Разработка распределенных децентрализованных безградиентных методов решения задач оптимизации

Александр Николаевич Безносиков

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Научный руководитель: д.ф.-м.н. А. В. Гасников

Москва 2020

Цель работы

- Для задачи композитной оптимизации разработать новый метод, который использует смешанный оракул: для одной из частей задачи – оракул нулевого порядка, а для другой – первого порядка.
- Применить полученный метод для практических задач.

Литература

- Lan G. Gradient sliding for composite optimization // Mathematical Programming.2016. Vol. 159, no. 1-2. P. 201–235.
- Lan G. Lectures on Optimization Methods for Machine Learning. H. Milton Stewart School of Industrial and Systems Engineering Georgia Institute of Technology,Atlanta, GA, 2019.
- Shamir O. An Optimal Algorithm for Bandit and Zero-Order Convex Optimizationwith Two-Point Feedback. // Journal of Machine Learning Research. 2017. Vol. 18,no. 52. P. 1–11.

Задача композитной оптимизации

- Задача композитной оптимизации

$$\Psi_0(x) = f(x) + g(x) \to \min_{x \in X}$$

- $X \subseteq \mathbb{R}^n$ – выпуклый компакт с диаметром D_X.
- Функция g выпуклая и L-гладкая на X,
- Функция f выпуклая и дифференцируемая на X с ограниченным градиентом.

Оракулы

- Доступен градиент $\nabla g(x)$.
- Для f имеется только оракул нулевого порядка

$$\tilde{f}(x) = f(x, \xi) + \Delta(x),$$

где $\Delta(x)$ – ограниченный шум неизвестной природы:

$$|\Delta(x)| \leq \Delta,$$

ξ отвечает за стохастический шум

$$\mathbb{E}[f(x,\xi)] = f(x), \quad \|\nabla f(x,\xi)\|_2 \leq M(\xi), \quad \mathbb{E}[M^2(\xi)] = M^2.$$

- Стохастическая аппроксимация $\nabla f(x)$:

$$\tilde{f}'_r(x) = \frac{n}{2r}(\tilde{f}(x+re) - \tilde{f}(x-re))e,$$

где e – случайный вектор, равномерно распределенный на евклидовой сфере, r – параметр сглаживания.

Новый метод

Algorithm 1 Zeroth-Order Sliding Algorithm (`zoSA`)

Input: Initial point $x_0 \in X$ and iteration limit N.

Let $\beta_k \in \mathbb{R}_{++}, \gamma_k \in \mathbb{R}_+$, and $T_k \in \mathbb{N}$, $k = 1, 2, \ldots$, be given and set $\bar{x}_0 = x_0$.

for $k = 1, 2, \ldots, N$ **do**

1. Set $\underline{x}_k = (1 - \gamma_k)\bar{x}_{k-1} + \gamma_k x_{k-1}$,
   and let $h_k(\cdot) \equiv g(\underline{x}_{k-1}) + \langle \nabla g(\underline{x}_{k-1}), \cdot - \underline{x}_{k-1} \rangle$.
2. Set
$$(x_k, \tilde{x}_k) = \mathrm{PS}(h_k, x_{k-1}, \beta_k, T_k);$$
3. Set $\bar{x}_k = (1 - \gamma_k)\bar{x}_{k-1} + \gamma_k \tilde{x}_k$.

end for

Output: $\bar{x}_N$.

Algorithm 2 The PS (prox-sliding) procedure

procedure $(x^+, \tilde{x}^+) = \mathrm{PS}(h,\, x,\, \beta,\, T)$

Let the parameters $p_t \in \mathbb{R}_{++}$ and $\theta_t \in [0, 1]$, $t = 1, \ldots,$ be given. Set $u_0 = \tilde{u}_0 = x$.

for $t = 1, 2, \ldots, T$ **do**

$$\begin{aligned} u_t &= \arg\min_{u \in X} \left\{ h(u) + \langle \tilde{f}_r'(u_{t-1}), u \rangle + \beta V(x, u) + \beta p_t V(u_{t-1}, u) \right\}, \\ \tilde{u}_t &= (1 - \theta_t)\tilde{u}_{t-1} + \theta_t u_t. \end{aligned}$$

end for

Set $x^+ = u_T$ and $\tilde{x}^+ = \tilde{u}_T$.

end procedure

Сходимость

Теорема

Для любого натурального $N \geq 1$:

$$\mathbb{E}[\Psi_0(\overline{x}_N) - \Psi_0(x^*)] \leq 2rM + \frac{12LD_X^2}{N(N+1)} + \frac{n\Delta D_X p_*}{r}.$$

Дополнительно, если $r = \Theta\left(\frac{\varepsilon}{M}\right), \Delta = O\left(\frac{\varepsilon^2}{nMD_X}\right)$, то количество вызовов оракулов ∇g и $\tilde{f}_r'$, чтобы найти ε-решение, может быть оценено следующим образом

$$O\left(\sqrt{\frac{LD_X^2}{\varepsilon}}\right), \quad O\left(\sqrt{\frac{LD_X^2}{\varepsilon}} + \frac{D_X^2 p_*^2 n M^2 (C_1^2+1)}{\varepsilon^2}\right),$$

где $\sqrt[4]{\mathbb{E}[\|e\|_*^4]} \leq p_*$, $\|x\|_* \leq C_1 \|x\|_2$.

Сходимость: специальные случаи

- Евклидов случай $\|\cdot\| = \|\cdot\|_2$, $p_* = C_1 = 1$. Количество вызовов оракула $\tilde{f}'_r$ ограничено:

$$O\left(\sqrt{\frac{LD_X^2}{\varepsilon}} + \frac{D_X^2 n M^2}{\varepsilon^2}\right)$$

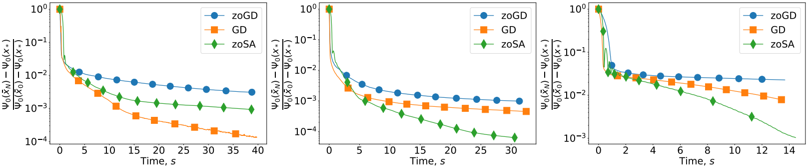
- Случай первой нормы $\|\cdot\| = \|\cdot\|_1$, $p_* = O\left(\ln(n)/n\right)$, а $C_1 = 1$. Количество вызовов оракула $\tilde{f}'_r(x)$ ограничено:

$$O\left(\sqrt{\frac{LD_X^2}{\varepsilon}} + \frac{D_X^2 M^2 \ln n}{\varepsilon^2}\right).$$

Эксперименты

- Логистическая регрессия:

$$
\begin{aligned}
\min_{x \in \mathbf{R}^n} \Psi_0(x) &= \overbrace{l_1 \|x\|_1}^{f(x)} + g(x) \\
g(x) &= \frac{1}{m} \sum_{i=1}^{m} \log\left(1 + \exp\left(-y_i \cdot (Ax)_i\right)\right).
\end{aligned}
$$



(a) `mushrooms`, $l_1 = 10^{-3}$

(b) `a5a`, $l_1 = 10^{-4}$

(c) `german.numer`, $l_1 = 10^{-4}$

Рис. 1: `zoSA`, `GD` и `zoGD` для задачи (1) с различными датасетами и параметрами l_1.

Заключение

Предложенный метод — `zoSA` — является первым $\frac{1}{2}$-методом для выпуклой композитной оптимизации: он использует оракул нулевого порядка для негладкого члена, и оракул первого порядка – для гладкого. Метод имеет хорошо изученную теорию и конкурентоспособен на практике даже с некоторыми методами первого порядка.

Публикации

- Beznosikov A., Gorbunov E., Gasnikov A. Derivative-Free Method For Composite Optimization With Applications To Decentralized Distributed Optimization. 2019. arXiv:math.OC/1911.10645.
- Beznosikov A., Sadiev A., Gasnikov A. Gradient-Free Methods for Saddle-Point Problem. 2020. arXiv:math.OC/2005.05913.