Gradient descent

Ковалев Алексей

1. Пусть для функции f для любых x, y выполняется неравенство

$$f(y) \leqslant f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} L \|y - x\|_2^2$$

Покажем, что градиентный спуск для нее достигает точки x, такой что $\|\nabla f(x)\|_2 \leqslant \varepsilon$ за $O(1/\varepsilon^2)$ итераций. Положим $y = x^{k+1} = x^k - \alpha \nabla f(x^k)$. Тогда

$$f(x^{k+1}) \leqslant f(x^k) - \alpha \nabla f(x^k)^\top \nabla f(x^k) + \frac{1}{2} L \|\alpha \nabla f(x^k)\|_2^2 = f(x^k) - \left(1 - \frac{1}{2} L\alpha\right) \alpha \|\nabla f(x^k)\|_2^2$$

Воспользуемся условием $\alpha \leqslant \frac{1}{L}$, получим $1 - \frac{1}{2} L\alpha \geqslant \frac{1}{2}$, а значит

$$\frac{1}{2} \alpha \|f(x^k)\|_2^2 \leqslant \left(1 - \frac{1}{2} L\alpha\right) \alpha \|\nabla f(x^k)\|_2^2 \leqslant f(x^k) - f(x^{k+1})$$

$$\|f(x^k)\|_2^2 \leqslant \frac{2}{\alpha} \left(f(x^k) - f(x^{k+1})\right)$$

Просуммировав последнее неравенство по k, получим

$$\sum_{k=0}^{n} \|\nabla f(x^k)\|_2^2 \leqslant \sum_{k=0}^{n} \frac{2}{\alpha} \left(f(x^k) - f(x^{k+1})\right) = \frac{2}{\alpha} \left(f(x^0) - f^*\right)$$

Сумма $n + 1$ неотрицательных слагаемых не превосходит $\frac{2}{\alpha} \left(f(x^0) - f^*\right)$, значит минимальное из них можно оценить как

$$\min_{k=0\ldots n} \|\nabla f(x^k)\|_2^2 \leqslant \frac{2}{\alpha(k+1)} \left(f(x^0) - f^*\right)$$

$$\varepsilon^2 \leqslant \frac{2}{\alpha(k+1)} \left(f(x^0) - f^*\right)$$

$$k + 1 \leqslant \frac{2}{\alpha \varepsilon^2} \left(f(x^0) - f^*\right)$$

То есть $k = O(1/\varepsilon^2)$. □

2. Код смотри ниже. Из графика можно сделать вывод, что сходимость метода и ее скорость не зависят от размерности задачи.

Subgradient descent

1. Рассмотрим метод субградиентого спуска $x^{k+1} = x^k - \alpha_k g_k$, $g_k \in \partial f(x^k)$. Отсюда получаем

$$x^{k+1} - x^* = x^k - x^* - \alpha_k g_k$$

$$\langle x^{k+1} - x^*, \, x^{k+1} - x^* \rangle = \langle x^k - x^* - \alpha_k g_k, \, x^k - x^* - \alpha_k g_k \rangle$$

$$\|x^{k+1} - x^*\|_2^2 = \|x^k - x^*\|_2^2 + \alpha_k^2 \|g_k\|_2^2 - 2\alpha_k g_k^\top (x^k - x^*)$$

По определению субградиента $g_k^\top (x^* - x^k) \leqslant f(x^*) - f(x^k) = f^* - f(x^k)$. Отсюда

$$-2\alpha_k g_k^\top (x_k - x^*) \leqslant -2\alpha_k \left(f(x_k) - f^*\right)$$

Откуда получаем

$$\|x^{k+1} - x^*\|_2^2 \leqslant \|x^k - x^*\|_2^2 - 2\alpha_k \left(f(x^k) - f^*\right) + \alpha_k^2 \|g_k\|_2^2$$

$$2\alpha_k \left(f(x^k) - f^*\right) \leqslant \|x^k - x^*\|_2^2 + \alpha_k^2 \|g_k\|_2^2 - \|x^{k+1} - x^*\|_2^2$$

Пусть $\|g_k\|_2 \leqslant G$ для любого k и $\|x^0 - x^*\|_2 \leqslant R$. Тогда суммируя последнее неравенство по k от 0 до n

$$2\sum_{k=1}^{n} \alpha_k \left(f(x^k) - f^*\right) \leqslant \|x^0 - x^*\|_2^2 - \|x^{n+1} - x^*\|_2^2 + G^2 \sum_{k=1}^{n} \alpha_k^2$$

$$\|x^{n+1} - x^*\|_2^2 \leqslant R^2 - 2\sum_{k=1}^{n} \alpha_k \left(f(x^k) - f^*\right) + G^2 \sum_{k=1}^{n} \alpha_k^2$$

Пусть $f_n^{\text{best}} = \min\limits_{k=1\ldots n} f(x^k)$. Тогда

$$\sum_{k=1}^{n} \alpha_k \left(f(x^k) - f^*\right) \geqslant \sum_{k=1}^{n} \alpha_k \left(f_n^{\text{best}} - f^*\right) = \left(f_n^{\text{best}} - f^*\right) \sum_{k=1}^{n} \alpha_k$$

$$\|x^{n+1} - x^*\|_2^2 \leqslant R^2 - 2\left(f_n^{\text{best}} - f^*\right) \sum_{k=1}^{n} \alpha_k + G^2 \sum_{k=1}^{n} \alpha_k^2$$

$$f_n^{\text{best}} - f^* \leqslant \frac{R^2 + G^2 \sum\limits_{k=1}^{n} \alpha_k^2}{2\sum\limits_{k=1}^{n} \alpha_k}$$

Рассмотрим теперь несколько возможных значений шага α_k

- $\alpha_k = \alpha$. Получаем

  $$f_n^{\text{best}} - f^* \leqslant \frac{R^2 + G^2\alpha^2 n}{2\alpha n} = \frac{R^2}{2\alpha n} + \frac{1}{2}G^2\alpha$$

  Минимизируя по α получим $\alpha^* = \frac{R}{G\sqrt{n}}$. При таком шаге субградиентный спуск сходится

  $$f_n^{\text{best}} - f^* \leqslant \frac{GR}{\sqrt{n}}$$

- $\alpha_k = \frac{\gamma}{\|g_k\|_2}$. Получаем

  $$f_n^{\text{best}} - f^* \leqslant \frac{R^2 + \sum\limits_{k=1}^{n} \|g_k\|_2^2 \alpha_k^2}{2\sum\limits_{k=1}^{n} \frac{\gamma}{\|g_k\|_2}} = \frac{GR^2 + G\gamma^2 n}{2\gamma n} = \frac{GR^2}{2\gamma n} + \frac{1}{2}G\gamma$$

  Минимизируя по γ получим $\gamma^* = \frac{R}{\sqrt{n}}$. При таком шаге субградиентный спуск сходится

  $$f_n^{\text{best}} - f^* \leqslant \frac{GR}{\sqrt{n}}$$

- $\alpha_k = \frac{R}{G\sqrt{k}}$. Получаем для некоторых C_1, C_2, C_3

  $$f_n^{\text{best}} - f^* \leqslant \frac{GR + GR\sum\limits_{k=1}^{n} \frac{1}{k}}{2\sum\limits_{k=1}^{n} \frac{1}{\sqrt{k}}} \leqslant \frac{RG + RGC_1 \log n}{C_2\sqrt{n}} \leqslant \frac{C_3 RG}{\sqrt{n}}$$

  То есть при таком шаге субградиентный спуск сходится.

- $\alpha_k = \frac{1}{k}$. Получаем

$$f_n^{\text{best}} - f^* \leqslant \frac{R^2 + G^2 \sum\limits_{k=1}^n \frac{1}{k^2}}{2 \sum\limits_{k=1}^n \frac{1}{k}} \leqslant \frac{R^2 + G^2 C_1}{C_2 \log n}$$

  То есть при таком шаге субградиентный спуск сходится.

- $\alpha_k = \frac{f(x^k) - f^*}{\|g_k\|_2^2}$. Получаем

$$f_n^{\text{best}} - f^* \leqslant \frac{R^2 + \sum\limits_{k=1}^n \|g_k\|_2^2 \alpha_k^2}{2 \sum\limits_{k=1}^n \alpha_k} = \frac{R^2 + \sum\limits_{k=1}^n \frac{\left(f(x^k) - f^*\right)^2}{\|g_k\|_2^2}}{2 \sum\limits_{k=1}^n \frac{f(x^k) - f^*}{\|g_k\|_2^2}} \leqslant \dots$$

  При таком шаге субградиентный спуск наверное сходится, иначе бы его не назвали шагом Поляка.

2.

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_1$$

$$\partial f(x) = \begin{cases} \{0\}, & Ax - b = 0 \\ \left\{\|Ax - b\|_2 \cdot A^\top s : \; s \in \mathbb{R}^n, \; \|s\|_2 = 1, \; \langle s, \, Ax - b\rangle = \|Ax - b\|_2\right\}, & Ax - b \neq 0 \end{cases}$$
$$+ \begin{cases} B_{\|\cdot\|_\infty}(0, \lambda), & x = 0 \\ \{\lambda s : \; s \in \mathbb{R}^n, \; \|s\|_\infty = 1, \; \langle s, \, x\rangle = \|x\|_1\}, & x \neq 0 \end{cases}$$

3. Воспользуемся следующим алгоритмом: будем минимизировать методом субградиентного спуска функцию f

$$f(x) = \|A^{1/2}(x - y)\|_2 - 1 + \|\Sigma x\|_\infty - 1$$

Реализацию алгоритма для заданых значений смотри ниже

Accelerated methods

1. При $x_0 = 3.4$ наблюдаем, что точка перемещается по циклической траектории по треугольнику, перестает сходиться. При изменении α^* и β^* метод снова начинает сходиться. Сам ход смотри ниже

2.