Uniform-in-Time Propagation of Chaos for Consensus-Based Optimization

SIAM Conference on Uncertainty Quantification (UQ)

Dohyeon Kim

Joint work with
Nicolai Gerber (University of Ulm), **Franca Hoffmann** (Caltech), **Urbain Vaes** (INRIA)

22 March 2026

Caltech

Let's optimize

Consider $f : \mathbb{R}^d \to \mathbb{R}$.

Goal: Find x_* such that $f(x_*) = \min_x f(x)$.

Challenges:

- f is non-convex
- f is only given as a black-box oracle
- f is not differentiable, or derivatives are too costly
- Additional constraints may apply

Interacting Particle Systems

Approach: Evolve a cloud of particles instead of one point.

Consider $J-$particles $x_1(t), \dots, x_J(t) \in \mathbb{R}^d$ satisfying

$$\dot{x}_j(t) = g_j\big(t, x_1(t), \dots, x_J(t)\big),$$

(and stochastic variants) such that $x_j(t) \to x_*$ as $t \to \infty$.

- Goal: Design dynamics where group self-organizes to regions of lower function-values
- Features: Parallelizable, amenable to Mean-field Analysis ($J \to \infty$)

Mean-field viewpoint: Analyze algorithm by tracking probability density instead of individual particles

Interacting particle methods: Gradient-based vs. Derivative-free

Gradient-based

- Stein variational gradient descent (SVGD)[a]
- Swarm-Based Gradient Descent[b]
- Other particle methods using ∇f or score information

Require gradient, score information to guide particle motion.

[a] Q. Liu and D. Wang. In Advances in Neural Information Processing Systems 29, 2016.
[b] J. Lu, E. Tadmor, and A. Zenginoglu. 2022.

Derivative-free

- **Particle Swarm Optimization (PSO)**[a]: classical swarm-based global optimizer
- Sequential Monte Carlo (SMC)
- Ensemble Kalman inversion (EKI)
- **Consensus-Based Optimization (CBO)**[b]
- Consensus-Based Sampling (CBS)[c]

Rely on **particle interaction, weighting, or function values** rather than direct gradient access.

[a] J. Kennedy and R. Eberhart. In Proceedings of ICNN'95-international conference on neural networks. IEEE, 1995.
[b] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. Math. Models Methods Appl. Sci., 2017.
[c] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. Stud. Appl. Math., 2022.

Interacting Particle System

$$\mathrm{d}X_t^{(j)} = \underbrace{-\left(X_t^{(j)} - \mathcal{M}_{\alpha}(\mu_t^J)\right)\mathrm{d}t}_{\text{drift term}} + \underbrace{\sqrt{2}\sigma\left|X_t^{(j)} - \mathcal{M}_{\alpha}(\mu_t^J)\right|\mathrm{d}W_t^{(j)}}_{\text{noise term}}, \qquad j = 1, \ldots, J.$$

where $W_t^{(j)}$ is a d-dimensional Brownian motion, and

$$\mathcal{M}_\alpha(\mu) = \frac{\int x\, e^{-\alpha f(x)}\,\mu(\mathrm{d}x)}{\int e^{-\alpha f(x)}\,\mu(\mathrm{d}x)}, \qquad \mu \in \mathcal{P}(\mathbb{R}^d).$$

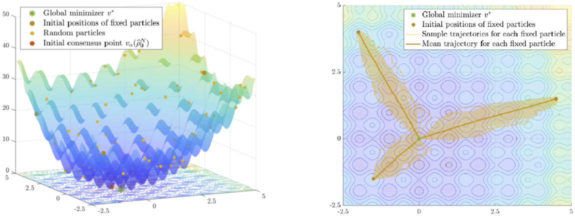
Balance of two forces:

- drift toward consensus $\to$ **exploitation**
- multiplicative noise $\to$ **exploration**

Consensus-Based Optimization (CBO)

Interacting Particle System

$$\mathrm{d}X_t^{(j)} = \underbrace{-\left(X_t^{(j)} - \mathcal{M}_{\color{red}\alpha}(\mu_t^J)\right)\mathrm{d}t}_{\text{drift term}} + \underbrace{\sqrt{2}\sigma\left|X_t^{(j)} - \mathcal{M}_{\color{red}\alpha}(\mu_t^J)\right|\mathrm{d}W_t^{(j)}}_{\text{noise term}}, \qquad j = 1, \dots, J.$$



(a) The Rastrigin function $\mathcal{E}$ and an exemplary initialization for one run of the experiment

(b) Individual agents follow, on average, the gradient flow of the map $v \mapsto \|v - v^*\|_2^2$

Figure: Excerpted from[a]

[a] M. Fornasier, T. Klock, and K. Riedl. SIAM J. Optim., 2024.

Consensus-Based Optimization (CBO)

$$\mathrm{d}X_t^{(j)} = -\left(X_t^{(j)} - \mathcal{M}_{\alpha}(\mu_t^J)\right)\mathrm{d}t + \sqrt{2}\sigma\left|X_t^{(j)} - \mathcal{M}_{\alpha}(\mu_t^J)\right|\mathrm{d}W_t^{(j)}. \quad j = 1, \dots, J.$$

Advantages:

- Derivative-free
- Theoretical guarantees via mean-field analysis

Key tool: **Laplace's principle** for any $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathcal{M}_{\alpha}(\mu) \to x_* \quad \text{as} \quad \alpha \to \infty \quad \text{if} \quad x_* \in \operatorname{supp} \mu\,.$$

Mean-field equations:

- **SDE:** $\overline{X}_t \in \mathbb{R}^d$ solves

$$\mathrm{d}\overline{X}_t = -\left(\overline{X}_t - \mathcal{M}_\alpha(\mu_t)\right)\mathrm{d}t + \sqrt{2}\sigma\left|\overline{X}_t - \mathcal{M}_\alpha(\mu_t)\right|\mathrm{d}W_t, \qquad \mu_t = \mathrm{Law}(\overline{X}_t)$$

- **PDE**: $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ solves

$$\partial_t \mu_t = \nabla\cdot\left(\left(x - \mathcal{M}_\alpha(\mu_t)\right)\mu_t\right) + \sigma^2 \triangle\left(\left|x - \mathcal{M}_\alpha(\mu_t)\right|^2 \mu_t\right).$$

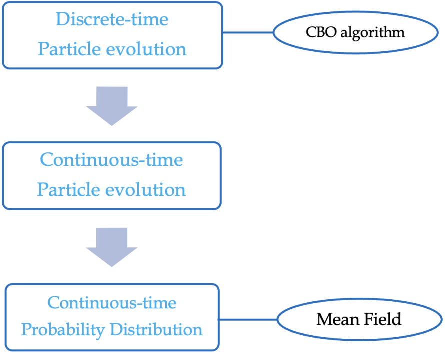
Convergence of the mean-field solution[1,2]

(1) **Consensus formation:** $\mathcal{M}_\alpha(\mu_t) \to \widehat{x}(\alpha)$ as $t \to \infty$ (exponential rate).

(2) **Laplace's method:** $\widehat{x}(\alpha) \to \arg\min_{x\in\mathbf{R}^d} f(x)$ as $\alpha \to \infty$.

[1] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. Mathematical Models and Methods in Applied Sciences, 2018.
[2] H. Huang, J. Qiu, and K. Riedl. Appl. Math. Optim., 2023.

- **Analysis** assumes a **Continuous Mean-Field PDE**
- **Algorithm** uses **Discrete-time, Finite Particles**
  - **time-discretization gap**
  - **finite-particle gap**



Paper	Result Type	Limitations
[1]	Qualitative MFL	No Rate
[2,3]	Finite-time MFL	Time-dependent rate
[4]	Uniform-time MFL	Modified dynamics
[5]	Uniform-time Weak PoC	Modified dynamics
This Work[6]	**Uniform-time Strong PoC**	**Original dynamics**

- **Finite-in-time limits**: For $J \in \mathbb{N}$ particles,

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|X_t^{(j)} - \overline{X}_t\right|^p\right] \le CJ^{-\frac{p}{2}}, \text{ where } C = C(T) \to \infty \text{ as } T \to \infty$$

- **Our Result**: Remove time-dependence of C for the unmodified algorithm.

[1] H. Huang and J. Qiu. Math. Methods Appl. Sci., 2022.
[2] H. Huang, J. Qiu, and K. Riedl. Appl. Math. Optim., 2023.
[3] N. J. Gerber, F. Hoffmann, and U. Vaes. 2024.
[4] H. Huang and H. Kouhkouh. arXiv preprint, 2024.
[5] E. Bayraktar, I. Ekren, and H. Zhou. arXiv preprint, 2025.
[6] N. Gerber, F. Hoffmann, D. Kim, and U. Vaes. 2026.

Main Result 1

Uniform-in-time propagation of chaos [Gerber, Hoffmann, K., Vaes (2025)][1]

Assume:

- f is bounded and globally Lipschitz
- $\overline{\rho}_0$ has finite moments of all orders
- Noise coefficient σ is sufficiently small

Then there exists $C_{\mathrm{MFL}}\left(\alpha, \underline{f}, \overline{f}, L_f, \sigma, d\right)$ such that for $J \in \mathbb{N}$ particles:

$$\sup_{t \geq 0} \mathbf{E}\left[\frac{1}{J}\sum_{j=1}^{J} \left|X_t^{(j)} - \overline{X}_t^{(j)}\right|^2\right] \leq \frac{C_{\mathrm{MFL}}}{J} .$$

- Optimal Monte Carlo rate $\mathcal{O}(J^{-1})$.
- Can track how C_{MFL} depends on other parameters explicitly
- **Small-noise condition** $\Rightarrow$ **contraction towards consensus** $>$ **noise dispersion**

[1] N. Gerber, F. Hoffmann, D. Kim, and U. Vaes. 2026.

Propagation of Chaos (PoC) via Coupling

Set Up

- $(W_t^j)_{j=1}^J$ are i.i.d. Brownian motions in $\mathbb{R}^d$.
- $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \qquad \sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$.
- $(X_t^j)_{j=1}^J$: interacting particle system, $(\overline{X}_t^j)_{j=1}^J$: i.i.d. copies of the mean-field system.

  For $j = 1, \dots, J$, consider

$$X_t^j = X_0^j + \int_0^t \frac{1}{J} \sum_{k=1}^J b(X_s^j, X_s^k)\, ds + \int_0^t \frac{1}{J} \sum_{k=1}^J \sigma(X_s^j, X_s^k)\, dW_s^j,$$

(microscopic system)

$$\overline{X}_t^j = X_0^j + \int_0^t \int_{\mathbb{R}^d} b(\overline{X}_s^j, z)\, \rho_s(dz)\, ds + \int_0^t \left(\int_{\mathbb{R}^d} \sigma(\overline{X}_s^j, z)\, \rho_s(dz) \right) dW_s^j,$$

(i.i.d. mean-field copies)

where $\rho_t = \mathrm{Law}(\overline{X}_t^j)$ for all $t \geq 0$.

Propagation of Chaos:

$$\mu_t^J := \frac{1}{J} \sum_{j=1}^J \delta_{X_t^j} \longrightarrow \rho_t \qquad \text{as } J \to \infty.$$

Propagation of Chaos (PoC) via Coupling

Theorem (PoC via Synchronous Coupling (McKean, Sznitman[1],[2]))

If b, σ are globally Lipschitz and bounded, and $X_0^j, \overline{X}_0^j$ are sampled i.i.d. from $\overline{\rho}_0$, then:

$$\mathbb{E}\left[W_2^2(\mu_t^J, \rho_t)\right] \leq \mathbb{E}\left[|X_t^j - \overline{X}_t^j|^2\right] \leq \frac{C_1}{J} e^{\varepsilon t}$$

Synchronous Coupling Principles

- Pair particle system and J mean-field copies under **shared noise**.
- Original argument requires b, σ to be globally Lipschitz (can be adapted).
- **Limitation**: Classic estimates blow up as $t \to \infty$ ($e^{\varepsilon t}$).
- Other couplings (e.g., reflection coupling[3], Optimal Coupling[4]) may lead to uniform-in-time bounds.

[1] H. P. McKean. In A. K. Aziz, editor, Lecture Series in Differential Equations, Volume 2. Van Nostrand Reinhold Company, 1969.

[2] A.-S. Sznitman. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 1984.

[3] A. Eberle. Probability Theory and Related Fields, Oct. 2015.

[4] S. Salem. 2020.

Toy example (with $\mathcal{M}(\mu)$ the usual mean under μ)

Interacting particle system:

$$\mathrm{d}X_t^j = -\Big(X_t^j - \mathcal{M}\big(\mu_t^J\big)\Big)\,\mathrm{d}t + \mathrm{e}^{-t}\,\mathrm{d}W_t^j, \qquad X_0^j = x_0^j \overset{\text{i.i.d.}}{\sim} \overline{\rho}_0 \qquad j = 1, \dots, J.$$

Mean field limit:

$$\begin{cases} \mathrm{d}\overline{X}_t = -\Big(\overline{X}_t - \mathcal{M}(\overline{\rho}_t)\Big)\,\mathrm{d}t + \mathrm{e}^{-t}\,\mathrm{d}\overline{W}_t, \\ \overline{\rho}_t = \mathrm{Law}(\overline{X}_t). \end{cases}$$

Synchronous Coupling

We couple the particle system and J copies of the mean field dynamics:

$$\mathrm{d}X_t^j = -\Big(X_t^j - \mathcal{M}\big(\mu_t^J\big)\Big)\,\mathrm{d}t + \mathrm{e}^{-t}\,\mathrm{d}W_t^j, \qquad X_0^j = x_0^j, \qquad j = 1, \dots, J,$$

$$\mathrm{d}\overline{X}_t^j = -\Big(\overline{X}_t^j - \mathcal{M}(\overline{\rho}_t)\Big)\,\mathrm{d}t + \mathrm{e}^{-t}\,\mathrm{d}W_t^j, \qquad \overline{X}_0^j = x_0^j, \qquad j = 1, \dots, J,$$

with same initial condition and driving Brownian motions.

Application of Synchronous Coupling to a Toy example

$$
\begin{aligned}
\mathrm{d}X_t^j &= -\left(X_t^j - \mathcal{M}\left(\mu_t^J\right)\right)\mathrm{d}t + \mathrm{e}^{-t}\,\mathrm{d}W_t^j, \qquad X_0^j = x_0^j, \\
\mathrm{d}\overline{X}_t^j &= -\left(\overline{X}_t^j - \mathcal{M}(\overline{\rho}_t)\right)\mathrm{d}t + \mathrm{e}^{-t}\,\mathrm{d}W_t^j, \qquad \overline{X}_0^j = x_0^j.
\end{aligned}
$$

Key Lemma

Global Lipschitz continuity of $\mathcal{M}\colon \mathcal{P}_1(\mathbf{R}^d) \to \mathbf{R}^d$:

$$
\forall(\mu,\nu), \qquad \left|\mathcal{M}(\mu) - \mathcal{M}(\nu)\right| \leq \mathcal{W}_2(\mu,\nu).
$$

$$
\begin{aligned}
\mathbf{E}\left[\left|X_t^1 - \overline{X}_t^1\right|^2\right] &\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 + \mathbf{E}\left|\mathcal{M}\left(\mu_s^J\right) - \mathcal{M}\left(\overline{\rho}_s\right)\right|^2 \mathrm{d}s \\
&\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 + \mathbf{E}\left|\mathcal{M}\left(\mu_s^J\right) - \mathcal{M}\left(\overline{\mu}_s^J\right)\right|^2 + \mathbf{E}\left|\mathcal{M}\left(\overline{\mu}_s^J\right) - \mathcal{M}\left(\overline{\rho}_s\right)\right|^2 \mathrm{d}s \\
&\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 + \mathbf{E}\left[\mathcal{W}_2\left(\mu_s^J, \overline{\mu}_s^J\right)^2\right] \mathrm{d}s + C_{\mathrm{MC}} J^{-1} \\
&\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 \mathrm{d}s + C_{\mathrm{MC}} J^{-1} \qquad \overset{\text{Grönwall}}{\rightsquigarrow} \qquad \mathbf{E}\left[|X_t^1 - \overline{X}_t^1|^2\right] \leq C(t) J^{-1}.
\end{aligned}
$$

Why the classical approach fails for CBO

Synchronous coupling for CBO, $j \in \{1, \dots, J\}$

$$\mathrm{d}X_t^j = -\left(X_t^j - \mathcal{M}_\alpha\left(\mu_t^J\right)\right)\mathrm{d}t + \sqrt{2}\sigma\left|X_t^j - \mathcal{M}_\alpha\left(\mu_t^J\right)\right|\mathrm{d}W_t^j$$
$$\mathrm{d}\overline{X}_t^j = -\left(\overline{X}_t^j - \mathcal{M}_\alpha(\overline{\rho}_t)\right)\mathrm{d}t + \sqrt{2}\sigma\left|\overline{X}_t^j - \mathcal{M}_\alpha(\overline{\rho}_t)\right|\mathrm{d}W_t^j$$

Technical difficulties:

- **Non-Lipschitz drift:** The weighted mean $\mathcal{M}_\alpha : \mathcal{P}_1(\mathbf{R}^d) \to \mathbf{R}^d$ is **not globally Lipschitz**.
- **Long-time explosion:** A standard Grönwall argument yields an $\mathcal{O}(e^{Ct})$ error bound.
- **Multiplicative noise:** The diffusion term depends on the state and the empirical measure μ_t^J.

Extending Synchronous Coupling approach to CBO

Synchronous coupling for CBO, $j \in \{1, \ldots, J\}$

$$\mathrm{d}X_t^j = -\left(X_t^j - \mathcal{M}_\alpha\left(\mu_t^J\right)\right)\mathrm{d}t + \sqrt{2}\sigma\left|X_t^j - \mathcal{M}_\alpha\left(\mu_t^J\right)\right|\mathrm{d}W_t^j$$
$$\mathrm{d}\overline{X}_t^j = -\left(\overline{X}_t^j - \mathcal{M}_\alpha(\overline{\rho}_t)\right)\mathrm{d}t + \sqrt{2}\sigma\left|\overline{X}_t^j - \mathcal{M}_\alpha(\overline{\rho}_t)\right|\mathrm{d}W_t^j$$

Ideas

- Requires **new tools** to avoid time-dependent blow-up.
- **Our Approach**: Classical coupling + **moment control** & **concentration inequality**

Assumptions

For simplicity, from now on we assume

- No noise ($\sigma = 0$)
- f bounded and globally Lipschitz
- Initialization in a compact set ($x_0^j \sim \overline{\rho}_0$) with $\overline{\rho}_0$ compactly supported

Extending Synchronous Coupling approach to CBO

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2J}\sum_{j=1}^{J}\mathbf{E}\left|X_t^j-\overline{X}_t^j\right|^2=\underbrace{-\frac{1}{J}\mathbf{E}\sum_{j=1}^{J}\left\langle X_t^j-\overline{X}_t^j,X_t^j-\mathcal{M}(\mu_t^J)-\overline{X}_t^j+\mathcal{M}(\overline{\mu}_t^J)\right\rangle}_{(\mathrm{I})}$$

$$\underbrace{-\frac{1}{J}\mathbf{E}\sum_{j=1}^{J}\left\langle X_t^j-\overline{X}_t^j,\mathcal{M}(\mu_t^J)-\mathcal{M}_\alpha(\mu_t^J)-\mathcal{M}(\overline{\mu}_t^J)+\mathcal{M}_\alpha(\overline{\mu}_t^J)\right\rangle}_{(\mathrm{II})}$$

$$\underbrace{-\frac{1}{J}\mathbf{E}\sum_{j=1}^{J}\left\langle X_t^j-\overline{X}_t^j,\mathcal{M}_\alpha\left(\overline{\rho}_t\right)-\mathcal{M}_\alpha(\overline{\mu}_t^J)\right\rangle}_{(\mathrm{III})}.$$

- **(I) good contractive term**
- **(II) weighted-mean perturbation**
- **(III) empirical sampling error**

Extending Synchronous Coupling approach to CBO

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2J}\sum_{j=1}^{J}\mathbf{E}|X_t^{(j)} - \overline{X}_t^{(j)}|^2 = (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}).$$

- **(I) good contractive term**: favorable sign after pivoting around $\mathcal{M}$
- **(II) weighted-mean perturbation**: handled by a **novel stability estimate**
- **(III) empirical sampling error**: handled by **Monte Carlo estimate** + **decay of centered moments**
  - Monte Carlo estimate: For any $p \geq 2$,

$$\mathbf{E}\left|\mathcal{M}_\alpha\left(\mu_{\mathcal{Z}^J}\right) - \mathcal{M}_\alpha\left(\pi\right)\right|_p^p \leq C_{\mathrm{WM},p}\mathbf{E}\left|Z^1 - \mathbf{E}Z^1\right|_p^p J^{-\frac{p}{2}},$$

$$\mu_{\mathcal{Z}^J} := \frac{1}{J}\sum_{j=1}^{J}\delta_{Z^j}, \qquad \left\{Z^j\right\}_{j\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} \pi\,.$$

  - Recall (decay of centered moments):

$$\mathbf{E}\left|\overline{X}_t - \mathbf{E}\overline{X}_t\right|^p \leq \mathbf{E}\left|\overline{X}_0 - \mathbf{E}\overline{X}_0\right|^p \mathrm{e}^{-\lambda_p t}\,.$$

The proof is about exploiting (I) and making (II)–(III) small uniformly in time.

Key Ingredients (1/2)

For a probability measure $\mu \in \mathcal{P}_1(\mathbf{R}^d)$, let:

$$\mathfrak{M}_p(\mu) := \int_{\mathbf{R}^d} \left| x - \mathcal{M}(\mu) \right|^p \mu(\mathrm{d}x)\,.$$

1. **Novel Stability Estimate** for the weighted mean:

$$\left| \mathcal{M}_\alpha(\mu) - \mathcal{M}(\mu) - \mathcal{M}_\alpha(\nu) + \mathcal{M}(\nu) \right| \leq C_{\mathcal{M}} \left(\sqrt{\mathfrak{M}_2(\mu)} + \sqrt{\mathfrak{M}_2(\nu)} \right) \mathcal{W}_2(\mu,\nu).$$

2. **Decay of Centered Moments** ($\lambda_p := \lambda_p(\sigma, \alpha, \underline{f}, \overline{f})$):

$$\mathbf{E}\left[\mathfrak{M}_p(\mu_t^J)\right] \leq \mathbf{E}\left[\mathfrak{M}_p(\mu_0^J)\right] \mathrm{e}^{-\lambda_p t}, \quad \mathbf{E}\left|\overline{X}_t - \mathbf{E}\overline{X}_t\right|^p \leq \mathbf{E}\left|\overline{X}_0 - \mathbf{E}\overline{X}_0\right|^p \mathrm{e}^{-\lambda_p t}$$

3. Uniform-in-time **Raw Moment Control**:

$$\mathbf{E}\left[\sup_{t\geq 0}\left|X_t^j\right|^p\right]^{\frac{1}{p}} \leq C_{\mathrm{Raw},p}\mathbf{E}\left[\left|X_0^j\right|^p\right]^{\frac{1}{p}}, \quad \mathbf{E}\left[\sup_{t\geq 0}\left|\overline{X}_t\right|^p\right]^{\frac{1}{p}} \leq C_{\mathrm{Raw},p}\mathbf{E}\left[\left|\overline{X}_0\right|^p\right]^{\frac{1}{p}}$$

4. **Concentration of Measure (Bounding Rare Excursions):**
   We control the probability of falling into a bad set where centered moments become too large:

$$\mathbf{P}\left[\sup_{t\geq 0} \mathrm{e}^{\kappa t}\,\mathfrak{M}_2(\mu_t^J) \geq \mathfrak{M}_2(\overline{\rho}_0) + 1\right] \lesssim J^{-\frac{q}{2}},$$

$$\mathbf{P}\left[\sup_{t\geq 0} \mathrm{e}^{\kappa t}\,\mathfrak{M}_2(\overline{\mu}_t^J) \geq \mathfrak{M}_2(\overline{\rho}_0) + 1\right] \lesssim J^{-\frac{q}{2}}.$$

Why concentration is needed

The difficult term in the stability estimate has the form

$$Q_t := \mathbf{E}\Big[\Big(\mathfrak{M}_2(\mu_t^J) + \mathfrak{M}_2(\overline{\mu}_t^J)\Big)\, W_2^2(\mu_t^J, \overline{\mu}_t^J)\Big]\,.$$

- This is a **product** of moment factors and Wasserstein error
- The expectation is taken **outside** the whole product
- Average moment control alone is not enough

Key Ingredients (2/2)

4. **Concentration of Measure (Bounding Rare Excursions):**
   We control the probability of falling into a bad set where centered moments become too large:

$$\mathbf{P}\left[\sup_{t\geq 0} \mathrm{e}^{\kappa t}\, \mathfrak{M}_2(\mu_t^J) \geq \mathfrak{M}_2(\overline{\rho}_0) + 1\right] \lesssim J^{-\frac{q}{2}},$$

$$\mathbf{P}\left[\sup_{t\geq 0} \mathrm{e}^{\kappa t}\, \mathfrak{M}_2(\overline{\mu}_t^J) \geq \mathfrak{M}_2(\overline{\rho}_0) + 1\right] \lesssim J^{-\frac{q}{2}}.$$

Good set / bad set strategy

We split the expectation into:

$$Q_t = \mathbf{E}\Big[\mathbf{1}_{\Omega_\kappa^\star}(\cdots)\Big] + \mathbf{E}\Big[\mathbf{1}_{\Omega\setminus\Omega_\kappa^\star}(\cdots)\Big]\,.$$

- On the $\Omega_\kappa^\star$ **(good set)**: centered moments decay + raw moments bound
- On the $\Omega \setminus \Omega_\kappa^\star$ **(bad set)**: concentration inequality

4. **Concentration of Measure (Bounding Rare Excursions):**
   We control the probability of falling into a bad set where centered moments become too large:

$$\mathbf{P}\left[\sup_{t\geq 0} \mathrm{e}^{\kappa t}\, \mathfrak{M}_2(\mu_t^J) \geq \mathfrak{M}_2(\overline{\rho}_0) + 1\right] \lesssim J^{-\frac{q}{2}},$$

$$\mathbf{P}\left[\sup_{t\geq 0} \mathrm{e}^{\kappa t}\, \mathfrak{M}_2(\overline{\mu}_t^J) \geq \mathfrak{M}_2(\overline{\rho}_0) + 1\right] \lesssim J^{-\frac{q}{2}}.$$

5. **Monte Carlo estimate**: weighted-mean sampling error is $\mathcal{O}(J^{-p/2})$

$$\mathbf{E}\left|\mathcal{M}_{\alpha}\left(\mu_{\overline{\mathcal{X}}_t^J}\right) - \mathcal{M}_{\alpha}\left(\overline{\rho}_t\right)\right|_p^p \leq C_{\mathrm{MC}} J^{-\frac{p}{2}}\,.$$

How the estimate closes

- Stability $\rightarrow$ Moment Control + Concentration inequality
- Monte Carlo estimate

Main Result 2: Stability

Uniform-in-time stability of particle dynamics

Let $(X_t^{(j)})_{j=1}^J$ and $(\widetilde{X}_t^{(j)})_{j=1}^J$ solve the CBO system with the **same Brownian motions** but different initial data. Then for all $q \geq 2$, for small enough σ,

$$\sup_{t\geq 0} \mathbf{E}\left[\frac{1}{J}\sum_{j=1}^J |X_t^{(j)} - \widetilde{X}_t^{(j)}|^2\right] \leq C_{\mathrm{Stab},1}\mathbf{E}\left[\frac{1}{J}\sum_{j=1}^J |X_0^{(j)} - \widetilde{X}_0^{(j)}|^2\right] + \frac{C_{\mathrm{Stab},2}}{J^q}.$$

Proof features:

- Same proof architecture as PoC
  - stability estimate, contractivity, moment bounds, concentration inequalities
- Monte Carlo term is not needed (particle-to-particle)
  - Residual error decays faster: J^{-q}

Outlook and future directions

Main takeaway

- **Quantitative, Uniform-in-time** bridge between finite-particle CBO and its mean-field model
- Works for the **original, unmodified dynamics**
- **Key challenge resolved**: Uniform-in-time **Moment Control**
- Leveraged contractivity and stochastic analysis (concentration, BDG)

Next questions

- Can we close the **time-discretization gap** uniformly in time?
- Can we analyze **Kinetic CBO** (second-order model connection to Particle Swarm Optimization)? **(on arXiv soon)**
- Can we extend this to Consensus-Based Sampling (CBS)?
- Can we study the method with **adaptive** α?
- Can we unify the framework for derivative-free optimization algorithms?

Acknowledgements

(a) Prof. Seung-Yeal Ha

(b) Prof. Franca Hoffmann

(c) Prof. Urbain Vaes

(d) Nicolai Gerber

Figure: Mentors and Collaborators

Thank you for your attention!

For questions, contact Dohyeon Kim via dohyeon@caltech.edu

References I

E. Bayraktar, I. Ekren, and H. Zhou. Uniform-in-time weak propagation of chaos for consensus-based optimization. arXiv preprint, 2502.00582, 2025.

J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. An analytical framework for consensus-based global optimization method. Mathematical Models and Methods in Applied Sciences, 28(6):1037–1066, 2018.

J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. Consensus-based sampling. Stud. Appl. Math., 148(3):1069–1140, 2022.

A. Eberle. Reflection couplings and contraction rates for diffusions. Probability Theory and Related Fields, 166(3–4):851–886, Oct. 2015.

M. Fornasier, T. Klock, and K. Riedl. Consensus-Based Optimization Methods Converge Globally. SIAM J. Optim., 34(3):2973–3004, 2024.

N. Gerber, F. Hoffmann, D. Kim, and U. Vaes. Uniform-in-time propagation of chaos for consensus-based optimization. 2026. arXiv: 2505.08669 [math.PR].

N. J. Gerber, F. Hoffmann, and U. Vaes. Mean-field limits for consensus-based optimization and sampling. 2024. arXiv: 2312.07373 [math.PR].

References II

H. Huang and H. Kouhkouh. Uniform-in-time mean-field limit estimate for the consensus-based optimization. arXiv preprint, 2411.03986, 2024.

H. Huang and J. Qiu. On the mean-field limit for the consensus-based optimization. Math. Methods Appl. Sci., 45(12):7814–7831, 2022.

H. Huang, J. Qiu, and K. Riedl. On the global convergence of particle swarm optimization methods. Appl. Math. Optim., 88(2):Paper No. 30, 44, 2023.

J. Kennedy and R. Eberhart. Particle swarm optimization. In Proceedings of ICNN'95-international conference on neural networks, volume 4, pages 1942–1948. IEEE, 1995.

Q. Liu and D. Wang. Stein variational gradient descent: a general purpose bayesian inference algorithm. In Advances in Neural Information Processing Systems 29, 2016.

J. Lu, E. Tadmor, and A. Zenginoglu. Swarm-based gradient descent method for non-convex optimization. 2022. eprint: `arXiv:2211.17157`.

H. P. McKean. Propagation of chaos for a class of non-linear parabolic equations. In A. K. Aziz, editor, Lecture Series in Differential Equations, Volume 2, number 19 in Van Nostrand Mathematical Studies, pages 177–194. Van Nostrand Reinhold Company, 1969.

References III

R. Pinnau, C. Totzeck, O. Tse, and S. Martin. A consensus-based model for global optimization and its mean-field limit. Math. Models Methods Appl. Sci., 27(1):183–204, 2017.

S. Salem. A gradient flow approach of propagation of chaos. 2020.

A.-S. Sznitman. Équations de type de boltzmann, spatialement homogènes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 66:559–592, 1984.

Laplace's method [APPENDIX]

Laplace's method can be employed for studying the limit as $\beta \to \infty$ of the integral

$$I_\beta(\varphi) = \frac{\displaystyle\int_{\mathbf{R}^d} \varphi(x)\, \mathrm{e}^{-\beta J(x)}\, \mu(\mathrm{d}x)}{\displaystyle\int_{\mathbf{R}^d} \mathrm{e}^{-\beta J(x)}\, \mu(\mathrm{d}x)} =: \int_{\mathbf{R}^d} \varphi\, \mathrm{d}(\mathcal{R}_\beta \mu), \qquad \mathcal{R}_\beta : \mu \mapsto \frac{\mu\, \mathrm{e}^{-\beta J}}{\int \mu\, \mathrm{e}^{-\beta J}}.$$

Let $x_* = \arg\min J$. Under appropriate assumptions, it holds

$$I_\beta(\varphi) = \int_{\mathbf{R}^d} \varphi\, \mathrm{d}g_\beta + \mathcal{O}\left(\frac{1}{\beta^2}\right) \qquad \text{as } \beta \to \infty.$$

where $g_\beta = \mathcal{N}\left(x_*, \beta^{-1} \left(\mathrm{Hess}\, J(x_*)\right)^{-1}\right)$. In other words $\mathcal{R}_\beta \mu \approx g_\beta$ for large β.

Motivation:

$$\mathrm{e}^{-\beta J(x)} \approx \mathrm{e}^{-\beta\left(J(x_*) + \frac{1}{2}\, \mathrm{Hess}\, J(x_*) : \left((x - x_*) \otimes (x - x_*)\right)\right)}$$

From Implementation to Mean Field Limit [APPENDIX]

Particle-level, Discrete time: $x_h^{(j)} \in \mathbb{R}^d$ for $j = 1, \dots, J$ solves

$$x_h^{(j)}(n+1) = x_h^{(j)}(n) - h\Big(x_h^{(j)} - \mathcal{M}_{\beta}(X_h)\Big) + \sqrt{2h}\sigma\Big|x_h^{(j)} - \mathcal{M}_{\beta}(X_h)\Big|Z_h^{(j)}(n)\,,$$

where

$$\mathcal{M}_{\beta}(X_h) = \frac{\sum_{k=1}^J x_h^{(k)} \exp\Big(-\beta J(x_h^{(k)})\Big)}{\sum_{k=1}^J \exp\Big(-\beta J(x_h^{(k)})\Big)}, \qquad X_h := \Big(x_h^{(j)}\Big)_{j=1}^J \in \mathbb{R}^{dJ}.$$

Particle-level, Continuous time: $X_t^{(j)} \in \mathbb{R}^d$ for $j = 1, \dots, J$ solves

$$\mathrm{d}X_t^{(j)} = -\Big(X_t^{(j)} - \mathcal{M}_{\beta}(\mu_t^J)\Big)\,\mathrm{d}t + \sqrt{2}\sigma\Big|X_t^{(j)} - \mathcal{M}_{\beta}(\mu_t^J)\Big|\mathrm{d}W_t^{(j)}\,,$$

where

$$\mathcal{M}_{\beta}(\mu_t^J) = \frac{\int x\,\mathrm{e}^{-\beta f(x)}\,\mu_t^J(\mathrm{d}x)}{\int \mathrm{e}^{-\beta f(x)}\,\mu_t^J(\mathrm{d}x)} = \frac{\sum_{k=1}^J X_t^{(k)} \exp\Big(-\beta f(X_t^{(k)})\Big)}{\sum_{k=1}^J \exp\Big(-\beta f(X_t^{(k)})\Big)}, \qquad \mu_t^J = \frac{1}{J}\sum_{k=1}^J \delta_{X_t^{(k)}}.$$

Mean-field limit: $\overline{x} \in \mathbb{R}^d$ solves

$$\mathrm{d}\overline{x}_t = -\Big(\overline{x}_t - \mathcal{M}_{\beta}(\mu_t)\Big)\,\mathrm{d}t + \sqrt{2}\sigma\Big|\overline{x}_t - \mathcal{M}_{\beta}(\mu_t)\Big|\mathrm{d}W_t\,,$$

where

$$M_{\beta}(\mu_t) = \frac{\int_{\mathbb{R}^d} \overline{x} e^{-\beta f(\overline{x})}\mu(d\overline{x})}{\int_{\mathbb{R}^d} e^{-\beta f(\overline{x})}\mu(d\overline{x})}, \qquad \mu = \text{Law}(\overline{x}_t) \in \mathcal{P}(\mathbb{R}^d).$$

Synchronous Coupling Strategy

Proof Sketch

Define $\alpha_t^j := X_t^j - \overline{X}_t^j$.

$$\frac{d}{dt}\mathbb{E}\left[\frac{1}{2}|\alpha_t^j|^2\right] \leq \text{Term (A)} + \text{Term (B)}$$

$$\text{Term (A)} := \frac{1}{J}\sum_{k=1}^{J}\mathbb{E}[\langle b(X_t^j, X_t^k) - b(\overline{X}_t^j, \overline{X}_t^k), \alpha_t^j\rangle]$$

$$\text{Term (B)} := \mathbb{E}\left[\left\langle \frac{1}{J}\sum_{k=1}^{J} b(\overline{X}_t^j, \overline{X}_t^k) - \int b(\overline{X}_t^j, z)\rho_t(dz), \alpha_t^j \right\rangle\right]$$

- Use Young's inequality:

$$\text{Term (B)} \leq \frac{1}{2\varepsilon}\mathbb{E}[|\alpha_t^j|^2] + \frac{\varepsilon}{2}\mathbb{E}\left[\left|\frac{1}{J}\sum_{k=1}^{J} b(\overline{X}_t^j, \overline{X}_t^k) - \int b(\overline{X}_t^j, z)\rho_t(dz)\right|^2\right]$$

- The second part is $\mathcal{O}(1/J)$ by Law of Large Numbers.

Key ingredients

For $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, define the centered moment

$$\mathfrak{M}_p(\mu) := \int_{\mathbb{R}^d} |x - \mathcal{M}(\mu)|^p \, \mu(\mathrm{d}x).$$

1. **Stability estimate for the weighted mean**

$$|\mathcal{M}_\alpha(\mu) - \mathcal{M}(\mu) - \mathcal{M}_\alpha(\nu) + \mathcal{M}(\nu)| \le C_{\mathcal{M}} \Big(\sqrt{\mathfrak{M}_2(\mu)} + \sqrt{\mathfrak{M}_2(\nu)} \Big) W_2(\mu, \nu)$$

2. **Decay of centered moments** and **uniform raw moment bounds**
3. **Concentration Inequalities**: bad events where moments get too large have probability $\lesssim J^{-q/2}$
4. **Monte Carlo estimate**: weighted-mean sampling error is $\mathcal{O}(J^{-p/2})$

Takeaway

- Bypasses non-Lipschitz issues
- Large errors occur with vanishingly small probability

BDG inequality [APPENDIX]

Theorem (Burkholder–Davis–Gundy inequality)

Let $(W_t)_{t\geq 0}$ *denote a standard Brownian motion in* $\mathbf{R}^m$ *and let* $(\mathcal{F}_t)_{t\geq 0}$ *be the induced filtration. Let* $(g_t)_{t\geq 0}$ *be a* $\mathbf{R}^{n\times m}$*-valued* $\mathcal{F}_t$*-adapted process such that for every time* $T \geq 0$*, it holds that* $\int_0^T \left\|g(t)\right\|_{\mathrm{F}}^2 \,\mathrm{d}t < +\infty$ *almost surely. Denote*

$$X_t := \int_0^t g(s)\,\mathrm{d}W_s \qquad \textit{and} \qquad \langle X \rangle_t := \int_0^t \left\|g(s)\right\|_{\mathrm{F}}^2 \,\mathrm{d}s.$$

Then for all $p > 0$*, there exist positive constants* $c_{\mathrm{BDG,p}}, C_{\mathrm{BDG,p}} < +\infty$ *such that*

$$\forall t \geq 0, \qquad c_{\mathrm{BDG,p}}\mathbb{E}\left[\langle X \rangle_t^{\frac{p}{2}}\right] \leq \mathbb{E}\left[\sup_{0\leq s\leq t} |X_s|^p\right] \leq C_{\mathrm{BDG,p}}\mathbb{E}\left[\langle X \rangle_t^{\frac{p}{2}}\right]. \tag{5.1}$$

Connection to Particle Swarm Optimization

Particle Swarm Optimization (PSO)

$$
\begin{aligned}
dx_i(t) &= v_i(t)dt\,, \\
m dv_i(t) &= -\gamma v_i(t)dt + \lambda_1(y_i(t) - x_i(t))dt + \lambda_2(\overline{y}(t) - x_i(t))dt \\
&\quad + \sigma_1 S(y_i(t) - x_i(t))dW_i^2(t) + \sigma_2 S(\overline{y}(t) - x_i(t))dW_i^2(t)\,.
\end{aligned}
$$

- $y_i(t)$: local best (personal best position for each $x_i(t)$),
  $\overline{y}(t)$: global best (min among all y_i's)
- 1) Eliminate local best (memory) effect

  2) Regularize $\overline{y}$ via softmin ($\mathcal{M}_\beta$)

  $\longrightarrow$ Second order CBO (: Kinetic CBO)

  3) Zero inertia limit ($m \to 0$)

  $\longrightarrow$ First order CBO
- Despite its wide implementation, its convergence analysis is yet nascent

Kinetic CBO

Let $S : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a Lipschitz function with Lipschitz constant L.

Particle Evolution

Let given J particles and their velocity vectors $(x_j(t), v_j(t))_{j=1}^J$ solve

$$
\begin{aligned}
dx_j(t) &= v_j(t)dt\,,\\
mdv_j(t) &= -\gamma v_j(t)dt - (x_j(t) - \mathcal{M}_\beta(\mu_t^J))dt + \sigma S(x_j(t) - \mathcal{M}_\beta(\mu_t^J))dW_j(t)\,,
\end{aligned}
$$

Liouville Equation

$$
\begin{aligned}
\partial_t f_J = -\sum_{j=1}^J \nabla_{x_j}\cdot(v_j f_J) + \frac{1}{m}\nabla_{v_j}\cdot\Big((\gamma v_j + x_j - \mathcal{M}_\beta(f_J))f_J\Big)\\
+ \frac{\sigma^2}{2m^2}\nabla_v\cdot\nabla_v\left(S\left(x_j - \mathcal{M}_\beta(f_J)\right)^2 f_J\right)\,.
\end{aligned}
$$

Mean Field Limit SDE and PDE

When $J \to \infty$,

$$\begin{aligned} d\overline{X}(t) &= \overline{V}(t)\,, \\ md\overline{V}(t) &= -\gamma V(t) - (\overline{X}(t) - \mathcal{M}_{\beta}(\mu_t))dt + \sigma S(\overline{X}(t) - \mathcal{M}_{\beta}(\mu_t))dW(t)\,. \end{aligned} \tag{5.2}$$

The corresponding PDE to (5.2) is

$$\partial_t f + v \cdot \nabla_x f - \frac{1}{m} \nabla_v \cdot ((\gamma v + x - \mathcal{M}_{\beta}(f))\, f) = \frac{\sigma^2}{2m^2} \nabla_v \cdot \nabla_v \left(S\left(x - \mathcal{M}_{\beta}(f)\right)^2 f \right)\,. \tag{5.3}$$