A Variational Approach for Global Optimization

Gaëtan Serré, Argyris Kalogeratos & Nicolas Vayatis
École Normale Supérieure Paris-Saclay, Centre Borelli, AISTATS 2025

Global Optimization

We aim to minimize a Sobolev and continuous function f over Ω, a compact subset of $\mathbb{R}^d$:

$$f \in C^0(\Omega \to \mathbb{R}) \cap W^{1,4}(\Omega \to \mathbb{R})$$
$$x \in \underset{x \in \Omega}{\operatorname{argmin}}\ f(x).$$

We propose a variational approach to solve this problem by sampling from the **Boltzmann distribution** using **SVGD** [1].

Boltzmann distribution

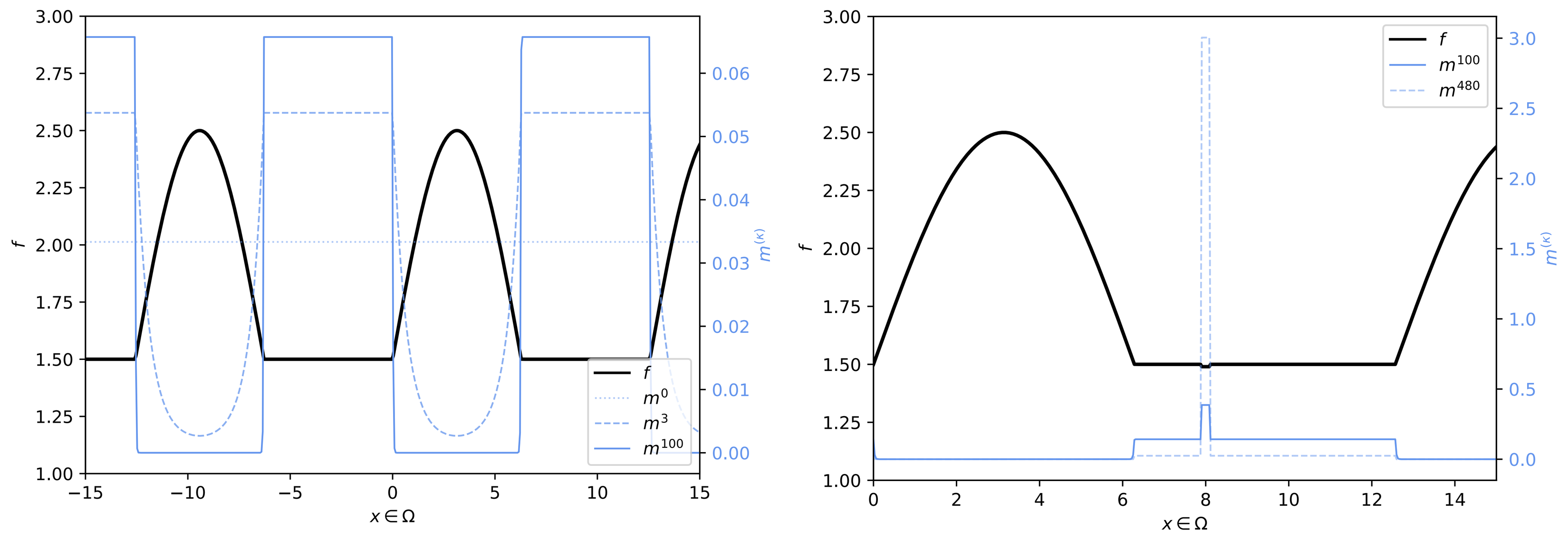


Figure 1: Illustration of the Boltzmann density for different function geometries and values of κ.

Definition

Given a function $f \in C^0(\Omega, \mathbb{R})$, the Boltzmann distribution over Ω is induced by the probability density function $m^{(\kappa)}_{f,\Omega}(x) : \Omega \to \mathbb{R}_{\geq 0}$ defined by:

$$m^{(\kappa)}_{f,\Omega}(x) = m^{(\kappa)}(x) = \frac{e^{-\kappa f(x)}}{\int_\Omega e^{-\kappa f(t)}\,\mathrm{d}t}.$$

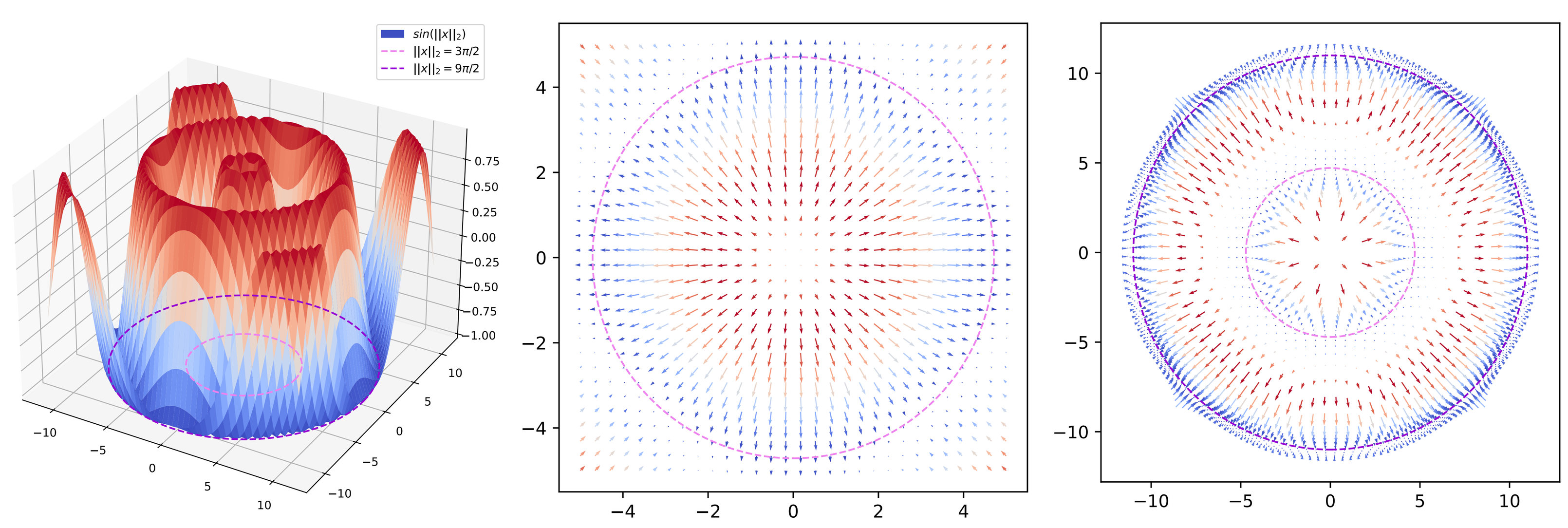
Stein Boltzmann Sampling

SBS is the application of **SVGD** to the **Boltzmann distribution** (or any distribution whose density is asymptotically supported over X^*) on **Sobolev** functions (more suitable for global optimization than the original SVGD framework).

Select a large κ and define $\pi \triangleq m^\kappa$. Sample N particles uniformly and moves them according to the following differential equation:

$$\frac{\mathrm{d}X_t^{(i)}}{\mathrm{d}t} = \frac{1}{N}\sum_{j=1}^{N} \nabla \log \pi\left(X_t^{(j)}\right) k\left(X_t^{(i)}, X_t^{(j)}\right) + \nabla_{X_t^{(j)}} k\left(X_t^{(i)}, X_t^{(j)}\right), \quad (1)$$

where k is a continuous and positive definite kernel. This process is known as **Stein Variational Gradient Descent** [1].



$x \mapsto \sin\|x\|_2$ $\quad t = 0$ $\quad t = 700$

Figure 2: Illustration of $\frac{\mathrm{d}X_t^{(i)}}{\mathrm{d}t}$ (Eq. 1).

We prove the **asymptotic convergence** in this new setting:

Theorem

Let $f \in C^0(\Omega) \cap W^{1,4}(\Omega)$. Let μ_0 be the initial continuous measure of the particles and $\pi_{f,\kappa}$ be the Boltzmann measure. Let $(T_t)_{0 \le t}$ be the trajectories associated with Eq. 1, such that $T_0 = I_d$ and $\mu_t = T_{t\#}\mu$. Then,

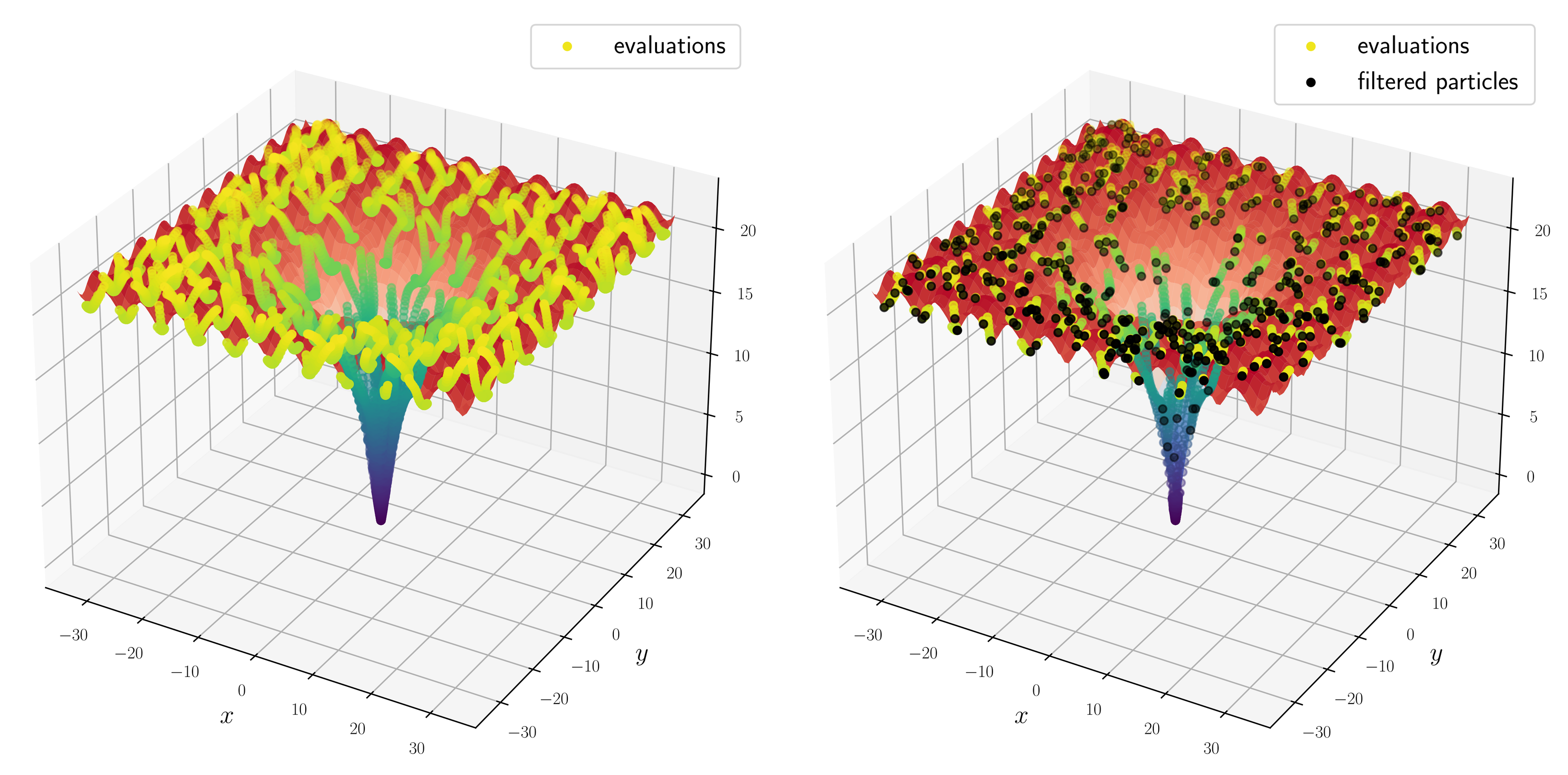
$$\mu_t \underset{t}{\rightharpoonup} \pi_{f,\kappa}.$$

We formalize some key results using the L∃∀N theorem prover [2].

Algorithm: STEIN BOLTZMANN SAMPLING

Input:
$f : \Omega \to \mathbb{R}$; N; κ; ε; n; μ_0
Output:
$\hat{x}$, an approximation of a minimizer of f.

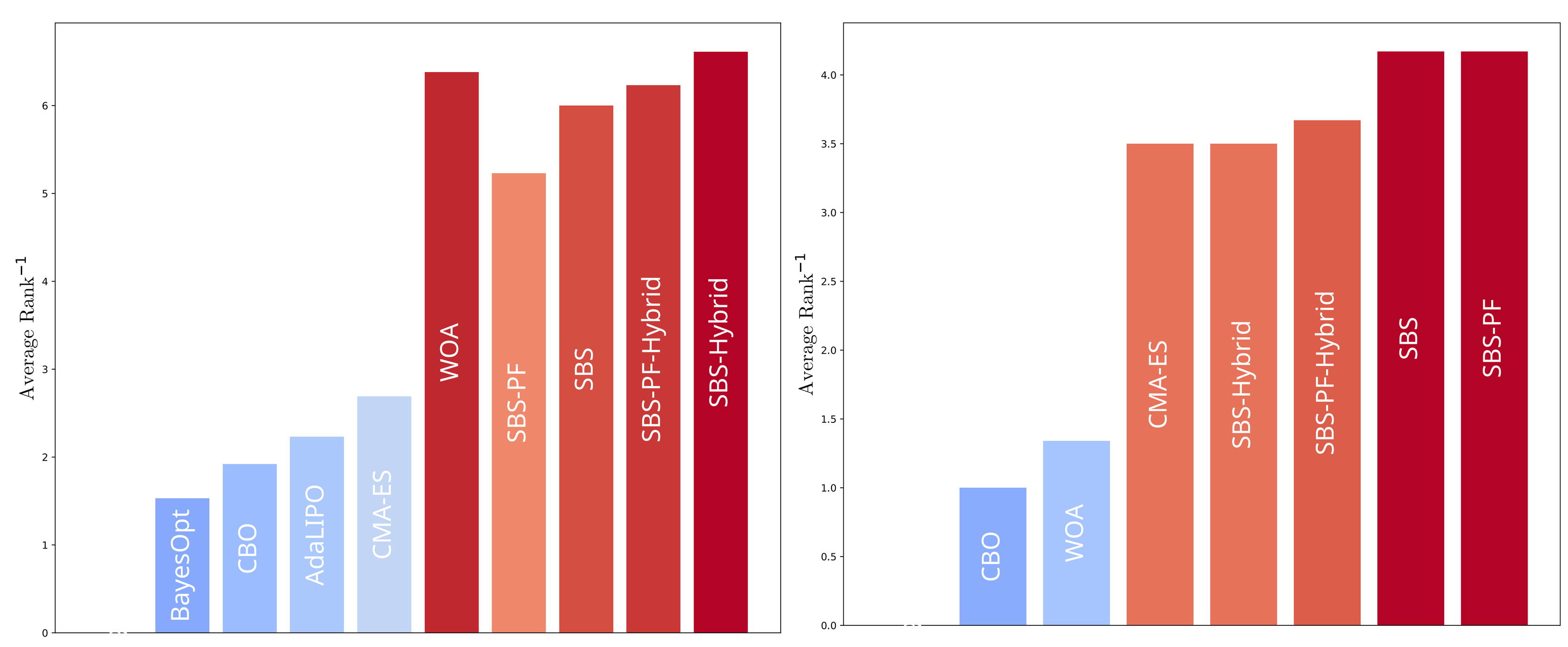
Sample N particles: $X_1 \leftarrow (x^{(1)}, ..., x^{(N)}) \sim \mu_0^{\otimes N}$
for $i = 1$ **to** n **do**
  Compute the vector field $\phi_{\hat{\mu}_i}$ described in Eq. 1
  $X_{i+1} \leftarrow X_i + \varepsilon \phi_{\hat{\mu}_i}$ *update the particles*
  $\hat{\mu}_{i+1} \leftarrow \frac{1}{N}\sum_{j=1}^{N} \delta_{X_{i+1}^{(j)}}$ *update the empirical measure*
end for
$\hat{x} \leftarrow \underset{1 \le j \le N}{\operatorname{argmin}} f\left(X_{n+1}^{(j)}\right)$
return $\hat{x}$



SBS SBS-PF

Figure 3: Illustration of the behavior of SBS and its particle-filtering variant on the Ackley function.

Experimental results



2-dimensional benchmark $\quad$ 50-dimensional benchmark

Figure 4: Illustration of the performance of SBS and its variants on a 2-dimensional and a 50-dimensional benchmark.

[1] Q. Liu and D. Wang, "Stein variational gradient descent: A general purpose bayesian inference algorithm," 2016.
[2] L. de Moura and S. Ullrich, "The Lean 4 Theorem Prover and Programming Language," *Automated Deduction – CADE 28*. Springer International Publishing, 2021.