Stochastic Subgradient Descent Escapes Active Strict Saddles on Weakly Convex Functions

- **Sholom Schechtman** (CMAP, Ecole Polytechnique)
- Pascal Bianchi (Télécom Paristech)
- Walid Hachem (CNRS, LIGM, Université Gustave Eiffel)

Keywords: nonsmooth optimization, stochastic gradient descent, avoidance of traps, Clarke subdifferential, stratification

Abstract:

Reinforced projection formula. In the first part of this work we bring to the fore the fact that semialgebraic (or more generally definable in an o-minimal structure) functions admit stratifications of the Verdier type. These are more refined than the Whitney stratifications which were popularized in the optimization literature by [1]. While such stratifications are well-known in the literature on o-minimal structures, up to our knowledge, they have not been used yet in the field of non smooth optimization. To illustrate their interest in this field, we refine the so-called projection formula to the case of definable, locally Lipschitz continuous functions by establishing a Lipschitz-like condition on the (Riemannian) gradients of two adjacent stratas.

Escaping active strict saddles. In the second part of this work we analyze the *stochastic subgradient descent* (SGD), that produces the iterates as follows:

$$x_{n+1} \in x_n - \gamma_n \partial f(x_n) + \gamma_n \eta_{n+1} \,, \tag{1}$$

where (γ_n) is a sequence of step-sizes, (η_n) are the perturbations, $f : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipshitz continuous function and $\partial f(x_n)$ is the Clarke subgradient of f at x_n.

In this context, an *active strict saddle*, a notion introduced in [2], is a Clarke critical point x^* that is lying on a manifold M such that: *i)* f varies smoothly on M, *ii)* f restricted to M admits a direction of negative curvature, *iii)* f varies sharply outside of M. A typical example of such a point is the origin for the function $(y, z) \mapsto -y^2 + |z|$. In [2] the authors have shown that *generically* (in the sense of linear perturbations) the only Clarke critical points that a semialgebraic (or more generally definable), weakly convex function might have are local minima and active strict saddles.

We introduce two additional assumption on the active manifold M: the Verdier and the angle condition. Under these, and assuming that the perturbation sequence (η_n) is omnidirectional, we show that $\mathbb{P}[x_n \to x^*] = 0$, where x^* is an active strict saddle. Furthermore, we show that both of our conditions are generic in the class of weakly convex, semialgebraic functions. As a consequence, we can interpret our results as the fact that the SGD on a generic, semialgebraic and weakly convex function converges to a local minimum.

References:

[1] Bolte, J. and Daniilidis, A. and Lewis, A. and Shiota, M. Clarke subgradients of stratifiable functions *SIAM Journal on Optimization* 18,2:556–572, 2007.

[2] D. Davis and D. Drusvyatskiy Proximal methods avoid active strict saddles of weakly convex functions *Foundations of Computational Mathematics* 2021