Применение стохастической аппроксимации нулевого порядка с техникой запоминания в алгоритме Франка-Вульфа

Богданов Александр Иванович

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

Московский физико-технический институт
(национальный исследовательский университет)
Физтех-школа прикладной математики и информатики
Кафедра «Интеллектуальные системы»

Цель исследования

Проблема: Ставится выпуклая задача оптимизации на ограниченном выпуклом множестве с доступом только к зашумленному нулевому оракулу.

Цель: Предложить робастый алгоритм аппроксимации градиента, использующий $\mathcal{O}(1)$ вызовов оракула на каждой итерации.

Решение: Предлагается модификация `JAGUAR`, которая использует технику запоминания.

Постановка задачи

Рассматривается две оптимизационные задачи:

- Нестохастическая

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

  Доступ только к $f_\delta(x) := f(x) + \delta(x)$, где $\delta(x)$ – шум.
  $f(x)$ – выпуклая на Q функция.

- Стохастическая

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

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

$Q \subseteq \mathbb{R}^d$ – произвольное выпуклое ограниченное множество.

Обычный алгоритм Франка-Вульфа

Общие допущения:

1. Ограниченность множества Q:

$$\forall\ x, y \in Q : \|x - y\|^2 \le D^2.$$

2. Выпуклость множества Q:

$$\forall\ 0 \le \alpha \le 1, \forall\ x, y \in Q : \alpha x + (1-\alpha) y \in Q.$$

Алгоритм 1 Франк-Вульф

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

Франк-Вульф с JAGUAR в нестохастическом случае

Схема аппроксимации:

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

где e_i – i-ый базисный вектор, τ – параметр сглаживания.

Алгоритм 2 ФВ с JAGUAR в нестохастическом случае

Вход: $x^0 \in Q$, $h^0 = \widetilde{\nabla} f_\delta(x^0)$, γ_k, τ

for $k = 0, 1, 2, ..., N$ **do**

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

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

$h^{k+1} = h^k - \langle h^k, e_i \rangle e_i + \widetilde{\nabla}_i f_\delta(x^k)$

$s^k = \arg\min\limits_{x \in Q} \left\langle s, h^{k+1} \right\rangle$

$x^{k+1} = x^k + \gamma_k (s^k - x^k)$

end for

Выход: x^{N+1}

Франк-Вульф с JAGUAR в нестохастическом случае

Допущения:

1. Функция $f(x)$ L-гладкая на множестве Q:
$$\forall\ x, y \in Q : \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|.$$
2. Функция $f(x)$ выпукла на множестве Q:
$$\forall\ x, y \in Q : f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$
3. Ограниченность оракульного шума:
$$\exists\ \Delta > 0 \forall\ x \in Q : |\delta(x)|^2 \leq \Delta^2.$$

Франк-Вульф с JAGUAR в нестохастическом случае

Теорема (Богданов, 2023) При шаге оптимизатора:

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

получается оценка на сходимость:

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

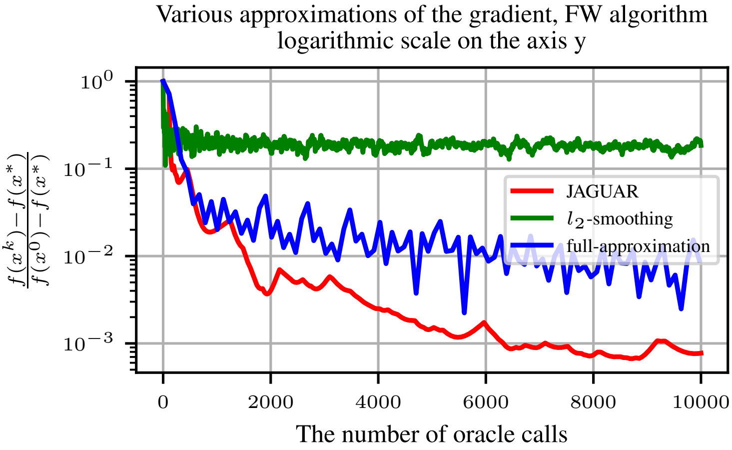
Следствие Пусть ε определяет точность: $\mathbb{E}\left[f(x^N) - f(x^*)\right] \leq \varepsilon$:

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

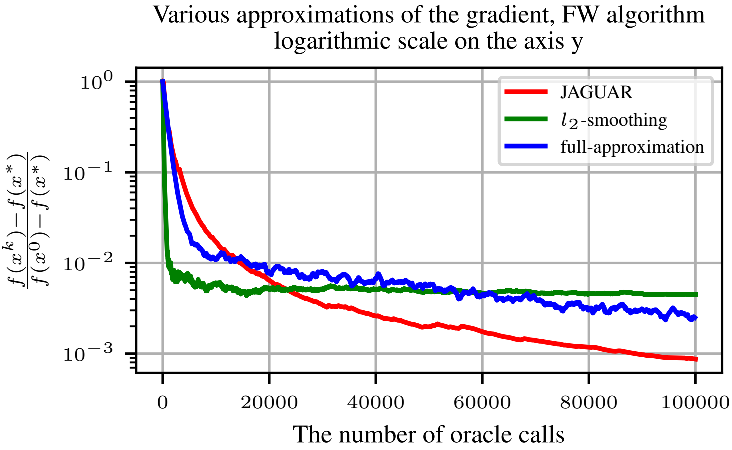
$$\gamma = \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}LD}\right), \quad \Delta = \mathcal{O}\left(\frac{\varepsilon^2}{dLD^2}\right).$$

Эксперимент для нестохастического случая

Эксперимент проводился на симплексном множестве на датасете "mushrooms".



Эксперимент проводился на симплексном множестве на квадратичной задаче.



Франк-Вульф с JAGUAR в стохастическом случае

Схемы аппроксимации:

- Двухточечная обратная связь (ДОС):

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

- Одноточечная обратная связь (ООС):

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

где e_i – i-ый базисный вектор, τ – параметр сглаживания.

Франк-Вульф с JAGUAR в стохастическом случае

Алгоритм 3 ФВ с JAGUAR в стохастическом случае

Вход: $x^0 \in Q$, $h^0 = g^0 = \widetilde{\nabla} f_\delta(x^0, \xi^{\pm}_{\overline{1,d}})$, γ_k, η_k, τ

for $k = 0, 1, 2, ..., N$ **do**

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

Сэмплируем 2 реализации ξ: ξ_k^+ и ξ_k^- независимо (в ДОС $\xi_k^+ = \xi_k^-$)

Считаем $\widetilde{\nabla}_{i_k} f_\delta(x^k, \xi_k^+, \xi_k^-) = \frac{f_\delta(x^k + \tau e_{i_k}, \xi_k^+) - f_\delta(x^k - \tau e_{i_k}, \xi_k^-)}{2\tau} e_{i_k}$

$h^{k+1} = h^k - \langle h^k, e_{i_k} \rangle e_{i_k} + \widetilde{\nabla}_{i_k} f_\delta(x^k, \xi_k^+, \xi_k^-)$

$\rho^k = h^k - d \cdot \langle h^k, e_{i_k} \rangle e_{i_k} + d \cdot \widetilde{\nabla}_{i_k} f_\delta(x^k, \xi_k^+, \xi_k^-)$

$g^{k+1} = (1 - \eta_k) g^k + \eta_k \rho^k$

$s^k = \arg\min\limits_{s \in Q} \left\langle s, g^{k+1} \right\rangle$

$x^{k+1} = x^k + \gamma_k (s^k - x^k)$

end for

Выход: x^{N+1}

Франк-Вульф с JAGUAR в стохастическом случае

Допущения:

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

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

2. Функция $f(x, \xi)$ выпукла на множестве Q:

$$\forall\, x, y \in Q : f(y, \xi) \geq f(x, \xi) + \langle \nabla f(x, \xi), y - x \rangle.$$

3. Ограниченность оракульного шума:

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

4. Ограниченность второго момента градиента:

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

5. Ограниченность второго момента оракула (для ООС):

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

Франк-Вульф с `JAGUAR` в стохастическом случае

Теорема (Богданов, 2023) 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 оценка на сходимость:

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

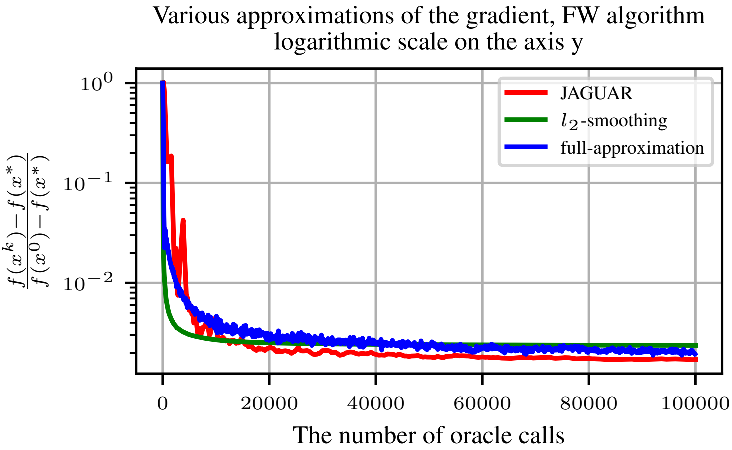
$$\left. + \sqrt{d}LD\gamma + \frac{\sqrt{d}\Delta D}{\gamma} \right)$$

Следствие Пусть ε определяет точность: $\mathbb{E}\left[f(x^N) - f(x^*)\right] \le \varepsilon$:

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

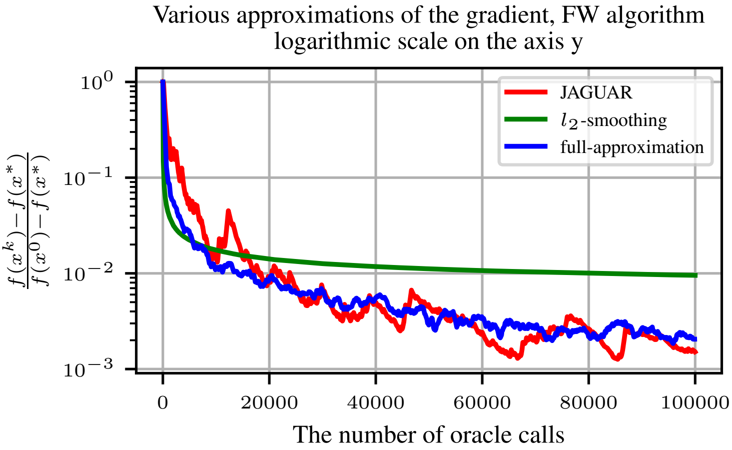
$$\gamma = \mathcal{O}\left(\frac{\varepsilon}{\sqrt{d}LD} \right), \quad \Delta = \mathcal{O}\left(\frac{\varepsilon^2}{dLD^2} \right).$$

Эксперимент проводился на $L2$-шаре на квадратичной задаче.



Эксперимент

Эксперимент проводился на симплексном множестве на квадратичной задаче.



Выносится на защиту

1. Для выпуклой задачи оптимизации на ограниченном выпуклом множестве с доступом только к зашумленному нулевому оракулу предложен и исследован метод `JAGUAR` для алгоритма Франка-Вульфа.
2. Доказаны теоретические оценки сходимости использования данного метода для нестохастического и стохастического случаев в алгоритме Франка-Вульфа. Показано теоретическое превосходство над l_2-сглаживанием и полной аппроксимаций.
3. Проведены вычислительные эксперименты, в которых сравнивается `JAGUAR`-аппроксимация с l_2-сглаживанием и полной аппроксимаций на различных задачах минимизации. Показано практическое превосходство.

Литература

- Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. Naval research logistics quarterly, 3(1-2):95–110, 1956.
- Darina Dvinskikh, Vladislav Tominin, Iaroslav Tominin, and Alexander Gasnikov. Noisy zeroth-order optimization for non-smooth saddle point problems. In International Conference on Mathematical Optimization Theory and Operations Research, pages 18–33. Springer, 2022.
- Anit Kumar Sahu, Manzil Zaheer, and Soummya Kar. Towards gradient free and projection free stochastic optimization. In The 22nd International Conference on Artificial Intelligence and Statistics, pages 3468–3477. PMLR, 2019.