Consensus-based optimization

Claudia Totzeck

University of Wuppertal

Mascot NUM, Clermont-Ferrand

June 7-9, 2022

MATHEMATICAL MODELLING, ANALYSIS AND COMPUTATIONAL MATHEMATICS

BERGISCHE UNIVERSITÄT WUPPERTAL

Motivation - interacting particle systems

(smoothed) Hegselmann-Krause (2002)

Cucker-Smale dynamics (2007)

Motivation - interacting particle systems

Let $f: \mathbb{R}^D \to \mathbb{R}^+$, $D \geq 1$, be a function. The problem we consider is given by

$$x^* := \underset{x \in \mathbb{R}^D}{\operatorname{argmin}} f(x).$$

Task: Construct an algorithm that finds the global minimizer of f!

What's on the market?

- Genetic Algorithms
- Particle Swarm optimization
- Stochastic Gradient descent
- Ensemble methods (Kalman)
- Simulated annealing
- ....

What was observed:

- local best / global best information
- indistiguishable particles
- no structure (O/S/PDE)
- few convergence proofs

$\Rightarrow$ **What can we do?**

Let $f : \mathbb{R}^D \to \mathbb{R}^+$, $D \geq 1$, be a function. The problem we consider is given by

$$x^* := \operatorname*{argmin}_{x \in \mathbb{R}^D} f(x).$$

Task: Construct an algorithm that finds the global minimizer of f!

What's on the market?

- Genetic Algorithms
- Particle Swarm optimization
- Stochastic Gradient descent
- Ensemble methods (Kalman)
- Simulated annealing
- ....

What was observed:

- local best / global best information
- indistiguishable particles
- no structure (O/S/PDE)
- few convergence proofs

$\Rightarrow$ **What can we do?**

Let $f: \mathbb{R}^D \to \mathbb{R}^+$, $D \geq 1$, be a function. The problem we consider is given by

$$x^* := \operatorname*{argmin}_{x \in \mathbb{R}^D} f(x).$$

Task: Construct an algorithm that finds the global minimizer of f!

What's on the market?

- Genetic Algorithms
- Particle Swarm optimization
- Stochastic Gradient descent
- Ensemble methods (Kalman)
- Simulated annealing
- ....

What was observed:

- local best / global best information
- indistiguishable particles
- no structure (O/S/PDE)
- few convergence proofs

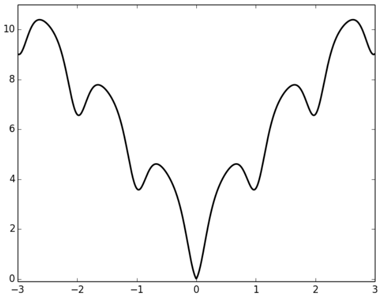
$\Rightarrow$ **What can we do?**

Motivation

Let $f : \mathbb{R}^D \to \mathbb{R}^+$, $D \geq 1$, be a function. The problem we consider is given by

$$\operatorname*{argmin}_{x \in \mathbb{R}^D} f(x).$$

Task: Construct an algorithm that finds the global minimizer of f!



Features we like:

- global minimum
- particle scheme
- no gradient information
- indistiguishable particles
- mean-field equation for analysis

Outline

- Introduction to CBO
  - motivation of weighted mean
  - importance of scaled stochastics
  - numerical illustration
- Overview & some comments
  - machine learning
  - component-wise common noise
  - towards particle swarm optimization - Vlasov dynamic
  - towards Boltzmann
  - dynamics constrained to sphere

Combine **swarm optimization** and **opinion dynamics** to obtain:

$$dX_t^i = -\lambda(X_t^i - v_f)\, H[f(X_t^i) - f(v_f)]dt + \sigma|X_t^i - v_f|dB_t^i, \quad i = 1, \ldots, N,$$

where the **weighted average** v_f is given by

$$v_f = \frac{\sum_{i=1}^N X_t^i \omega_f^\alpha(X_t^i)}{\sum_{i=1}^N \omega_f(\alpha)(X_t^i)}, \quad \text{with} \quad \omega_f^\alpha(x) = \exp(-\alpha f(x)),$$

supplemented with random initial data $\rho_0 = \text{law}(X_0^i)$. We use the notation

f	objective function,	α	weight parameter,
λ	drift parameter,	B_t^i	Brownian Motion,
σ	diffusion parameter,	H	Heaviside function.

Note:

- v_f is our approximation of x^*
- interactions scale **linear** in N
- analysis with $H \equiv 1$
- assumption: unique x^*

Combine **swarm optimization** and **opinion dynamics** to obtain:

$$dX_t^i = -\lambda(X_t^i - v_f)\, H[f(X_t^i) - f(v_f)]dt + \sigma|X_t^i - v_f|dB_t^i, \quad i = 1, \dots, N,$$

where the **weighted average** v_f is given by

$$v_f = \frac{\sum_{i=1}^N X_t^i \omega_f^\alpha(X_t^i)}{\sum_{i=1}^N \omega_f(\alpha)(X_t^i)}, \quad \text{with} \quad \omega_f^\alpha(x) = \exp(-\alpha f(x)),$$

supplemented with random initial data $\rho_0 = \text{law}(X_0^i)$. We use the notation

f	objective function,	α	weight parameter,
λ	drift parameter,	B_t^i	Brownian Motion,
σ	diffusion parameter,	H	Heaviside function.

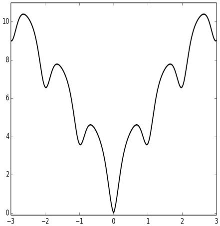
Note:

- v_f is our approximation of x^*
- interactions scale **linear** in N
- analysis with $H \equiv 1$
- assumption: unique x^*

Let $f : \mathbb{R}^D \to \mathbb{R}^+$, $D \geq 1$, be a function. The problem we consider is given by

$$\operatorname*{argmin}_{x \in \mathbb{R}^D} f(x).$$

<u>Task:</u> Construct an algorithm that finds the global minimizer of f!



<u>Features we like:</u>

- global minimum
- particle scheme
- no gradient information
- indistiguishable particles
- mean-field equation for analysis

Global optimization

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}
\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x) \\
&= -\int f(x)(f(x) - f_*)d\eta^\alpha(x) + \int f(y)d\eta^\alpha(y)\int (f(z) - f_*)d\eta^\alpha(z) \\
&= -\int f(x)^2 d\eta^\alpha(x) + \int f(x)d\eta^\alpha(x)\int f(y)d\eta^\alpha(y) \\
&= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x)d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}
\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}
\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x) \\
&= -\int f(x)(f(x) - f_*) d\eta^\alpha(x) + \int f(y) d\eta^\alpha(y) \int (f(z) - f_*) d\eta^\alpha(z) \\
&= -\int f(x)^2 d\eta^\alpha(x) + \int f(x) d\eta^\alpha(x) \int f(y) d\eta^\alpha(y) \\
&= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x) d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}
\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}
\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x) \\
&= -\int f(x)(f(x) - f_*)d\eta^\alpha(x) + \int f(y)d\eta^\alpha(y)\int (f(z) - f_*)d\eta^\alpha(z) \\
&= -\int f(x)^2 d\eta^\alpha(x) + \int f(x)d\eta^\alpha(x)\int f(y)d\eta^\alpha(y) \\
&= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x)d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}
\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x)\\ &= -\int f(x)(f(x) - f_*)d\eta^\alpha(x) + \int f(y)d\eta^\alpha(y)\int (f(z) - f_*)d\eta^\alpha(z)\\ &= -\int f(x)^2 d\eta^\alpha(x) + \int f(x)d\eta^\alpha(x)\int f(y)d\eta^\alpha(y)\\ &= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x)d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x) \\ &= -\int f(x)(f(x) - f_*)d\eta^\alpha(x) + \int f(y)d\eta^\alpha(y)\int (f(z) - f_*)d\eta^\alpha(z) \\ &= -\int f(x)^2 d\eta^\alpha(x) + \int f(x)d\eta^\alpha(x)\int f(y)d\eta^\alpha(y) \\ &= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x)d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}
\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x) \\
&= -\int f(x)(f(x) - f_*)d\eta^\alpha(x) + \int f(y)d\eta^\alpha(y)\int (f(z) - f_*)d\eta^\alpha(z) \\
&= -\int f(x)^2 d\eta^\alpha(x) + \int f(x)d\eta^\alpha(x)\int f(y)d\eta^\alpha(y) \\
&= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x)d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}
\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 1 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \mathrm{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof: (Lyapunov argument) By construction $\eta^\alpha = e^{-\alpha f(x)}\rho(x)/\|e^{-\alpha f}\|_{L^1(\rho)} \in \mathcal{P}^{ad}(\mathbb{R}^d)$. We first show that the functional

$$\mathcal{E}_\alpha(f) := \int_{\mathbb{R}^d} f d\eta^\alpha \longrightarrow f_* = f(x_*) \quad \text{as} \quad \alpha \to \infty.$$

Indeed, it holds

$$\begin{aligned}
\frac{d}{d\alpha}\mathcal{E}_\alpha(f - f_*) &= \frac{d}{d\alpha}\int f - f_* d\eta^\alpha(x) \\
&= -\int f(x)(f(x) - f_*)d\eta^\alpha(x) + \int f(y)d\eta^\alpha(y)\int (f(z) - f_*)d\eta^\alpha(z) \\
&= -\int f(x)^2 d\eta^\alpha(x) + \int f(x)d\eta^\alpha(x)\int f(y)d\eta^\alpha(y) \\
&= -\frac{1}{2}\int |f(x) - f(y)|^2 d\eta^\alpha(x)d\eta^\alpha(y) < 0 \text{ (strictly as } f \text{ nonconst)}
\end{aligned}$$

Since $\mathcal{E}_\alpha(f) \geq f_*$ for all α this implies $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.

Proposition 2 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

Proof (continued): We have $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.
In the second step we show that $\eta^\alpha \to \delta_{x_*}$ in the sense of distributions.
Let $\epsilon > 0$. By Chebyshev inequality we obtain

$$\begin{aligned}\eta^\alpha(\{x \in \mathbb{R}^d : f(x) - f_* \geq \epsilon\}) &\leq \frac{1}{\epsilon}\int_{\{f-f_*\geq\epsilon\}} (f - f_*) d\eta^\alpha(x) \\ &\leq \frac{1}{\epsilon}\mathcal{E}_\alpha(f - f_*) \longrightarrow 0 \quad \text{as } \alpha \to \infty.\end{aligned}$$

Hence, $\eta^\alpha \to \delta_{x_*}$ which implies $v_f \to x_*$.

Remark: This is the reason why the weights in the weighted mean of CBO methods are usually chosen as $e^{-\alpha f(x)}$.

Proposition 2 (Approximation of the global minimizer)

Assume that $f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}^+)$ attains a unique global minimum at x_* and let $\rho \in \mathcal{P}^{ac}(\mathbb{R}^d)$ with $x_* \in \operatorname{supp}(\rho)$. Then we have that $\lim_{\alpha\to\infty} v_f^\alpha = x_*$.

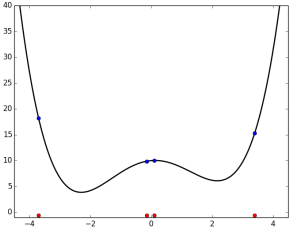
Proof (continued): We have $\mathcal{E}_\alpha(f) \to f_*$ as $\alpha \to \infty$.
In the second step we show that $\eta^\alpha \to \delta_{x_*}$ in the sense of distributions.
Let $\epsilon > 0$. By Chebyshev inequality we obtain

$$\begin{aligned}\eta^\alpha(\{x \in \mathbb{R}^d \colon f(x) - f_* \geq \epsilon\}) &\leq \frac{1}{\epsilon}\int_{\{f-f_*\geq\epsilon\}} (f - f_*)d\eta^\alpha(x)\\ &\leq \frac{1}{\epsilon}\mathcal{E}_\alpha(f - f_*) \longrightarrow 0 \quad \text{as } \alpha \to \infty.\end{aligned}$$

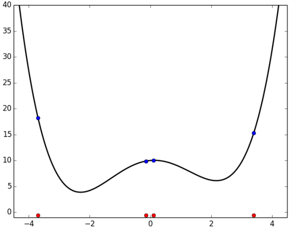
Hence, $\eta^\alpha \to \delta_{x_*}$ which implies $v_f \to x_*$.

Remark: This is the reason why the weights in the weighted mean of CBO methods are usually chosen as $e^{-\alpha f(x)}$.

Importance of Scaled Stochasticity



$N = 4,\ \lambda = 1,\ \sigma = 0.0,\ \alpha = 30$
No stochastics!
level set at 9.7



$N = 4,\ \lambda = 1,\ \sigma = 0.7,\ \alpha = 30$
with stochastics

no scaling $\rightarrow$ no consensus!

Importance of Scaled Stochasticity

There is no local best or global best involved in the dynamic:

Mean-Field system

As $N \to \infty$ the SDE system turns into the McKean non-linear Process

$$d\bar{X}_t = -\lambda(\bar{X}_t - v_f)H[f(X_t) - f(v_f)]dt + \sqrt{2}\sigma|\bar{X}_t - v_f|dB_t,$$

with initial data $\text{law}(\bar{X}_0) = \rho_0$. Here the weighted average is given by

$$v_f = \frac{1}{\int_{\mathbb{R}^d} \omega_f^\alpha d\rho_t} \int_{\mathbb{R}^D} x\omega_f^\alpha d\rho_t, \qquad \rho_t = \text{law}(\bar{X}_t),$$

which can be expressed equivalently by the **non-local**, **non-linear** PDE

$$\partial_t \rho_t = \Delta(\kappa\rho_t) + \nabla \cdot (\mu\rho_t),$$

with $\kappa = \sigma^2|x - v_f|^2$ and $\mu = -\lambda(x - v_f)H[f(x) - f(v_f)]$.

Remark: Derivation is standard argument - Itô's formula and $\rho_t = \text{law}(\bar{X}_t)$.

For the analysis we consider the scheme **without Heaviside function.**

Theorem 1 (Uniform Consensus)

Let $\rho_0 \in \mathcal{P}^2(\mathbb{R}^D)$ and $f \in Lip_{loc}(R^D)$, $\inf f > 0$ satisfy the following additional conditions:

(i) *there exist constants L_f and $c_u, c_l > 0$ such that*

$$|f(x) - f(y)| \leq L_f(|x| + |y|)|x - y| \quad \text{for all } x, y \in \mathbb{R}^d,$$

$$f(x) - \inf f \leq c_u(1 + |x|^2) \quad \text{for all } x \in \mathbb{R}^d, \text{ (bounded } f\text{)}$$

$$f(x) - \inf f \geq c_l|x|^2 \quad \text{for all } |x| > M \text{ (unbounded } f\text{)}$$

(ii) *There exist constants $M, c_0, c_1, c_f > 0$ such that*

$$\|\nabla f\|_\infty \leq c_f, \qquad \Delta f \leq c_0 + c_1|\nabla f|^2 \quad \text{in } \mathbb{R}^D.$$

Then there exist constants α and $\lambda > 0$ such that we obtain uniform consensus for ρ_t as $t \to \infty$ arbitrary close to the global minimizer.

In other words: The invariant measure of this process is a $\delta_{\hat{x}}(x)$ positioned close to the global minimum of the objective, $\hat{x} \in B_\epsilon(x^*)$.

Strategy of the proof:

(i) show concentration at $\hat{x}$ for some $\hat{x} \in \mathbb{R}^d$:
compute evolution of $V(\rho_t)$ and $E(\rho_t)$ and use Chebyshev's inequality

(ii) show $\hat{x} \in B_\epsilon(x_*)$.
similar idea to Laplace principle

No rigorous result for the mean-field limit

- „cured" by an compactness result
- quantitative estimate in N still open
- different amount of information on SDE / PDE level

Numerics

$N = 10,\ \lambda = 2,\ \sigma = 0.7,\ \alpha = 30$

Numerics

Some comments on variants

So far we had the dynamics given by

$$dX_t^i = -\lambda(X_t^i - v_f)\ H[f(X_t^i) - f(v_f)]dt + \sigma|X_t^i - v_f|dB_t^i.$$

Variant by Carrillo, Jin, Lei, Zhu (2021)

The isotropic independent diffusion is replaced by an **anisotropic** independent diffusion leading to

$$dX_t^i = -\lambda(X_t^i - v_f)dt + \sigma \mathrm{diag}(X_t^i - v_f)dB_t^i, \quad i = 1, \ldots, N.$$

Moreover, they propose to use **mini-batches** for the computation of v_f and the updates.

Advantages:

- dimension independent estimates $\Rightarrow$ robust in high dimensions
- mini-batches significantly reduce the computational cost
- mini-batches are another stochastic influence
- stay in the (sub)space of the initial particle crowd

Based on the previous version is the following variant

Variant by Ha, Jin, Kim (2020)

The anisotropic independent diffusion is replaced by an anisotropic **common** diffusion leading to

$$dX_t^i = -\lambda(X_t^i - v_f)dt + \sigma \mathrm{diag}(X_t^i - v_f)dB_t, \quad i = 1, \dots, N.$$

Moreover, the article states a **time discrete version** of the common noise scheme.

Advantages:

- thanks to the common noise, it is easier to study the distance of two particles. In fact, it holds

$$\mathbb{E}|X^i(t) - X^j(t)|^2 = e^{-(2\lambda - \sigma^2)t}\mathbb{E}|X_0^i - X_0^j|^2, \quad t > 0.$$

- proof of convergence on the particle level
- convergence and error analysis for the discrete scheme (elementary arguments)

Particle swarm optimization:

$$x^i_{n+1} = x^i_n + v^i_{n+1}, \qquad v^i_{n+1} = v^i_n + c_1 R_1(y^i - x^i_n) + c_2 R_2(\bar{y} - x^i_n)$$

y^i local best, $\bar{y}$ global best

Step 1: second order without memory

$$dX^i_t = V^i_t dt,$$
$$m\, dV^i_t = -\gamma(V^i_t) + \lambda(X^i_t - v_f) + \sigma D(X^i_t - v_f) dB^i_t$$

$$\Rightarrow \partial_t f + v \cdot \nabla_x f = \nabla_v \cdot \left(\frac{\gamma}{m} v f + \frac{\lambda}{m}(x - v_f[\rho]) f + \frac{\sigma^2}{2m^2} D(x - v_f[\rho])^2 \nabla_v f \right)$$

Step 2: approximate local and global best

$$dY^i_t = \nu(X^i_t - Y^i_t) S^\beta(X^i_t, Y^i_t) dt, \qquad S^\beta(x, y) = 1 + \tanh(\beta(f(y) - f(x))),$$

$$\bar{Y}_t = \frac{1}{\sum_{i=1}^N e^{-\alpha f(Y^i_t)}} \sum_{i=1}^N Y^i_t e^{-\alpha f(Y^i_t)}$$

Particle swarm optimization:

$$x^i_{n+1} = x^i_n + v^i_{n+1}, \qquad v^i_{n+1} = v^i_n + c_1 R_1 (y^i - x^i_n) + c_2 R_2 (\bar{y} - x^i_n)$$

$$y^i \text{ local best}, \quad \bar{y} \text{ global best}$$

Step 1: second order without memory

$$\begin{aligned} dX^i_t &= V^i_t dt, \\ m\, dV^i_t &= -\gamma(V^i_t) + \lambda(X^i_t - v_f) + \sigma D(X^i_t - v_f) dB^i_t \end{aligned}$$

$$\Rightarrow \partial_t f + v \cdot \nabla_x f = \nabla_v \cdot \left(\frac{\gamma}{m} v f + \frac{\lambda}{m}(x - v_f[\rho]) f + \frac{\sigma^2}{2m^2} D(x - v_f[\rho])^2 \nabla_v f \right)$$

Step 2: approximate local and global best

$$dY^i_t = \nu(X^i_t - Y^i_t) S^\beta(X^i_t, Y^i_t) dt, \qquad S^\beta(x, y) = 1 + \tanh(\beta(f(y) - f(x))),$$

$$\bar{Y}_t = \frac{1}{\sum_{i=1}^N e^{-\alpha f(Y^i_t)}} \sum_{i=1}^N Y^i_t e^{-\alpha f(Y^i_t)}$$

Particle swarm optimization:

$$x^i_{n+1} = x^i_n + v^i_{n+1}, \qquad v^i_{n+1} = v^i_n + c_1 R_1 (y^i - x^i_n) + c_2 R_2 (\bar{y} - x^i_n)$$

y^i local best, $\quad \bar{y}$ global best

Step 1: second order without memory

$$dX^i_t = V^i_t dt,$$
$$m\, dV^i_t = -\gamma(V^i_t) + \lambda(X^i_t - v_f) + \sigma D(X^i_t - v_f) dB^i_t$$

$$\Rightarrow \partial_t f + v \cdot \nabla_x f = \nabla_v \cdot \left(\frac{\gamma}{m} v f + \frac{\lambda}{m}(x - v_f[\rho]) f + \frac{\sigma^2}{2m^2} D(x - v_f[\rho])^2 \nabla_v f \right)$$

Step 2: approximate local and global best

$$dY^i_t = \nu(X^i_t - Y^i_t) S^\beta(X^i_t, Y^i_t) dt, \qquad S^\beta(x,y) = 1 + \tanh(\beta(f(y) - f(x))),$$

$$\bar{Y}_t = \frac{1}{\sum_{i=1}^N e^{-\alpha f(Y^i_t)}} \sum_{i=1}^N Y^i_t e^{-\alpha f(Y^i_t)}$$

For simplicity we focus on the sphere Γ

$$dX_t^i = -\lambda P(X_t^i)(X_t^i - v_{\alpha,\mathcal{E}}(\rho_t^N))dt + \sigma|X_t^i - v_{\alpha,\mathcal{E}}(\rho_t^N)|P(X_t^i)dB_t^i$$
$$-\frac{\sigma^2}{2}(X_t^i - v_{\alpha,\mathcal{E}}(\rho_t^N))^2\Delta\gamma(X_t^i)\nabla\gamma(X_t^i)dt\,,$$

with projection operator $P(x) = I - \frac{xx^T}{|x|^2}$

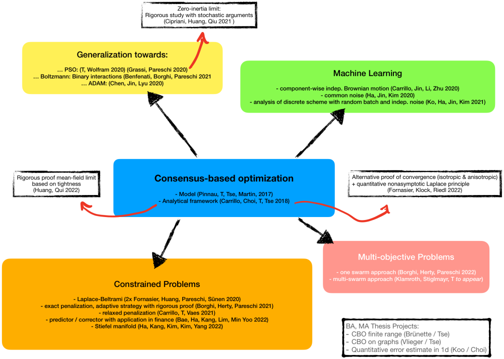
$$\partial_t\rho_t = \lambda\nabla_\Gamma\cdot(P(v)(v - v_{\alpha,\mathcal{E}}(\rho_t))\rho_t) + \frac{\sigma^2}{2}\Delta_\Gamma(|v - v_{\alpha,\mathcal{E}}(\rho_t)|^2\rho_t), \quad t > 0,\ v \in \Gamma\,,$$

where $\nabla_\Gamma, \Delta_\Gamma$ are the divergence and Laplace-Beltrami operator corresponding to the hypersurface.

Remark:

By compactness of Γ, the mean-field limit is not a big issue.

Outlook & open problems

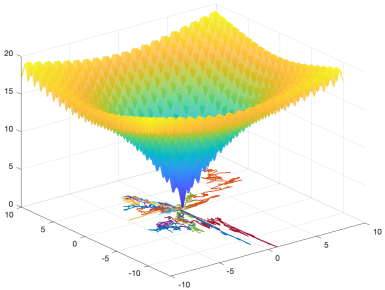


- multi-modal objectives
- rate in N
- multi-objective problems
- discrete settings
- uncertainties
- …

Thank you for your attention!

References:

- A consensus-based model for global optimization and its mean-field limit R. Pinnau, CT, O. Tse, S. Martin Mathematical Models and Methods in Applied Sciences 27 (1), pp. 183-204, 2017.
- An analytical framework for a consensus-based global optimization method J. A. Carrillo, Y.-P. Choi, CT, O. Tse Mathematical Models and Methods in Applied Sciences 28 (06), pp. 1037-1066, 2018.
- Survey: Trends in consensus-based optimization CT Active Particles, Volume 3, Springer International Publishing, 2022.



✉ totzeck@uni-wuppertal.de

Advertisement: **Consensus-Based Sampling** (Studies Appl. Math 2022)