§8 МЕТОД ПРОЕКЦИИ ГРАДИЕНТА

Пусть $X \subset \mathbb{R}^n$ — непустое выпуклое замкнутое множество. Для произвольной точки $a \in \mathbb{R}^n$ через $\Pi_X a$ обозначим проекцию точки a на множество X, т.е. такую точку $b \in X$, что $\|a - b\| = \inf\limits_{x \in X} \|a - x\|$. Так как X — замкнутое множество, то такая точка b действительно существует. Единственность проекции следует из выпуклости множества X. Действительно, если $a \in X$, то в силу определения имеем $\Pi_X a = a$. Предположим, что $a \notin X$ и нашлись две различные ближайшие точки b_1 и $b_2 \in X$. Тогда несложно видеть, что $\dfrac{b_1 + b_2}{2} \in X$ и

$$\|a - (b_1 + b_2)/2\| = \|(a - b_1)/2 + (a - b_2)/2\| < \frac{\|a - b_1\| + \|a - b_2\|}{2},$$

где последнее неравенство строгое, так как a не принадлежит прямой, проходящей через точки b_1 и b_2. Далее, если это не вызывает разночтений, нижний индекс у отображения Π_X будем опускать, т.е. будем писать Π вместо Π_X.

Пусть $f\colon \mathbb{R}^n \to \mathbb{R}$ — дважды непрерывно дифференцируемая выпуклая функция. Рассмотрим следующую задачу выпуклой оптимизации

$$f(x) \to \inf_X. \tag{8.1}$$

Будем предполагать, что $f(x)$ достигает минимума на X (последнее, например, будет следовать из теоремы Вейерштрасса, если множество X ограничено). Для решения задачи (8.1) в некоторых частных случаях множества X может быть применён метод проекции градиента, к описанию и анализу которого мы сейчас перейдём.

Пусть $x^0 \in X$ — заданное начальное приближение. Последовательность точек $(x^k)_{k \in \mathbb{N}_0}$ строится рекуррентным образом: $x^{k+1} = \Pi\bigl(x^k - t^k \nabla f(x^k)\bigr)$, где $(t^k)_{k \in \mathbb{N}_0}$ — некоторая последовательность положительных чисел.

Через $x^* = \arg\min\limits_{x \in X} f(x)$ обозначим решение задачи (8.1). Вообще говоря, не для всякой последовательность $(t^k)_{k \in \mathbb{N}_0}$ будет иметь место сходимость $x^k \to x^*$ (или $f(x^k) \to f(x^*)$). Для построения последовательности $(t^k)_{k \in \mathbb{N}_0}$ будем использовать следующую модификацию процедуры BLS. Пусть $x \in X$ и $\beta \in (0, 1)$. Для действительного числа t положим $x^+ = \Pi\bigl(x - t\nabla f(x)\bigr)$ и $\Delta x^+ = (x^+ - x)/t$, т.е. $x^+ = x + t\Delta x^+$. Начиная с $t = 1$, будем последовательно уменьшать t, умножением на β, до тех пор пока не будет выполнено неравенство:

$$f(x + t\Delta x^+) \le f(x) + t\nabla f(x)^{\mathsf{T}} \Delta x^+ + \frac{t}{2}\|\Delta x^+\|^2. \tag{8.2}$$

Покажем, что t, удовлетворяющее неравенству (8.2), будет найдено за конечное число итераций. Действительно, пусть $K \overset{\text{def}}{=} \max\{\|\nabla^2 f(y)\| : y \in \overline{B}_1(x)\}$, где через $\overline{B}_1(x)$ обозначен единичный замкнутый шар с центром в x. При $t \in \bigl(0, \|\nabla f(x)\|^{-1}\bigr)$

справедливо включение $x - t\nabla f(x) \in \overline{B}_1(x)$, а значит, и $x^+ \in \overline{B}_1(x)$. Следовательно, верно неравенство

$$f(x + t\Delta x^+) \le f(x) + t\nabla f(x)^\mathsf{T}\Delta x^+ + \frac{t^2 K}{2}\|\Delta x^+\|^2.$$

Таким образом, неравенство (8.2) будет выполнено при $t \le \min\bigl(\|\nabla f(x)\|^{-1}, K^{-1}\bigr)$.

Предположим, что существуют такие неотрицательные числа m и M, что для всех $x \in X$ верно $mI \preceq \nabla^2 f(x) \preceq MI$. Так как матрица $\nabla^2 f(x)$ неотрицательно определена, то левая часть неравенства выполнена, например, при $m = 0$. Если X — ограниченное множество, то существование постоянной M следует из непрерывности $\nabla^2 f$. Из проведённых рассуждений, в частности, следует, что t, найденное согласно BLS, удовлетворяет неравенству $t \ge \min(1, \beta/M)$. В силу сделанных предположений при x_1, $x_2 \in X$ справедливы неравенства

$$f(x_2) \ge f(x_1) + \nabla f(x_1)^\mathsf{T}(x_2 - x_1) + \frac{m}{2}\|x_2 - x_1\|^2, \tag{8.3}$$

$$f(x_2) \le f(x_1) + \nabla f(x_1)^\mathsf{T}(x_2 - x_1) + \frac{M}{2}\|x_2 - x_1\|^2. \tag{8.4}$$

Для произвольных $z \in X$ и $t \in \mathbb{R}$ имеет место неравенство

$$0 \ge (z - x^+, x - t\nabla f(x) - x^+). \tag{8.5}$$

Действительно, если $x - t\nabla f(x) \in X$, т.е. $x^+ = x - t\nabla f(x)$, то скалярное произведение в (8.5) равно 0. В противном случае точки z и $x - t\nabla f(x)$ находятся по разные стороны от опорной гиперплоскости, проходящей через x^+ перпендикулярно вектору $x - t\nabla f(x) - x^+$. Следовательно угол между векторами $z - x^+$ и $x - t\nabla f(x) - x^+$ не меньше 90°.

Преобразуем неравенство (8.5), подставив $x^+ = x + t\Delta x^+$. После подстановки получаем $0 \ge \bigl(z - x - t\Delta x^+, -t\Delta x^+ - t\nabla f(x)\bigr)$. Сокращая на $-t$, имеем

$$0 \le (z - x, \Delta x^+) + \nabla f(x)^\mathsf{T}(z - x) - t\|\Delta x^+\|^2 - t\nabla f(x)^\mathsf{T}\Delta x^+. \tag{8.6}$$

Перейдём к оценке с верху значения $f(x^+)$:

$$\begin{aligned} f(x + t\Delta x^+) &\overset{(8.2)}{\le} f(x) + t\nabla f(x)^\mathsf{T}\Delta x^+ + \frac{t}{2}\|\Delta x^+\|^2 \le \\ &\overset{(8.3)}{\le} f(z) - \nabla f(x)^\mathsf{T}(z - x) - \frac{m}{2}\|z - x\|^2 + t\nabla f(x)^\mathsf{T}\Delta x^+ + \frac{t}{2}\|\Delta x^+\|^2 \le \\ &\overset{(8.6)}{\le} f(z) + (z - x, \Delta x^+) - \frac{t}{2}\|\Delta x^+\|^2 - \frac{m}{2}\|z - x\|^2. \end{aligned} \tag{8.7}$$

При $z = x$ из (8.7) следует, что $f(x^+) \le f(x) - \dfrac{t}{2}\|\Delta x^+\|^2$. При $z = x^*$ из (8.7) следует, что

$$f(x^+) - f(x^*) \le (x^* - x, \Delta x^+) - \frac{t}{2}\|\Delta x^+\|^2 - \frac{m}{2}\|x^* - x\|^2.$$

Так как $(x^* - x, \Delta x^+) - \dfrac{t}{2}\|\Delta x^+\|^2 = \dfrac{1}{2t}(\|x - x^*\|^2 - \|x^+ - x^*\|^2)$. То

$$f(x^+) - f(x^*) \le \frac{1}{2t_{\min}}(\|x - x^*\|^2 - \|x^+ - x^*\|^2) - \frac{m}{2}\|x - x^*\|^2,$$

где $t_{\min} = \min(1, \beta/M)$. Если $m > 0$, то в силу того, что $0 \le f(x^+) - f(x^*)$, следует неравенство $\|x^+ - x^*\|^2 \le (1 - mt_{\min})\|x - x^*\|^2$. Поэтому для элементов последовательности $(x^k)_{k\in\mathbb{N}_0}$ справедлива оценка:

$$\|x^k - x^*\| \le c^k\|x^0 - x^*\|, \quad \text{где} \quad c = \sqrt{1 - mt_{\min}},$$

из которой следует, что $x^k \to x^*$.

Если же $m = 0$, то имеют место неравенства:

$$f(x^k) - f(x^*) \le \frac{1}{2t_{\min}}(\|x^{k-1} - x^*\|^2 - \|x^k - x^*\|^2),$$
$$\ldots$$
$$f(x^1) - f(x^*) \le \frac{1}{2t_{\min}}(\|x^0 - x^*\|^2 - \|x^1 - x^*\|^2).$$

Таким образом, $f(x^k) - f(x^*) \le \dfrac{1}{k}\sum\limits_{i=1}^{k}\big(f(x^i) - f(x^*)\big) \le \dfrac{\|x^0 - x^*\|^2}{2t_{\min}k}$, а значит, $f(x^k) \to f(x^*)$.

Вообще говоря, нахождение проекций точек на выпуклое множество является сложной задачей. К счастью, для некоторых выпуклых множеств, которые часто встречаются на практике, эта задача может быть решена эффективно и даже аналитически. Рассмотрим такие множества.

Пример 8.1 (Шар)**.** Пусть $X = \{x\colon \|x - a\| \le R\}$, тогда

$$\Pi y = \begin{cases} y, & \text{если} \quad y \in X, \\ a + \dfrac{R}{\|y - a\|}(y - a) & \text{иначе.} \end{cases} \tag{8.8}$$

Пример 8.2 (Параллелепипед)**.** Пусть $X = \{x\colon a_i \le x_i \le b_i, 1 \le i \le n\}$, тогда i-я координата вектора Πy, которую обозначим как $[\Pi y]_i$, определяется равенством:

$$[\Pi y]_i = \begin{cases} a_i, & \text{если} \quad y_i < a_i; \\ y_i, & \text{если} \quad a_i \le y_i \le b_i; \\ b_i & \text{иначе.} \end{cases}$$

Пример 8.3 (Симплекс)**.** Пусть $X = \{x\colon x_i \ge 0, \sum\limits_{i=1}^{n} x_i \le 1\}$. Нетрудно видеть, что если $y_i < 0$, то $[\Pi y]_i = 0$, а значит, $\Pi y = \Pi\widetilde{y}$, где вектор $\widetilde{y}$ получен из y обнулением отрицательных компонент. Если $\sum\limits_{i=1}^{n} \widetilde{y}_i \le 1$, то $\Pi\widetilde{y} = \widetilde{y}$. Предположим, что $\sum\limits_{i=1}^{n} \widetilde{y}_i > 1$. Тогда $x = \Pi\widetilde{y}$ — решение следующей задачи выпуклой оптимизации:

$$\begin{cases} \frac{1}{2}\|x-\widetilde{y}\|^2 \to \min, \\ \sum\limits_{i=1}^{n} x_i \le 1, \\ x_i \ge 0, \quad 1 \le i \le n. \end{cases} \tag{8.9}$$

Для задачи (8.9) рассмотрим функцию Лагранжа $\mathcal{L}(x;\lambda,\mu) = \frac{1}{2}\|x-\widetilde{y}\|^2 + \lambda^{\mathsf{T}}(-x) + {} + \mu\left(\sum\limits_{i=1}^{n} x_i - 1\right)$. Подберём такие x, λ и $\mu \ge 0$, что

$$\begin{gathered} x_i - \widetilde{y}_i - \lambda_i + \mu = 0, \quad 1 \le i \le n; \quad \lambda^{\mathsf{T}} x = 0; \\ \mu\left(\sum_{i=1}^{n} x_i - 1\right) = 0; \quad \sum_{i=1}^{n} x_i \le 1. \end{gathered} \tag{8.10}$$

Тогда из теоремы Куна - Таккера будет следовать, что x — решение задачи (8.9). Выберем μ как решение уравнения $\sum\limits_{i=1}^{n} (\widetilde{y}_i - \mu)_+ = 1$, где $(z)_+ \stackrel{\text{def}}{=} \max(0,z)$. Указанное уравнение может быть решено бинарным поиском. Положим $x_i = (\widetilde{y}_i - \mu)_+$ и $\lambda_i = x_i - (\widetilde{y}_i - \mu)$. Нетрудно видеть, что тройка $(x;\lambda,\mu)$ удовлетворяет уравнениям (8.10). Таким образом,

$$\Pi y = \begin{cases} (y)_+, & \text{если} \quad (y)_+ \in X; \\ (y-\mu)_+ & \text{иначе.} \end{cases}$$

Упражнения

20. Пусть $X \subset \mathbb{R}^n$ — выпуклое замкнутое множество, а $f\colon X \to \mathbb{R}$ — непрерывно дифференцируемая функция. Докажите, что
    (a) если $x^* \in X$ — локальный минимум фукнции f, то для произвольного $\alpha \ge 0$ справедливо равенство $x^* = \Pi\big(x^* - \alpha\nabla f(x^*)\big)$;
    (b) если фукнция f выпуклая, то точка $x^* \in X$ — локальный минимум функции f тогда и только тогда, когда для произвольного $\alpha \ge 0$ справедливо равенство $x^* = \Pi\big(x^* - \alpha\nabla f(x^*)\big)$.
21. Пусть Λ — множество допустимых точек задачи (3.11). Разработайте алгоритм вычисления проекции $\Pi_\Lambda \lambda$ для произвольной точки $\lambda \in \mathbb{R}^N$, использующий не более $O(N \log N)$ арифметических операций.