Исследования вариаций субградиентного метода на классе слабо выпуклых функций

Федор Сергеевич Стонякин[1], **Есения Андреевна Лушко**[2]

[1] Московский физико-технический институт, e-mail: fedyor@mail.ru

[2] Московский физико-технический институт, e-mail: lushko.ea@phystech.edu

На практике сейчас возникает много задач, где встречаются функции, которые не являются выпуклыми, но удовлетворяют свойствам слабой выпуклости и острого минимума. Среди самых известных из них — задача восстановления фазы или задача восстановления матрицы. Такого типа задачи имеют множество приложений в обработке речи, визуализации, кристаллографии и т.д. [1, 2, 3]. Известно [1], что на классе слабо выпуклых липшицевых задач с острым минимумом субградиентный метод (субградиент можно понимать, например, в смысле Кларка) с шагом Б.Т. Поляка может локально сходиться со скоростью геометрической прогрессии. Цель нашей работы — обобщить этот результат на случай доступности неточной информации о субградиенте и целевой функции. В некотором смысле мы распространяем на слабо выпуклые задачи известный для выпуклых задач подход [4]. Как показано в [4], для неускоренного метода с неточным оракулом параметры неточности не накаливаются в оценке качества выдаваемого решения.

Исследование состоит в том, чтобы выяснить, как неточность информации о субградиенте и/или целевой функции может влиять на качество выдаваемого субградиентным методом приближённого решения для задач минимизации слабо выпуклых функций с острым минимумом. Точнее, рассматриваются модификации субградиентного метода с шагом Б.Т. Поляка (Proj_Q — евклидова проекция точки на множество Q):

$$x_{k+1} := \mathrm{Proj}_Q\left(x_k - h_k \nabla f\left(x_k\right)\right), \quad k = 0, 1, 2, \ldots$$

где норма $\|\cdot\|_2$ евклидова (известно [1], что в достаточно малой окрестности множества минимумом не существует стационарных точек, не доставляющих минимум)

$$h_k := \frac{f\left(x_k\right) - f^*}{\|\nabla f\left(x_k\right)\|_2^2} \left(\text{при } \nabla f\left(x_k\right) \neq 0\right).$$

Выделяются две вариации метода при условиях:

1. На каждом шаге метода доступен субградиент целевой функции с аддитивной неточностью:
$$\left\|\widetilde{\nabla} f\left(x_i\right) - \nabla f\left(x_i\right)\right\|_2 \leq \Delta,$$
где Δ положительно, но достаточно мало.
2. На каждом шаге погрешность присутствует как для значения целевой функции, так и для значении субградиента:
$$\left\|\widetilde{\nabla} f\left(x_i\right) - \nabla f\left(x_i\right)\right\|_2 \leq \Delta, \quad \left|\widetilde{f(x_k)} - f(x_k)\right| \leq \delta,$$

где $\delta, \Delta > 0$, но достаточно малые.

Цель — выяснить, как влияют значения δ и $\Delta > 0$ на качество выдаваемого решения. При этом в случае молой нормы субградиента целевой функции его погрешность может существенно повлиять на качество решения. С целью преодоления этой проблемы был использован подход, похожий на градиентный клиппинг [5]. Так, для постановки 1 выше (с аддитивной неточностью градиента) имеем (M — константа Липшица f):

(a) Если $\left\|\widetilde{\nabla} f\left(x_k\right)\right\|_2 \leq M\Delta^{\lambda}$ (где $\lambda \in (0,1)$), то действуем по такому алгоритму:

$$x_{k+1} := \mathrm{Proj}_Q\left(x_k - \frac{f\left(x_k\right) - f^*}{M^2}\widetilde{\nabla} f\left(x_k\right)\right).$$

(b) Если $\left\|\widetilde{\nabla} f\left(x_k\right)\right\|_2 > M\Delta^{\lambda}$ (где $\lambda \in (0,1)$), то действуем по такому алгоритму:

$$x_{k+1} := \mathrm{Proj}_Q\left(x_k - \frac{f\left(x_k\right) - f^*}{\|\widetilde{\nabla} f\left(x_k\right)\|_2^2}\widetilde{\nabla} f\left(x_k\right)\right).$$

Доказано, что в таком случае случае после достаточно большого количества итераций можно гарантировать

$$\|x_N - x_*\|_2^2 = O\left(\Delta^{2-2\lambda}\right)$$

для достаточно большого N. Подобно [6] обосновывается близость данной оценки к оптимальной на классе слабо выпуклых липшицевых задач с острым минимумом.

Список литературы

[1] Davis D., Drusvyatskiy D., Kellie M., Paquette C.: Subgradient methods for sharp weakly convex functions // Journal of Optimization Theory and Applications — 2018.— Vol. 179, P. 962– 982

[2] Damek Davis and Dmitriy Drusvyatskiy and Courtney Paquette: The nonsmooth landscape of phase retrieval // IMA Journal of Numerical Analysis – 2020, Vol. 40, P. 2652–2695

[3] Li, Xiao and Zhu, Zhihui and Man-Cho So, Anthony and Vidal, René: Nonconvex Robust Low-Rank Matrix Recovery // SIAM Journal on Optimization – 2020, Vol. 30, P.660-686

[4] Devolder, O., Glineur, F. Nesterov, Y.: First-order methods of smooth convex optimization with inexact oracle. // Math. Program. 146, 37–75 (2014).

[5] Jingzhao Zhang, Tianxing He: Why gradient clipping accelerates training

[6] Fedyor Stonyakin, Ilya Kuruzov, Boris Polyak: Stopping rules for gradient methods for non-convex problems with additive noise in gradient