Reviewer's report

Manuscript ID	JMLMC-59453
Article Title	NUMERICAL SPLITTING SCHEMES AS THE CORNERSTONE FOR MINI-BATCH OPTIMIZATION: ON THE IMPORTANCE OF WELL-BALANCED METHODS
Reviewer	Reviewer
Review Date	06/09/2025 16:44
Report	1. **Reviewer's recommendations: Reject** ◦ Author should consider the suggestions of the reviewers and re-submit paper 2. **This paper is average** ◦ Is the importance such that publication procedures should be accelerated? **NO**

Comments

Topic & Contribution – The manuscript reinterprets mini-batch optimizers (such as SGD and random-reshuffle GD) through the lens of operator splitting and shows that conventional splitting schemes (e.g., Lie–Trotter, Strang) are unbalanced. This imbalance causes steady states to depend on the learning rate and can result in convergence to “artificial” stationary points when the interpolation condition is not met. The authors adapt a balanced splitting method from Speth (2020), generalize it to multiple batches, and propose a low-memory variant.

Strengths

• Offers a fresh operator-splitting perspective that clearly explains the origin of optimization pathologies under common training regimes.
• Proposes a balanced update rule that aligns the fixed points of each sub-flow, eliminating the mismatch seen in standard methods.
• One-dimensional toy examples effectively illustrate how the proposed method avoids artificial stationary distributions and converges to true minima.

Main Technical Weaknesses

• Lacks theoretical guarantees: no convergence or stability theorem is provided for the proposed method in nonlinear or high-dimensional settings.
• Experimental validation is limited entirely to simple 1D benchmarks; no results are shown for neural networks or practical datasets.
• Extension to adaptive optimizers like Adam is only briefly discussed and not analyzed in depth.

Empirical Limitations

• Results do not include runtime, memory usage, or accuracy comparisons on realistic tasks (e.g., MNIST, CIFAR).
• No benchmarking against variance reduction methods that also operate outside the interpolation regime, such as STORM or RR-SARAH.
• It is unclear how robust the learning rate and stopping criteria are across diverse problem settings.

Presentation Issues

• Section and subsection numbering is broken throughout (e.g., “1.0.0.4 Objectives”; “3.0.0.3 Construction of Balanced Schemes”).
• Contains numerous grammatical and typographical errors (e.g., splitted, litterature, frequence).
• Many figures are hard to read due to missing or tiny axis labels and legends.
• Equations occasionally omit definitions or proper notation.
• The reference list is not consistent, many entries lack publication venues, or follow different formatting styles..

Recommendations for Revision

• Include at least one high-dimensional experiment using a real neural network (e.g., MLP on MNIST) comparing standard and proposed methods in terms of training loss, accuracy, runtime, and robustness to initialization.
• Provide a formal (even local) convergence result for the balanced splitting scheme under reasonable assumptions.
• Include clear pseudocode and computational cost analysis (both memory and FLOPs) for the proposed algorithm and its variants.
• Substantially revise the manuscript to correct grammar, improve figure quality, and ensure references are complete and properly formatted.

The paper introduces a promising and insightful idea that could have practical impact. However, the current manuscript is more of a conceptual proposal than a fully substantiated research article. Major additions to the theory, experiments, and presentation are needed before it is suitable for publication.

Back