Smoothing and its Implications for Optimization and Generalization in Machine Learning

November 16, 2025

Abstract

Many neural network architectures induce non-smooth, non-convex optimization problems. These networks are typically trained with stochastic first-order methods such as SGD or ADAM. However, theoretical results fail to explain the empirical success of such methods: step sizes used in practice are much larger than theory supports, yet performance is often better than predicted.

This project aims to investigate this discrepancy using tools from online learning, smoothing, and learning theory.

Supervisors:
Mathieu Gerber (mathieu.gerber@bristol.ac.uk), Sarah Sachs (sarah.sachs@bristol.ac.uk)

1 Background

A machine learning method has access to a finite number of samples S. Based on these training samples, the learner solves the *empirical risk minimization problem* $\hat{w} = \operatorname{argmin}_{w \in \mathcal{W}} \hat{R}_S(w)$. This is a substitute for the *expected risk minimization problem* $w^\star = \operatorname{argmin}_{w \in \mathcal{W}} R(w)$. Solving the latter is typically not possible, since the expected risk R depends on an unknown distribution. The discrepancy between the empirical and expected risk can be decomposed into *optimization* and *generalization* errors.

While these errors are well understood for convex and smooth settings, real-world problems are often non-convex and non-smooth, where theoretical understanding remains limited. This project seeks to address this gap. Two key results provide a starting point:

Optimization Error. [Cutkosky et al., 2023] proved that an ϵ-stationary point can be found in $O(\epsilon^{-4})$ iterations for non-convex, non-smooth problems. This result is tight up to constants. Their analysis relies on stochastic smoothing and online-to-batch conversion.

Generalization Error. [Lei, 2023] established generalization bounds for non-smooth, non-convex stochastic optimization using Moreau envelope smoothing, a technique distinct from stochastic smoothing.

Interestingly, there is a mismatch in the smoothing techniques applied: to obtain guarantees for the optimization error, stochastic smoothing is used; conversely, for guarantees on the generalization error, the Moreau envelope is used as a smooth approximation. This project aims to close this gap.

2 Project Proposal

The project may focus on one or several of the following directions, depending on the student's background and interests.

Generalization and Stochastic Smoothing. The results of [Lei, 2023] rely on Moreau envelope smoothing. However, this approach is computationally expensive in practice. Stochastic smoothing provides a more tractable alternative. A promising direction is to derive generalization guarantees under stochastic smoothing, potentially leveraging the online-to-PAC framework of [Lugosi and Neu, 2023].

Optimization and Moreau Smoothing. Another direction is to extend online-to-batch techniques from [Cutkosky et al., 2023] for non-convex functions to the Moreau envelope setting. The main challenge is that the gradient of the Moreau envelope (aka the proximal operator) is non-linear, introducing bias in gradient estimates.

Simulations. Empirical studies can guide theory. The lower bound in [Cutkosky et al., 2023] is based on a constructed optimization instance, and it is unclear whether such cases occur in practice. Experiments could clarify this and inspire tighter guarantees when pathological cases are excluded. Similarly, [Lei, 2023] assumes *uniform stability*, which is often violated in real-world learning. Testing how sensitive their bounds are to this assumption would provide valuable insight.

References

[Cutkosky et al., 2023] Cutkosky, A., Mehta, H., and Orabona, F. (2023). Optimal, stochastic, non-smooth, non-convex optimization through online-to-non-convex conversion. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

[Lei, 2023] Lei, Y. (2023). Stability and generalization of stochastic optimization with nonconvex and nonsmooth problems. In Neu, G. and Rosasco, L., editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 191–227. PMLR.

[Lugosi and Neu, 2023] Lugosi, G. and Neu, G. (2023). Online-to-pac conversions: Generalization bounds via regret analysis. *ArXiv*, abs/2305.19674.