Исследование методов высших порядков в выпуклой оптимизации

Евсей Обжерин

Московский физико-технический институт (ФПМИ)

11 марта 2025 г.

План

1. Постановка задачи
2. Методы второго порядка: кубическая регуляризация (CRN)
3. Методы третьего порядка: тензорный подход
4. Quasi-Newton методы (приближённая аппроксимация гессиана) и ускорение
5. Шаг методов высших порядков
6. (L0, L1)-гладкость
7. Обобщение (L0, L1)-условия на методы высших порядков
8. Итог и направления развития
9. Источники

Постановка задачи

Рассматриваем задачу выпуклой оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x),$$

где f — выпуклая и достаточно гладкая функция (например, $f \in C^2$).
Возможны дополнительные условия:

- $\|\nabla f(x) - \nabla f(y)\| \le L_1 \|x - y\|$,
- $\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L_2 \|x - y\|$.

Также используются условия (L0, L1)-гладкости (см. [2]).

Кубическая регуляризация (CRN), $p = 2$

Шаг метода (см. [1]):

$$x_{k+1} = \arg\min_s \left\{ \langle \nabla f(x_k), s \rangle + \frac{1}{2} s^\top \nabla^2 f(x_k) s + \frac{M}{6} \|s\|^3 \right\}.$$

При $M \geq L_2$ метод обеспечивает глобальную сходимость с числом итераций $O(\varepsilon^{-1/2})$.

Методы третьего порядка (тензорный подход, [3])

Разложение Тейлора порядка 3:

$$\Phi_{x,3}(y) = f(x)+\langle\nabla f(x), y-x\rangle+\frac{1}{2}(y-x)^\top\nabla^2 f(x)(y-x)+\frac{1}{6}D^3 f(x)[y-x]^3.$$

Tensor-метод:

$$x_{k+1} = \arg\min_y\left\{\Phi_{x_k,3}(y) + \frac{M}{24}\|y-x_k\|^4\right\}.$$

Quasi-Newton методы и ускорение ([1])

Идея: приближённая аппроксимация гессиана

$$\nabla^2 f(x_k) \approx B_k \quad \Longrightarrow \quad [\nabla^2 f(x_k)]^{-1} \approx H_k,$$

где B_k или H_k уточняются по накопленной информации. **Общее QN-обновление:**

$$x_{k+1} = x_k - H_k \nabla f(x_k).$$

Cubic QN: $O(k^{-1})$.

Accelerated Cubic QN: $O(k^{-2})$ (иногда $O(k^{-3})$).

Шаг методов высших порядков

- **CRN ($p = 2$), [1]:**

$$x_{k+1} = \arg\min_s \Big\{ \langle \nabla f(x_k), s \rangle + \tfrac{1}{2} s^\top \nabla^2 f(x_k) s + \tfrac{M}{6} \|s\|^3 \Big\}.$$

- **Tensor-метод ($p = 3$), [3]:**

$$x_{k+1} = \arg\min_y \Big\{ \Phi_{x_k,3}(y) + \tfrac{M}{24} \|y - x_k\|^4 \Big\}.$$

- **QN-обновление, [1]:**

$$x_{k+1} = x_k - H_k \nabla f(x_k).$$

(L0, L1)-гладкость ([2])

Определение:

$$\|\nabla f(y) - \nabla f(x)\| \leq \left(L_0 + L_1 \|\nabla f(x)\|\right) \|y - x\|.$$

Следствие: Если $\|\nabla f(x_k)\| \geq \frac{L_0}{L_1}$, то GD-модификации сходятся линейно.

Обобщение (L0, L1)-условия на методы высших порядков

Идея:

$$\|D^p f(x) - D^p f(y)\| \leq \left(L_0 + L_1 \|D^{p-1} f(x)\|\right) \|y - x\|.$$

Цель:

- Сохранить режим линейной или суперлинейной сходимости.
- Адаптивно переключаться, когда нормы высших производных малы.

Итог и направления развития

- CRN обеспечивает сходимость $O(\varepsilon^{-1/2})$ (см. [1]).
- QN-аппроксимация с кубической регуляризацией даёт $O(k^{-1})$ и $O(k^{-2})$ (ускоренный вариант).
- (L0, L1)-гладкость демонстрирует линейный режим для GD (по [2]).
- **Обобщение на высшие порядки**: исследовать аналогичные свойства для $p \geq 2$ (тензорные методы, см. [3]).

Источники

[1] Kamzolov, D., Ziu, K., Agafonov, A., Takáč, M.
"Cubic Regularization is the Key! The First Accelerated Quasi-Newton Method with a Global Convergence Rate of $O(k^{-2})$ *for Convex Functions"*
arXiv:2302.04987.pdf.

[2] Lobanov, A., Gasnikov, A., Gorbunov, E., Takáč, M.
"Linear Convergence Rate in Convex Setup is Possible! Gradient Descent Method Variants under (L0, L1)-Smoothness"
arXiv:2412.17050.pdf.

[3] *"OPTAMI: Global Superlinear Convergence of High-Order Methods"*
OpenReview: Cpr6Wv2tfr (или arXiv:3177_OPTAMI_Global_Superlinear.pdf).