Универсальные методы для стохастических вариационных неравентсв

Климза Антон Алексеевич

Московский физико-технический институт

Курс: Научный трек Иннпрака ФПМИ

Эксперт: д.ф.-м.н. А. В. Гасников

26 марта 2024

Введение

Вариационное неравенство
Пусть дано выпуклое множество $Z \in \mathbb{R}^n$ и оператор $F : \mathbb{R}^n \to \mathbb{R}^n$. Тогда хотим найти $z^* \in Z$, такую что:

$$\langle F(z^*), z - z^* \rangle \geq 0, \ \forall z \in Z$$

Стохастический случай

$$F(z) = \mathbb{E}_\xi F(z, \xi)$$

$$\mathbb{E}_\xi \| F(z) - F(z, \xi) \|^2 \leq \sigma^2$$

Градиентный спуск

При условии липшицевости $\|\nabla f(y) - \nabla f(x)\|_2 \leq L\|y - x\|_2$ и шаге $\gamma_k = \frac{1}{L}$:

$$x^{k+1} = arg \min_{x \in \mathbb{R}^n} \left\{ f(x^k) + \left\langle \nabla f(x^k), x - x^k \right\rangle + \frac{L}{2}\|x - x^k\|_2^2 \right\}$$

Универсальный градиентный спуск

При условии Гёльдера
$\|\nabla f(y) - \nabla f(x)\|_* \leq L_\nu \|y - x\|^\nu, \ \ \nu \in [0, 1]$:

Универсальный градиентный спуск

$L^{k+1} = \frac{L^k}{2}$

While True Do

$x^{k+1} = \arg\min_{x \in Q} \left\{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + L^{k+1} V(x, x^k) \right\}$

If $\left\{ f(x^{k+1}) \leqslant f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + L^{k+1} V(x^{k+1}, x^k) + \frac{\varepsilon}{2} \right\}$

Перейти на следующую итерацию: $k \to k+1$

Else

$L^{k+1} := 2L^{k+1}$

Где $L^{k+1} \leq L_\nu \left(\frac{L_\nu}{\varepsilon} \frac{1-\nu}{1+\nu} \right)^{(1-\nu)/(1+\nu)}$ (см. [2])

Универсальный стохастический градиентный спуск

Задача минимизации функции
$F(x) = f(x) + \psi(x),\ x \in dom(\psi)$, для выпуклых функций
$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, где вычисление градиента ψ - значительно более простая задача:

Algorithm 4.1 Universal Stochastic Gradient Method

1: **Initialize:** $x_0 \in \operatorname{dom}\psi,\ D > 0,\ H_0 := 0,\ g_0 \sim \hat{g}(x_0)$.
2: **for** $k = 0, 1, \ldots$ **do**
3: $\quad x_{k+1} = \operatorname{argmin}_{x \in \operatorname{dom}\psi}\{\langle g_k, x\rangle + \psi(x) + \frac{H_k}{2}\|x - x_k\|^2\}$.
4: $\quad g_{k+1} \sim \hat{g}(x_{k+1})$.
5: $\quad H_{k+1} := H_k + \dfrac{[\hat{\beta}_{k+1} - \frac{1}{2}H_k r_{k+1}^2]_+}{D^2 + \frac{1}{2}r_{k+1}^2}$,

$\quad$ where $r_{k+1} = \|x_{k+1} - x_k\|$, $\hat{\beta}_{k+1} = \langle g_{k+1} - g_k, x_{k+1} - x_k\rangle$.
6: **end for**

Где $f'(x) = \mathbb{E}_\xi g(x, \xi),\ D = \sup\limits_{x,y \in dom(\psi)} \|x - y\|$

Сравнение с другими методами

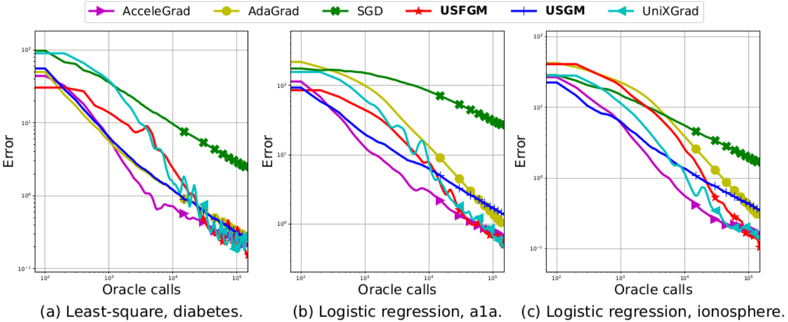

Рис.: сравнение работы оптимизаторов для задач выпуклой оптимизации (см. [1])

Сравнение с другими методами

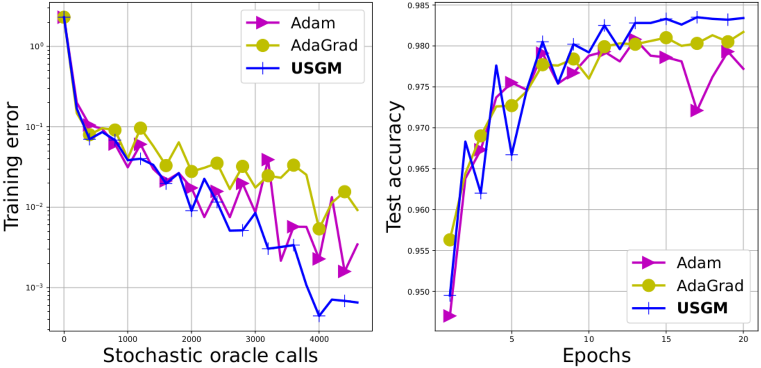

Рис.: сравнение работы оптимизаторов для нейронных сетей на датасете MNIST (см. [1])

Источники

[1] Anton Rodomanov Ali Kavis Yongtao Wu Kimon Antonakopoulos Volkan Cevher Universal Gradient Methods for Stochastic Convex Optimization. 2024.
[2] А. В. Гасников Современные численные методы оптимизации 2021
[3] Fedor Stonyakin Alexander Gasnikov Pavel Dvurechensky Alexander Titov Mohammad Alkousa Generalized Mirror Prox Algorithm for Monotone Variational Inequalities: Universality and Inexact Oracle 2022