# Dear Editor-in-Chief,

# We sincerely thank you and the reviewers for the valuable feedback provided on our earlier submission to *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*. Following the rejection decision on our earlier submission TNNLS-2024-P-35335 on 1 December 2024, we have undertaken a comprehensive revision of the manuscript to address every comment in detail, substantially improving both the theoretical and experimental components.

# The revised paper, now entitled “Enhanced Adaptive Stochastic Gradient Descent: Convergence Analysis and Its Application in Single-Cell Perturbation Analysis, “is being submitted as a new manuscript for consideration.

# In this new version, we have:

# Expanded the theoretical foundation providing complete proofs for all theorems and lemmas in convex and non-ergodic/ergodic nonconvex settings.

# Strengthened the theoretical and experimental section adding new and comprehensive concept and statistical analyses demonstrating significant improvements over baseline optimizers.

# Clarified the intrinsic motivation for curvature-sensitive learning-rate adjustment with theoretical justification and quantify the improvement over existing optimizers

# Improved manuscript structure and readability by revising grammar, standardizing notation, and removing redundant content.

# Released full source code and datasets via an open-access repository to ensure transparency and reproducibility.

https://github.com/TNNLS-PeerReview/ASGD-ADAM/tree/main

# We respectfully request that the manuscript be reconsidered as a new submission to *TNNLS*. We believe that the substantial theoretical and empirical improvements presented in this version now meet the journal’s standards and contribute meaningfully to adaptive optimization and biological learning applications.

# Thank you for your time and consideration. We would be honored if our revised work could again be reviewed by *TNNLS*. Please, find attached the comments and response to reviewers.

# Sincerely, Francis Boabang, Ph.D. Concordia Institute for Information and Systems Engineering, Montreal, Quebec, Canada

# Samuel Asante Gyamerah, Ph.D.

Department of Mathematics, Toronto Metropolitan University, Toronto, Ontario, Canada.

# Response to Reviewers

## Reviewer 1

1. 1. Intrinsic motivation for different learning rates

We appreciate this observation. The motivation behind is introduced in motivation subsection III.A, and by stating verbatim in the introduction section that,

Compared to the work in [7], we improved the mechanism for selecting the base learning rate by introducing curvature-sensitive switching factors that determine whether the optimization dynamics are dominated by high variability () or smoother, more stable curvature (). These indicators guide adaptive adjustments to the parameter update direction, thereby enhancing stability and convergence in complex, high-dimensional loss landscapes.

Furthermore, in our formulation, the coordinate magnitude correlates with local curvature estimated from the second-order momentum H\_t. When H\_t is small (flat region), a larger base learning rate accelerates convergence; when H\_t is large (sharp curvature), a smaller rate prevents divergence. To clarify, we have expanded Section III-C to include the intuition derived from the curvature-sensitive switching factors f\_min and f\_max (Eqs. 2–5)

2. Missing theoretical proofs for theorems and lemmas

We have now included the full theoretical derivations in Appendix A (Convex Setting), Appendix B (Nonconvex Setting), and Appendix C (Non-Ergodic Convergence). These proofs follow the methodology of AMSGrad [13] and Zhou et al. [11] but substitute the fixed base learning rate with the proposed linear function α\_base = u f(H) + C. Each theorem now contains detailed assumptions, bounding steps, and intermediate lemmas for clarity.

1. Marginal performance gains

Compared to the work in [6], we improve the mechanism for selecting the base learning rate by introducing curvature-sensitive switching factors that determine whether the optimization dynamics are dominated by high variability () or smoother, more stable curvature (). These indicators guide adaptive adjustments to the parameter update direction, thereby enhancing stability and convergence in complex, high-dimensional loss landscapes leading to improved performance gain compared to the state-of-the-art optimizers. We have quantified performance improvement theoretically in section III. H.

We have also experimentally found that the ASGD-Adam/ASGDAMSgrad has better generalization done state-of-the-art optimizers in the single-cell perturbation experiments.

[6] F. Boabang, “Refining optimization methods for training machine learn-

ing models: A case study in robotic surgical procedures,” Ph.D. dissertation, Concordia University, 2024.

4. Demonstration on remote surgery recognition

We acknowledge the limitation of not including domain-specific evaluations. In this revision, we have focused on stem cell differentiation using graph attention mechanism with reinforcement with improve base learning rate selection concept. This improves model convergence compared to state-of-the-art optimizers, demonstrating strong generalization.

1. Excessive introduction to CNN and LSTM

We have removed the introductory description of CNNs and LSTMs and focus on a different problem which is machine learning for stem cell differentiation. This change streamlines Section II and better highlights the optimizer’s contribution.

1. . Minor corrections (symbols, references, grammar)

All notational inconsistencies (e.g., “p-norm”) have been standardized. Repetitive references were merged; the incorrect name “adabrief” has been corrected to AdaBelief; symbol definitions (e.g., y\_1, y) have been clarified. We also restructured sentences for grammatical accuracy and layout consistency, especially in Section III.

1. Definition of α\_base (Equation 11)

We have clarified that “small” and “large” refer to thresholds on the second-order moment H\_t. Specifically, H\_t < mean(H\_t) corresponds to “small” (flat curvature, α\_max), and H\_t > mean(H\_t) corresponds to “large” (sharp curvature, α\_min). This definition is now explicitly stated in Section III-C and illustrated in Algorithm 1 and 2.

## Reviewer 2

1. Code availability

The full implementation, including training scripts and datasets, has been made publicly available at:

https://github.com/TNNLS-PeerReview/ASGD-ADAM/tree/main

A README now details environment setup, parameter settings, and reproduction steps.

1. Figures 1–8 readability

Not applicable

1. More datasets and models

We have added experiments on two additional datasets ATAC and scRNA-seq to demonstrate robustness. The proposed optimizer consistently higher F1 and lower MSE across these datasets, validating generalizability compared to state-of-the-art optimizers.

## Reviewer 3

We thank the reviewer for recognizing the potential of this work. We have implemented all requested minor revisions specifically improving figure clarity, proof completeness, and notation consistency to strengthen the manuscript’s presentation and rigor.