

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

Александр Иванович Богданов

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

Московский физико-технический институт
ФПМИ
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, аппроксимирующий градиент, использующий $\mathcal{O}(1)$ вызовов оракула на каждой итерации.

Решение: Предлагается модификация JAGUAR, которая использует технику запоминания.

- ▶ Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. Naval research logistics quarterly, 3(1-2):95–110, 1956.



Постановка задачи

Рассматриваются 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

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

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 Франк-Вульф

Общие допущения:

1. Ограниченность множества Q :

$$\forall x, y \in Q \hookrightarrow \|x - y\|^2 \leq D^2.$$

2. Выпуклость множества Q :

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

Алгоритм 1 ФВ

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

Франк-Вульф с JAGUAR в нестохастическом случае

Схема аппроксимации:

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

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

Алгоритм 2 ФВ с JAGUAR в нестохастическом случае

- 1: **Вход:** $x^0 \in Q$, $h^0 = \tilde{\nabla} f_\delta(x^0)$, γ_k , τ
- 2: **for** $k = 0, 1, 2, \dots, N$ **do**
- 3: Сэмплируем $i \in \overline{1, d}$ независимо и равномерно
- 4: Считаем $\tilde{\nabla}_i f_\delta(x^k) = \frac{f_\delta(x^k + \tau e_i) - f_\delta(x^k - \tau e_i)}{2\tau} e_i$
- 5: $h^{k+1} = h^k - \langle h^k, e_i \rangle e_i + \tilde{\nabla}_i f_\delta(x^k)$
- 6: $s^k = \arg \min_{x \in Q} \langle s, h^{k+1} \rangle$
- 7: $x^{k+1} = x^k + \gamma_k (s^k - x^k)$
- 8: **end for**
- 9: **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:

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

Франк-Вульф с JAGUAR в нестохастическом случае

Теорема При шаге оптимизатора:

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

получается оценка на сходимость:

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

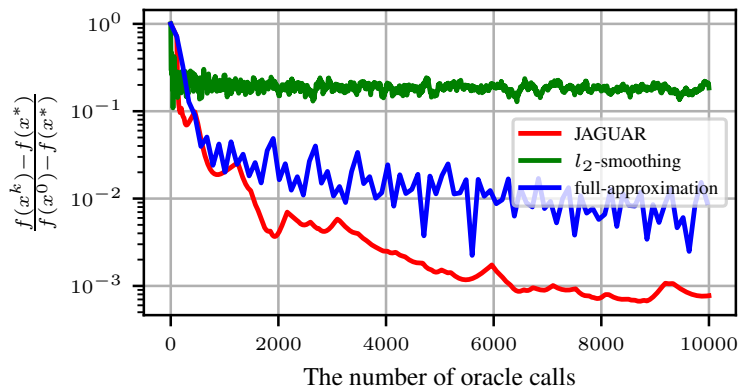
Следствие Пусть ε определяет точность: $\mathbb{E} [f(x^N) - f(x^*)] \leq \varepsilon$:

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

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

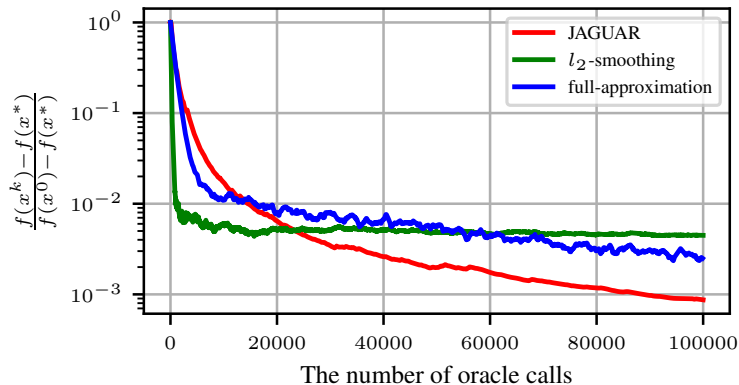
Эксперимент проводился на симплексном множестве на датасете "mushrooms".

Various approximations of the gradient, FW algorithm
logarithmic scale on the axis y



Эксперимент проводился на симплексном множестве на квадратичной задаче.

Various approximations of the gradient, FW algorithm
logarithmic scale on the axis y



Франк-Вульф с JAGUAR в стохастическом случае

Схемы аппроксимации:

- ▶ Двухточечная обратная связь:

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

- ▶ Одноточечная обратная связь:

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

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

Франк-Вульф с JAGUAR в стохастическом случае

Алгоритм 3 ФВ с JAGUAR в стохастическом случае

- 1: **Вход:** $x^0 \in Q$, $h^0 = g^0 = \tilde{\nabla} f_\delta(x^0, \xi_{1,d}^\pm)$, γ_k , η_k , τ
 - 2: **for** $k = 0, 1, 2, \dots, N$ **do**
 - 3: Сэмплируем $i \in \overline{1, d}$ независимо и равномерно
 - 4: Сэмплируем 2 реализации ξ : ξ_k^+ и ξ_k^- независимо (в ДОС $\xi_k^+ = \xi_k^-$)
 - 5: Считаем $\tilde{\nabla}_{i_k} f_\delta(x^k, \xi_k^+, \xi_k^-) = \frac{f_\delta(x^k + \tau e_{i_k}, \xi_k^+) - f_\delta(x^k - \tau e_{i_k}, \xi_k^-)}{2\tau} e_{i_k}$
 - 6: $h^{k+1} = h^k - \langle h^k, e_{i_k} \rangle e_{i_k} + \tilde{\nabla}_{i_k} f_\delta(x^k, \xi_k^+, \xi_k^-)$
 - 7: $\rho^k = h^k - d \cdot \langle h^k, e_{i_k} \rangle e_{i_k} + d \cdot \tilde{\nabla}_{i_k} f_\delta(x^k, \xi_k^+, \xi_k^-)$
 - 8: $g^{k+1} = (1 - \eta_k) g^k + \eta_k \rho^k$
 - 9: $s^k = \arg \min_{s \in Q} \langle s, g^{k+1} \rangle$
 - 10: $x^{k+1} = x^k + \gamma_k (s^k - x^k)$
 - 11: **end for**
 - 12: **Выход:** x^{N+1}
-

Франк-Вульф с JAGUAR в стохастическом случае

Допущения:

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

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

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:

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

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 При шаге оптимизатора и шаге импульса:

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

получается оценка на сходимость:

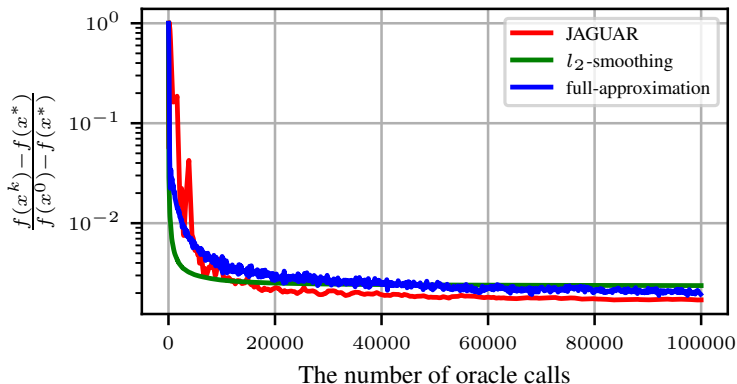
$$\mathbb{E} [f(x^N) - f(x^*)] = \mathcal{O} \left(\frac{LD^2 + d\sigma_f D/\gamma + d\sigma_{\nabla} D + \sqrt{d}(f(x^0) - f(x^*))}{(N + 8d^{3/2})^{1/3}} + \sqrt{d}LD\gamma + \frac{\sqrt{d}\Delta D}{\gamma} \right)$$

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

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

Эксперимент проводился на L_2 -шаре на квадратичной задаче.

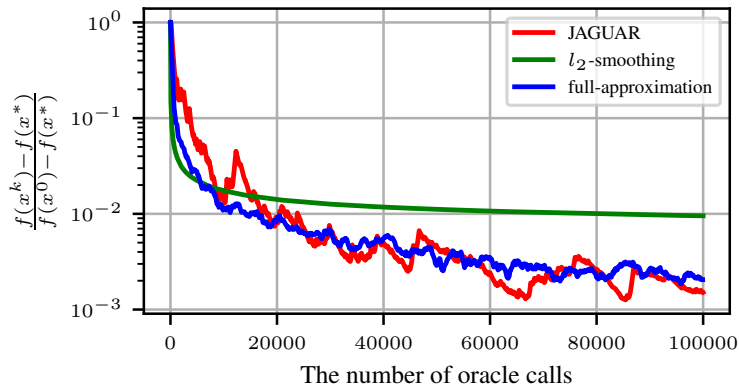
Various approximations of the gradient, FW algorithm
logarithmic scale on the axis y



Эксперимент

Эксперимент проводился на симплексном множестве на квадратичной задаче.

Various approximations of the gradient, FW algorithm
logarithmic scale on the axis y



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