Аппроксимации градиента с помощью оракула нулевого порядка и техники запоминания

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

Кафедра интеллектуальных систем ФПМИ МФТИ
Специализация: Интеллектуальный анализ данных
Направление: 03.03.01 Прикладные математика и физика

2024

Аппроксимация градента

Проблема: Ставится задача оптимизации с доступом только к зашумленному нулевому оракулу.

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

Решение: Предлагается аппроксимация градиента `JAGUAR`, которая использует технику запоминания.

Постановка задачи

Рассматривается две оптимизационные задачи на произвольном множестве $Q \subseteq \mathbb{R}^d$:

1. Детерминированная

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

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

2. Стохастическая

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

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

Аппроксимации `JAGUAR-d` и `JAGUAR-s`

Алгоритм 1

Вход: $x \in Q$; $h \in \mathbb{R}^d$; $\tau \in \mathbb{R}$
Сэмплируем $i \in \overline{1, d}$ равномерно и независимо
Считаем $\widetilde{\nabla}_i f_\delta(x) = \frac{f_\delta(x+\tau e_i) - f_\delta(x-\tau e_i)}{2\tau} e_i$
$h = h - \langle h, e_i \rangle e_i + \widetilde{\nabla}_i f_\delta(x)$
Выход: h

Алгоритм 2

Вход: $x \in Q$; $h, g \in \mathbb{R}^d$; $\tau \in \mathbb{R}$; $\eta \in [0, 1]$
Сэмплируем $i \in \overline{1, d}$ равномерно и независимо
Сэмплируем $\xi^+, \xi^- \sim \pi$ независимо (в случае ДОС $\xi^+ = \xi^-$)
Считаем $\widetilde{\nabla}_i f_\delta(x, \xi^\pm) = \frac{f_\delta(x+\tau e_i, \xi^+) - f_\delta(x-\tau e_i, \xi^-)}{2\tau} e_i$
$\rho = h - d \cdot \langle h, e_i \rangle e_i + d \cdot \widetilde{\nabla}_i f_\delta(x, \xi^\pm)$
$h = h - \langle h, e_i \rangle e_i + \widetilde{\nabla}_i f_\delta(x, \xi^\pm)$
$g = (1 - \eta) g + \eta \rho$
Выход: g, h

Обычный алгоритм Франка-Вульфа

1. Ограниченность множества Q:

$$\exists D > 0 : \forall x, y \in Q \hookrightarrow \|x - y\| \leq D.$$

2. Выпуклость множества Q:

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

3. Функция $f(x)$ выпукла на множестве Q:

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

Алгоритм 3

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

---

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

$$\exists L > 0 : \forall x, y \in Q \hookrightarrow \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| .$$

2. Шум оракула ограничен:

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

Алгоритм 4

1: **Вход:** $x^0 \in Q$, $h^0 = \widetilde{\nabla} f_\delta(x^0)$, γ_k, τ
2: **for** $k = 0, 1, 2, ..., N$ **do**
3: $\quad h^{k+1} = \texttt{JAGUAR-d}\left(x^k, h^k\right)$
4: $\quad s^k = \arg\min\limits_{x \in Q} \left\langle s, h^{k+1} \right\rangle$
5: $\quad x^{k+1} = x^k + \gamma_k (s^k - x^k)$
6: **end for**
7: **Вход:** x^{N+1}

Сходимость алгоритма Франка-Вульфа с `JAGUAR-d`

Теорема (Богданов А., 2023) При шаге оптимизатора:

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

получается оценка на сходимость:

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

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

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

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

Допущения для аппроксимации JAGUAR-s

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

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

2. Шум оракула ограничен:

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

3. Второй момент градиента оракула ограничен:

$$\exists \sigma_\nabla \geq 0 : \forall x \in Q \hookrightarrow \mathbb{E}\left[\|\nabla f(x, \xi) - \nabla f(x)\|^2\right] \leq \sigma_\nabla^2.$$

4. Второй момент оракула ограничен (для ООС):

$$\exists \sigma_f \geq 0 : \forall x \in Q \hookrightarrow \mathbb{E}\left[|f(x, \xi) - f(x)|^2\right] \leq \sigma_f^2.$$

Алгоритм Франка-Вульфа с `JAGUAR-s`

Алгоритм 5

Вход: $x^0 \in Q$, $h^0 = g^0 = \widetilde{\nabla} f_\delta(x^0)$, γ_k, η_k, τ
for $k = 0, 1, 2, ..., N$ **do**
    $g^{k+1}, h^{k+1} = \texttt{JAGUAR-s}\left(x^k, h^k, g^k, \eta_k\right)$
    $s^k = \arg\min\limits_{x \in Q} \left\langle s, g^{k+1} \right\rangle$
    $x^{k+1} = x^k + \gamma_k (s^k - x^k)$
end for
Вход: x^{N+1}

Сходимость алгоритма Франка-Вульфа с JAGUAR-s

Теорема (Богданов А., 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:

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

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

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

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

Постановка эксперимента

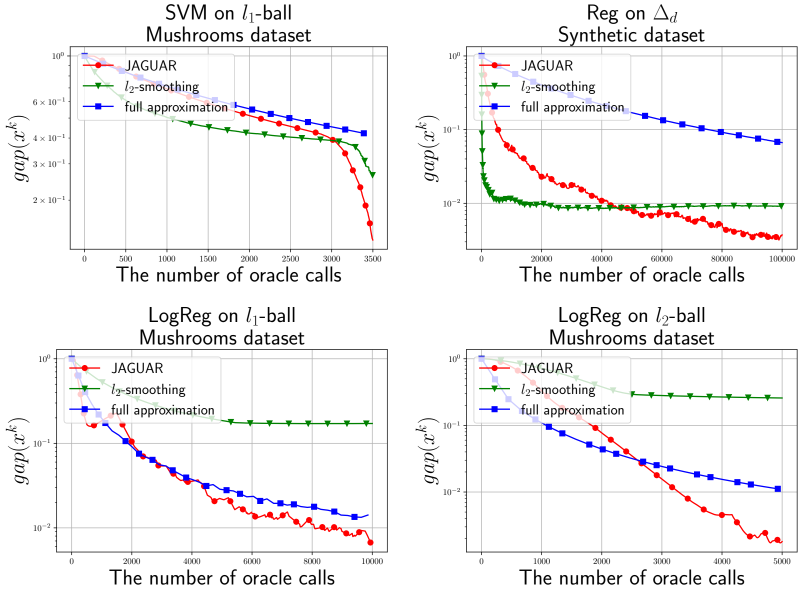
1. Модели: SVM, LogReg и Reg;
2. Множества: симплекс Δ_d, l_1-шар и l_2-шар;
3. Датасеты: MNIST, Mushrooms и синтетический;
4. Шум:
   4.1 Детерминированный: $f_\delta(x) = \text{round}(f(x), 5)$;
   4.2 Стохастический:

$$f_\delta(x, \xi) = f(x) + \xi^T x; \xi_i = clip(\tilde{\xi}_i, -1, 1), \tilde{\xi}_i \sim \mathcal{N}(0, 1);$$

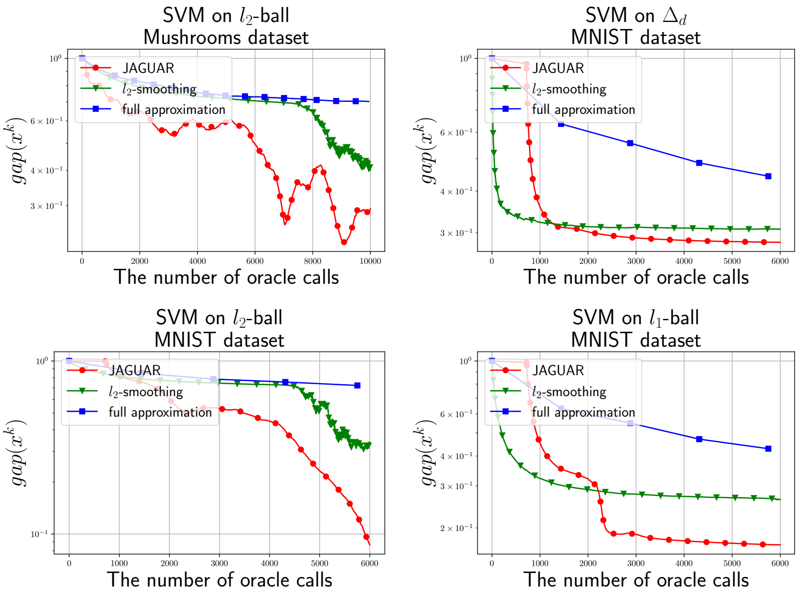
5. Сравнение: с l_2-*сглаживанием* и *полной аппроксимацией*
6. Метрика:

$$gap(x^k) = \max_{y \in Q} \langle \nabla f(x^k), x^k - y \rangle$$

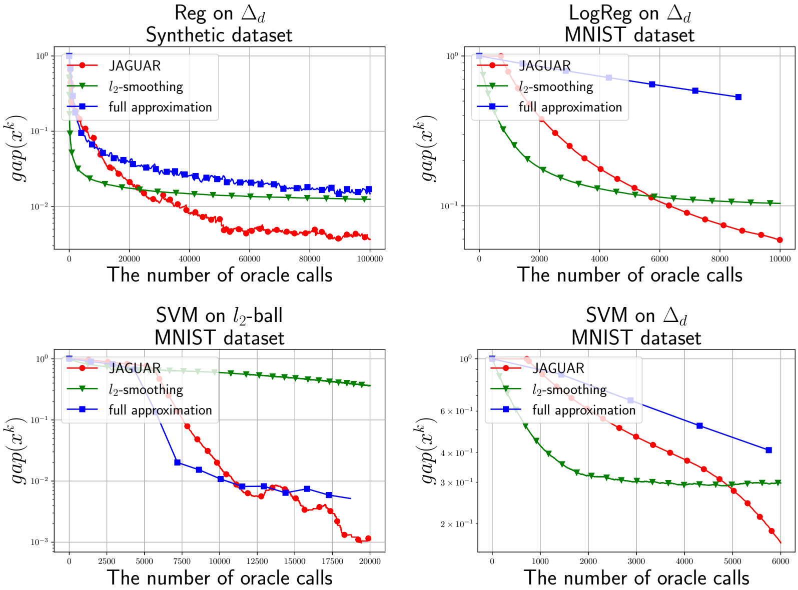
Эксперимент для детерминированного случая



Эксперимент для одноточечной обратной связи



Эксперимент для двухточечной обратной связи



Выносится на защиту

1. Предложен робастый алгоритм аппроксимация градиента `JAGUAR`, использующий $\mathcal{O}(1)$ вызовов оракула на каждой итерации;
2. Вычислительные эксперименты подтвердили теоретические оценки сходимости метода в детерминированных и стохастических условиях для алгоритма Франка-Вульфа, продемонстрировав его превосходство над *l_2-сглаживанием* и *полной аппроксимаций*.

Список работ автора по теме диплома

1. New Aspects of Black Box Conditional Gradient: Variance Reduction and One Point Feedback // Neural Information Processing Systems (на рецензировании).
2. Применение стохастической аппроксимации нулевого порядка с техникой запоминания в алгоритме Франка-Вульфа // 66-я Всероссийская научная конференция МФТИ, 2024.