Uniform-in-time weak propagation of chaos for consensus-based optimization

Joint work with Ibrahim Ekren and Hongyi Zhou

To appear in the Annals of Applied Probability

Erhan Bayraktar

Department of Mathematics
University of Michigan

IMSI, May 2026

Talk roadmap

Contents

Consensus-based optimization algorithm that finds the minimizer

$$x^* = \arg\min_{x \in \mathbb{R}^d} \mathcal{E}(x)$$

by "collective intelligence":

1. Large number of participants
2. Mean-field interaction
3. Consensus mechanism.

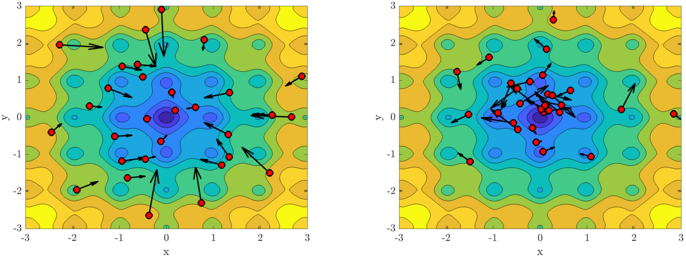


Figure: Snapshots of the PSO method, cr. [GHPQ21]

Consensus-Based Optimization algorithms

We have a system of interacting particles

$$dX_t^{N,i} = b(X_t^{N,i}, \nu_t^N)dt + a^{\frac{1}{2}}(X_t^{N,i}, \nu_t^N)dW_t^i\,, \qquad i \in [N]\,,$$

with $\nu_t^N \stackrel{\text{def}}{=} \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{N,j}}$.

- Drift b pulls each $X^{N,i}$ towards the provisional optimizer m_t^N (consensus) to be constructed from ν_t^N,
- Volatility $\|a\|$ helps avoid getting stuck at local optima and it degenerates when consensus is attained.
  - Advantage: We stay at the minimum.
  - Disadvantage: Lack of noise at one point generates new difficulties.

Consensus-based optimization: Intuition

Given the cost function $\mathcal{E}(x)$ to minimize, one can define consensus as

- best participant: $m_t^N = \arg\min\{\mathcal{E}(X_t^{N,i})\}_{i\in[N]}$, or
- a weighted average (α-regularized best):

$$m_t^N = \frac{\int x e^{-\alpha\mathcal{E}(x)}\nu_t^N(dx)}{\int e^{-\alpha\mathcal{E}(x)}\nu_t^N(dx)} = \frac{\sum_{j=1}^N X_t^{N,j} e^{-\alpha\mathcal{E}(X_t^{N,j})}}{\sum_{j=1}^N e^{-\alpha\mathcal{E}(X_t^{N,j})}}.$$

- Laplace's principle: We find the global minimizer of $\mathcal{E}$,

$$\lim_{\alpha\to+\infty}\left(-\frac{1}{\alpha}\log\int_{\mathbb{R}^d} e^{-\alpha\mathcal{E}(x)}\rho(dx)\right) = \inf_{x\in\mathrm{supp}(\rho)}\mathcal{E}(x)$$

and for any full support probability measure ρ on $\mathbb{R}^d$

$$\int e^{-\alpha\mathcal{E}(x)}\rho(dx) \approx e^{-\alpha\mathcal{E}(x^*)} \quad \text{when} \quad \alpha\to\infty.$$

- In the dynamical system $\{X^{N,i}\}_{i\in[N]}$, we use consensus

$$m_t^N = \frac{\sum_{j=1}^N X_t^{N,j} e^{-\alpha\mathcal{E}(X_t^{N,j})}}{\sum_{j=1}^N e^{-\alpha\mathcal{E}(X_t^{N,j})}}\,.$$

- As proposed by Pinnau et al. [PTTM17], this leads to

$$dX_t^{N,i} = -\lambda(X_t^{N,i} - m_t^N)dt + \sigma D(X_t^{N,i} - m_t^N)dW_t^i$$

  with initial distribution $\nu_{\text{init}}^{\otimes N}$.
- One can choose $D(x) = \text{diag}(x)$ or $D(x) = |x|\, I_{d\times d}$.
- For technical reasons we will study

$$dX_t^{N,i} = -\lambda(X_t^{N,i} - m_t^N)dt + \sigma \left|X_t^{N,i} - m_t^N\right| \phi(X_t^{N,i})dW_t^i$$

  for a cutoff function ϕ that is 0 outside of a large ball.

Consensus-Based Optimization: Particle System

- In the dynamical system $\{X^{N,i}\}_{i\in[N]}$, we use consensus

$$m_t^N = \frac{\sum_{j=1}^N X_t^{N,j} e^{-\alpha\mathcal{E}(X_t^{N,j})}}{\sum_{j=1}^N e^{-\alpha\mathcal{E}(X_t^{N,j})}}.$$

- As proposed by Pinnau et al. [PTTM17], this leads to

$$dX_t^{N,i} = -\lambda(X_t^{N,i} - m_t^N)dt + \sigma D(X_t^{N,i} - m_t^N)dW_t^i$$

  with initial distribution $\nu_{\text{init}}^{\otimes N}$.
- One can choose $D(x) = \text{diag}(x)$ or $D(x) = |x|\, I_{d\times d}$.
- For technical reasons we will study

$$dX_t^{N,i} = -\lambda(X_t^{N,i} - m_t^N)dt + \sigma \left|X_t^{N,i} - m_t^N\right| \phi(X_t^{N,i})dW_t^i$$

  for a cutoff function ϕ that is 0 outside of a large ball.
- We need to know that the minimum is inside the large ball.

Mean-field limit dynamics

- The existence and uniqueness of solutions to the system for finite N is simple; see [CCTT18].
- When $N \to \infty$, the solution on $[0, T]$ behaves as

$$d\bar{X}_t = -\lambda(\bar{X}_t - \bar{m}_t)dt + \sigma \left|\bar{X}_t - \bar{m}_t\right| \phi(\bar{X}_t)dW_t \,,$$

  where

$$\bar{m}_t = \frac{\mathbf{E}[\bar{X}_t e^{-\alpha\mathcal{E}(\bar{X}_t)}]}{\mathbf{E}e^{-\alpha\mathcal{E}(\bar{X}_t)}} = \frac{\int x e^{-\alpha\mathcal{E}(x)}\bar{\nu}_t(dx)}{\int e^{-\alpha\mathcal{E}(x)}\bar{\nu}_t(dx)} \,.$$

- [CCTT18] shows that $\lim_{t\to\infty} \mathrm{Var}(\bar{X}_t) \to 0$, so that

$$\bar{\nu}_t = \mathcal{L}(\bar{X}_t) \to \delta_{\tilde{x}}.$$

Consensus-based Optimization: Pros and Cons

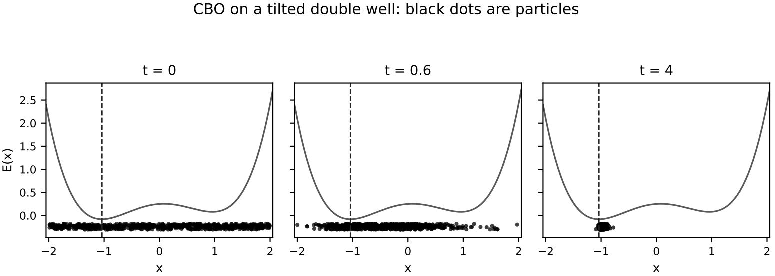
Advantages:

1. Gradient free optimization method
2. No need for convexity
3. Exponential convergence
4. Escapes local minima

Disadvantages:

1. limit point not optimal ($\tilde{x} \neq x^*$)

$$E(x) = \frac{1}{4}(x^2 - 1)^2 + 0.08x, \qquad x_\star \approx -1.04.$$



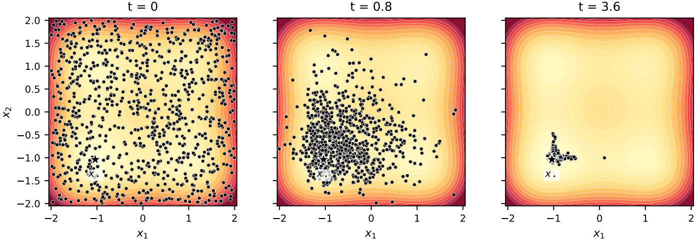
Message

The colored background is the objective heat map; the black dots are particles. The Gibbs-weighted consensus is biased toward the lower well, and the cloud contracts near the global minimizer.

Numerical illustration in two dimensions

$$E(x_1, x_2) = \frac{1}{4}\{(x_1^2 - 1)^2 + (x_2^2 - 1)^2\} + 0.08(x_1 + x_2), \qquad x_\star \approx (-1.04, -1.04).$$

Tilted four-well landscape: heat map shows objective value; black dots are particles



Message

The heat map shows the nonconvex objective landscape. The black particles first explore several wells, then the weighted consensus drives the population toward the tilted global well.

Accuracy/Error of CBO method

- We run the algorithm for finite N and finite t and we need to estimate $d(\nu_t^N, \delta_{x^*})$.
- Sources of error:
  1. Finite running time and systemic error
$$d(\bar{\nu}_t, \delta_{\tilde{x}}) \lesssim e^{-\kappa t} \quad \text{and} \quad \tilde{x} \neq x^*$$
  2. Propagation of chaos
$$\mathbf{E} d(\nu_t^N, \bar{\nu}_t) \lesssim N^{-\gamma}$$
  3. time discretization

- Large-population limit

$$d\bar{X}_t = -\lambda(\bar{X}_t - \bar{m}_t)dt + \sigma \left|\bar{X}_t - \bar{m}_t\right| \phi(\bar{X}_t)dW_t$$

  admits weak solution $(\bar{\nu}_t)_{t\geq 0}$ such that $\bar{\nu}_0 = \nu_{\text{init}}$ and

$$\|\bar{\nu}_t - \delta_{\tilde{x}}\|_{(1,\infty)'} \leq Ce^{-\kappa t}$$

  by [CCTT18].
- The limit point $\tilde{x}$ does not need be the global optimizer x^*, but $|\tilde{x} - x^*|$ is small for large α.
- Thus, for large α,

$$\lim_{t\to\infty} \bar{\nu}_t = \lim_{t\to\infty} \lim_{N\to\infty} \sim \delta_{x^*}.$$

Accuracy of CBO method: Finite particle

- Propagation of chaos (main concern of this project)
- By an abstract convergence result, Huang and Qiu [HQ22] shows that

$$(\nu_t^N)_{t\in[0,T]} \rightharpoonup (\bar{\nu}_t)_{t\in[0,T]} \quad \text{as} \quad N \to \infty$$

- Uniform convergence on $[0,T]$ is shown in Gerber, Hoffmann and Vaes [GHV23],

$$\mathbf{E}\left[\sup_{t\in[0,T]} \left|X_t^{N,i} - \bar{X}_t^i\right|^2\right] \leq C_T N^{-1}\,,$$

where $\bar{X}^i$ are i.i.d. copies of $\bar{X}$.

Accuracy of CBO method: Sum of the errors

Errors from items 1 and 2 lead to

$$\mathbf{E}\left\|\nu_t^N - \delta_{x^*}\right\|_{-s,2}^2 \leq C_2(t)N^{-\gamma} + Ce^{-\kappa t} + C_1(|\tilde{x} - x^*|)\,.$$

The procedure of parameter setting:

1. Choose large enough α so that $C_1\,|\tilde{x} - x^*| < \varepsilon$,
2. Choose t so that $Ce^{-\kappa t} < \varepsilon$,
3. Choose N so that $C_2(t)N^{-\gamma} < \varepsilon$.

Major disadvantage: choice of N highly depends on t.

Using the the centered negative Sobolev metric which allow us to ignore the issue $\tilde{x} \neq x^*$, we have

$$\mathbf{E} \left\| \nu_t^N \circ \tau_{\langle \mathsf{Id}, \nu_t^N \rangle} - \delta_0 \right\|_{-s,2}^2 \leq C_2(t) N^{-\gamma} + C e^{-\kappa t} \, .$$

The procedure of parameter setting:

1. Choose t so that $Ce^{-\kappa t} < \varepsilon$,
2. Choose N so that $C_2(t) N^{-\gamma} < \varepsilon$.

Major disadvantage: Choice of N highly depends on t.

Accuracy: Centered Fourier-Wasserstein metric

Using the the centered negative Sobolev metric which allow us to ignore the issue $\tilde{x} \neq x^*$, we have

$$\mathbf{E} \left\| \nu_t^N \circ \tau_{\langle \mathrm{Id}, \nu_t^N \rangle} - \delta_0 \right\|_{-s,2}^2 \leq C_2(t) N^{-\gamma} + C e^{-\kappa t} .$$

The procedure of parameter setting:

1. Choose t so that $C e^{-\kappa t} < \varepsilon$,
2. Choose N so that $C_2(t) N^{-\gamma} < \varepsilon$.

Major disadvantage: Choice of N highly depends on t.

Goal: Improve propagation of chaos to make C_2 uniform in t.

Uniform in time results for Langevin descent

The convergence properties of the mean-field Langevin descent are well understood. [CRW22] shows the convergence of

$$dX_t^{i,N} = -D_m F(\nu_t^N, X_t^{i,N})dt + \sigma dW_t^i$$

whose mean field limit is $dX_t = -D_m F(\bar{\nu}_t, X_t)dt + \sigma dW_t$.

1. With enough convexity, there is an exponential decay of an energy as $t \to \infty$.
2. $\bar{\nu}_t$ converges to a unique invariant measure.
3. The convergence of the finite particle system and the uniform in time convergence can also be obtained.

The convergence properties of the mean-field Langevin descent are well understood. [CRW22] shows the convergence of

$$dX_t^{i,N} = -D_m F(\nu_t^N, X_t^{i,N})dt + \sigma dW_t^i$$

whose mean field limit is $dX_t = -D_m F(\bar{\nu}_t, X_t)dt + \sigma dW_t$.

1. With enough convexity, there is an exponential decay of an energy as $t \to \infty$.
2. $\bar{\nu}_t$ converges to a unique invariant measure.
3. The convergence of the finite particle system and the uniform in time convergence can also be obtained.
4. $\bar{\nu}_t = \delta_x$ for any x is an invariant measure for the CBO algorithm, and we cannot use the methodology of mean-field Langevin descent for CBO.

Contents

Weak propagation of chaos: Objective

For a family of (nonlinear) functions $\Phi : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$, we expect

$$\sup_{t\geq 0} \left|\Phi(\bar{\nu}_t) - \mathbf{E}\Phi(\nu_t^N)\right| \to 0$$

as $N \to \infty$. This gives a constant C_2 independent of t.
Typical functions Φ of concern:

- $\Phi(\mu) = \|\mu - \rho_{\mathsf{ref}}\|_{-s,2}^2$,
- $\Phi(\mu) = \mathcal{W}_2^2(\mu, \rho_{\mathsf{ref}})$,

where ρ_{ref} is usually related to the limit $\bar{\nu}_\infty$. We will use their centered versions:

- $\Phi(\mu) = \left\|\mu \circ \tau_{\langle \mathsf{Id}, \mu\rangle} - \rho_{\mathsf{ref}} \circ \tau_{\langle \mathsf{Id}, \rho_{\mathsf{ref}}\rangle}\right\|_{-s,2}^2$,
- $\Phi(\mu) = \mathcal{W}_2^2\left(\mu \circ \tau_{\langle \mathsf{Id}, \mu\rangle}, \rho_{\mathsf{ref}} \circ \tau_{\langle \mathsf{Id}, \rho_{\mathsf{ref}}\rangle}\right)$,

We use the methodology developed by [DT25].

Theorem (uniform-in-time weak propagation of chaos)

Suppose $\Phi : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ *satisfies the following conditions:*

1. *There exists some constant* C_Φ *such that*

$$\sup_{\mu \in \mathcal{P}(\mathbb{R}^d)} \left\| \frac{\delta \Phi}{\delta m}(\mu) \right\|_{6,\infty} < C_\Phi \,, \qquad \sup_{\mu \in \mathcal{P}(\mathbb{R}^d)} \left\| \frac{\delta^2 \Phi}{\delta m^2}(\mu) \right\|_{6,\infty} < C_\Phi \,,$$

2. *and*

$$\left.\frac{d}{dz}\right|_{z=0} \Phi(\mu \circ \tau_z^{-1}) = 0 \qquad \forall \mu \in \mathcal{P}(\mathbb{R}^d) \,,$$

   where $\tau_z : x \mapsto x + z$ *is the translation operator.*

Then for all $N \geq 2$ *we have*

$$\sup_{t \geq 0} \left| \mathbf{E}[\Phi(\nu_t^N)] - \Phi(\bar{\nu}_t) \right| \leq \frac{C_{\mathsf{main}}}{N} \,.$$

Corollary

Let $s > \frac{d+7}{2}$. There exist some constants $C_{FW}, C_W > 0$ such that

$$\mathbf{E}\left\|\nu_t^N \circ \tau_{\langle \mathsf{Id}, \nu_t^N \rangle} - \delta_0\right\|_{-s,2}^2 \leq C_{FW}(N^{-1} + e^{-2\kappa t})$$

and

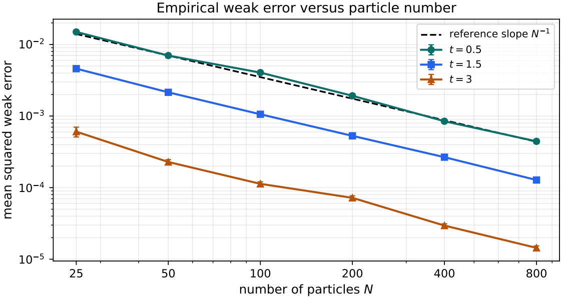
$$\mathbf{E}\mathcal{W}_2^2(\nu_t^N \circ \tau_{\langle \mathsf{Id}, \nu_t^N \rangle}, \delta_0) \leq C_W(N^{-1} + e^{-\kappa t})$$

for all $N \geq 2$ and $t \geq 0$.

- Decay of variance of $\bar{\nu}_t$ leads to decay of variance of ν_t^N up to a uniform error of size N^{-1}.

Numerical check: convergence in N

$$\mathrm{Err}_N(t) := \frac{1}{M}\sum_{m=1}^{M}\left|\left\langle \varphi, \nu_t^{N,m}\right\rangle - \langle \varphi, \bar{\nu}_t\rangle\right|^2, \qquad \varphi(x) = \tanh(x) + 0.15\cos(3x).$$



Takeaway

For several time horizons, the empirical squared weak error decreases approximately like N^{-1}.

Contents

Toolbox: Nonlinear Fokker-Planck Equation

Recall that $(\bar{\nu}_t)_{t\geq 0}$ uniquely solves the nonlinear Fokker-Planck equation (FPE)

$$\partial_t \mu_t = \mathcal{L}^*_{\mu_t} \mu_t \stackrel{\mathsf{def}}{=} -\nabla \cdot (b(\cdot, \mu_t)\mu_t) + \frac{1}{2}\operatorname{tr}[\nabla^2 (a(\cdot, \mu_t)\mu_t)]$$

with initial data ν_{init}, where

$$b(x,\mu) = -\lambda(x - M(\mu))\,, \qquad a(x,\mu) = \sigma^2 \phi(x)^2 \left|x - M(\mu)\right|^2\,,$$

with the consensus operator defined via

$$M(\mu) \stackrel{\mathsf{def}}{=} \frac{\int y e^{-\alpha \mathcal{E}(y)} \mu(dy)}{\int e^{-\alpha \mathcal{E}(y)} \mu(dy)}\,.$$

- We need to analyze the temporal error

$$\mathbf{E}\Phi(\nu_t^N) - \Phi(\bar{\nu}_t)\,.$$

- Define the solution operator of FPE by $\bar{\nu}_t$ as $P_t\nu_{\text{init}} := \bar{\nu}_t$..
- Define $\mathcal{U} : [0,\infty) \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ by $\mathcal{U}(t,\mu) = \Phi(P_t\mu)$ which solves the master equation

$$\partial_t \mathcal{U}(t,\mu) = \int_{\mathbb{R}^d} \Big(b(x,\mu)^\top \nabla_x \frac{\delta \mathcal{U}}{\delta m}(t,\mu,x) + \frac{1}{2}\operatorname{tr}\Big[a(x,\mu) \nabla_x^2 \frac{\delta \mathcal{U}}{\delta m}(t,\mu,x) \Big] \Big) \mu(dx)$$

where the derivatives of function $f : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ in a measure argument defined as in [CDLL19, CD18]:

$$\int \frac{\delta f}{\delta m}(\mu,x)(\tilde{\mu} - \mu)(dx) = \frac{d}{d\varepsilon}\left[f(\mu + \varepsilon(\tilde{\mu} - \mu)) \right]_{\varepsilon = 0^+}\,.$$

Analysis of temporal error

Following the methodology of [DT25, CD18], the temporal error

$$\mathbf{E}\Phi(\nu_t^N) - \Phi(\bar{\nu}_t) = \mathbf{E}[\mathcal{U}(0, \nu_t^N)] - \mathcal{U}(t, \nu_{\text{init}}) \,, \tag{3.1}$$

is decomposed as

$$\begin{aligned} &\mathbf{E}\left[\mathcal{U}(0, \nu_t^N)\right] - \mathcal{U}(t, \nu_{\text{init}}) \\ &= \underbrace{\mathbf{E}\left[\mathcal{U}(t, \nu_0^N)\right] - \mathcal{U}(t, \nu_{\text{init}})}_{\text{error due to initial condition}} + \underbrace{\mathbf{E}\left[\mathcal{U}(0, \nu_t^N) - \mathcal{U}(t, \nu_0^N)\right]}_{\text{error due to difference in dynamics}} \quad . \end{aligned}$$

Objective function: Static Part

Thanks to Theorem 2.14 of [CST22] the error due to the difference in initial conditions can be written as

$$\mathbf{E}[\mathcal{U}(t,\nu_0^N)-\mathcal{U}(t,\nu_{\mathsf{init}})]=\frac{1}{N}\int_0^1\int_0^1$$
$$\mathbf{E}\left[s_1\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\tilde{\mu}^N_{s_1,s_2},\tilde{\eta},\tilde{\eta})-s_1\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\tilde{\mu}^N_{s_1,s_2},\tilde{\eta},X_0^{N,1})\right]ds_1 ds_2\,,$$

with $\tilde{\eta}\sim\nu_{\mathsf{init}}$ independent of $(X_0^{N,j})_{j=1}^N$ and

$$\tilde{\mu}^N_{s_1,s_2}\overset{\mathsf{def}}{=}(1-s_1)\nu_{\mathsf{init}}+s_1\left(\nu_0^N+\frac{s_2}{N}(\delta_{\tilde{\eta}}-\delta_{X_0^{N,1}})\right)\,.$$

Objective function: Dynamic Part

In order to control the error due to initial conditions, we apply Ito's formula to $\mathcal{U}(0,\nu_t^N)-\mathcal{U}(t,\nu_0^N)$ and using the master equation obtain

$$\mathbf{E}[\mathcal{U}(0,\nu_t^N)-\mathcal{U}(t,\nu_0^N)] = \frac{1}{2N}\sum_{i=1}^{d}\int_0^t$$
$$\mathbf{E}\left[\int_{\mathbb{R}^d} a(z,\nu_s^N)\partial_{(x_2)_i}\partial_{(x_1)_i}\frac{\delta^2\mathcal{U}}{\delta m^2}(t-s,\nu_s^N,z,z)\nu_s^N(dz)\right]ds\,.$$

Thus, we need uniform (in t,μ,z) bound of

$$\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\mu,z,z)$$

and uniform (in μ,z) exponential decay (in t) of

$$\partial_{(x_2)_i}\partial_{(x_1)_i}\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\mu,z,z).$$

Linearized Fokker-Planck equation

- Fokker-Planck equation is associated to the linear operator

$$\mathcal{L}_\mu f(x) = \int \left(\frac{1}{2}\,\mathrm{tr}[a(x,\mu)\nabla^2 f(x)] + b(x,\mu)^\top \nabla f(x)\right)\mu(dx)\,.$$

- As in [Tse21], we linearize the nonlinearity in μ as

$$\begin{aligned} q \mapsto & \frac{1}{2}\int \mathrm{tr}(a(x,\mu)\nabla^2 f(x))q(dx) + \int b(x,\mu)^\top \nabla f(x) q(dx) \\ & + \frac{1}{2}\int \mathrm{tr}\left(\frac{\delta a}{\delta m}(x,\mu)(q)\nabla^2 f(x)\right)\mu(dx) \\ & + \int \frac{\delta b}{\delta m}(x,\mu)(q)\cdot \nabla f(x)\mu(dx). \end{aligned}$$

- Define the linearized Fokker-Planck equation (L-FPE) via

$$\partial_t q_t = \mathcal{L}^*_{\mu_t} q_t + \mathcal{A}^*_{\mu_t} q_t + r_t\,.$$

and denote by $(q_t)_{t\geq 0} = \text{Linear-}[\mu, q_0, (r_t)_{t\geq 0}]$.

- We can decompose $\frac{\delta^2 \mathcal{U}}{\delta m^2}$ as in [DT25]

$$\frac{\delta^2 \mathcal{U}}{\delta m^2}(t, \mu_0, z_1, z_2) = \frac{\delta^2 \Phi}{\delta m^2}(\mu_t)(m^{(1)}(t; \mu, \delta_{z_1}), m^{(1)}(t; \mu, \delta_{z_2})) + \frac{\delta \Phi}{\delta m}(\mu_t)(m^{(2)}(t; \mu, \delta_{z_1}, \delta_{z_2})),$$

where

$$m^{(1)}(\cdot; \mu, \nu) \overset{\text{def}}{=} \text{Linear-}[\mu, \nu - \mu_0, 0],$$
$$m^{(2)}(\cdot; \mu, \nu_1, \nu_2) \overset{\text{def}}{=} \text{Linear-}[\mu, \mu_0 - \nu_2, \text{Source-}[\mu, m^{(1)}(\cdot; \mu, \nu_1), m^{(1)}(\cdot; \mu, \nu_2)]].$$

- We can also decompose $\partial_{(z_1)_i}\partial_{(z_2)_j}\frac{\delta^2\mathcal{U}}{\delta m^2}$ by [DT25]

$$
\begin{aligned}
&\partial_{(z_2)_j}\partial_{(z_1)_i}\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\mu_0,z_1,z_2) = \\
&\quad \frac{\delta^2\Phi}{\delta m^2}(\mu_t)(d_i^{(1)}(t;\mu,z_1), d_j^{(1)}(t;\mu,z_2)) + \frac{\delta\Phi}{\delta m}(\mu_t)(d_{i,j}^{(2)}(t;\mu_0,z_1,z_2)),
\end{aligned}
$$

where

$$
\begin{aligned}
&d_j^{(1)}(\cdot;\mu,z) \overset{\text{def}}{=} \text{Linear-}[\mu, -\partial_{x_j}\delta_z, 0], \\
&d_{i,j}^{(2)}(\cdot;\mu,z_1,z_2) \overset{\text{def}}{=} \\
&\qquad \text{Linear-}[\mu, 0, \text{Source-}[\mu, d_i^{(1)}(\cdot;\mu,z_1), d_j^{(1)}(\cdot;\mu,z_2)]].
\end{aligned}
$$

- Thus, uniform in time convergence can be obtained by only studying the decay properties of the (L-FPE).

- The proof relies on considering the adjoint backward equation

$$\begin{cases} \partial_s w_s + \mathcal{L}_{\mu_s} w_s + \mathcal{A}_{\mu_s} w_s = 0\,, \quad s \in [0,t] \\ w_t = \xi\,, \end{cases}$$

and the duality argument

$$\|q_t\|_{(n,\infty)'} = \sup_{\|\xi\|_{n,\infty}\le 1} \langle \xi, q_t\rangle = \sup_{\|\xi\|_{n,\infty}\le 1} \langle w_0^{t,\xi}, q_0\rangle + \int_0^t \langle w_s^{t,\xi}, r_s\rangle ds\,.$$

- Regularity assumptions: Φ is C^2 with bounded derivatives; the linearized semigroup has spectral gap κ.
- Mechanism: Adjoint PDE $\Rightarrow \kappa$ (spectral gap) $\Rightarrow$ derivative decay $\Rightarrow O(N^{-1})$
- With regularity on Φ, one gets the desired upper bound and exponential decay

$$\left|\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\mu_0,z_1,z_2)\right| \leq C,\ \left|\partial_{(z_2)_j}\partial_{(z_1)_i}\frac{\delta^2\mathcal{U}}{\delta m^2}(t,\mu_0,z_1,z_2)\right| \leq Ce^{-\kappa t}\,.$$

- Given the decomposition of the temporal error this leads to

$$\sup_{t\geq 0}\left|\mathbf{E}[\Phi(\nu_t^N)] - \Phi(\bar{\nu}_t)\right| \leq O(N^{-1})\,.$$

Contents

- For appropriate choice of parameters, the CBO algorithm enjoys uniform in time propagation of chaos

$$\mathbf{E}\left\|\nu_t^N \circ \tau_{\langle \mathsf{Id}, \nu_t^N\rangle} - \delta_0\right\|_{-s,2}^2 \le C_{FW}(N^{-1} + e^{-2\kappa t}).$$

- We decompose $\Phi(\bar{\nu}_t) - \Phi(\nu_t^N)$ via the master equation.
- The decomposition can be written with only (L-FPE).
- Due to non-degeneracy, (L-FPE) only enjoys exponential decays up to a projection.
- We use a duality argument to show this decay.

Summary

- For appropriate choice of parameters, the CBO algorithm enjoys uniform in time propagation of chaos

$$\mathbf{E}\left\|\nu_t^N \circ \tau_{\langle \mathsf{Id}, \nu_t^N \rangle} - \delta_0\right\|_{-s,2}^2 \leq C_{FW}(N^{-1} + e^{-2\kappa t}).$$

- We decompose $\Phi(\bar{\nu}_t) - \Phi(\nu_t^N)$ via the master equation.
- The decomposition can be written with only (L-FPE).
- Due to non-degeneracy, (L-FPE) only enjoys exponential decays up to a projection.
- We use a duality argument to show this decay.
- The estimate provides a clear reference to balance between the size of the system of particle and the length of simulation.
- CBO can solve high-dimensional non-convex optimization problem and one can rigorously prove the convergence.

Reference I

José A. Carrillo, Young-Pil Choi, Claudia Totzeck, and Oliver Tse.
An analytical framework for consensus-based global optimization method.
Mathematical Models and Methods in Applied Sciences, 28(06):1037–1066, 2018.

René Carmona and Francois Delarue.
Probabilistic Theory of Mean Field Games with Applications I. Mean Field FBSDEs, Control, and Games.
Probability Theory and Stochastic Modelling. Springer Cham, New York, NY, 2018.

Pierre Cardaliaguet, Francois Delarue, Jean-Michel Lasry, and Pierre-Louis Lions.
The Master Equation and the Convergence Problem in Mean Field Games.
Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, 2019.

Reference II

Fan Chen, Zhenjie Ren, and Songbo Wang.
Uniform-in-time propagation of chaos for mean field langevin dynamics.
arXiv preprint arXiv:2212.03050, 2022.

Jean-Francois Chassagneux, Lukasz Szpruch, and Alvin Tse.
Weak quantitative propagation of chaos via differential calculus on the space of measures.
The Annals of Applied Probability, 32(3):1929 – 1969, 2022.

Francois Delarue and Alvin Tse.
Uniform in time weak propagation of chaos on the torus.
Annales de l'Institut Henri Poincaré, 2025.

Sara Grassi, Hui Huang, Lorenzo Pareschi, and Jinniao Qiu.
Mean-field particle swarm optimization, 2021.

Reference III

Nicolai Jurek Gerber, Franca Hoffmann, and Urbain Vaes.
Mean-field limits for consensus-based optimization and sampling, 2023.

Hui Huang and Jinniao Qiu.
On the mean-field limit for the consensus-based optimization.
Mathematical Methods in the Applied Sciences, 45(12):7814–7831, 2022.

René Pinnau, Claudia Totzeck, Oliver Tse, and Stephan Martin.
A consensus-based model for global optimization and its mean-field limit.
Mathematical Models and Methods in Applied Sciences, 27(01):183–204, 2017.

Alvin Tse.
Higher order regularity of nonlinear fokker-planck pdes with respect to the measure component.
Journal de Mathématiques Pures et Appliquées, 150:134–180, 2021.

Thank you!

Figure: paper@arxiv: 2502.00582, to appear in Annals of Applied Probability