

# Стохастический метод Ньютона с различным семплингом

Денис Швейкин

Московский физико-технический институт

*Курс:* Моя первая научная статья

*Эксперт:* Рустем Исламов

2023

# Цель исследования

## Задача

Использовать различные стратегии семплинга для стохастического метода Ньютона, чтобы получить наилучшую скорость сходимости

## Мотивация

- ▶ Задача минимизации функции потерь, имеющей структуру конечной суммы, возникает повсеместно в машинном обучении. Методы первого порядка для этой задачи хорошо изучены
- ▶ Методы второго порядка типа Ньютон лучше адаптируются к кривизне задачи и имеют квадратичную сходимость. Однако эти методы изучены менее подробно
- ▶ Применяются различные стратегии семплинга для стохастического варианта, поскольку для методов первого порядка их подбор ведет к улучшениям скорости сходимости

# Минимизируемая функция

## Структура функции потерь

$$\min_{x \in \mathbb{R}^d} \left[ f(x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n f_i(x) \right]$$

## Предположения

- ▶ Сильная выпуклость

$$f(x) \geq f(y) + \langle \nabla f, x - y \rangle + \frac{\mu}{2} \|x - y\|^2$$

- ▶ Липшицевы Гессианы

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq H \|x - y\|$$

**Initialize:** Задать начальные приближения  $w_1^0, w_2^0, \dots, w_n^0 \in \mathbb{R}^d$

**for**  $k = 0, 1, 2, \dots$  **do**

$$x^{k+1} = \left[ \frac{1}{n} \sum_{i=1}^n \nabla^2 f_i(w_i^k) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \nabla^2 f_i(w_i^k) w_i^k - \nabla f_i(w_i^k) \right]$$

Выбрать множество  $S^k \subseteq \{1, 2, \dots, n\}$  одной из стратегий семплинга

$$w_i^{k+1} = \begin{cases} x^{k+1} & i \in S^k \\ w_i^k & i \notin S^k \end{cases}$$

**end for**

## Определение

Семплингом называется  $\hat{S} : [n] \rightarrow 2^{[n]}$

## Метрика качества

Качество стратегий измеряется в скорости сходимости

# Стратегии семплинга

## Определение

Семплингом называется  $\hat{S} : [n] \rightarrow 2^{[n]}$

## Однородные стратегии

Любое множество размера  $j$  выбирается с одинаковой вероятностью,  $P(|\hat{S}| = j) = q_j$

- ▶  $\tau$ -nice семплинг  
 $q_\tau = 1$
- ▶  $\tau$ -независимый семплинг  
 $q_k = \binom{n}{k} c_k, k = 1, 2, \dots, \tau$
- ▶  $(\tau, p_b)$ -биномиальный семплинг  
 $q_k = \binom{\tau}{k} p_b^k (1 - p_b)^{\tau - k}, k \leq \tau$

## Независимые стратегии

Теперь каждый объект  $i$  включается в множество  $S$  независимо с вероятностью  $p_i$ .

## Importance sampling

Каждая  $f_i$  имеет свою константу Липшица  $L_i$

$$p_i = \frac{L_i}{\sum_{i=1}^n L_i}$$

# Результаты: оценки скорости сходимости

## Средний квадрат невязки

$$\mathcal{W}^k \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \|w_i^k - x^*\|^2$$

## Однородные стратегии

Пусть  $\rho = \frac{\mathbb{E}[|\hat{S}|]}{n}$ . Тогда  $\mathbb{E}_k[\mathcal{W}^{k+1}] \leq (1 - \frac{3}{4}\rho) \mathcal{W}^k$

Результат получается одинаковым для любой стратегии. Поэтому при равных  $\rho$  различие будет только в удобстве реализаций стратегий

## Независимые стратегии

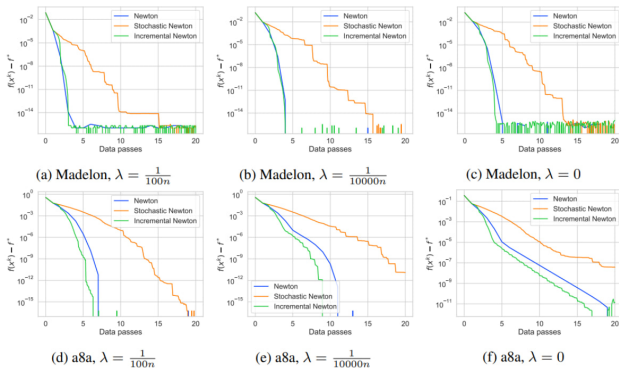
Пусть фиксировано матожидание размера батча  $\mathbb{E}_k[|\hat{S}^k|] = \tau$ . Тогда  $\mathbb{E}_k[\mathcal{W}^{k+1}]$  будет минимально, если некоторые  $p_i$  равны 1, возможно одно  $p_i \in (0, 1)$ , а остальные равны 0.

## Importance sampling

Предстоит доработать

# Вычислительный эксперимент

График сходимости для однородного семплинга, полученный авторами алгоритма



Видно, что соотношение скоростей сходимости стохастического и детерминированного алгоритмов получается примерно одинаковым на разных задачах. **Предстоит доработать**

# Заключение

- ▶ Рассмотрены различные стратегии семплинга для стохастического метода Ньютона - однородные и независимые
- ▶ Получены линейные оценки скорости сходимости для однородных стратегий в общем случае
- ▶ Для независимой стратегии importance sampling - **предстоит доработать**