

ИССЛЕДОВАНИЕ УГЛОВОГО ШУМА В МЕТОДЕ LION

*Елистратов Семен Юрьевич*¹

*Подivilов Андрей Евгеньевич*²

*Южаков Тимофей Алексеевич*³

1: Аспирант, факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

2: Аспирант, Санкт-Петербургский государственный университет,

Санкт-Петербург, Россия

3: Constructor University, Бремен, Германия

E-mail: semenelist@gmail.com, andrey.podivilov@gmail.com,

tiuzhakov@constructor.university

*Научный руководитель — Ветров Дмитрий Петрович, профессор,
Constructor University, Бремен, Германия.*

Метод оптимизации Lion в последние годы привлёк значительное внимание благодаря высокому качеству работы при обучении глубоких нейронных сетей [1]. Несмотря на это, механизмы, определяющие его поведение и объясняющие его эффективность, до сих пор изучены не полностью [4]. В связи с этим представляет интерес изучение, как именно правила обновления весов в Lion влияют на свойства процесса обучения нейронных сетей и чем данный метод отличается от более стандартных методов оптимизации [3,4,5].

Метод Lion задаётся соотношениями

$$\begin{aligned} m_{t+1} &= \beta_2 m_t + (1 - \beta_2) \nabla L(w_t), \\ w_{t+1} &= (1 - \eta \lambda) w_t - \eta \operatorname{sign}(\beta_1 m_{t+1} + (1 - \beta_1) \nabla L(w_t)). \end{aligned} \quad (1)$$

Здесь w_t — параметры модели, m_t — вектор момента, η — шаг обучения, λ — коэффициент регуляризации. В отличие от традиционных методов, направление обновления в Lion определяется не самим градиентом, а его покомпонентным знаком [1,3]. Такое преобразование меняет направление шага и вносит в динамику метода специфическое возмущение, которое в работе интерпретируется как угловой шум.

Цель работы состоит в исследовании роли этого эффекта. Для этого рассматривается угол θ между градиентом и его sign-преобразованием. Показывается, что в модели со случайным выбором базиса данный угол в пространствах большой размерности концентрируется около постоянной величины. В прикладных задачах

его значение определяется структурой градиента и зависит от шага обучения η : вычислительные эксперименты показывают, что при увеличении η угол поворота θ , возникающий в Lion, возрастает.

Для выделения этого механизма вводится вспомогательный метод Enim, в котором sign-обновление заменяется случайным поворотом направления на фиксированный угол:

$$\text{sign} \rightarrow \text{LNI} \quad (2)$$

Здесь оператор LNI осуществляет случайный поворот нормированного направления обновления. Такое построение позволяет отделить эффект углового возмущения от остальных особенностей Lion. Сравнение Lion и Enim показывает, что при согласовании угла поворота оба метода демонстрируют близкое качество, а в ряде режимов Enim оказывается более устойчивым к выбору шага обучения. Это указывает на существенную роль углового шума в преимуществах Lion.

Кроме того, теоретический анализ позволил показать, как именно угловой шум влияет на свойства аттракторов: наряду с кривизной вдоль градиента начинают учитываться и ортогональные направления. Тем самым предлагается геометрическая интерпретация Lion, объясняющая его устойчивость и склонность к выбору более пологих областей ландшафта функции потерь.

Литература

1. Chen X., Liang C., Huang D., Real E., Wang K., Pham H., Dong X., Luong T., Hsieh C.-J., Lu Y., Le Q. V. Symbolic Discovery of Optimization Algorithms // Advances in Neural Information Processing Systems. 2023. Vol. 36.
2. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization // In 3rd International Conference on Learning Representations (ICLR). 2015.
3. Balles L., Pedregosa F., Le Roux N. The Geometry of Sign Gradient Descent // CoRR. abs/2002.08056. 2020.
4. Dong Y., Li H., Lin Z. Convergence Rate Analysis of LION // CoRR. abs/2411.07724. 2024.
5. Smith S. L., Dherin B., Barrett D. G. T., DeS. On the Origin of Implicit Regularization in Stochastic Gradient Descent // In 9th International Conference on Learning Representations (ICLR). 2021.