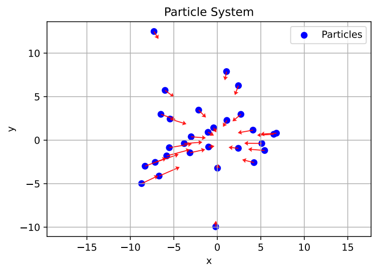
Convergence of stochastic particle systems and applications

PhD Defense: Victor Priser
Supervisors: Pascal Bianchi (Telecom Paris) and François Portier (ENSAI)
Funded by the DSAIDIS Chair

Thesis outline

Interacting particle systems:



- **Long-time convergence in a general setting**
- Stein Variational Gradient Descent
- **Consensus Based Optimization**

Importance Sampling

Sample particles from a proposal and assign weights to fit a target.

- Adaptive Importance Sampling which mitigates weight degeneracy
- **Importance Sampling for optimization**

Outline

- **Interacting particle systems oveview**
- Contribution: Long-run convergence of discrete-time interacting particle systems of the McKean–Vlasov type Bianchi, Hachem, and Priser (2025)
- Contribution: Consensus-Based Optimization beyond finite-time analysis Bianchi, Dragomir, and Priser (2025)
- Contribution: Importance Sampling Optimization with Laplace Principle Dragomir, Portier, and Priser (2026)
- Contribution: Long-time asymptotics of noisy SVGD outside the population limit Priser, Bianchi, and Salim (2024)
- Contribution: Stochastic mirror descent for nonparametric adaptive importance sampling Bianchi, Delyon, Portier, and Priser (2024)

Interacting particle systems

Interacting particle systems: n $\mathbb{R}^d$-valued stochastic particles $(X_k^{i,n})_{i\in[n],k\geq 0}$:

For every $i \in [n]$: $\quad X_{k+1}^{i,n} = X_k^{i,n} + \eta_k b(X_k^{i,n}, \mu_k^n) + \sqrt{\eta_k}\sigma(X_k^{i,n}, \mu_k^n)\xi_k^{i,n}$

Empirical measure

$$\mu_k^n := \frac{1}{n}\sum_{i\in[n]} \delta_{X_k^i} \in \mathcal{P}(\mathbb{R}^d) := \{\text{Probability measures on } \mathbb{R}^d\}$$

- Continuous with linear growth/bounded $b, \sigma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$
- $(\xi_k^{i,n}) \sim_{i.i.d.} \mathcal{N}(0, I_d)$
- $(X_0^{i,n})_i \sim_{i.i.d.} \rho_0 \in \mathcal{P}(\mathbb{R}^d)$
- η_k: step-size with $\sum_k \eta_k = \infty$ and $\eta_k \to 0$

Interacting particle systems

Interacting particle systems: n $\mathbb{R}^d$-valued stochastic particles $(X_k^{i,n})_{i\in[n],k\geq 0}$:

$$\text{For every } i \in [n]: \quad X_{k+1}^{i,n} = X_k^{i,n} + \eta_k b(X_k^{i,n}, \mu_k^n) + \sqrt{\eta_k}\sigma(X_k^{i,n}, \mu_k^n)\xi_k^{i,n}$$

Empirical measure

$$\mu_k^n := \frac{1}{n}\sum_{i\in[n]} \delta_{X_k^i} \in \mathcal{P}(\mathbb{R}^d) := \{\text{Probability measures on } \mathbb{R}^d\}$$

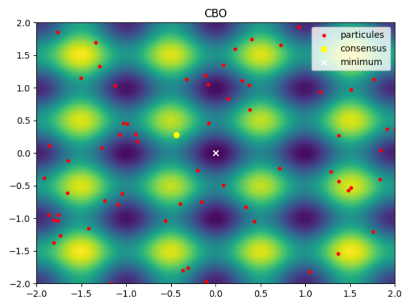
- Continuous with linear growth/bounded $b, \sigma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$
- $(\xi_k^{i,n}) \sim_{i.i.d.} \mathcal{N}(0, I_d)$
- $(X_0^{i,n})_i \sim_{i.i.d.} \rho_0 \in \mathcal{P}(\mathbb{R}^d)$
- η_k: step-size with $\sum_k \eta_k = \infty$ and $\eta_k \to 0$

- Goes back to the 19th century
- Physics, Biology, Social Sciences, Finance, Machine Learning

Some applications

- **Particle Swarm Optimization (PSO)**

> **Goal**
>
> Find the minimizer of a non-convex function f with unknown gradients.



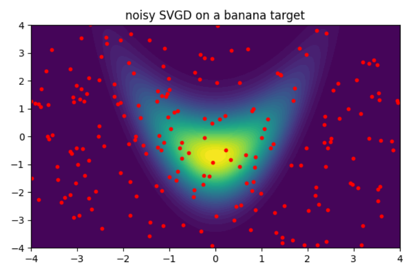
Consensus Based Optimization (CBO) Pinnau et al. (2017):

$$X_{k+1}^{i,n} = X_k^{i,n} + \eta_k \overbrace{(X_k^{*,n} - X_k^{i,n})}^{\text{Interaction}} + \sqrt{\eta_k}\text{Noise} \quad (X_k^{*,n} \approx \text{global best particle})$$

Question: μ_k^n approaches $\delta_{\arg\min f}$ when k grows?

Some applications

- Particle Swarm Optimization (PSO)
- **Sampling**



Goal

Sample a target distribution π with access to informations like the score $\nabla \log \pi$.

Stein Variational Gradient Descent (SVGD) Liu and Wang (2016):

$$X_{k+1}^{i,n} = X_k^{i,n} + \eta_{k+1} \sum_{j\in[n]} \left(\underbrace{\nabla \log \pi(X_k^{j,n}) \overbrace{K(X_k^{i,n}, X_k^{j,n})}^{\text{Kernel}}}_{\text{Attracting term}} + \underbrace{\nabla_2 K(X_k^{i,n}, X_k^{j,n})}_{\text{Repulsive term}} \right)$$

Question: μ_k^n approaches π as k grows?

Our main focus: long-time results

Question

Can we characterize the limits of μ_k^n in any regime $(k, n) \to (\infty, \infty)$?

Interesting regime (long-time): $k \to \infty$, $n \to \infty$ where $k \gg n$.
Usual results: $k \to \infty$, $n \to \infty$ where $k \leq \log(n)$

In applications

- **Particle Swarm Optimization:** $W_2(\mu_k^n, \delta_{\arg\min f}) \xrightarrow[(k,n)\to(\infty,\infty)]{\mathbb{P}} 0$?
- **Sampling:** $W_2(\mu_k^n, \pi) \xrightarrow[(k,n)\to(\infty,\infty)]{\mathbb{P}} 0$?

Outline

- Interacting particle systems oveview
- **Contribution: Long-run convergence of discrete-time interacting particle systems of the McKean–Vlasov type Bianchi, Hachem, and Priser (2025)**
- Contribution: Consensus-Based Optimization beyond finite-time analysis Bianchi, Dragomir, and Priser (2025)
- Contribution: Importance Sampling Optimization with Laplace Principle Dragomir, Portier, and Priser (2026)
- Contribution: Long-time asymptotics of noisy SVGD outside the population limit Priser, Bianchi, and Salim (2024)
- Contribution: Stochastic mirror descent for nonparametric adaptive importance sampling Bianchi, Delyon, Portier, and Priser (2024)

Usual way to deal with particle systems

Discrete time particle system ($\eta_k \equiv \eta$)

$$X_{k+1}^{i,n} = X_k^{i,n} + \eta b(X_k^{i,n}, \mu_k^n) + \sqrt{\eta}\sigma(X_k^{i,n}, \mu_k^n)\xi_k^{i,n}$$

$\downarrow$ (i)

Continuous time particle system

$$\boxed{d\tilde{X}_t^{i,n} = b(\tilde{X}_t^{i,n}, \tilde{\mu}_t^n)dt + \sigma(\tilde{X}_t^{i,n}, \tilde{\mu}_t^n)dB_t^{i,n}}$$

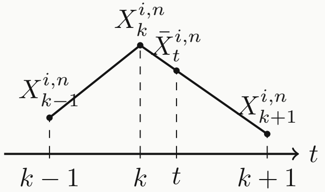
$\tilde{\mu}_t^n = n^{-1}\sum_{i\in[n]} \delta_{\tilde{X}_t^{i,n}}$; $(B_t^{i,n})$: Brownians

i) **Time discretization:** Linear interpolation $(\bar{X}_t^{i,n})_{t\geq 0}$ of $(X_k^{i,n})_{k\in\mathbb{N}}$

$$\bar{\mu}^n := \frac{1}{n}\sum_{i\in[n]} \delta_{t\mapsto \bar{X}_t^{i,n}} \in \mathcal{P}(C([0,T]))$$

$\tilde{\mu}^n$: pathwise empirical measure of $(\tilde{X}_t^{i,n})$

$$W_2(\tilde{\mu}^n, \bar{\mu}^n) \xrightarrow[\eta\to 0]{\mathbb{P}} 0$$



Usual way to deal with particle systems

Discrete time particle system ($\eta_k \equiv \eta$)
$X_{k+1}^{i,n} = X_k^{i,n} + \eta b(X_k^{i,n}, \mu_k^n) + \sqrt{\eta}\sigma(X_k^{i,n}, \mu_k^n)\xi_k^{i,n}$

$\downarrow$ (i)

Continuous time particle system
$\boxed{d\tilde{X}_t^{i,n} = b(\tilde{X}_t^{i,n}, \tilde{\mu}_t^n)dt + \sigma(\tilde{X}_t^{i,n}, \tilde{\mu}_t^n)dB_t^{i,n}}$
$\tilde{\mu}_t^n = n^{-1}\sum_{i\in[n]} \delta_{\tilde{X}_t^{i,n}}$; $(B_t^{i,n})$: Brownians

$\downarrow$ (ii)

McKean–Vlasov equation
$\boxed{d\mathcal{X}_t = b(\mathcal{X}_t, \mathrm{Law}(\mathcal{X}_t))dt + \sigma(\mathcal{X}_t, \mathrm{Law}(\mathcal{X}_t))dB_t}$
(B_t): Brownian

i) **Time discretization:** Linear interpolation $(\bar{X}_t^{i,n})_{t\geq 0}$ of $(X_k^{i,n})_{k\in\mathbb{N}}$

$$\bar{\mu}^n := \frac{1}{n}\sum_{i\in[n]} \delta_{t\mapsto \bar{X}_t^{i,n}} \in \mathcal{P}(C([0,T]))$$

$\tilde{\mu}^n$: pathwise empirical measure of $(\tilde{X}_t^{i,n})$

$$W_2(\tilde{\mu}^n, \bar{\mu}^n) \xrightarrow[\eta\to 0]{\mathbb{P}} 0$$

ii) **Propagation of chaos:** Sznitman (1991)
$\rho = \mathrm{Law}(t \mapsto \mathcal{X}_t) \in \mathcal{P}(C([0,T]))$

$$W_2(\rho, \tilde{\mu}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

Usual way to deal with particle systems

Discrete time particle system ($\eta_k \equiv \eta$)
$$X_{k+1}^{i,n} = X_k^{i,n} + \eta b(X_k^{i,n}, \mu_k^n) + \sqrt{\eta}\sigma(X_k^{i,n}, \mu_k^n)\xi_k^{i,n}$$

$\downarrow$ (i)

Continuous time particle system
$$\boxed{d\tilde{X}_t^{i,n} = b(\tilde{X}_t^{i,n}, \tilde{\mu}_t^n)dt + \sigma(\tilde{X}_t^{i,n}, \tilde{\mu}_t^n)dB_t^{i,n}}$$
$\tilde{\mu}_t^n = n^{-1}\sum_{i\in[n]}\delta_{\tilde{X}_t^{i,n}}$; $(B_t^{i,n})$: Brownians

$\downarrow$ (ii)

McKean–Vlasov equation
$$\boxed{d\mathcal{X}_t = b(\mathcal{X}_t, \mathrm{Law}(\mathcal{X}_t))dt + \sigma(\mathcal{X}_t, \mathrm{Law}(\mathcal{X}_t))dB_t}$$
(B_t): Brownian

i) **Time discretization:** Linear interpolation $(\bar{X}_t^{i,n})_{t\geq 0}$ of $(X_k^{i,n})_{k\in\mathbb{N}}$

$$\bar{\mu}^n := \frac{1}{n}\sum_{i\in[n]}\delta_{t\mapsto \bar{X}_t^{i,n}} \in \mathcal{P}(C([0,T]))$$

$\tilde{\mu}^n$: $(\tilde{X}_t^{i,n})_{t\in[0,T]}$ pathwise empirical measure

$$W_2(\tilde{\mu}^n, \bar{\mu}^n) \xrightarrow[\eta\to 0]{\mathbb{P}} 0$$

ii) **Propagation of chaos:** Sznitman (1991)
$\rho = \mathrm{Law}(t \mapsto \mathcal{X}_t) \in \mathcal{P}(C([0,T]))$

$$W_2(\rho, \tilde{\mu}^n) \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

$T < \infty$: **particle systems don't inherit the long-time behavior of McKean–Vlasov SDE.**

How to obtain long-time results

Uniform-in-time propagation of chaos (Malrieu; Durmus et al. (2001; 2020)):

- Particle system ≈ McKean–Vlasov equation on $[0, \infty)$
- Require a strong contraction property
- A unique McKean–Vlasov stationary measure (ρ_∞)
- With Karimi Jaghargh, Hsieh, and Krause (2024)

$$W_2(\mu_k^n, \rho_\infty) \xrightarrow[(k,n)\to(\infty,\infty)]{\mathbb{P}} 0$$

Goal

Provide a long-time result when uniform-in-time propagation of chaos doesn't hold:

- the case where there exist several McKean–Vlasov stationary measures
- the case of CBO or SVGD

Our result

Time shift: $\Phi_t : \mathcal{P}(C([0,T])) \to \mathcal{P}(C([0,T]))$ (i.e. $\Phi_t(\mathrm{Law}(s \mapsto X_s)) = \mathrm{Law}(s \mapsto X_{t+s})$)
Birkhoff center:

$$\mathrm{BC} := \overline{\{\rho \in \text{pathwise McKean–Vlasov distributions} : \exists (t_k)\; s.t.\; \Phi_{t_k}(\rho) \to \rho\}}$$

Theorem Bianchi, Hachem, and Priser (2025)

$$\frac{1}{t}\int_0^t W_2(\Phi_s(\bar{\mu}^n), \mathrm{BC})ds \xrightarrow[(t,n)\to(\infty,\infty)]{\mathbb{P}} 0$$

With a marginalization argument:

$$\frac{\sum_{i\in[k]} \eta_i W_2(\mu_i^n, \overbrace{(\pi_0)_\# \mathrm{BC}}^{\text{marginale}})}{\sum_{i\in[k]} \eta_i} \xrightarrow[(k,n)\to(\infty,\infty)]{\mathbb{P}} 0$$

In most applications: $(\pi_0)_\#\mathrm{BC} = \{\text{McKean–Vlasov stationary distributions}\}$

Main steps of the proof

Step 1: Particle systems behave like a McKean–Vlasov equation as $(t,n) \to (\infty,\infty)$:

- Set of random measures $\{\Phi_t(\bar{\mu}^n) : (t,n) \in \mathbb{R}_+ \times \mathbb{N}\} \subset \mathcal{P}(C([0,T]))$ is tight
- Any limit point (in law) of $\Phi_t(\bar{\mu}^n)$ as $(t,n) \to (\infty,\infty)$ is a McKean–Vlasov distribution
- **Key point**: McKean–Vlasov distributions are characterized by the martingale problem

Step 2: Specification of the limiting McKean–Vlasov distributions:

- Analysis of the ergodic measure of $(\Phi_t(\bar{\mu}^n))_{t\geq 0}$
- Accumulation points of the ergodic measure of $\Phi_t(\bar{\mu}^n)$ as $(t,n) \to (\infty,\infty)$:
  - Time-shift (Φ_t) invariant (by the nature of ergodic measures)
  - Supported by McKean–Vlasov distributions (by Step 1)
- **Poincaré recurrence theorem**: accumulation points supported by the Birkhoff center

Outline

- Interacting particle systems oveview
- Contribution: Long-run convergence of discrete-time interacting particle systems of the McKean–Vlasov type Bianchi, Hachem, and Priser (2025)
- **Contribution: Consensus-Based Optimization beyond finite-time analysis Bianchi, Dragomir, and Priser (2025)**
- Contribution: Importance Sampling Optimization with Laplace Principle Dragomir, Portier, and Priser (2026)
- Contribution: Long-time asymptotics of noisy SVGD outside the population limit Priser, Bianchi, and Salim (2024)
- Contribution: Stochastic mirror descent for nonparametric adaptive importance sampling Bianchi, Delyon, Portier, and Priser (2024)

Objective

Zeroth order optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

where

- $f : \mathbb{R}^d \to \mathbb{R}$ non-convex
- f has unknown gradient
- f has a unique minimizer x^*

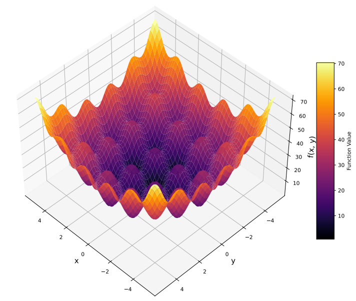


Figure: Example of a function f

Several algorithms: Particle Swarm Optimization, Genetic and Evolutionary algorithms, Bayesian Optimization, Consensus Based Optimization ...

Consensus Based Optimization

Consensus Based Optimization (CBO) Pinnau et al. (2017) (with $\alpha \gg 1$, $\eta_k \equiv \eta$):

$$\text{For every } i \in [n], \quad X_{k+1}^{i,n} = X_k^{i,n} + \eta \left(\frac{\sum_{i\in[n]} X_k^{i,n} \mathrm{e}^{-\alpha f(X_k^{i,n})}}{\sum_{i\in[n]} \mathrm{e}^{-\alpha f(X_k^{i,n})}} - X_k^{i,n} \right) + \sqrt{\eta}\sigma_k \xi_k^i$$

Smooth approximation with the Laplace principle/Softmin

$$\frac{\sum_{i\in[n]} X_k^{i,n} \mathrm{e}^{-\alpha f(X_k^{i,n})}}{\sum_{i\in[n]} \mathrm{e}^{-\alpha f(X_k^{i,n})}} \underset{\alpha\to\infty}{\to} \operatorname*{arg\,min}_{x\in\{X_k^1,\dots,X_k^n\}} f(x)$$

Choice: $\boxed{\sigma_k = \sqrt{\frac{\gamma}{\alpha}}}$

Context:

- Particle Swarm Optimization (PSO) Kennedy and Eberhart (1995)
- Several theoretical analyses: Carrillo et al.; Fornasier, Klock, and Riedl (2018; 2024)
- Many variants: multiple minima, constraints, closer to PSO

Theoretical results

Laplace principle consequence: we can at best expect that particles reach $B(x^*, \mathcal{O}(\alpha^{-1/2}))$.

For which value of n, k it arises?

	$k < \log(\alpha)$	$k > \log(\alpha)$
$n = \infty$	Fornasier, Klock, and Riedl (2024)	Carrillo et al. (2018)
$n < \infty$	Fornasier, Klock, and Riedl (2024)	Gerber et al.; Bayraktar et al. (2025; 2026)

- Results when $k > \log(\alpha)$ hold under restrictive conditions: the initial particles are close to x^*
- Results when $k < \log(\alpha)$ hold under reasonable conditions but few iterations can be done
- Results obtain with other choices of diffusion: $\sigma_k \neq \sqrt{\frac{\gamma}{\alpha}}$

Our result

Theorem Bianchi, Dragomir, and Priser (2025)

Assumptions: f is locally convex around x^*; γ is large enough depending on f

$$\mathbb{E}\left\|X_k^{i,n} - x^*\right\|^2 \leq C(\underbrace{\frac{1}{\alpha}}_{\text{Laplace error}} + \underbrace{e^{-Ck}}_{\text{mean-field convergence}}) + C_\alpha(\underbrace{\frac{1}{n^{\frac{2}{d}}}}_{\text{propagation of chaos}} + \underbrace{\eta}_{\text{time-discretization}})$$

- Restrictive conditions on previous long-time results is removed
- Condition $k \leq \log(\alpha)$ is dropped
- $\mathbb{E}\left\|X_k^{i,n} - x^*\right\|^2 \leq \frac{C}{\alpha}$ when $k \gtrsim \log \alpha$, $n \gtrsim \alpha^{\frac{d}{2}}$ and $\eta \lesssim \alpha^{-1}$

Our result

Theorem Bianchi, Dragomir, and Priser (2025)

Assumptions: f is locally convex around x^*; γ is large enough depending on f

$$\mathbb{E}\left\|X_k^{i,n} - x^*\right\|^2 \leq C\Big(\underbrace{\frac{1}{\alpha}}_{\text{Laplace error}} + \underbrace{e^{-Ck}}_{\text{mean-field convergence}}\Big) + C_\alpha\Big(\underbrace{\frac{1}{n^{\frac{2}{d}}}}_{\text{propagation of chaos}} + \underbrace{\eta}_{\text{time-discretization}}\Big)$$

- Restrictive conditions on previous long-time results is removed
- Condition $k \leq \log(\alpha)$ is dropped
- $\mathbb{E}\left\|X_k^{i,n} - x^*\right\|^2 \leq \frac{C}{\alpha}$ when $k \gtrsim \log\alpha$, $n \gtrsim \alpha^{\frac{d}{2}}$ and $\eta \lesssim \alpha^{-1}$

CBO as a proximal point descent

Mean Field Equation: $C_\alpha(\mu) := \frac{\int x e^{-\alpha f(x)} d\mu(x)}{\int e^{-\alpha f(x)} d\mu(x)}$

$$d\mathcal{X}_t = (C_\alpha(\mathrm{Law}(\mathcal{X}_t)) - \mathcal{X}_t)dt + \sqrt{\frac{\gamma}{\alpha}} dB_t$$

$$\mathcal{X}_t \approx \mathcal{N}(x_t, \tfrac{\gamma}{\alpha} I_d) \text{ where } \dot{x}_t = C_\alpha(\mathrm{Law}(\mathcal{X}_t)) - x_t$$

Laplace principle Kirwin (2010):

$$C_\alpha(\mathcal{N}(x_t, \frac{\gamma}{\alpha} I_d)) = \underbrace{\operatorname*{arg\,min}_{x \in \mathbb{R}^d} \left(f(x) + \frac{\|x - x_t\|^2}{2\gamma} \right)}_{\mathrm{prox}_{\gamma f}(x_t) \text{ defined for } \gamma \gg 1} + \underbrace{\mathcal{O}(\alpha^{-1})}_{\text{Laplace error}}$$

Proximal point descent: $\dot{x}_t = \mathrm{prox}_{\gamma f}(x_t) - x_t + \mathcal{O}(\alpha^{-1})$
Mean-field result:

$$\boxed{\mathbb{E} \|\mathcal{X}_t - x^*\|^2 \leq \mathbb{E} \|\mathcal{X}_0 - x^*\|^2 e^{-Ct} + \mathcal{O}(\alpha^{-1})}$$

Finite particle analysis

Propagation of chaos + discretization error

$$\mathbb{E}\left\|\mathcal{X}_{k\eta} - X_k^i\right\|^2 \leq \mathrm{e}^{Ck}(\eta + n^{-2/d})$$

- Cannot be applied directly because of $\mathrm{e}^{Ck} \to \infty$
- We apply this result on sliding time-windows
- Extension of our previous work with convergence rates

Outline

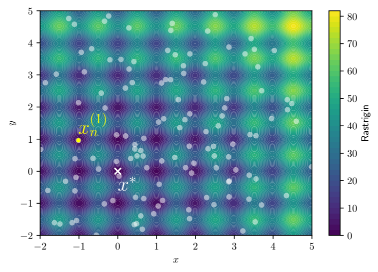
- Interacting particle systems oveview
- Contribution: Long-run convergence of discrete-time interacting particle systems of the McKean–Vlasov type Bianchi, Hachem, and Priser (2025)
- Contribution: Consensus-Based Optimization beyond finite-time analysis Bianchi, Dragomir, and Priser (2025)
- **Contribution: Importance Sampling Optimization with Laplace Principle Dragomir, Portier, and Priser (2026)**
- Contribution: Long-time asymptotics of noisy SVGD outside the population limit Priser, Bianchi, and Salim (2024)
- Contribution: Stochastic mirror descent for nonparametric adaptive importance sampling Bianchi, Delyon, Portier, and Priser (2024)

Settings and idea

Particles: $(X^i)_{i\in[n]} \sim_{i.i.d.} q_0$

How to estimate the minimizer x^* of a non-convex function f using $(f(X^i))_{i\in[n]}$?

Random Search: $x_n^{(1)} := \arg\min\{f(x) \,:\, x \in \{X^1, \ldots, X^n\}\}$

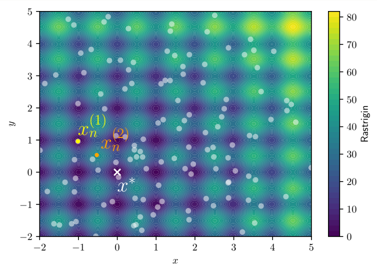


Settings and idea

Particles: $(X^i)_{i\in[n]} \sim_{i.i.d.} q_0$

How to estimate the minimizer x^* of a non-convex function f using $(f(X^i))_{i\in[n]}$?

Random Search: $x_n^{(1)} := \arg\min\{f(x) : x \in \{X^1, \dots, X^n\}\}$
Our weighted average: $x_n^{(2)} := \frac{1}{\sum_{i\in[n]} w_i} \sum_{i\in[n]} w_i X^i$ with $w_i := \frac{\exp(-\alpha f(X^i))}{q_0(X^i)}$



Contributions:

Context:

- Random search: very simple zeroth-order method with low complexity
- Laplace principle used in many zeroth-order methods (CBO, Simulated Annealing, ZOPPA)
  - Smooth the problem
  - Smoothing parameter (α) of the Laplace principle not theoretically optimized

Contributions

- Importance Sampling together with a Laplace principle
- Optimize α with respect to number of particles
- Better convergence rates than Random Search (under additional assumptions on f)
- Motivate the Laplace-based method: Laplace principle accelerates convergence

Our result

Idea: approximate the distribution $\pi_\alpha \propto \exp(-\alpha f)$ with importance sampling:

$$x_n^{(2)} := \frac{1}{\sum_{i\in[n]} w_i} \sum_{i\in[n]} w_i X^i \to_{n\to\infty} \int x \pi_\alpha(dx) \text{ with } w_i = \frac{\exp(-\alpha f(X^i))}{q_0(X^i)}$$

Laplace principle: $\int x\pi_\alpha(dx) \to_{\alpha\to\infty} x^*$

Theorem Dragomir, Portier, and Priser (2026)

Assume f strongly convex in a neighborhood of x^*:

$$\mathbb{E}\left\|x_n^{(2)} - x^*\right\|^2 \lesssim n^{-\frac{4}{d+2}} \quad \text{with } \alpha = n^{\frac{2}{d+2}}$$

Random Search: $\mathbb{E}\left\|x_n^{(1)} - x^*\right\|^2 \lesssim n^{-\frac{2}{d}}$ with $x_n^{(1)} = \arg\min\{f(x) : x \in \{X^1, \dots, X^n\}\}$

A bias/variance analysis

Bias/variance decomposition:

$$\mathbb{E}\left\|x_n^{(2)} - x^*\right\|^2 \leq 2\,\underbrace{\mathbb{E}\left\|x_n^{(2)} - \int x\pi_\alpha(dx)\right\|^2}_{\text{variance}} + 2\,\underbrace{\mathbb{E}\left\|\int x\pi_\alpha(dx) - x^*\right\|^2}_{\text{bias}}$$

Variance term controlled by importance sampling technique:

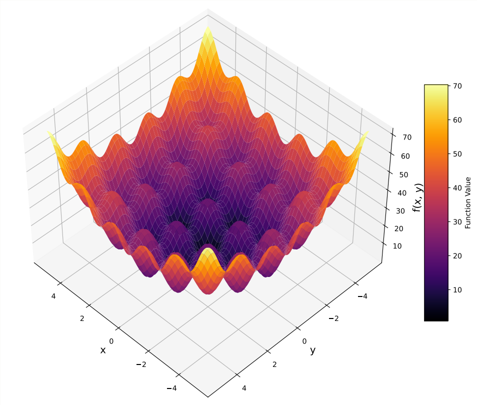
$$\mathbb{E}\left\|x_n^{(2)} - \int x\pi_\alpha(dx)\right\|^2 \lesssim \frac{1}{n}\alpha^{\frac{d}{2}-1}$$

Bias controlled by quantitative Laplace principle:

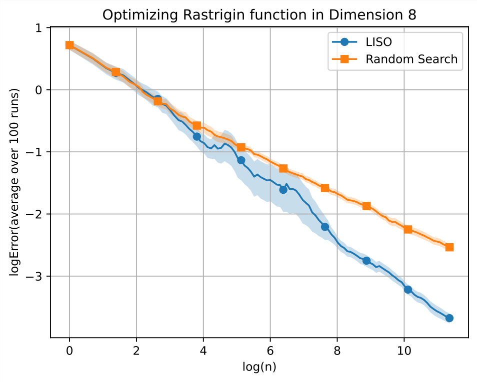
$$\mathbb{E}\left\|\int x\pi_\alpha(dx) - x^*\right\|^2 \lesssim \frac{1}{\alpha^2}$$

Optimal choice $\alpha = n^{\frac{2}{d+2}}$: Bias $=$ Variance $= n^{-\frac{4}{d+2}}$

Numerical simulation



(a) Target function: Rastrigin



(b) Comparison of our method (LISO) with random search

Perspectives on interacting particle systems

Other applications of interacting particle systems

- Wide Neural Networks: SGD as interacting particle system Chizat and Bach (2018)
- Simplified transformers: clustering of tokens Geshkovski et al. (2025)

Other type of interacting particle systems

Self-Interacting Diffusions

- Converges to McKean–Vlasov stationary measures Kurtzmann (2010)
- Link with SVGD?

Absorbed McKean–Vlasov diffusions

- Long-time behaviour in the finite particle regime Assadeck and Panloup (2025)?
- Constraint optimisation inspired by CBO?

Perspectives on studied algorithms

CBO

- Propagate our analysis to the PSO algorithm and other variants of CBO
- Tighten our analysis to reduce the exponential constants in α
- Using an adaptive diffusion but lower bounded by $\sqrt{\frac{\gamma}{\alpha}}$

Laplace Importance Sampling Optimization

- Use adaptive importance sampling methods for optimization. What are the rates?
- Extend optimal α to other Laplace-based methods
- Expore other weighted averages (beyond exponential weights)

High-level overview of the thesis

- In most of the analysis of the thesis: we study algorithms with multiple stationary points
- We show that we converge to one of the stationary points
- When a stationary point is unwanted, the convergence result fails
- **Key idea of the thesis:** Use noise to avoid unwanted stationary points

Perspectives

- Quantify the probability to fall into one stationary point rather than another
- Show that unwanted stationary points are reached with low probability
- Chizat and Bach; Monmarché (2018; 2025)

Contributions

- (*Stochastic Processes and Applications*) Long run convergence of discrete-time interacting particle systems of the McKean–Vlasov type Bianchi, Hachem, and Priser (2025)
- (*ICLR*): Long-time asymptotics of noisy SVGD outside the population limit Priser, Bianchi, and Salim (2024)
- (*Preprint*): Consensus-Based Optimization Beyond Finite-Time Analysis Bianchi, Dragomir, and Priser (2025)
- (*Preprint*): Stochastic mirror descent for nonparametric adaptive importance sampling Bianchi, Delyon, Portier, and Priser (2024)
- (*Preprint*): Importance Sampling Optimization with Laplace Principle Dragomir, Portier, and Priser (2026)