Improving Gradient Descent Optimization through Guided Exploration of Loss Landscape

Ayushya Pare
Mülheim, Germany
ayushya.pare9@gmail.com

***Abstract*—In machine learning, optimizing models in high-dimensional, non-convex loss landscapes presents significant challenges due to the prevalence of local minima and saddle points which hinder convergence. Gradient descent based optimization algorithms often get stuck in suboptimal regions of the loss function and converge poorly. This paper presents Guided Exploration (GE) algorithm that combines local and global search strategies to navigate these suboptimal regions by leveraging information about the loss landscape's topology. It employs guided exploration to identify promising regions in the loss function landscape and then circumnavigates these regions, consequently improving convergence and boosting performance. Experiments show an improved accuracy by 4.39% over SGD and 1.57% over Adam algorithms respectively on fashion-MNIST dataset.**

***Index Terms*—Machine learning, Optimization Algorithms, Gradient Descent, Convergence Improvement**

I. Introduction

Optimizing machine learning models can become challenging due to high-dimensional, non-convex loss landscapes filled with local minima and saddle points that hinder optimization algorithms to locate the true global minimum. Popular optimization algorithms such as Stochastic Gradient Descent (SGD) and Adam often encounter poor convergence in non-linear loss landscapes. This paper introduces a novel optimization algorithm, called the Guided Exploration (GE) algorithm, that combines local and global search approaches to identify promising regions within the loss landscape, and then employ gradient descent approach for efficient navigation and convergence to the optimum. Moreover, this algorithm can be integrated with any gradient descent method, such as SGD or Adam, thereby enhancing their performance.

Major contributions of this paper include:

- An algorithm that improves gradient descent optimization through guided exploration to better navigate complex loss landscapes.
- Experiments demonstrating significant improvement in performance on benchmark datasets. (Improved model accuracy by 4.39% over SGD and 1.57% over Adam on fashion-MNIST).
- It can be employed as an add-on with any gradient descent algorithm.

II. Background and Related Work

The optimization of neural networks critically hinges on the characteristics of the loss landscape, which significantly influence both training dynamics and generalization [1]. The dynamic adaptation of loss functions, as explored by Raymond et al. [2], provides a method for real-time adjustments during training, enhancing model accuracy and training efficiency [3]. Moreover, considering the curvature of the loss landscape can lead to more efficient optimization strategies [4]. Notably, sharp minima have been shown to potentially enhance generalization under certain conditions [5]. These insights make way for the development of a new algorithm that integrates curvature-aware optimization with adaptive loss function updates to navigate the loss landscape more effectively and improve generalization and convergence speeds.

III. Proposed Algorithm

This paper introduces a novel optimization algorithm designed to improve convergence of gradient descent based algorithms through guided exploration and navigation of the loss function landscape. The algorithm integrates local and global search strategies to identify promising regions and then navigates these areas using gradient-based methods. The process involves three main steps:

1) **Sampling the Landscape:** Initially, the algorithm samples the vicinity of the current parameters using strategies like random perturbations to explore the local topology and to build a detailed map of the landscape.
2) **Identifying Key Features:** With the landscape mapped, the algorithm identifies key features such as local minima and saddle points using methods

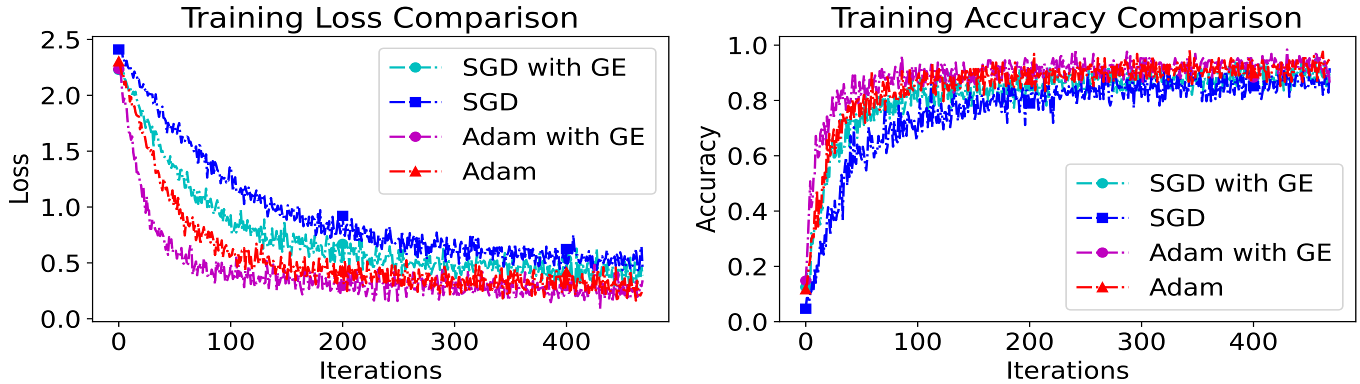

Fig. 1. Comparison of model accuracy and loss trajectories for the Guided Exploration (GE), SGD, and Adam algorithm.

like Hessian matrix analysis or finite differences to determine promising regions.

3) **Guided Navigation:** Finally, the optimizer steers towards promising regions identified in the previous step, adjusting parameters through conventional gradient descent combined with strategic adjustments based on the identified features.

This approach aims to improve the traditional methods by integrating direct landscape information into the optimization process. The detailed implementation is shown in Algorithm 1 below.

Algorithm 1 Guided Exploration Algorithm

Input:
initial model-parameters $\boldsymbol{\theta}_0$
learning-rate η
Iterations T
Procedure:

for $k = 1, \ldots, T$ **do**
  Sampling and Mapping:
  Generate a set of perturbations $\Delta\theta$ around θ
  Evaluate loss $L(\theta + \Delta\theta)$ for each perturbation
  Feature Identification:
  Compute gradient $\nabla L(\cdot)$ and Hessian $H(\cdot)$ at $(\theta + \Delta\theta)$
  Identify critical points by analyzing ∇L and the eigenvalues of H
  Guided Optimization:
  Determine promising regions based on identified features and loss values
  Compute the effective gradient $\nabla_e L(\theta)$: where $\nabla_{\text{effective}} L(\theta)$ is the gradient adjusted for directional biases towards promising regions
  Update θ: $\theta \leftarrow \theta - \eta \cdot \nabla_{\text{e}} L(\theta)$
end for
return θ s.t. $\nabla_e L(\theta)$ =0

IV. Methodology and Evaluation

The proposed Guided Exploration algorithm was implemented for training the fashion-MNIST dataset using the TensorFlow framework. The experiments employed Guided Exploration with SGD and Adam algorithms, comparing the performance to traditional SGD and Adam. An extensive hyperparameter search was performed to ensure optimal settings for each algorithm. The results (as shown in Fig. 1 and Table I) demonstrate that the algorithm improved the accuracy by 4.39% when employed with SGD algorithm, and 1.57% when employed with Adam, These results emphasize the algorithm's superior performance and applicability, suggesting its potential for broader applications.

TABLE I
Model Performance

Optimizer	Accuracy	Loss
SGD with GE	0.9197	0.2894
SGD	0.8949	0.3957
Adam with GE	0.9382	0.2359
Adam	0.9251	0.2773

References

[1] Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T. (2018). Visualizing the loss landscape of neural nets. Advances in neural information processing systems, 31.

[2] Christian Raymond and others, "Online Loss Function Learning," *arXiv preprint arXiv:2301.13247*, 2023.

[3] S. Gonzalez and R. Miikkulainen, "Improved Training Speed, Accuracy, and Data Utilization Through Loss Function Optimization," in *2020 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1-8, 2019.

[4] Yangyang Li, "Curvature-aware Manifold Learning," *arXiv preprint arXiv:1706.07167*, 2017.

[5] L. Dinh, "Sharp Minima Can Generalize For Deep Nets," in *Proceedings of the International Conference on Machine Learning*, 2017.