Оптимальные методы оптимизации первого порядка с относительным шумом

Денис Николаевич Рубцов

Московский физико-технический институт

Курс: Автоматизация научных исследований (практика, В. В. Стрижов)/Группа 105

Эксперт: к.ф.-м.н. Э. А. Горбунов

Консультант: Н. М. Корнилов

2024

Цель исследования

Цели

- разработать оптимальные алгоритмы оптимизации первого порядка, в которых градиенты известны лишь с некоторой относительной погрешностью
- исследовать скорость сходимости этих алгоритмов

Оптимизация первого порядка

Решается задача

$$x_\star = \arg\min_{x \in \mathbf{R}^d} f(x)$$

$$f(x) \in \mathcal{F}$$

Анализ наихудшего случая

$$\max_{f \in \mathcal{F},\ (x_t)_{t \le T} \in (\mathbf{R}^d)^{T+1}} P(f, (x_t)_{t \le T})$$

$$\text{s.t.} \begin{cases} x_0 \in \mathcal{N}(x_\star) \\ (x_t)_{t \le T} = \mathcal{A}(f, T, x_0) \end{cases}$$

$\mathcal{F}$	некоторый класс функций
$\mathcal{F}_{\mu,L}$	класс μ-сильно выпуклых L-гладких функций
f	исследуемая функция
$x_\star$	точка минимума
x_0	начальная точка
$\mathcal{O}^{(f)}$	оракул
$\mathcal{A}$	алгоритм
$(x_t)_{t \le T}$	последовательность точек
T	число итераций
$P(f, (x_t)_{t \le T})$	метрика точности

Постановка задачи

Решается задача

$$x_\star = \arg\min_{x \in \mathbf{R}^d} f(x)$$

$$f(x) \in \mathcal{F}_{\mu,L}$$

Жадный алгоритм

$$\max_{f \in \mathcal{F}_{\mu,L},\ (x_t)_{t \le T} \in (\mathbf{R}^d)^{T+1}} f(x_t) - f_\star$$

$$\text{s.t.} \begin{cases} x_0 \in U_R(x_\star) \\ (x_t)_{t \le T} = \text{span}\{x_t, (\mathcal{O}_t^{(f)})_{t \le T}\} \end{cases}$$

Неточный градиент

$$\mathcal{O}^{(f)}(x) = \widetilde{\nabla} f(x)$$

$$\|\widetilde{\nabla} f(x) - \nabla f(x)\|_2 \le \hat{\varepsilon} \|\nabla f(x)\|_2$$

$$\forall \hat{\varepsilon} \in [0, 1].$$

Гипотеза

$$f(x_t) - f(x_\star) \le q^t (f(x_0) - f(x_\star))$$

GD	NAG, HB
$q \sim \frac{\mu}{L}$	$q \sim \sqrt{\frac{\mu}{L}}$

$$q \sim (\tfrac{\mu}{L})^{g(\varepsilon)}$$

Поясняющие картинки

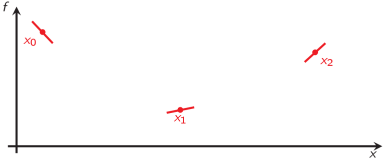

Рис.: Плодотворные идеи интерполяции в PEP. Можно ли по значениям точек, значениям точек и производным в этих точках восстановить функцию нужного класса?

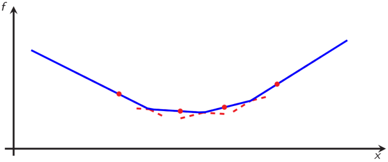

Рис.: Плодотворные идеи интерполяции в PEP. Ответ - да. Можно брать максимум значений красных пунктирных линий (поточечный максимум)

Вычислительные эксперименты

$$\tau_t = f(x_t) - f(x_\star) \leq q^t(f(x_0) - f(x_\star)) \Rightarrow \log \tau_t \sim t$$
$$q \sim \left(\frac{\mu}{L}\right)^{g(\varepsilon)}$$
$$\log q \sim g(\varepsilon) \log \mu$$

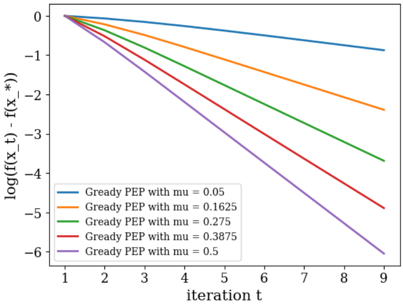

Рис.: График зависимости лонарифма невязки от номера итерации $\log \tau_t \sim t$ при $\varepsilon = 0.5$. Доказывает линейную сходимость.

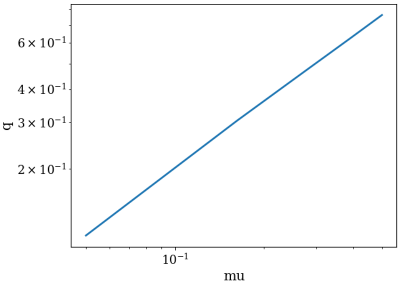

Рис.: График зависимости $q(\mu)$ при $\varepsilon = 0.5$ в лог-лог-масштабе. Показывает, что степень сходимости зависит лишь от шума

Основной результат вычислительных экспериментов

$$q \sim \left(\frac{\mu}{L}\right)^{g(\varepsilon)}$$

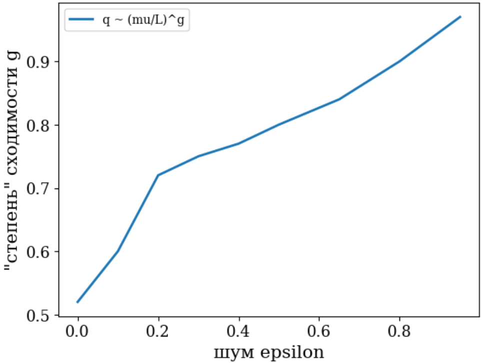

Рис.: График зависимости $g(\varepsilon)$. При нулевом шуме, сходимость как у ускоренных методов (NAG, HB). При больших шумах сходимость как у GD

Заключение

Результаты

- предложена гипотеза сходимости оптимального (жадного) алгоритма оптимизации первого порядка в случае, когда градиенты известны лишь с некоторой погрешностью
- гипотеза доказана численно

Планы

- теоретическое доказательсто полученных экспериментальных результатов

Литература

- Baptiste Goujaud, Aymeric Dieuleveut, and Adrien Taylor. On fundamental proof structures in first-order optimization. In 2023 62nd IEEE Conference on Decision and Control (CDC), pages 3023–3030. IEEE, 2023.
- Nikita Kornilov, Eduard Gorbunov, Mohammad Alkousa, Fedor Stonyakin, Pavel Dvurechensky, and Alexander Gasnikov. Intermediate gradient methods with relative inexactness. arXiv preprint arXiv:2310.00506, 2023
- Adrien B Taylor. Convex interpolation and performance estimation of first-order methods for convex optimization. PhD thesis, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2017
- Adrien B Taylor, Julien M Hendrickx, and Fran cois Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. Mathematical Programming, 161:307–345, 2017.