Жадный метод 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 жадный метод оптимизации первого порядка, в котором градиенты известны лишь с некоторой относительной погрешностью
- исследовать скорость сходимости этого метода с помощью численной техники PEP

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.
- 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.

NoisyGFOM

$$x_\star = \arg\min_{x\in\mathbb{R}^n} f(x) \qquad \|\widetilde{\nabla} f(x) - \nabla f(x)\|_2 \leq \varepsilon\|\nabla f(x)\|_2$$

$$f(x) \in \mathcal{F}_{\mu,L} \qquad \varepsilon \in [0,1]$$

Итерация метода	$f(x_N) - f(x_\star) =$
Gradient descent (GD)	
$x_{k+1} = x_k - \alpha\nabla f(x_k)$	$O\left(\left(\frac{1-\frac{\mu}{L}}{1+\frac{\mu}{L}}\right)^N\right)$
Greedy first-order method (GFOM)	
$x_k = \arg\min_{x\in\mathbb{R}^n}\{f(x) : x \in x_0 + \text{span}\{\nabla f(x_0), \ldots, \nabla f(x_{k-1})\}\}$	$O\left(\left(\frac{1-\sqrt{\frac{\mu}{L}}}{1+\sqrt{\frac{\mu}{L}}}\right)^N\right)$
Noisy greedy first-order method (**NoisyGFOM**)	
$x_k = \arg\min_{x\in\mathbb{R}^n}\{f(x) : x \in x_0 + \text{span}\{\widetilde{\nabla} f(x_0), \ldots, \widetilde{\nabla} f(x_{k-1})\}\}$	

NoisyGFOM

$$x_\star = \arg\min_{x \in \mathbb{R}^n} f(x) \qquad \|\widetilde{\nabla} f(x) - \nabla f(x)\|_2 \le \varepsilon \|\nabla f(x)\|_2$$

$$f(x) \in \mathcal{F}_{\mu,L} \qquad \varepsilon \in [0,1]$$

Итерация метода	$f(x_N) - f(x_\star) =$
Gradient descent (GD)	
$x_{k+1} = x_k - \alpha \nabla f(x_k)$	$O\left(\left(\frac{1-\frac{\mu}{L}}{1+\frac{\mu}{L}}\right)^N\right)$
Greedy first-order method (GFOM)	
$x_k = \arg\min_{x \in \mathbb{R}^n} \{f(x) : x \in x_0 + \mathrm{span}\{\nabla f(x_0), \ldots, \nabla f(x_{k-1})\}\}$	$O\left(\left(\frac{1-\sqrt{\frac{\mu}{L}}}{1+\sqrt{\frac{\mu}{L}}}\right)^N\right)$
Noisy greedy first-order method (**NoisyGFOM**)	
$x_k = \arg\min_{x \in \mathbb{R}^n} \{f(x) : x \in x_0 + \mathrm{span}\{\widetilde{\nabla} f(x_0), \ldots, \widetilde{\nabla} f(x_{k-1})\}\}$	$O\left(\left(\frac{1-\left(\frac{\mu}{L}\right)^{\alpha(\varepsilon)}}{1+\left(\frac{\mu}{L}\right)^{\alpha(\varepsilon)}}\right)^N\right)$

Скорость сходимости NoisyGFOM

Анализ наихудшего случая

$$\max_{f \in \mathcal{F}_{\mu,L},\ (x_k)_{k \le N} \in (\mathbf{R}^n)^{N+1}} \|f(x_N) - f_\star\|$$

$$\text{s.t.} \begin{cases} x_0 \in \{x : \|x - x_\star\|_2^2 \le R^2\} \\ (x_k)_{k \le N} = x_0 + \operatorname{span}\{(\widetilde{\nabla} f(x_k)\}_{k \le N}\} \\ \|\widetilde{\nabla} f(x) - \nabla f(x)\|_2 \le \varepsilon \|\nabla f(x)\|_2 \end{cases}$$

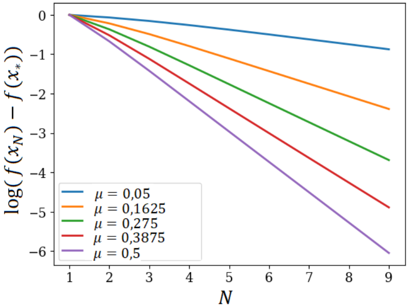
PEP

Бесконечномерная задача оптимизации с помощью плодотворных идей интерполяции может быть сведена к конечномерной задаче полуопределенного программирования. Такая техника называется PEP. Используемый солвер - PEPit.

Вычислительные эксперименты

При фиксированном ε и $\frac{\mu}{L}$ $\log|f(x_N) - f_\star| \sim N \cdot k(\varepsilon, \mu)$, где $k(\varepsilon, \mu) = \log\left(1 - 2\left(\frac{\mu}{L}\right)^{\alpha(\varepsilon)}\right) \sim \left(\frac{\mu}{L}\right)^{\alpha(\varepsilon)}$. Тогда $\log k(\varepsilon, \mu) = \alpha(\varepsilon)\log\left(\frac{\mu}{L}\right)$

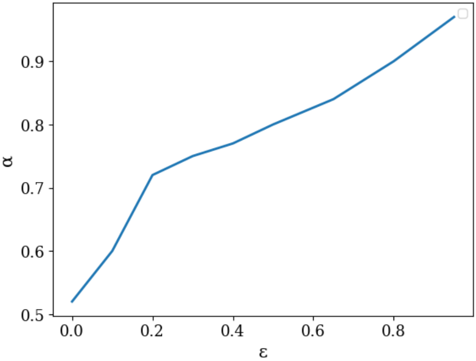
Рис.: График зависимости логарифма невязки от номера итерации $\log\left(f(x_N) - f(x_\star)\right) \sim N$ при $\varepsilon = 0.5$.



Результат проверки гипотезы

$$f(x_N) - f(x_\star) = O\left(\left(\frac{1-\left(\frac{\mu}{L}\right)^{\alpha(\varepsilon)}}{1+\left(\frac{\mu}{L}\right)^{\alpha(\varepsilon)}}\right)^N\right), \ \alpha|_{\varepsilon=0} = \frac{1}{2}, \ \alpha|_{\varepsilon=1} = 1$$

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



Заключение

Результаты

- предложена гипотеза сходимости жадного алгоритма оптимизации первого порядка в случае, когда градиенты известны лишь с некоторой погрешностью
- численные результаты не противоречат гипотезе

Планы

- теоретическое доказательство полученных экспериментальных результатов