}x_{k+1} := A_k x_k + \alpha_{k+1}x_k^+.$$

Значит

$$A_k\Big[f(x_{k+1}) + \langle v(x_{k+1}), x_k - x_{k+1}\rangle\Big] + \alpha_{k+1}\Big[f(x_{k+1}) + \langle v(x_{k+1}), x - x_{k+1}\rangle\Big] =$$
$$= A_{k+1}f(x_{k+1}) + \alpha_{k+1}\langle v(x_{k+1}), x - x_k^+\rangle.$$

Отсюда получаем, что неравенство (14) преобразуется к виду

$$\psi_{k+1}^* \geqslant A_{k+1}f(x_{k+1}) - B_k + \min_{x\in Q}\left(\alpha_{k+1}\langle v(x_{k+1}), x - x_k^+\rangle + \frac{\gamma_k}{2}\|x - x_k^+\|^2\right). \quad (15)$$

Заметим, что функция

$$g_k(x) := \alpha_{k+1}\langle v(x_{k+1}), x - x_k^+\rangle + \frac{\gamma_k}{2}\|x - x_k^+\|^2$$

достигает глобального минимума на $\mathbb{R}^n$ в точке

$$\widehat{x} = x_k^+ - \frac{\alpha_{k+1}}{\gamma_k}v(x_{k+1}),$$

причём

$$g_k(\widehat{x}) = -\frac{\alpha_{k+1}^2}{\gamma_k}\|v(x_{k+1})\|^2.$$

Следовательно, с учётом неравенства (15) получаем нижнюю оценку для ψ_{k+1}^* вида

$$\psi_{k+1}^* \geqslant A_{k+1}f(x_{k+1}) - B_k + \min_{x\in Q} g_k(x) \geqslant$$
$$\geqslant A_{k+1}f(x_{k+1}) - B_k - \frac{\alpha_{k+1}^2}{\gamma_k}\|v(x_{k+1})\|^2 := A_{k+1}f(x_{k+1}) - B_{k+1}.$$

Таким образом,

$$A_{k+1}f(x_{k+1}) \leqslant \psi_{k+1}^* + B_{k+1},$$

что и требовалось доказать. □

Введём несколько дополнительных обозначений. Положим

$$s_k = \frac{1}{A_k}\sum_{i=0}^{k}\alpha_i v(x_i).$$

Для произвольного множества $C \subset Q$ обозначим

$$\xi_C(s) = \max_{x\in C}\langle s, x\rangle \quad \forall s \in \mathbb{R}^n.$$

ЛЕММА 2. *Пусть последовательность $\{x_k\}$ построена по субградиентному методу с двойным усреднением. Тогда для любого $k \in \mathbb{N}$ справедливо неравенство*

$$f(x_k) + f^*(s_k) + \xi_C(-s_k) \leqslant \frac{1}{A_k}(B_k + \gamma_k D_C), \tag{16}$$

где $f^*(s) = \sup_x\{\langle s, x\rangle - f(x)\}$ — функция, сопряжённая (в смысле выпуклого анализа) с функцией f и

$$D_C = \frac{1}{2}\max_{x\in C}\|x - x_0\|^2.$$

Доказательство. По Лемме 1 для любого $x \in Q$ справедливо неравенство

$$A_k f(x_k) \leqslant \sum_{i=0}^{k}\alpha_i\big[f(x_i) + \langle v(x_i), x - x_i\rangle\big] + \frac{\gamma_k}{2}\|x - x_0\|^2 + B_k.$$

Преобразуем данное неравенство. Перенесём слагаемые $\alpha_i\langle v(x_i, x-x_i\rangle$ в левую часть. Получим

$$A_k f(x_k) + \sum_{i=0}^{k} \alpha_i \langle v(x_i), x_i - x\rangle \leqslant \sum_{i=0}^{k} \alpha_i f(x_i) + \frac{\gamma_k}{2}\|x - x_0\|^2 + B_k.$$

Прибавляя и отнимая $\alpha_i\langle v(x_i), y\rangle$ для всех $i \in \{0, \ldots, k\}$, где $y \in \mathbb{R}^n$ — произвольный вектор, имеем

$$A_k f(x_k) + A_k\langle s_k, y - x\rangle + \sum_{i=0}^{k} \alpha_i\langle v(x_i), x_i - y\rangle \leqslant$$
$$\leqslant \sum_{i=0}^{k} \alpha_i f(x_i) + \frac{\gamma_k}{2}\|x - x_0\|^2 + B_k.$$

По определению субградиента будет

$$f(y) - f(x_i) \geqslant \langle v(x_i), y - x_i\rangle.$$

Поэтому для любых $x \in Q$ и $y \in \mathbb{R}^n$ справедливо неравенство

$$A_k f(x_k) + A_k\langle s_k, y - x\rangle + \sum_{i=0}^{k} \alpha_i f(x_i) - A_k f(y) \leqslant$$
$$\leqslant \sum_{i=0}^{k} \alpha_i f(x_i) + \frac{\gamma_k}{2}\|x - x_0\|^2 + B_k$$

или, что эквивалентно,

$$f(x_k) + \big[\langle s_k, y\rangle - f(y)\big] + \langle -s_k, x\rangle \leqslant \frac{1}{A_k}\left(\frac{\gamma_k}{2}\|x - x_0\|^2 + B_k\right).$$

Поскольку данное неравенство выполнено для всех $y \in \mathbb{R}^n$, то

$$f(x_k) + f^*(s_k) + \langle -s_k, x\rangle \leqslant \frac{1}{A_k}\left(\frac{\gamma_k}{2}\|x - x_0\|^2 + B_k\right) \quad \forall x \in Q.$$

Беря супремум по всем $x \in C \subset Q$, получаем, что справедливо неравенство (16). $\square$

Теперь мы можешь получить достаточные условия сходимости субградиентного метода с двойным усреднением.

ТЕОРЕМА 3. *Пусть последовательность $\{x_k\}$ построена по субградиентному методу с двойным усреднением. Тогда для любого $k \in \mathbb{N}$ справедливо неравенство*

$$f(x_k) - f(x^*) \leqslant \frac{1}{A_k}\left(\frac{\gamma_k}{2}\|x_0 - x^*\|^2 + \frac{1}{2}\sum_{i=0}^{k} \frac{\alpha_i^2}{\gamma_{i-1}}\|v(x_i)\|^2\right),$$

где $\gamma_{-1} = \gamma_0$. *Отсюда, в частности, следует, что если последовательность* $\{v(x_k)\}$ *ограничена и*

$$\lim_{k\to\infty}\left(\frac{\gamma_k}{A_k}+\frac{1}{A_k}\sum_{i=0}^{k}\frac{\alpha_i^2}{\gamma_{i-1}}\right)=0,$$

то $\lim\limits_{k\to\infty} f(x_k) = f(x^*)$.

Доказательство. По Лемме 2 для любого множества C и для всех $k \in \{0\} \cup \mathbb{N}$ справедливо неравенство

$$f(x_k)+f^*(s_k)+\xi_C(-s_k)\leqslant\frac{1}{A_k}(B_k+\gamma_k D_C), \tag{17}$$

где

$$D_C=\frac{1}{2}\max_{x\in C}\|x-x_0\|^2.$$

Положим $C = \{x^*\}$. Тогда неравенство (17) преобразуется к виду

$$f(x_k)+f^*(s_k)-\langle s_k,x^*\rangle\leqslant\frac{1}{A_k}\left(\frac{\gamma_k}{2}\|x_0-x^*\|^2+B_k\right).$$

Заметим, что

$$f^*(s_k)-\langle s_k,x^*\rangle=\sup_{x\in\mathbb{R}^n}\big(\langle s_k,x\rangle-f(x)\big)-\langle s_k,x^*\rangle\geqslant$$
$$\geqslant\langle s_k,x^*\rangle-f(x^*)-\langle s_k,x^*\rangle=-f(x^*).$$

Поэтому

$$f(x_k)-f(x^*)\leqslant\frac{1}{A_k}\left(\frac{\gamma_k}{2}\|x_0-x^*\|^2+B_k\right),$$

что и требовалось доказать. □

Замечание 3. Заметим, что достаточные условия сходимости субградиентного метода, указанные в предыдущей теореме, будут выполнены, если положить $\alpha_k \equiv 1$ и $\gamma_k = \sqrt{k+1}$. Следует отметить, что при таком выборе параметров α_k и γ_k будет

$$f(x_k)-f(x^*)=O\left(\frac{1}{\sqrt{k}}\right),$$

если предполагать, что последовательность $\{v(x_k)\}$ ограничена.

ЛИТЕРАТУРА

1. Ю.Е. Нестеров "Convergent subgradient methods for nonsmooth convex minimization". URL: www.youtube.com/watch?v=YucrsxdwgTY

2. Шор Н.З. *Методы минимизации недифференцируемых функций и их приложения.* Киев: Наукова думка, 1979. 200 с.

3. Немировский А.С., Юдин Д.Б. *Сложность задач и эффективность методов оптимизации.* М.: Наука, 1979. 384 с.

4. Nesterov Y. *Primal-dual subgradient methods for convex problems* // Mathematical Programming, 2009, vol. 120, no. 1, pp. 221–259.

5. Nesterov Y., Shikhman V. *Convergent subgradient methods for nonsmooth convex minimization.* Université catholique de Louvain, Center for Operations Research and Econometric (CORE), 2014. no. 2014/5. (URL:http://www.uclouvain.be/cps/ucl/doc/core/documents/coredp2014_5web.pdf)