Новые грани концепции черного ящика в методе условного градиента

Александр Иванович Богданов

Научный руководитель:
к.ф.-м.н. Безносиков А. Н.

Московский физико-технический институт
ФПМИ
Кафедра «Интеллектуальные системы»

Цель исследования

Цель: Ставится выпуклая задача оптимизации на ограниченном множестве с доступом только к нулевому оракулу.

Решение: Предлагается модификация JAGUAR, которая использует аппроксимацию градиента для подсчета условного градиента.

Литература

- Aryan Mokhtari, Hamed Hassani, and Amin Karbasi. Stochastic conditional gradient methods: From convex minimization to submodular maximization. The Journal of Machine Learning Research, 21(1):4232–4280, 2020.
- Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. The Journal of Machine Learning Research, 18(1):1703–1713, 2017.
- Alexander Gasnikov, Darina Dvinskikh, Pavel Dvurechensky, Eduard Gorbunov, Aleksander Beznosikov, and Alexander Lobanov. Randomized gradient-free methods in convex optimization. arXiv preprint arXiv:2211.13566, 2022.

Постановка задачи

Рассматривается две оптимизационные задачи:

- Нестохастическая

$$\min_{x \in Q} \quad f(x)$$

  Доступ только к $f_\delta(x) := f(x) + \delta(x)$, где $\delta(x)$ - шум.

- Стохастическая

$$\min_{x \in Q} \quad f(x) := \mathbb{E}_\xi \left[f(x, \xi) \right],$$

  Доступ только к $f_\delta(x, \xi) := f(x, \xi) + \delta(x, \xi)$, где $\delta(x, \xi)$ - шум.

$Q \subseteq \mathbb{R}^d$ - произвольное выпуклое ограниченное множество, $f(x)$ - выпуклая на Q функция.

JAGUAR. Нестохастический случай

Разностная схема, которая будет использоваться в алгоритме:

$$\widetilde{\nabla}_i f_\delta(x) := \frac{f_\delta(x+\gamma e_i) - f_\delta(x-\gamma e_i)}{2\gamma} e_i,$$

где e_i - i-ый базисный вектор, γ - параметр сглаживания.

Algorithm JAGUAR. Нестохастический случай

1: **Вход:** $x, h \in \mathbb{R}^d$
2: Сэмплируем $i \in \overline{1, d}$ независимо и равномерно
3: Считаем $\widetilde{\nabla}_i f_\delta(x) = \dfrac{f_\delta(x+\gamma e_i) - f_\delta(x-\gamma e_i)}{2\gamma} e_i$
4: $g = g - \langle h, e_i \rangle e_i + \widetilde{\nabla}_i f_\delta(x)$
5: **Выход:** g

Франк-Вульф с JAGUAR в нестохастическом случае

Algorithm ФВ с JAGUAR в нестохастическом случае

Вход: $x_0 \in Q$, $g^0 = \widetilde{\nabla} f(x^0)$
for k = 0, 1, 2, ... , N **do**
    $g^{k+1} = \text{JAGUAR}\left(x^k, g^k\right)$
    $s^k = \underset{x \in Q}{\arg\min} \left\{ \left\langle s, g^{k+1} \right\rangle \right\}$
    $x^{k+1} = x^k + \gamma_k (s^k - x^k)$
end for
Выход: x^{N+1}

JAGUAR. Нестохастический случай

Допущения:

1. Функция $f(x)$ L-гладкая на множестве Q, то есть:

$$\forall x, y \in Q \hookrightarrow \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| .$$

2. Ограниченность оракульного шума:

$$\exists \Delta > 0 : \ \forall x \in Q \hookrightarrow |\delta(x)|^2 \leq \Delta^2.$$

3. Ограниченность:

$$\|x - y\|^2 \leq D^2,$$

   для любых $x, y \in Q$.

JAGUAR. Нестохастический случай

Теорема При шаге оптимизатора:

$$\gamma_k = \frac{4}{k + 8d}$$

получается следующее:

$$\mathbb{E}\left[\left\|h^k - \nabla f(x^k)\right\|^2\right] = \mathcal{O}\Bigg(dL^2\gamma^2 + \frac{d\Delta^2}{\gamma^2} + \frac{\max\{d^2L^2D^2, \left\|h^0 - \nabla f(x^0)\right\|^2 \cdot d^2\}}{(k+d)^2}\Bigg).$$

Если $h_0 = \widetilde{\nabla} f_\delta(x^0)$, то можно упростить:

$$\mathbb{E}\left[\left\|h^k - \nabla f(x^k)\right\|^2\right] = \mathcal{O}\left(dL^2\gamma^2 + \frac{d\Delta^2}{\gamma^2} + \frac{d^2L^2D^2}{(k+8d)^2}\right).$$

Франк-Вульф с JAGUAR в нестохастическом случае

Теорема При шаге оптимизатора:

$$\gamma_k = \frac{4}{k + 8d}$$

получается следующее:

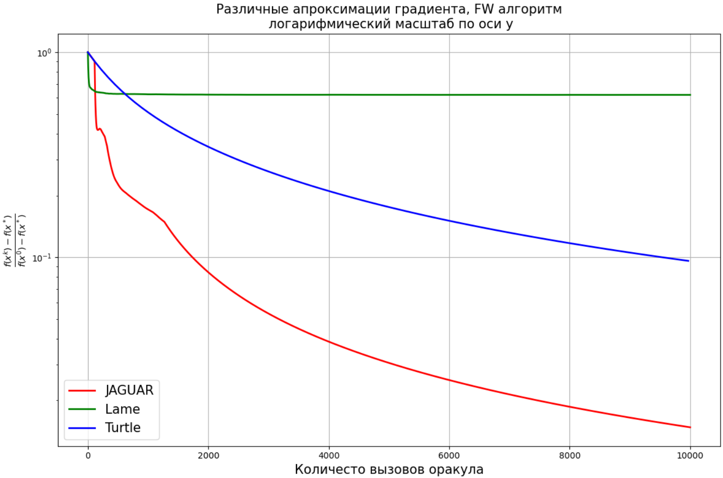
$$\mathbb{E}\left[f(x^N) - f(x^*)\right] = \mathcal{O}\left(\frac{d \max\{LD^2, f(x^0) - f(x^*)\}}{N + 8d} + \sqrt{d}LD\gamma + \frac{\sqrt{d}\Delta D}{\gamma}\right).$$

Следствие Пусть ε определяет точность, то есть $\mathbb{E}\left[f(x^N) - f(x^*)\right] \leq \varepsilon$. Тогда:

$$N = \mathcal{O}\left(\frac{d \max\{LD^2, f(x^0) - f(x^*)\}}{\varepsilon}\right),$$

$$\gamma = \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}LD}\right), \quad \Delta = \mathcal{O}\left(\frac{\varepsilon^2}{dLD^2}\right).$$

Эксперимент проводился на тренировочном датасете "mushrooms".



Разностные схемы, которые будут использоваться в алгоритме:

- Двухточечная обратная связь:

$$\widetilde{\nabla}_i f_\delta(x, \xi) := \frac{f_\delta(x + \gamma e_i, \xi) - f_\delta(x - \gamma e_i, \xi)}{2\gamma} e_i,$$

- Одноточечная обратная связь:

$$\widetilde{\nabla}_i f_\delta(x, \xi^+, \xi^-) := \frac{f_\delta(x + \gamma e_i, \xi^+) - f_\delta(x - \gamma e_i, \xi^-)}{2\gamma} e_i,$$

где e_i - i-ый базисный вектор, γ - параметр сглаживания.

JAGUAR. Стохастический случай

Algorithm JAGUAR. Стохастический случай

Вход: $x, h, g \in \mathbb{R}^d$, $0 \leq \eta \leq 1$

Сэмплируем $i \in \overline{1, d}$ независимо и равномерно

Сэмплируем 2 реализации ξ: ξ^+ и ξ^- независимо (в случае одноточечной обратной связи $\xi^+ = \xi^-$)

Считаем $\widetilde{\nabla}_i f_\delta(x, \xi^+, \xi^-) = \frac{f_\delta(x+\gamma e_i, \xi^+) - f_\delta(x-\gamma e_i, \xi^-)}{2\gamma} e_i$

$h = h - \langle h, e_i \rangle e_i + \widetilde{\nabla}_i f_\delta(x, \xi^+, \xi^-)$

$\rho = h - d \cdot \langle h, e_i \rangle e_i + d \cdot \widetilde{\nabla}_i f_\delta(x, \xi^+, \xi^-)$

$g = (1 - \eta) g + \eta \rho$

Выход: h и g

JAGUAR. Стохастический случай

Допущения:

1. Функция $f(x, \xi)$ $L(\xi)$-гладкая на множестве Q, то есть:

$$\forall x, y \in Q \hookrightarrow \|\nabla f(x, \xi) - \nabla f(y, \xi)\| \leq L(\xi) \|x - y\|.$$

2. Ограниченность оракульного шума:

$$\exists \Delta > 0: \ \forall x \in Q \hookrightarrow \mathbb{E}\left[|\delta(x, \xi)|^2\right] \leq \Delta^2$$

3. Ограниченность второго момента градиента:

$$\exists \sigma_\nabla^2 : \mathbb{E}\left[\|\nabla f(x, \xi) - \nabla f(x)\|^2\right] \leq \sigma_\nabla^2$$

4. Ограниченность второго момента градиента:

$$\exists \sigma_f^2 : \mathbb{E}\left[|f(x, \xi) - f(x)|^2\right] \leq \sigma_f^2$$

Франк-Вульф с JAGUAR в стохастическом случае

Algorithm ФВ с JAGUAR в стохастическом случае

Вход: $x_0 \in Q$, $h^0 = g^0 = \widetilde{\nabla} f_\delta(x^0, \xi_1^+, ..., \xi_d^-)$, $\{\eta_k\}_{k=0}^N \subset [0; 1]$
for k = 0, 1, 2, ... , N **do**
    $h^{k+1}, g^{k+1} = \text{JAGUAR}\left(x^k, h^k, g^k, \eta_k\right)$
    $s^k = \underset{x \in Q}{\arg\min}\left\{\left\langle s, g^{k+1}\right\rangle\right\}$
    $x^{k+1} = x^k + \gamma_k (s^k - x^k)$
end for
Выход: x^{N+1}

JAGUAR. Стохастический случай

Теорема При шаге оптимизатора и шаге моментума:

$$\gamma_k = \frac{4}{k + 8d^{3/2}}, \quad \eta_k = \frac{4}{(k + 8d^{3/2})^{2/3}}$$

получается следующее:

$$\mathbb{E}\left[\left\|g^k - \nabla f(x^k)\right\|^2\right] = \mathcal{O}\left(\frac{d^4\left\|h^0 - \nabla f(x^0)\right\|^2}{(k + 8d^{3/2})^{8/3}} + dL^2\gamma^2 + \frac{d\Delta^2}{\gamma^2}\right.$$
$$\left. + \frac{L^2D^2 + \max\{d^2\sigma_f^2/\gamma^2 + d^2\sigma_\nabla^2, d\left\|g^0 - \nabla f(x^0)\right\|^2\}}{(k + 8d^{3/2})^{2/3}}\right)$$

Если $h^0 = g^0 = \widetilde{\nabla} f_\delta(x^0, \xi_1^+, \xi_1^-, ..., \xi_d^+, \xi_d^-)$, то можно упростить:

$$\mathbb{E}\left[\left\|g^k - \nabla f(x^k)\right\|^2\right] = \mathcal{O}\left(\frac{L^2D^2 + d^2\sigma_f^2/\gamma^2 + d^2\sigma_\nabla^2}{(k + 8d^{3/2})^{2/3}} + dL^2\gamma^2 + \frac{d\Delta^2}{\gamma^2}\right)$$

Франк-Вульф с JAGUAR в стохастическом случае

Теорема При шаге оптимизатора и шаге моментума:

$$\gamma_k = \frac{4}{k + 8d^{3/2}}, \quad \eta_k = \frac{4}{(k + 8d^{3/2})^{2/3}}$$

получается следующее:

$$\mathbb{E}\left[f(x^N) - f(x^*)\right] = \mathcal{O}\Bigg(\frac{LD^2 + d\sigma_f D/\gamma + d\sigma_\nabla D + \sqrt{d}(f(x^0) - f(x^*))}{(N + 8d^{3/2})^{1/3}} + \sqrt{d}LD\gamma + \frac{\sqrt{d}\Delta D}{\gamma} \Bigg)$$

Следствие Пусть ε определяет точность, то есть $\mathbb{E}\left[f(x^N) - f(x^*)\right] \leq \varepsilon$. Тогда:

$$N = \mathcal{O}\left(\max\left\{ \left[\frac{LD^2 + d\sigma_\nabla D + \sqrt{d}(f(x^0) - f(x^*))}{\varepsilon} \right]^3, \frac{d^{9/2}\sigma_f^3 L^3 D^6}{\varepsilon^6} \right\} \right),$$

$$\gamma = \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}LD} \right), \quad \Delta = \mathcal{O}\left(\frac{\varepsilon^2}{dLD^2} \right).$$

Вывод

Для задачи выпуклой задачи оптимизации на ограниченном множестве с доступом только к нулевому оракулу предложена и исследована модификация JAGUAR для Франка-Вульфа. Получены теоретические оценки сходимости для нестохастического и стохастического случаев.

Дальнейшая работа

- Провести эксперименты на различных множествах, датасетах, функциях;
- Возможно, рассмотреть другие методы, например, градиентный спуск.