УДК 519.85

Адаптивный быстрый градиентный метод для задачи минимизации максимума выпуклых функций

А.И.Тюрин

Национальный исследовательский университет «Высшая школа экономики»

Задача:

$$f(x) = \max_i f_i(x) + h(x) \to \min_{x \in Q} \tag{1}$$

$f_i(x)$ - выпуклая функция с L липшицевым градиентом $\forall i$. $h(x)$ - выпуклая функция. Q - выпуклое, замкнутое множество.

Дано: x_0 - начальная точка, N - количество шагов и L_0 ($L_0 \leq L$)

0 - шаг:

$$y_0 = u_0 = x_0 \tag{2}$$

$$L_1 = \frac{L_0}{2} \tag{3}$$

$k+1$ - шаг:

$$\text{Найти } \alpha_{k+1} : A_k + \alpha_{k+1} = L_{k+1}\alpha_{k+1}^2$$

$$y_{k+1} = \frac{\alpha_{k+1}u_k + A_k x_k}{A_{k+1}} \tag{4}$$

$$\phi_{k+1}(x) = V(x, u_k) + \alpha_{k+1}(\max_j [f_j(y_{k+1}) + \langle \nabla f_j(y_{k+1}), x - y_{k+1} \rangle] + h(x)) \tag{5}$$

$$u_{k+1} = \operatorname*{argmin}_{x \in Q} \phi_{k+1}(x) \tag{6}$$

$$x_{k+1} = \frac{\alpha_{k+1}u_{k+1} + A_k x_k}{A_{k+1}} \tag{7}$$

Если выполнено условие

$$f(x_{k+1}) \leq \max_j \{f_j(y_{k+1}) + \langle \nabla f_j(y_{k+1}), x_{k+1} - y_{k+1} \rangle\} \tag{8}$$

$$+\frac{L_{k+1}}{2} \|x_{k+1} - y_{k+1}\|^2 + h(x_{k+1}), \tag{9}$$

то

$$L_{k+2} = \frac{L_{k+1}}{2} \tag{10}$$

и перейти к следующему шагу, иначе

$$L_{k+1} = 2L_{k+1} \tag{11}$$

и повторить текущий шаг.

Замечание:

1. Количество циклов в каждом шаге конечно. Это следует из того, что на каждом шаге цикла мы увеличиваем L_{k+1} в 2 раза, а значит L_{k+1} через конечное количество шагов станет больше L, поэтому из L - липшевости $f_j(x)$ для $\forall j$ будет выполнено условие 9.

2. $$L_k \leq 2L$$

   Это следует из того, что мы выйдем из цикла ранее, чем L_k станет больше $2L$.

3. Оценим общее число обращений за значениями функций. Пусть j_k - количество дополнительных циклов k-ого шага. Тогда общее количество обращений за значениями всех функций $f_j(x)$ равно

$$\sum_{k=1}^{N} 2(j_k+1) = \sum_{k=1}^{N} 2((j_k-1)+2) = \sum_{k=1}^{N} 2(\log_2(\frac{L_k}{L_{k-1}})+2) =$$
$$4N + \log_2(\frac{L_N}{L_0}) \leq 4N + \log_2(\frac{2L}{L_0})$$

   Второе равенство следует из того, что $L_k = 2^{j_k}\frac{L_{k-1}}{2}$. Поэтому мы получаем, что в среднем на каждом шаге мы будем считать значение всех функций 4 раза. Можно показать, что градиент всех функций $f_j(x)$ мы в среднем будем считать на каждом шаге 2 раза.

Лемма 1. *Пусть для последовательности α_k выполнено*

$$A_k = \sum_{i=0}^{k} \alpha_i$$
$$\alpha_0 = 0$$
$$A_k = L_k \alpha_k^2$$

Тогда верно следующее неравенство $\forall k \geq 1$

$$A_k \geq \frac{(k+1)^2}{8L}$$

Доказательство. Пусть $k = 1$.

$$\alpha_1 = L_1 \alpha_1^2$$

$$A_1 = \alpha_1 = \frac{1}{L_1} \geq \frac{1}{2L}$$

Пусть $k \geq 2$, тогда

$$L_{k+1}\alpha_{k+1}^2 = A_{k+1}$$

$$L_{k+1}\alpha_{k+1}^2 = A_k + \alpha_{k+1}$$
$$L_{k+1}\alpha_{k+1}^2 - \alpha_{k+1} - A_k = 0$$

Решая данное квадратное уравнение будем брать наибольший корень, поэтому

$$\alpha_{k+1} = \frac{1+\sqrt{1+4L_{k+1}A_k}}{2L_{k+1}}$$

По индукции, пусть нер-во верно для k, тогда:

$$\alpha_{k+1} = \frac{1}{2L_{k+1}} + \sqrt{\frac{1}{4L_{k+1}^2} + \frac{A_k}{L_{k+1}}} \geq \frac{1}{2L_{k+1}} + \sqrt{\frac{A_k}{L_{k+1}}} \geq$$
$$\frac{1}{4L} + \frac{1}{\sqrt{2L}}\frac{k+1}{2\sqrt{2L}} \geq \frac{k+2}{4L}$$
$$\alpha_{k+1} \geq \frac{k+2}{4L}$$

и

$$A_{k+1} = \sum_{i=0}^{k+1}\alpha_i \geq \sum_{i=1}^{k+1}\frac{i+1}{4L} \geq \frac{(k+2)^2}{8L}$$

□

Лемма 2. *Пусть $\psi(x)$ выпуклая и*

$$y = \underset{x}{\operatorname{argmin}}\{\psi(x) + V(x,z)\}$$

Тогда

$$\psi(x) + V(x,z) \geq \psi(y) + V(y,z) + V(x,y) \ \ \forall x \in Q$$

Доказательство. По критерию оптимальности:

$$\exists g \in \partial\psi(y), \ \ \langle g + \nabla_y V(y,z), x-y\rangle \geq 0 \ \ \forall x \in Q$$

Тогда

$$\psi(x) - \psi(y) \geq \langle g, x-y\rangle \geq \langle \nabla_y V(y,z), y-x\rangle$$

и

$$\langle \nabla_y V(y,z), y-x\rangle = \langle \nabla d(y) - \nabla d(z), y-x\rangle = d(y) - d(z) - \langle \nabla d(z), y-z\rangle$$
$$+d(x) - d(y) - \langle \nabla d(y), x-y\rangle - d(x) + d(z) + \langle \nabla d(z), x-z\rangle =$$
$$V(y,z) + V(x,y) - V(x,z)$$

□

Лемма 3. $\forall x \in Q$ *выполнено*

$$A_{k+1} f(x_{k+1}) - A_k f(x_k) + V(x, u_{k+1}) - V(x, u_k) \leq \alpha_{k+1} f(x)$$

Доказательство. Введем обозначение: $l_{f_j}(x; y) = f_j(y) + \langle \nabla f_j(y), x - y \rangle$.

$$f(x_{k+1}) \leq_{1} \max_j \{l_{f_j}(x_{k+1}; y_{k+1})\} + \frac{L_{k+1}}{2} \left\| x_{k+1} - y_{k+1} \right\|^2 + h(x_{k+1}) =$$

$$\max_j \{l_{f_j}(\frac{\alpha_{k+1} u_{k+1} + A_k x_k}{A_{k+1}}; y_{k+1})\} + \frac{L_{k+1}}{2} \left\| \frac{\alpha_{k+1} u_{k+1} + A_k x_k}{A_{k+1}} - y_{k+1} \right\|^2 +$$

$$h(\frac{\alpha_{k+1} u_{k+1} + A_k x_k}{A_{k+1}}) \leq$$

$$\max_j \{f_j(y_{k+1}) + \frac{\alpha_{k+1}}{A_{k+1}} \langle \nabla f_j(y_{k+1}), u_{k+1} - y_{k+1} \rangle +$$

$$\frac{A_k}{A_{k+1}} \langle \nabla f_j(y_{k+1}), x_k - y_{k+1} \rangle\} + \frac{L_{k+1} \alpha_{k+1}^2}{2 A_{k+1}^2} \left\| u_{k+1} - u_k \right\|^2$$

$$+ \frac{\alpha_{k+1}}{A_{k+1}} h(u_{k+1}) + \frac{A_k}{A_{k+1}} h(x_k) \leq$$

$$\frac{A_k}{A_{k+1}} (\max_j \{f_j(y_{k+1}) + \langle \nabla f_j(y_{k+1}), x_k - y_{k+1} \rangle\} + h(x_k)) +$$

$$\frac{\alpha_{k+1}}{A_{k+1}} (\max_j \{f_j(y_{k+1}) + \langle \nabla f_j(y_{k+1}), u_{k+1} - y_{k+1} \rangle\} + h(u_{k+1}))$$

$$+ \frac{L_{k+1} \alpha_{k+1}^2}{2 A_{k+1}^2} \left\| u_{k+1} - u_k \right\|^2 =_{2}$$

$$\frac{A_k}{A_{k+1}} (\max_j \{l_{f_j}(x_k; y_{k+1})\} + h(x_k)) +$$

$$\frac{\alpha_{k+1}}{A_{k+1}} (\max_j \{l_{f_j}(u_{k+1}; y_{k+1})\} + \frac{1}{2 \alpha_{k+1}} \left\| u_{k+1} - u_k \right\|^2 + h(u_{k+1})) \leq$$

$$\frac{A_k}{A_{k+1}} (\max_j \{l_{f_j}(x_k; y_{k+1})\} + h(x_k)) +$$

$$\frac{\alpha_{k+1}}{A_{k+1}} (\max_j \{l_{f_j}(u_{k+1}; y_{k+1})\} + \frac{1}{\alpha_{k+1}} V(u_{k+1}, u_k) + h(u_{k+1})) \leq_{3}$$

$$\frac{A_k}{A_{k+1}} f(x_k) +$$

$$\frac{\alpha_{k+1}}{A_{k+1}} (\max_j \{l_{f_j}(x; y_{k+1})\} + h(x) + \frac{1}{\alpha_{k+1}} V(x, u_k) - \frac{1}{\alpha_{k+1}} V(x, u_{k+1})) \leq_{4}$$

$$\frac{A_k}{A_{k+1}} f(x_k) + \frac{\alpha_{k+1}}{A_{k+1}} f(x) + \frac{1}{A_{k+1}} V(x, u_k) - \frac{1}{A_{k+1}} V(x, u_{k+1}))$$

□

1 - из условия 9
2 - из $A_k \geq L_k \alpha_k^2$
3 - из леммы 2 с $\psi(x) = \alpha_{k+1} \max_j \{f_j(y_{k+1}) + \langle \nabla f_j(y_{k+1}), x - y_{k+1} \rangle\} + h(x)$ и выпуклость $f_j(x)\ \forall j$
4 - выпуклость $f_j(x)\ \forall j$

Теорема 1.

$$f(x_N) - f(x_*) \leq \frac{8LR^2}{(N+1)^2}$$

Доказательство. Просуммируем нер-во из леммы 3 по $k = 0, ..., N-1$

$$A_N f(x_N) - A_0 f(x_0) + V(x, u_N) - V(x, u_0) \leq (A_N - A_0) f(x)$$
$$A_N f(x_N) + V(x, u_N) - V(x, u_0) \leq A_N f(u)$$

Возьмем $x = x_*$.

$$A_N (f(x_N) - f_*) \leq R^2$$

$\square$

Литература

[1] *Нестеров Ю. Е.* Введение в выпуклую оптимизацию //М.: МЦНМО. – 2010. – Т. 274.

[2] *Nesterov Y.* A method of solving a convex programming problem with convergence rate O (1/k2) //Soviet Mathematics Doklady. – 1983. – Т. 27. – №. 2. – С. 372-376.