Сходимость с оценкой вероятностей больших отклонений для задач выпуклой оптимизации и седловых задач в условиях повышенной гладкости

Выпускная квалификационная работа бакалавра

Рубцов Денис Николаевич
Научный руководитель: д.ф.-м.н. А. В. Гасников

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

2025

Цели

1. разработать быстрые алгоритмы для решения задач выпуклой стохастической оптимизации, обеспечивающие сходимость с высокой вероятностью
2. исследовать эти алгоритмы с помощью теоретического анализа и вычислительных экспериментов

Сходимость с высокой вероятностью

Задача стохастической оптимизации

$$\min_{x \in \mathbb{R}^d} f(x) := \mathbb{E} f(x, \xi),\ \ \xi \sim \mathcal{P}$$

Как правило, результатом стохастических градиентных методов является точка x_ε такая, что

$$\mathbb{E} f(x_\varepsilon) - \min f \leq \varepsilon$$

Мы рассматриваем алгоритмы, результатом которых являются точки $x_{\varepsilon,p}$, удовлетворяющие условию

$$P\{f(x_{\varepsilon,p}) - \min f \leq \varepsilon\} \geq 1 - p$$

$$\Updownarrow$$

$$P\{f(x_{\varepsilon,p}) - \min f \geq \varepsilon\} \leq p$$

Наивный подход

Наивное решение

Если решить задачу $\mathbb{E} f(x_\varepsilon) - \min f \le p\varepsilon$, то желаемое неравенство $P\{f(x_{\varepsilon,p}) - \min f \le \varepsilon\} \ge 1 - p$ следует автоматически по неравенству Маркова.

Проблема

Сложность решения задачи сходимости по матожиданию, обычно, порядка $\mathcal{O}(\frac{1}{\varepsilon})$. Тогда сложность наивного решения задачи сходимости с высокой вероятностью $\mathcal{O}(\frac{1}{p\varepsilon})$.

Хочется

уменьшить множитель $\frac{1}{p}$ до $\log(\frac{1}{p})$

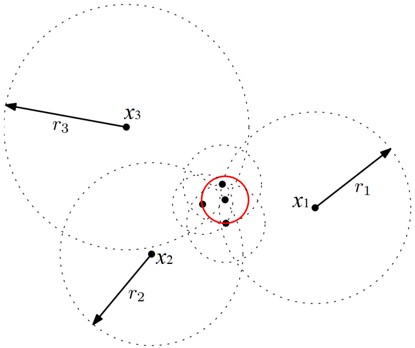
Ключевая идея №1 : Robust distance estimation (RDE)

Пусть имеется m точек $x_1, \dots, x_m$, для которых $\mathbb{E}x_i - x^* \leq \frac{\varepsilon}{3}$, т.е. $P[||x - x^*|| \leq \varepsilon] \geq \frac{2}{3}$. Тогда среди этих точек можно выбрать такую x_{i^*}, вокруг которой «группируются» остальные точки.

Theorem (Nemirovskiy, Yudin, 1983 [a])

Точка x_{i^*}, возвращаемая алгоритмом RDE удовлетворяет условию

$$P(||x_{i^*} - x^*|| \leq 3\varepsilon) \geq 1 - e^{-\frac{m}{18}}$$



[a] *Nemirovskij A.S., Yudin D.B.* Problem complexity and method efficiency in optimization. - 1983.

Применение RDE для обеспечения сходимости с высокой вероятностью: описание подхода

$$\frac{\mu}{2}||x - x^*||^2 \le f(x) - f(x^*) \le \frac{L}{2}||x - x^*||^2$$

Пусть мы имеем точки x_i $(i = 1, ..., m)$ такие, что

$$\mathbb{E}f(x_i) - \min f \le \frac{\varepsilon}{3} \xRightarrow{\text{неравенство Маркова}}$$

$$P(f(x_i) - f^* \le \varepsilon) \ge \frac{2}{3} \xRightarrow{\text{сильная выпуклость}}$$

$$P\left(||x_i - x^*|| < \sqrt{\frac{2\varepsilon}{\mu}} =: \delta\right) \ge \frac{2}{3} \xRightarrow{\text{RDE}}$$

$$P(||x_{i^*} - x^*|| < 3\delta) \ge 1 - e^{-\frac{m}{18}} \xRightarrow{\text{гладкость}}$$

$$P\left(f(x_{i^*}) - f^* \le 9\frac{L}{\mu}\varepsilon\right) \ge 1 - e^{-\frac{m}{18}}$$

Применение RDE для обеспечения сходимости с высокой вероятностью: проблема

$$\mathbb{E}f(x_i) - \min f \leq \frac{\varepsilon}{3} \overset{\dots}{\Longrightarrow}$$

$$P(f(x_{i^*}) - f^* \leq 9\frac{L}{\mu}\varepsilon) \geq 1 - e^{-\frac{m}{18}} = 1 - p$$

При $m \sim \log(1/p)$ получаем логарифмическую зависимость от p сложности решения задачи сходимости с высокой вероятностью. Однако, точность решения ухудшается в κ раз, где число обусловленности $\kappa = \frac{L}{\mu} \gg 1$ может быть достаточно большим!

Ключевая идея №2: Проксимальный метод

Как бороться с зависимостью от числа обусловленности κ?

Последовательно решать регуляризованные подзадачи:

$$\bar{x}_{i+1} := \arg\min_x \left\{ f^i(x) := f(x) + \frac{\lambda_i}{2}\|x - x_i\|^2 \right\}$$

1) Выбираем параметры растущими геометрически: $\lambda_i = \mu \cdot 2^i$.
2) **Эффект:** Новая функция f^i становится хорошо обусловленной.

$$\kappa_i = \frac{L + \lambda_i}{\mu + \lambda_i} = \mathcal{O}(1), \;\; i > \log \frac{L}{\mu}$$

3) При этом решение новых задач будет приближенным решением основной задачи

$$f(x_{j+1}) - f^* \le (f^j(x_{j+1}) - f^j(\bar{x}_{j+1})) + \sum_{i=0}^{j} \frac{\lambda_j}{2}||\bar{x}_i - x_i||^2.$$

Основной алгоритм и его сложность

Алгоритм `proxBoost`

Комбинация RDE и проксимального метода позволяет получить решение x_T такое, что:

$$P\{f(x_T) - \min f \leq \varepsilon\} \geq 1 - p$$

Об оракульной сложности

Все дальнейшие результаты будут выражены в терминах сложности $\mathcal{C}_{\mathcal{M}}(f, \varepsilon)$ оракула $\mathcal{M}(f, \varepsilon)$, возвращающего точку x_ε такую, что $\mathbb{E}f(x_\varepsilon) - \min f \leq \varepsilon$.

Сложность алгоритма proxBoost

Theorem (Davis, Drusvyatskiy, 2021[1])

Сложность алгоритма `proxBoost`*, решающего задачу* $P\{f(x_{\varepsilon,p}) - \min f \le \varepsilon\} \ge 1 - p$*, для* μ*-сильно выпуклых* L*-гладких функций:*

$$\mathcal{O}\left(\underbrace{\log\left(\frac{\log\kappa}{p}\right)\log(\kappa)}_{\textit{Доплата}}\cdot\mathcal{C}_{\mathcal{M}}\left(f, \frac{\epsilon}{\log\kappa}\right)\right)$$

Вывод: Доплата за сходимость с высокой вероятностью является логарифмической по параметрам задачи.

[1] From low probability to high confidence in stochastic convex optimization / D.Davis et al // JMLR - 2021

Метод сглаживания

На гладкий и негладкий случай можно смотреть единообразно:
Функция f - (L, γ)-гладкая, если

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 + \gamma. \qquad (1)$$

Если функция негладкая, но $M-$липшицева, то есть

$$\|\nabla f(y) - \nabla f(x)\| \leq M,$$

то (1) выполняется при

$$L = \frac{M^2}{2\gamma}$$

.

Сложность алгоритма в негладком случае

Theorem (Рубцов, 2025)

Сложность обобщенного алгоритма proxBoost для ***негладких*** *μ-сильно выпуклых функций, решающего задачу $P\{f(x_{\varepsilon,p}) - \min f \leq \varepsilon\} \geq 1 - p$:*

$$\mathcal{O}\left(\log\left(\frac{1}{p}\right)\log\left(\frac{M^2}{\mu\varepsilon}\right)\cdot\mathcal{C}_{\mathcal{M}}\left(f, \frac{\epsilon}{\log\left(\frac{M^2}{\mu\varepsilon}\right)}\right)\right).$$

Метод регуляризации для решения не сильно выпуклых задач

Theorem

Пусть функция $f(x)$ выпукла. Будем решать задачу минимизации функции

$$f^{\mu}(x) = f(x) + \frac{\mu}{2}||x - x_0||^2,$$

где $\mu \sim \frac{\varepsilon}{R^2}, R = ||x^* - x_0||$.
Пусть мы нашли точку x такую, что

$$f^{\mu}(x) - \min f^{\mu} < \frac{\varepsilon}{2}$$

Тогда

$$f(x) - \min f < \varepsilon$$

Сложность алгоритма proxBoost в выпуклом случае

Theorem (Рубцов, 2024)

Сложность обобщенного алгоритма proxBoost для ***выпуклых*** *L-гладких функций, решающего задачу $P\{f(x_{\varepsilon,p}) - \min f \le \varepsilon\} \ge 1 - p$:*

$$\mathcal{O}\left(\log(\frac{1}{p})\log\frac{LR^2}{\varepsilon}\cdot\mathcal{C}_{\mathcal{M}}(f, \frac{\varepsilon}{\log\frac{LR^2}{\varepsilon}})\right).$$

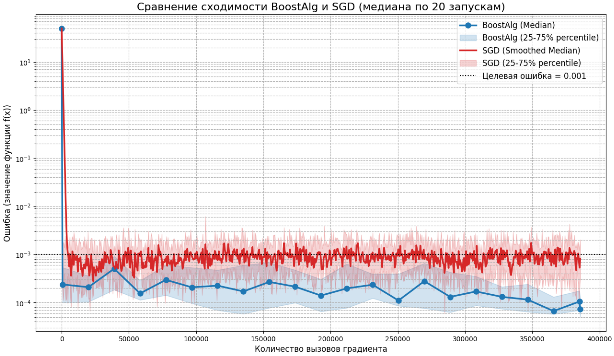
Обобщение на разные классы задач

Важно отметить, что предложенный подход является «оберткой» над методами и работает для разных классов функций.

Класс функции	Итоговая оракульная сложность
сильно выпуклая, гладкая	$\log\left(\frac{1}{p}\right)\log(\kappa)\cdot\mathcal{C}_{\mathcal{M}}\left(f, \frac{\epsilon}{\log\kappa}\right)$
сильно выпуклая, негладкая	$\log\left(\frac{1}{p}\right)\log\left(\frac{M^2}{\mu\varepsilon}\right)\cdot\mathcal{C}_{\mathcal{M}}\left(f, \frac{\epsilon}{\log\left(\frac{M^2}{\mu\varepsilon}\right)}\right)$
выпуклая, гладкая	$\log(\frac{1}{p})\log\frac{LR^2}{\varepsilon}\cdot\mathcal{C}_{\mathcal{M}}(f, \frac{\varepsilon}{\log\frac{LR^2}{\varepsilon}})$
выпуклая, негладкая	$\log\left(\frac{1}{p}\right)\log\left(\frac{MR}{\varepsilon}\right)\cdot\mathcal{C}_{\mathcal{M}}\left(f, \frac{\epsilon}{\log\left(\frac{MR}{\varepsilon}\right)}\right)$

Результаты: График сходимости

SGD vs ProxBoost



Ошибка $f(x) - f^*$ от количества вызовов оракула. Сплошная линия — медиана, область — 25-75% квантили.

SGD, обернутый в `proxBoost` (синий цвет), демонстрирует более надежную сходимость, чем SGD (красный).

Заключение

Выносится на защиту

1. **Универсальный мета-алгоритм** для преобразования сходимости по матожиданию в сходимость с высокой вероятностью.
2. **Новые оценки сложности** для широкого класса задач, включая негладкие и не сильно выпуклые случаи.
3. **Экспериментальное подтверждение** эффективности и надежности предложенного подхода.