Stochastic Methods for Global Optimization

Gaëtan Serré

École Normale Supérieure Paris-Saclay, Centre Borelli, MLMDA Team

Global optimization

Branch of mathematics who aim to find the global optima of an arbitrary function over a compact set. Also sometimes refered as **non-convex optimization**. This is a very hard problem and there is no general solution.

$$f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$$
$$x \in \underset{x \in \Omega}{\operatorname{argmin}} f(x),$$

We focus on sequential and stochastic global optimization methods:

Algorithm: Sequential Stochastic Global Optimization method

Input:
$f : \Omega \to \mathbb{R}$, the objective function.
Output:
$\hat{x}$, an approximation of a minimizer of f.

---

for i **in** $1 \ldots N$ **do**
  $x_i \sim d_\theta$ *Sample a point w.r.t. a distribution d parametrized by θ.*
  $\theta = F(x_1, \ldots, x_i, f(x_1), \ldots, f(x_i))$ *Update θ using the previous points and their evaluations.*
end for
return $\operatorname{argmin} \; f(x_i)$ *Return the best point found.*

Lipschitz continuity commonly assumed to design consistent methods.

The need for randomness

- Emergence of deep learning ⇒ resurgence of nonsmooth Lipschitz optimization (nonsmooth components, e.g. ReLU)
- Recent negative results for deterministic methods [1]
- M. I. Jordan and others [2] raises this fundamental question:

What is the role of randomization in dimension-free nonsmooth nonconvex optimization?

> **Theorem** – *Necessity of randomness*
>
> For any $0 < \Delta$, $0 < \kappa$, $3 \leq d$, $N \leq d - 2$, and any deterministic first-order algorithm $\mathcal{A}$, there exists a κ-Lipschitz function $f : \Omega \to \mathbb{R}$ such that $f(x_0) - \inf_{x \in \Omega} f(x) \leq \Delta$ for which any of the N first iterates produced by $\mathcal{A}$ applied to f is not a (δ, ε)-Goldstein stationary point of f, for any $\delta < \frac{\Delta}{\kappa}$ and $\varepsilon < \frac{\kappa}{252}$.

Randomness is necessary to achieve good performance on arbitrary Lipschitz functions.

Some stochastic methods

LIPO [3]

- Constructs **upper bound** and uses it to sample from an **uniform distribution**
- Adaptative version estimating the Lipschitz constant: AdaLIPO
- Both **consistent** over the class of Lipschitz functions

CMA-ES [4]

- Efficient method. Samples points from moving **Gaussian distribution**.
- **Not consistent**

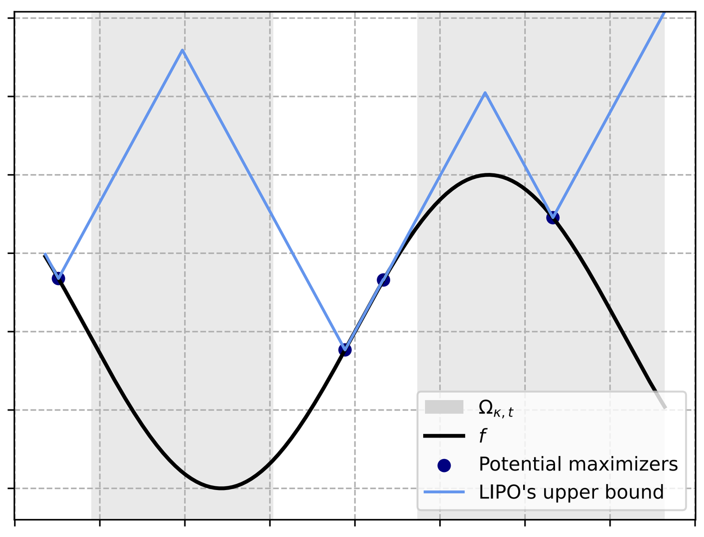

Figure 1: An example of upper bound constructed by LIPO. The set of point $\Omega_{\kappa,t}$ is the set of potential maximizers.

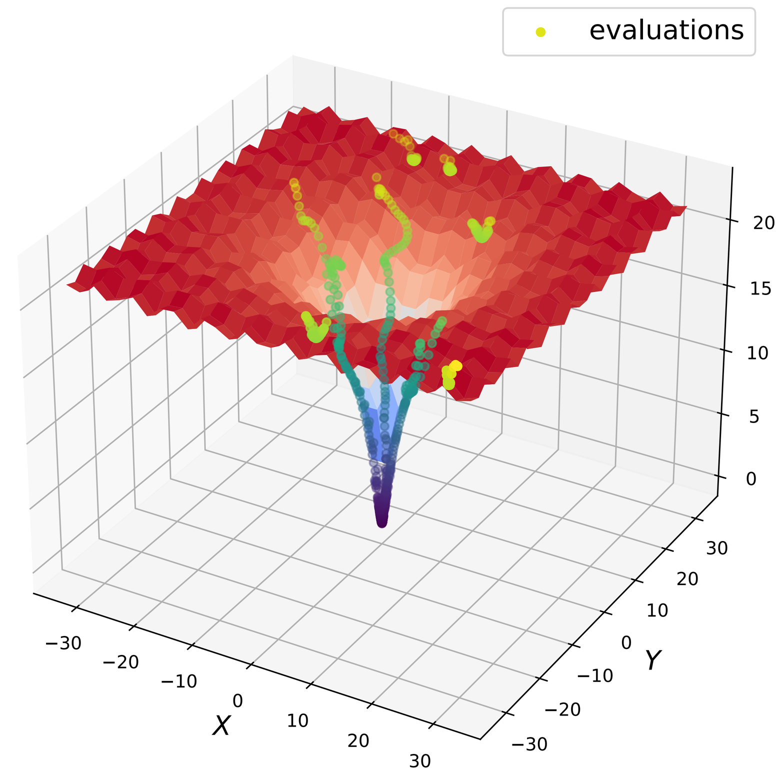

Figure 2: Movement of SBS' particles on the Ackley function.

Stein Boltzmann Sampling

- We propose the *Stein Boltzmann Sampling* method [5].
- Flow-based method. Samples from the **Boltzmann distribution**.
- **Consistent** over the class of Lipschitz functions
- Theory:
  - ↳ Multiple domains e.g. functional analysis, measure theory, flow theory

> **Theorem**
>
> Let k be a symmetric positive definite kernel, $\mathcal{H}$ the associated RKHS, and μ a measure such that $\mu \ll \lambda$. The following integral operator is a mapping in $\mathcal{H}$:
>
> $$T_k : L^2_{\mu(\Omega)} \to \mathcal{H}$$
> $$f \mapsto \int_\Omega k(x, \cdot) f(x) \, \mathrm{d}\mu(x).$$

  - ↳ Some results formalized using **Lean**

```
class RKHS {E F : Type*} [RCLike F] (H : Set (E → F))
    [NormedAddCommGroup H] [InnerProductSpace F H] where
  k : E → E → F
  memb : ∀ (x : E), k x ∈ H
  repro : ∀ (f : H), ∀ x, f.1 x = inner f ⟨k x, memb x⟩
```

- Easily adaptable for sampling in complex manifolds

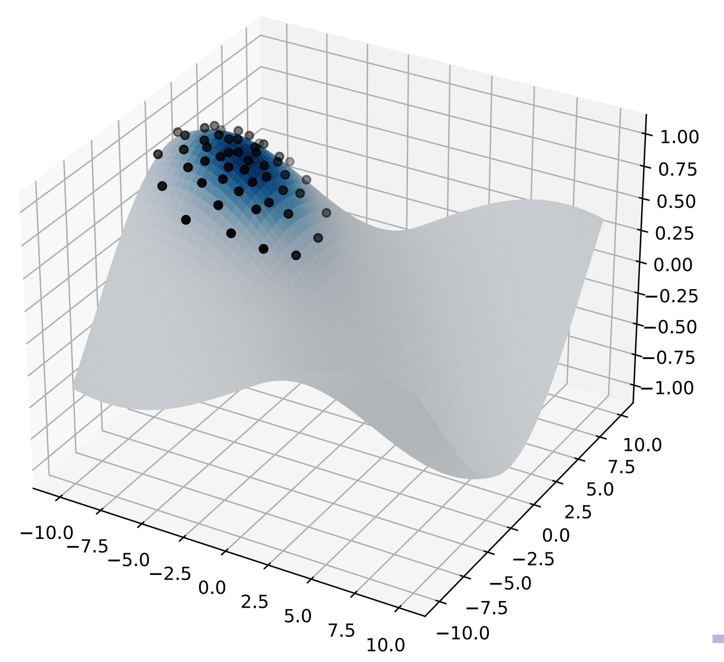

Figure 3: Sampling over a curve w.r.t. a Gaussian distribution using SVGD.

[1] G. Kornowski and O. Shamir, “Oracle complexity in nonsmooth nonconvex optimization,” 2021.
[2] M. I. Jordan and others, “Deterministic Nonsmooth Nonconvex Optimization.” 2023.
[3] C. Malherbe and N. Vayatis, “Global Optimization of Lipschitz Functions,” 2017.
[4] N. Hansen and A. Ostermeier, “Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation,” 1996.
[5] G. Serré, A. Kalogeratos, and N. Vayatis, “Stein Boltzmann Sampling: A Variational Approach for Global Optimization.” 2024.