Consensus Based Optimization and Sampling

Franca Hoffmann

Computing and Mathematical Sciences
California Institute of Technology

Lectures 6

Course on Gradient Flows in Data Science
ICTS, Bengaluru, India

Caltech

collaboration with:

José A. Carrillo, Andrew M. Stuart and Urbain Vaes

Consensus Based Sampling. Studies in Applied mathematics, 2022.

collaboration with:

Nicolai J. Gerber and Urbain Vaes

Mean-field limits for Consensus-Based Optimization and Sampling.

arXiv preprint, 2312.07373

collaboration with:

Nicolai J. Gerber, Dohyeon Kim and Urbain Vaes

Uniform-in-time propagation of chaos for Consensus-Based Optimization.

arXiv preprint, 2505.08669

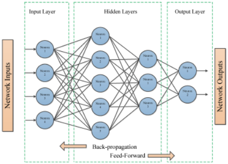
Outline

Let's optimize

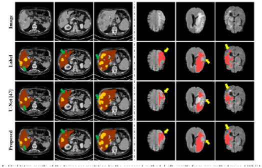
Consider $f : \mathbb{R}^d \to \mathbb{R}$. **Goal:** Find θ_* such that

$$f(\theta_*) = \min_{\theta} f(\theta)$$

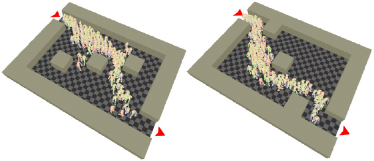
- High-dimensional (nonconvex) optimization problems are pervasive in many fields such as machine learning, signal/image processing and optimal control.



Training neural networks



Computer assisted tomography



Crowd evacuation control

Challenges:

- f is non-convex
- f is only given as a black-box oracle
- f is not differentiable, or derivatives are too costly to obtain
- additional constraints

Approaches

- Stochastic gradient descent-type methods (SGD, Adam, RMSProp, ...), are favored for their efficiency, scalability, ability to evade critical points, and their solid mathematical foundations.
- Metaheuristic (gradient-free) algorithms gained popularity due to the minimal assumptions on the optimization problem, making them versatile and applicable to a wider range of problems.
    - Metropolis-Hastings (1953,1970)
    - Simplex Heuristics (1965)
    - Evolutionary Programming (1966)
    - Genetic Algorithms (GA) (1975)
    - Simulated Annealing (SA) (1983)
    - Particle Swarm Optimization (PSO) (1995)
    - Ant Colony Optimization (ACO) (1997)
    - ...

$\Longrightarrow$ Despite the significant empirical success, most results are experimental in nature and lack a rigorous mathematical foundation.

Approaches

- Stochastic gradient descent-type methods (SGD, Adam, RMSProp, ...), are favored for their effciency, scalability, ability to evade critical points, and their solid mathematical foundations.
- Metaheuristic (gradient-free) algorithms gained popularity due to the minimal assumptions on the optimization problem, making them versatile and applicable to a wider range of problems.

Approach: Interacting Particle Systems

Consider evolution of $x_1(t), ..., x_J(t) \in \mathbb{R}^d$ solving

$$\dot{x}_i(t) = g_i(t, x_1(t), ..., x_J(t))$$

(and stochastic versions) such that $x_i(t) \to \theta_*$ (approximatively) as $t \to \infty$.

$\longrightarrow$ Often parallelizable, and some can be studied through mean-field equations.

Inverse problem formulation[1]

Find an unknown parameter $\theta \in \mathcal{U}$ from data $y \in \mathbf{R}^m$ where

$$y = \mathcal{G}(\theta) + \eta,$$

- $\mathcal{G}$ is the forward operator, $\mathcal{G} : \mathbb{R}^d \mapsto \mathbb{R}^K$.
- η is observational noise, $\eta \sim \mathsf{N}(0, \gamma^2 I)$.

In many PDE applications,

- Calibration & Uncertainty Quantification;
- $\mathcal{G}$ is expensive to evaluate;
- The derivatives of $\mathcal{G}$ are not available.

[1] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In Handbook of uncertainty quantification. Vol. 1, 2, 3. Springer, Cham, 2017.

Probabilistic approach

Bayesian approach to inverse problems[2]

Modeling step:

- Probability distribution on parameter: $\theta \sim \pi_0$, encoding our prior knowledge;
- Probability distribution for noise $\mathbb{P}(y|\theta)$ with $y - \mathsf{G}(\theta) \sim \mathsf{N}(0, \gamma^2 I)$ likelihood.

An application of Bayes' theorem gives the posterior distribution:

$$\rho^y(\theta) \propto \mathbb{P}(y|\theta)\,\pi_0(\theta) = \text{prior} \times \text{likelihood}.$$

In the Gaussian case where $\pi_0 = \mathcal{N}(m, \Sigma_0)$ and Gaussian noise,

$$\rho^y(\theta) \propto \exp\left(-\left(\frac{1}{2\gamma^2}\left|y - \mathcal{G}(\theta)\right|^2 + \frac{1}{2}\left|\theta - m\right|^2_{\Sigma_0}\right)\right) =: \exp\bigl(-f(\theta)\bigr).$$

Two approaches for extracting information:

- Find the maximizer of $\rho^y(\theta)$ (maximum a posteriori estimation);
- Sample the posterior distribution $\rho^y(\theta)$.

[2] A. M. Stuart. Inverse problems: a Bayesian perspective. Acta Numer., 2010.

Brief review of the recent literature on interacting particle methods

- 1995: Particle swarm optimization[3];
- 2006: Sequential Monte Carlo[4];
- 2010: Affine-invariant many-particle MCMC[5];
- 2013: Ensemble Kalman inversion[6];
- 2016: Stein variational gradient descent[7];
- 2017: Consensus-based optimization[8];
- 2020: Ensemble Kalman sampling[9];
- 2022: Consensus-based sampling[10];
- 2022: Swarm-Based Gradient Descent[11]

[3] James Kennedy and Russell Eberhart. Particle swarm optimization. In Proceedings of ICNN'95-international conference on neural networks. IEEE, 1995.

[4] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. J. R. Stat. Soc. Ser. B Stat. Methodol., 2006.

[5] J. Goodman and J. Weare. Ensemble samplers with affine invariance. Commun. Appl. Math. Comput. Sci., 2010.

[6] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Ensemble Kalman methods for inverse problems. Inverse Problems, 2013.

[7] Q. Liu and D. Wang. Stein variational gradient descent: a general purpose Bayesian inference algorithm. In Advances In Neural Information Processing Systems, 2016.

[8] 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., 2017.

[9] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. SIAM J. Appl. Dyn. Syst., 2020.

[10] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. Consensus-based sampling. Stud. Appl. Math., 2022.

[11] Jingcheng Lu, Eitan Tadmor, and Anil Zenginoglu. Swarm-based gradient descent method for non-convex optimization. 2022.

Outline

Consensus based optimization (CBO)[12]

Interacting Particle System

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

where $\mathcal{M}_\beta(\mu_t^J)$ is given by

$$\mathcal{M}_\beta(\mu_t^J) = \frac{\int \theta\, \mathrm{e}^{-\beta f(\theta)}\, \mu_t^J(\mathrm{d}\theta)}{\int \mathrm{e}^{-\beta f(\theta)}\, \mu_t^J(\mathrm{d}\theta)} = \frac{\sum_{j=1}^J \theta_t^{(j)} \exp\left(-\beta f(\theta_t^{(j)})\right)}{\sum_{k=1}^J \exp\left(-\beta f(\theta_t^{(k)})\right)}, \qquad \mu_t^J = \frac{1}{J}\sum_{j=1}^J \delta_{\theta_t^{(j)}}.$$

This is a weighted mean:

$$\mathcal{M}_\beta(\mu_t^J) = \sum_{j=1}^J \theta_t^{(j)} \omega_\beta^{(j)}, \qquad \omega_\beta^{(j)} = \frac{\exp\left(-\beta f(\theta_t^{(j)})\right)}{\sum_{k=1}^J \exp\left(-\beta f(\theta_t^{(k)})\right)} \in [0,1].$$

[12] 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., 2017.

Consensus based optimization (CBO)[13]

$$\mathrm{d}\theta_t^{(j)} = -\Big(\theta_t^{(j)} - \mathcal{M}_\beta(\mu_t^J)\Big)\,\mathrm{d}t + \sqrt{2}\sigma\Big|\theta_t^{(j)} - \mathcal{M}_\beta(\mu_t^J)\Big|\mathrm{d}W_t^{(j)}. \qquad j = 1, \ldots, J,$$

Advantages:

- derivative-free
- theoretical guarantees via mean field analysis
- convergence rate independent of Hessian(f)

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

$$M_\beta(\mu) \to \theta_* \quad \text{as} \quad \beta \to \infty \quad \text{if} \quad \theta_* \in \operatorname{supp} \mu\,.$$

[13] 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., 2017.

Laplace's method

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

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

Let $\theta_* = \arg\min f$. Under appropriate assumptions, it holds[14],[15]

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

where $g_\beta = \mathcal{N}\Big(\theta_*, \beta^{-1} \big(\mathrm{Hess}\, f(\theta_*)\big)^{-1}\Big)$. In other words $\mathcal{R}_\beta \mu \approx g_\beta$ for large β.

Motivation:

$$\mathrm{e}^{-\beta f(\theta)} \approx \mathrm{e}^{-\beta \left(f(\theta_*) + \frac{1}{2} \mathrm{Hess}\, f(\theta_*) : \left((\theta - \theta_*) \otimes (\theta - \theta_*)\right)\right)}$$

[14] P. D. Miller. Applied asymptotic analysis. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2006.

[15] 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, 2018.

CBO Variants I

- high-dimensional ML problems[16]: component-wise geometric Brownian motion

$$\sqrt{2}\sigma \sum_{k=1}^{d} \left(\theta_t^{(j)} - \mathcal{M}_{\beta}(\mu_t^J)\right)_k \mathrm{d}W_k^{(j)} \vec{e}_k \quad \text{instead of} \quad \sqrt{2}\sigma \left|\theta_t^{(j)} - \mathcal{M}_{\beta}(\mu_t^J)\right| \mathrm{d}\vec{W}^{(j)}$$

  and of including mini-batching[17];
- constrained optimization problems[18][19][20][21], including MirrorCBO[22];
- multi-objective optimization[23][24];

---

[16] José A. Carrillo, Shi Jin, Lei Li, and Yuhua Zhu. A consensus-based global optimization method for high dimensional machine learning problems. ESAIM Control Optim. Calc. Var., 2021.

[17] Dongnam Ko, Seung-Yeal Ha, Shi Jin, and Doheon Kim. Convergence analysis of the discrete consensus-based optimization algorithm with random batch interactions and heterogeneous noises. Math. Models Methods Appl. Sci., 2022.

[18] Massimo Fornasier, Hui-Lin Huang, Lorenzo Pareschi, and Philippe Sunnen. Consensus-based optimization on hypersurfaces: well-posedness and mean-field limit. Mathematical Models and Methods in Applied Sciences, 2020.

[19] Seung-Yeal Ha, Myeongju Kang, Dohyun Kim, Jeongho Kim, and Insoon Yang. Stochastic consensus dynamics for nonconvex optimization on the Stiefel manifold: mean-field limit and convergence. Math. Models Methods Appl. Sci., 2022.

[20] J.A. Carrillo, C. Totzeck, and U. Vaes. Consensus-based optimization and ensemble kalman inversion for global optimization problems with constraints. arXiv:2111.02970, 2021.

[21] José A. Carrillo, Shi Jin, Haoyu Zhang, and Yuhua Zhu. An interacting particle consensus method for constrained global optimization. 2024.

[22] **MirrorCBO**.

[23] Kathrin Klamroth, Michael Stiglmayr, and Claudia Totzeck. Consensus-based optimization for multi-objective problems: a multi-swarm approach. arXiv preprint, 2022.

CBO Variants II

- targets with multiple minimizers or distributions with many modes[25][26];
- including momentum[27][28][29];
- including memory[30][31] (effectively increasing the number of particles);
- truncated noise[32]
- applications to federated learning[33], finance[34], rare event estimation[35], stochastic optimization[36], closed-box adversarial attacks[37]...

[25] Leon Bungert, Tim Roith, and Philipp Wacker. Polarized consensus-based dynamics for optimization and sampling. arXiv preprint, 2022.

[26] Massimo Fornasier and Lukang Sun. A pde framework of consensus-based optimization for objectives with multiple global minimizers. 2024.

[27] Sara Grassi and Lorenzo Pareschi. From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit. Math. Models Methods Appl. Sci., 2021.

[28] Hui Huang, Jinniao Qiu, and Konstantin Riedl. On the global convergence of particle swarm optimization methods. Appl. Math. Optim., 2023.

[29] Jingrun Chen, Shi Jin, and Liyao Lyu. A consensus-based global optimization method with adaptive momentum estimation. Commun. Comput. Phys., 2022.

[30] Claudia Totzeck and Marie-Therese Wolfram. Consensus-based global optimization with personal best. Math. Biosci. Eng., 2020.

[31] Giacomo Borghi, Sara Grassi, and Lorenzo Pareschi. Consensus based optimization with memory effects: random selection and applications. Chaos Solitons Fractals, 2023.

[32] Massimo Fornasier, Peter Richtárik, Konstantin Riedl, and Lukang Sun. Consensus-based optimization with truncated noise. 2023.

[33] Jose A. Carrillo, Nicolas Garcia Trillos, Sixu Li, and Yuhua Zhu. FedCBO: reaching group consensus in clustered federated learning through consensus-based optimization. 2023.

[34] Hyeong-Ohk Bae et al. A constrained consensus based optimization algorithm and its application to finance. Appl. Math. Comput., 2022.

Convergence results in mean-field law of CBO

Mean-field limit: $\bar{\theta}_t$ with $\mu_t = \text{law}(\bar{\theta}_t)$ solves

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

$\longrightarrow$ Associated Fokker–Planck equation is nonlinear and nonlocal:

$$\partial_t \mu = \nabla\cdot\Big(\big(\theta - \mathcal{M}_\beta(\mu)\big)\mu\Big) + \sigma^2 \triangle\Big(\big|\theta - \mathcal{M}_\beta(\mu)\big|^2\mu\Big).$$

Let $W_2 : \mathcal{P}_2(\mathbf{R}^d) \times \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ denote the Wasserstein-2 metric.

Convergence of mean field CBO[38],[39]

Under mild conditions including existence of a unique minimizer, there is λ such that

$$\forall t \in [0, T_\beta], \qquad W_2(\overline{\rho}_t, \delta_{\theta_*}) \le W_2(\overline{\rho}_0, \delta_{\theta_*})\,\mathrm{e}^{-\lambda t}, \qquad \theta_* = \operatorname*{arg\,min}_{\theta\in\mathbf{R}^d} f.$$

Furthermore $T_\beta \to \infty$ as $\beta \to \infty$.

Proof Strategy

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

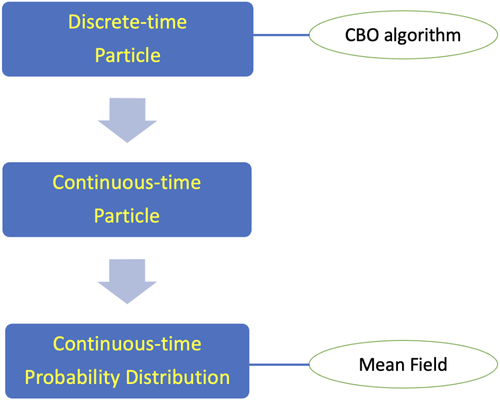
(2) **Laplace's method:** $\widehat{\theta}(\beta) \to \arg\min_{\theta\in\mathbf{R}^d} f(\theta)$ as $\beta \to \infty$.

[38] 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, 2018.

[39] Massimo Fornasier, Timo Klock, and Konstantin Riedl. Consensus-based optimization methods converge globally. Arxiv preprint, 2021.

From implementation to PDE analysis

- **Analysis** for **Mean Field PDE** mostly
- **Algorithm** is for **Discrete-time particle** evolution



Algorithm – Particle SDE – Mean-field PDE

(1) **Discrete-time to Continuous-time particle evolutions**

(2) **Continuous-time particle evolution to mean-field limit**

From implementation to PDE analysis

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\theta_{\Delta t}^{(i)}-\theta_*\right\|_2^2\right] \\
&\quad\leq \underbrace{\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\theta_{\Delta t}^{(i)}-\theta_t^{(i)}\right\|_2^2\right]}_{\text{(i) time-discretization error}} + \underbrace{\mathbb{E}\left[\left\|\theta_t^{(i)}-\overline{\theta}_t\right\|_2^2\right]}_{\text{(ii) propagation of chaos error}} + \underbrace{\mathbb{E}\left[\left\|\overline{\theta}_t-\theta_*\right\|_2^2\right]}_{\text{(iii) optimization error}}.
\end{aligned}
$$

Algorithm – Particle SDE – Mean-field PDE

(1) **Discrete-time to Continuous-time particle evolutions**

(2) **Continuous-time particle evolution to mean-field limit**

- Finite-in-time mean-field limit[40]:

$$
\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\theta_t^{(i)}-\bar{\theta}_t\right|^2\right]^{1/2} \leq C_T N^{-\frac{1}{2}}, \text{ where } C_T \to \infty \text{ as } T \to \infty
$$

- Uniform-in-time mean-field limit:[41] $C_T = C$ independent of time!

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

[41] Nicolai J. Gerber, Franca Hoffmann, Dohyeon Kim, and Urbain Vaes. Uniform-in-time propagation of chaos for consensus-based optimization. arXiv:2505.08669, 2025.

Convergence for the interacting particle system

By the triangle inequality,

$$\mathbf{E}\Big[W_2(\mu_t^J, \delta_{\theta^*})\Big] \leq \underbrace{\mathbf{E}\Big[W_2(\mu_t^J, \overline{\rho}_t)\Big]}_{\to 0 \text{ as } J\to\infty???} + \underbrace{W_2(\overline{\rho}_t, \delta_{\theta^*})}_{\leq C\,\mathrm{e}^{-\lambda t}} .$$

Pre-existing mean field results for CBO (i.i.d. initial condition and fixed t)

- [42] Based on a compactness argument, it was shown that

$$\mu_t^J \xrightarrow[J\to\infty]{\text{Law}} \overline{\rho}_t \qquad \text{(no rate)}.$$

- [43] For all $\varepsilon > 0$, there is $\Omega_\varepsilon \subset \Omega$ and $C_\varepsilon > 0$ such that for all J

$$\mathbf{P}[\Omega \setminus \Omega_\varepsilon] \leq \varepsilon \quad \text{and} \quad \mathbf{E}\left[W_2(\mu_t^J, \overline{\rho}_t) \,\middle|\, \Omega_\varepsilon\right] \leq C_\varepsilon J^{-\alpha}, \qquad C_\varepsilon \xrightarrow[\varepsilon\to 0]{} \infty$$

Our goal: obtain an estimate of the form $\mathbf{E}\left[W_2(\mu_t^J, \overline{\rho}_t)\right] \leq C J^{-\alpha}$.

[42] Hui Huang and Jinniao Qiu. On the mean-field limit for the consensus-based optimization. Math. Methods Appl. Sci., 2022.

[43] Massimo Fornasier, Timo Klock, and Konstantin Riedl. Consensus-based optimization methods converge globally. Arxiv preprint, 2021.

Why the classical Sznitman approach fails for CBO

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

$$
\begin{aligned}
\mathrm{d}X_t^j &= -\Big(X_t^j - \mathcal{M}_\beta\Big(\mu_t^J\Big)\Big)\,\mathrm{d}t + \sqrt{2}\sigma\Big|X_t^j - \mathcal{M}_\beta\Big(\mu_t^J\Big)\Big|\,\mathrm{d}W_t^j, \qquad & X_0^j &= x_0^j. \\
\mathrm{d}\overline{X}_t^j &= -\Big(\overline{X}_t^j - \mathcal{M}_\beta(\overline{\rho}_t)\Big)\,\mathrm{d}t + \sqrt{2}\sigma\Big|\overline{X}_t^j - \mathcal{M}_\beta(\overline{\rho}_t)\Big|\,\mathrm{d}W_t^j, \qquad & \overline{X}_0^j &= x_0^j.
\end{aligned}
$$

Technical difficulty:

- $\mathcal{M}_\beta : \mathcal{P}_1(\mathbf{R}^d) \to \mathbf{R}^d$ is **not globally Lipschitz** continuous in general.

Quantitative mean field limits

Assumption (focusing on the unbounded f setting for simplicity here)

- **Local Lischitz continuity.** f is bounded from below by $f_\star = \inf f$ and satisfies

$$\forall x, y \in \mathbf{R}^d, \qquad |f(x) - f(y)| \leq L_f \bigl(1 + |x| + |y|\bigr)^s |x - y|, \qquad s \geq -1.$$

- **Growth at infinity**. There are constants $c, u > 0$ and a compact $K \subset \mathbf{R}^d$ such that

$$\forall x \in \mathbf{R}^d \setminus K, \qquad \frac{1}{c}|x|^u \leq f(x) \leq c|x|^u.$$

Theorem (Finite-time Mean-Field Limit[44])

If f satisfies the above assumption and $\overline{\rho}_0$ has infinitely many moments, then

$$\forall J \in \mathbf{N}^+, \qquad \forall j \in \{1, \dots, J\}, \qquad \mathbf{E}\left[\sup_{t \in [0,T]} \left|X_t^j - \overline{X}_t^j\right|^p\right] \leq C J^{-\frac{p}{2}}.$$

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

Definition of $\mathcal{P}_{p,R}(\mathbf{R}^d)$

$$\mathcal{P}_{p,R}(\mathbf{R}^d) = \Big\{\mu \in \mathcal{P}_p(\mathbf{R}^d) : W_p(\mu, \delta_0) \leq R\Big\}.$$

- **Local Lipschitz continuity for $\mathcal{M}_\beta$.** For all $R > 0$ and for all $p \geq 1$, $\exists L$ s.t.

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_p(\mathbf{R}^d), \qquad \Big|\mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu)\Big| \leq L\, W_p(\mu, \nu).$$

- **Moment bound**: Suppose $\overline{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$. Then there is $\kappa > 0$ such that

$$\forall J \in \mathbf{N}^+, \qquad \mathbf{E}\left[\sup_{t\in[0,T]} \left|X_t^j\right|^q\right] \quad \vee \quad \mathbf{E}\left[\sup_{t\in[0,T]} \left|\overline{X}_t^j\right|^q\right] \leq \kappa.$$

Sketch of the proof: stopping time approach[45]

- Local Lipschitz continuity of $\mathcal{M}_\beta$ motivates stopping time

$$\theta_J = \inf\Big\{t \geq 0 : W_2(\overline{\mu}^J_t, \delta_0) \geq R\Big\}, \qquad \overline{\mu}^J_t := \frac{1}{J}\sum_{j=1}^{J} \delta_{\overline{X}^j_t}.$$

- Then decompose

$$\mathbf{E}\left[\left|X^1_t - \overline{X}^1_t\right|^p\right] = \mathbf{E}\left[\left|X^1_t - \overline{X}^1_t\right|^p \mathbf{1}_{\{\theta_J > T\}}\right] + \mathbf{E}\left[\left|X^1_t - \overline{X}^1_t\right|^p \mathbf{1}_{\{\theta_J \leq T\}}\right].$$

- First term can be shown to scale as $CJ^{-\frac{p}{2}}$ using classical approach;
- Second term requires handled as follows ($q > p$)

$$\mathbf{E}\left[\left|X^j_t - \overline{X}^j_t\right|^p \mathbf{1}_{\{\theta_J \leq T\}}\right] \leq \mathbf{E}\left[\left|X^j_t - \overline{X}^j_t\right|^q\right]^{\frac{p}{q}} \mathbf{P}\left[\theta_J \leq T\right]^{\frac{q-p}{q}}.$$

  - First factor bounded using moment bounds.
  - Second factor: for sufficiently large R, by generalized Chebyshev inequality,

$$\forall a > 0, \quad \exists C(a): \qquad \mathbf{P}\left[\theta_J \leq T\right] \leq C(a) J^{-a}$$

[45] Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM J. Numer. Anal., 2002.

Paper	Result Type	Limitations
[46]	Qualitative MFL	No Rate
[47],[48]	Finite-time MFL	Time-dependent rate
[49]	Uniform-time MFL	Modified dynamics
[50]	Uniform-time Weak PoC	Modified dynamics
[51]	Uniform-time Strong PoC	Original dynamics

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

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

- **Goal**: Remove time-dependence of C for the unmodified algorithm

[46] Hui Huang and Jinniao Qiu. On the mean-field limit for the consensus-based optimization. Math. Methods Appl. Sci., 2022.

[47] Hui Huang, Jinniao Qiu, and Konstantin Riedl. On the global convergence of particle swarm optimization methods. Appl. Math. Optim., 2023.

[48] Nicolai Jurek Gerber, Franca Hoffmann, and Urbain Vaes. Mean-field limits for consensus-based optimization and sampling. 2024.

[49] Hui Huang and Hicham Kouhkouh. Uniform-in-time mean-field limit estimate for the consensus-based optimization. arXiv preprint, 2024.

[50] Erhan Bayraktar, Ibrahim Ekren, and Hongyi Zhou. Uniform-in-time weak propagation of chaos for consensus-based optimization. arXiv preprint, 2025.

[51] Nicolai Gerber, Franca Hoffmann, Dohyeon Kim, and Urbain Vaes. Uniform-in-time propagation of chaos for consensus-based optimization. 2026.

Main Result 1

Uniform-in-time propagation of chaos [Gerber, H., Kim, Vaes (2025)][52]

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 $C_{\text{MFL}}\big(\alpha, \underline{f}, \overline{f}, L_f, \sigma, d\big)$ such that for $J \in \mathbb{N}$ particles:

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

- 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**

[52] Nicolai Gerber, Franca Hoffmann, Dohyeon Kim, and Urbain Vaes. Uniform-in-time propagation of chaos for consensus-based optimization. 2026.

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

Key Lemma

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

$$
\forall(\mu,\nu), \qquad \left|\mathcal{M}(\mu) - \mathcal{M}(\nu)\right| \leq_2 (\mu,\nu).
$$

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

Why the classical approach fails for CBO

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

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

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

Technical difficulties:

- **Non-Lipschitz drift:** The weighted mean $\mathcal{M}_\beta : \mathcal{P}_1(\mathbf{R}^d) \to \mathbf{R}^d$ 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 $_t$.

Extending Synchronous Coupling approach to CBO

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

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

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

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2J}\sum_{j=1}^{J}\mathbf{E}\left|X_t^j-\overline{X}_t^j\right|^2 &= \underbrace{-\frac{1}{J}\mathbf{E}\sum_{j=1}^{J}\left\langle X_t^j-\overline{X}_t^j, X_t^j-(\,_t)-\overline{X}_t^j+(\,_t)\right\rangle}_{\color{blue}{(\mathrm{I})}} \\
&\underbrace{-\frac{1}{J}\mathbf{E}\sum_{j=1}^{J}\left\langle X_t^j-\overline{X}_t^j, (\,_t)-\mathcal{M}_{\color{red}{\beta}}(\,_t)-(\,_t)+\mathcal{M}_{\color{red}{\beta}}(\,_t)\right\rangle}_{\color{red}{(\mathrm{II})}} \\
&\underbrace{-\frac{1}{J}\mathbf{E}\sum_{j=1}^{J}\left\langle X_t^j-\overline{X}_t^j, t-\mathcal{M}_{\color{red}{\beta}}(\,_t)\right\rangle}_{\color{green}{(\mathrm{III})}}\,.
\end{aligned}
$$

- **(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
- **(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 $p \geq 2$,

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

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

  - Recall (decay of centered moments):

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

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:

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

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

$$|-\mathcal{M}(\mu) - +\mathcal{M}(\nu)| \leq C_{\mathcal{M}} \left(\sqrt{2\mu} + \sqrt{2\nu} \right)_2 (\mu, \nu).$$

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

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

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

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

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

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

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

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:

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

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

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

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:

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

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

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

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

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**

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?

Outline

Can we construct a sampling method using ideas from CBO?

Notation: $\mathcal{M}_\beta$ weighted mean, $\mathcal{C}_\beta$ weighted covariance, $\mathcal{R}_\beta$ reweighting:

$$\mathcal{M}_\beta(\mu) = \mathcal{M}(\mathcal{R}_\beta \mu)\,, \quad \mathcal{C}_\beta(\mu) = \mathcal{C}(\mathcal{R}_\beta \mu)\,, \quad \mathcal{R}_\beta : \mu \mapsto \frac{\mu\, \mathrm{e}^{-\beta f}}{\int \mu\, \mathrm{e}^{-\beta f}}\,,$$

$$\mathcal{M}(\mu) = \int \theta \mu(\mathrm{d}\theta)\,, \quad \mathcal{C}(\mu) = \int \big(\theta - \mathcal{M}(\mu)\big) \otimes \big(\theta - \mathcal{M}(\mu)\big)\mu(\mathrm{d}\theta)\,.$$

Discrete-time consensus based sampling ($\beta \geq 0$)[53]

$$\begin{cases} \theta_{n+1} = \mathcal{M}_\beta(\mu_n) + \alpha\big(\theta_n - \mathcal{M}_\beta(\mu_n)\big) + \sqrt{\gamma \mathcal{C}_\beta(\mu_n)}\,\xi_n, \qquad \xi_n \sim \mathcal{N}(0, I_d), \\ \mu_n = \mathrm{Law}(\theta_n). \end{cases}$$

Question: Are there choices of (α, β, γ) such that e^{-f} is a steady state?

[53] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and UV. Consensus Based Sampling. Studies In Applied Mathematics, 2021.

Determining the parameters $\mathrm{e}^{-f} = \mathcal{N}(a, A)$

Discrete-time consensus-based sampling ($\beta \geq 0$)

$$\begin{cases} \theta_{n+1} = \mathcal{M}_\beta(\mu_n) + \alpha\big(\theta_n - \mathcal{M}_\beta(\mu_n)\big) + \sqrt{\gamma \mathcal{C}_\beta(\mu_n)}\, \xi_n, & \xi_n \sim \mathcal{N}(0, I_d), \\ \mu_n = \mathrm{Law}(\theta_n). \end{cases}$$

A simple explicit calculation shows that

$$\mathcal{M}_\beta(\mathrm{e}^{-f}) = a,$$
$$\mathcal{C}_\beta(\mathrm{e}^{-f}) = (1+\beta)^{-1} A.$$

If $\theta_n \sim \mathcal{N}(a, A)$, then

$$\theta_{n+1} \sim \mathcal{N}(a, \alpha^2 A + \gamma(1+\beta)^{-1} A).$$

Therefore $\mathrm{e}^{-f} = \mathcal{N}(a, A)$ is a steady state if

$$\alpha \in [-1, 1], \qquad \gamma = (1-\alpha^2)(1+\beta).$$

For what parameters is the target $\mathcal{N}(a, A)$ an attractor?

If $\theta_n \sim \mathcal{N}(m_n, C_n)$, then a calculation shows $\theta_{n+1} \sim \mathcal{N}(m_{n+1}, C_{n+1})$ with

$$\begin{aligned} m_{n+1} &= \alpha m_n + (1-\alpha)\left(C_n^{-1} + \beta A^{-1}\right)^{-1}\left(\beta A^{-1} a + C_n^{-1} m_n\right), \\ C_{n+1} &= \alpha^2 C_n + \gamma\left(C_n^{-1} + \beta A^{-1}\right)^{-1}, \end{aligned}$$

For e^{-f} to be an attractor for Gaussian initial conditions, we need in fact $\alpha \in (-1, 1)$.

Convergence result for target $\mathcal{N}(a, A)$ and Gaussian initial condition

If $\alpha \in (-1, 1)$ and $\gamma = (1-\alpha^2)(1+\beta)$, then

$$|m_n - a|_A + \|C_n - A\|_A \leq C\left(\frac{1-|\alpha|}{1+\beta} + |\alpha|\right)^n$$

Questions:

- Is $\mathcal{N}(a, A)$ an attractor for non-Gaussian initial conditions?
- What if the target e^{-f} is not Gaussian?

Interpretation as discretization of McKean SDE

When $\alpha = \mathrm{e}^{-\Delta t}$ with $\Delta t \ll 1$, the CBS dynamics

$$\begin{cases} \theta_{n+1} = \mathcal{M}_\beta(\mu_n) + \alpha\big(\theta_n - \mathcal{M}_\beta(\mu_n)\big) + \sqrt{(1-\alpha^2)(1+\beta)\mathcal{C}_\beta(\mu_n)}\,\xi_n, \quad \xi_n \sim \mathcal{N}(0, I_d), \\ \mu_n = \mathrm{Law}(\theta_n). \end{cases}$$

may be viewed as a discretization with time step Δt of the McKean SDE

$$\begin{cases} \mathrm{d}\theta_t = -\big(\theta_t - \mathcal{M}_\beta(\mu_t)\big)\,\mathrm{d}t + \sqrt{2(1+\beta)\mathcal{C}_\beta(\mu_t)}\,\mathrm{d}W_t, \\ \mu_t = \mathrm{Law}(\theta_t) \end{cases}$$

$\rightarrow$ Continuous-time sampling method.

Continuous-time CBS dynamics

Notation: $\mathcal{M}_\beta$ weighted mean, $\mathcal{C}_\beta$ weighted covariance, $\mathcal{R}_\beta$ reweighting:

$$\mathcal{M}_\beta(\mu) = \mathcal{M}(\mathcal{R}_\beta\mu)\,, \quad \mathcal{C}_\beta(\mu) = \mathcal{C}(\mathcal{R}_\beta\mu)\,, \quad \mathcal{R}_\beta : \mu \mapsto \frac{\mu\,\mathrm{e}^{-\beta f}}{\int \mu\,\mathrm{e}^{-\beta f}}\,,$$

$$\mathcal{M}(\mu) = \int \theta\mu(\mathrm{d}\theta)\,, \quad \mathcal{C}(\mu) = \int \big(\theta - \mathcal{M}(\mu)\big) \otimes \big(\theta - \mathcal{M}(\mu)\big)\mu(\mathrm{d}\theta)\,.$$

Consensus Based Sampling (CBS)[54]

$$\Big\{ \mathrm{d}\theta_t^{(j)} = -\big(\theta_t^{(j)} - \mathcal{M}_\beta(\mu_t^J)\big)\,\mathrm{d}t + \sqrt{2(1+\beta)\mathcal{C}_\beta(\mu_t^J)}\,\mathrm{d}W_t^{(j)}$$

Convergence of mean field CBS

If $\pi \propto \mathrm{e}^{-f}$ is Gaussian and $\overline{\rho}_0$ is Gaussian, then

$$\forall t \geq 0, \qquad W_2(\overline{\rho}_t, \pi) \leq C\,\mathrm{e}^{-\left(\frac{\beta}{1+\beta}\right)t}\,.$$

[54] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. Consensus-based sampling. Stud. Appl. Math., 2022.

Mean Field Limit: The law μ of θ_t evolves according to

$$\partial_t \mu = \nabla \cdot \Big(\big(\theta - \mathcal{M}_\beta(\mu)\big)\mu + (1+\beta)\mathcal{C}_\beta(\mu)\, \nabla \mu \Big).$$

- This dynamic propagates Gaussians even when e^{-f} is non-Gaussian;
- Any steady state must satisfy

$$\mu_\infty = \mathcal{N}\big(\mathcal{M}_\beta(\mu_\infty), (1+\beta)\mathcal{C}_\beta(\mu_\infty)\big).$$

$\rightarrow$ Recovers the Laplace approximation instead.

Convergence of the solution

Let us introduce

$$\widehat{f}(\theta) = f(\theta_*) + \frac{1}{2} \operatorname{Hess} f(\theta_*) : \big((\theta - \theta_*) \otimes (\theta - \theta_*)\big).$$

The distribution $\mathrm{e}^{-\widehat{f}} \propto \mathcal{N}(\theta_*, C_*)$ is the Laplace approximation of e^{-f}.

Convergence result

Under appropriate assumptions (one-dimensional, convex),

- There exists a unique steady-state $\mathcal{N}\big(m_\infty(\beta), C_\infty(\beta)\big)$ satisfying

$$\big|m_\infty(\beta) - \theta_*\big| + \big\|C_\infty(\beta) - C_*\big\| = \mathcal{O}(\beta^{-1}).$$

- If the initial condition is Gaussian, then

$$\big|m(t) - m_\infty(\beta)\big| + \big\|C(t) - C_\infty(\beta)\big\| \leq C \exp\left(-\left(1 - \frac{k}{\beta}\right) t\right).$$

Idea of the proof: Laplace's method and contraction argument.

With the parameter choice $\gamma = (1 - \alpha^2)$, we obtain an optimization method.

Discrete-time optimization variant:

$$\begin{cases} \theta_{n+1} = \mathcal{M}_\beta(\mu_n) + \alpha\big(\theta_n - \mathcal{M}_\beta(\mu_n)\big) + \sqrt{(1-\alpha^2)\mathcal{C}_\beta(\mu_n)}\,\xi_n, & \xi_n \sim \mathcal{N}(0, I_d), \\ \mu_n = \mathrm{Law}(\theta_n). \end{cases}$$

Continuous-time optimization variant:

$$\begin{cases} \mathrm{d}\theta_t = -\big(\theta_t - \mathcal{M}_\beta(\mu_t)\big)\,\mathrm{d}t + \sqrt{2\mathcal{C}_\beta(\mu_t)}\,\mathrm{d}W_t, \\ \mu_t = \mathrm{Law}(\theta_t) \end{cases}$$

Convergence result for the optimization method

If $\theta_0 \sim \mathcal{N}(m_0, C_0)$ and under appropriate assumptions (one-dimenisonal, convex),

$$W_2(\mu_n, \delta_{\theta_*}) \leq C n^{-p}, \qquad W_2(\mu_t, \delta_{\theta_*}) \leq C t^{-p}, \qquad p \in (0,1).$$

Outline

Example 1: one-dimensional elliptic BVP – Sampling

Find $(\theta_1, \theta_2) \in \mathbf{R}^2$ from noisy observations of $(p(.25), p(.75)) \in \mathbf{R}^2$, where $p(x)$ solves

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{\theta_1}\frac{\mathrm{d}p}{\mathrm{d}x}\right) = 1, \qquad x \in [0,1],$$

with boundary conditions $p(0) = 0$ and $p(1) = \theta_2$.

- Explicit solution $p(x, \theta)$ is available
- We define

$$G(\theta) = \begin{pmatrix} p(x_1, \theta) \\ p(x_2, \theta) \end{pmatrix}.$$

- Contour plots: $f(\theta) = \frac{1}{2\gamma^2}|y - G(\theta)|^2 + \frac{1}{2\sigma^2}|\theta|^2$

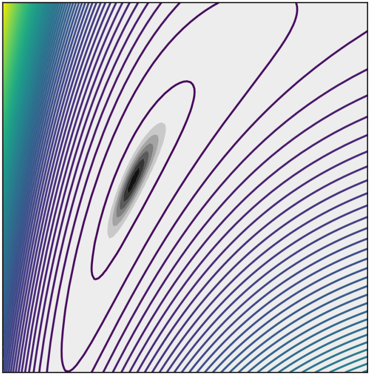


Figure: Contour plots.

Optimization: objective functions

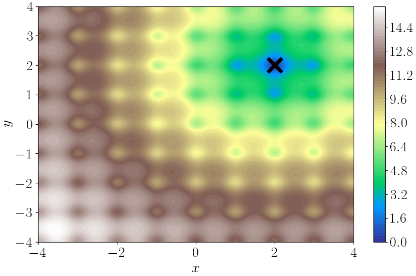
- the Ackley function, defined for $x \in \mathbf{R}^d$ by

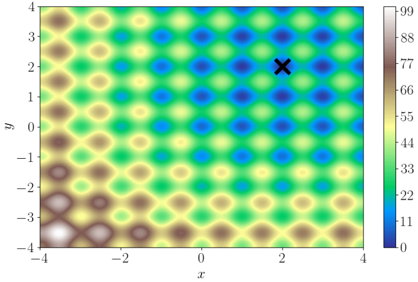
$$f_A(x) = -20 \exp\left(-\frac{1}{5}\sqrt{\frac{1}{d}\sum_{i=1}^{d} |x_i - b|^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d} \cos\bigl(2\pi(x_i - b)\bigr)\right) + \mathrm{e} + 20,$$

- the Rastrigin function, defined by

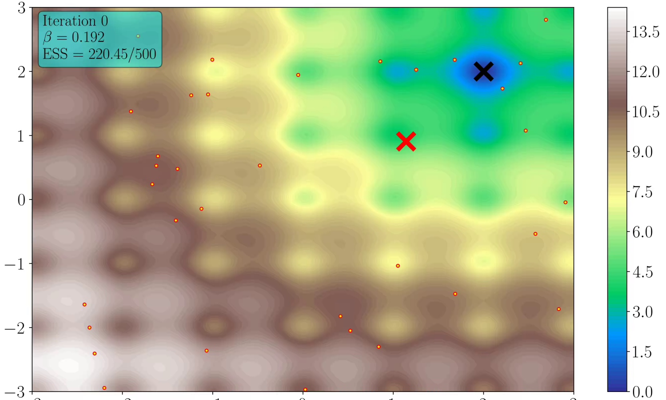
$$f_R(x) = \sum_{i=1}^{d} \Bigl((x_i - b)^2 - 10\cos\bigl(2\pi(x_i - b)\bigr) + 10\Bigr).$$

Minimizer: $x_* = (b, \dots, b)$, where $b \in \mathbf{R}$. Below $b = 2$.





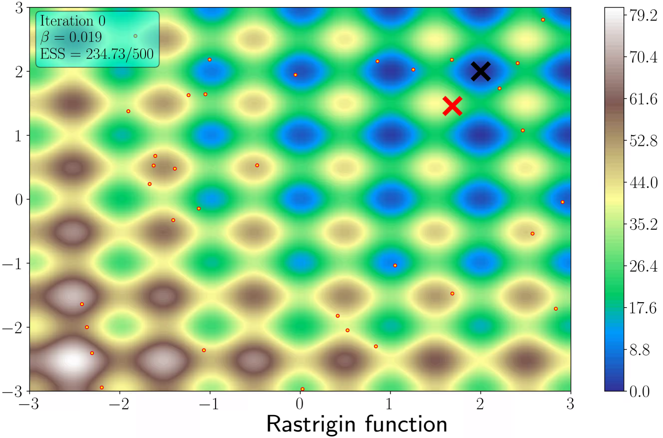
Convergence for $\alpha = .1$, adaptive β with $J_{\text{eff}}/J = .5$, and $J = 100$.



Ackley function

Optimization: illustration of the convergence

Convergence for $\alpha = .1$, adaptive β with $J_{\text{eff}}/J = .5$, and $J = 100$.



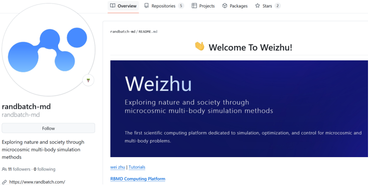
Outline

Software Implementation[55]

- standard CBO
- CBO with mini-batching
- polarised CBO
- CBO with memory effects
- consensus-based sampling (CBS)
- parallelisation tools
- performance evaluation methods
- visualisation tools

Random batch methods for molecular dynamics:

- order $O(N)$
- highly scalable



Github: `https://pdips.github.io/CBXpy/index.html`
Github: `https://github.com/randbatch-md`

[55] Rafael Bailo et al. Cbx: python and julia packages for consensus-based interacting particle methods. 2024.

Outline

- Gaussian initial condition with strictly positive definite covariance.

	Convergence	
	Mean	Covariance
$\alpha = 0$	$\left(\frac{1}{1+\beta}\right)^n$	$\left(\frac{1}{1+\beta}\right)^n$
$\alpha \in (0,1)$	$\left(\frac{1+\alpha\beta}{1+\beta}\right)^n$	$\left(\frac{1+\alpha^2\beta}{1+\beta}\right)^n$
$\alpha = 1$	$\mathrm{e}^{-\left(\frac{\beta}{1+\beta}\right)t}$	$\mathrm{e}^{-\left(\frac{2\beta}{1+\beta}\right)t}$

- $f \in C^2(\mathbf{R}^d)$ and
$$\ell I_d \leq L \leq \mathrm{Hess}\, f(\theta) \leq U \leq u I_d \,,$$
for all $\theta \in \mathbf{R}^d$ and some $\ell, u > 0$.
- Let $k > 0$ independent of n, t, α and β.
- Gaussian initial condition with strictly positive definite covariance.

	Convergence	
	Mean ($d = 1$)	Covariance ($d = 1$)
$\alpha = 0$	$\left(\frac{k}{\beta}\right)^n$	$\left(\frac{k}{\beta}\right)^n$
$\alpha \in (0,1)$	$\left(\alpha + (1-\alpha^2)\frac{k}{\beta}\right)^n$	$\left(\alpha + (1-\alpha^2)\frac{k}{\beta}\right)^n$
$\alpha = 1$	$\mathrm{e}^{-\left(1-\frac{2k}{\beta}\right)t}$	$\mathrm{e}^{-\left(1-\frac{2k}{\beta}\right)t}$

Take-Aways

CBO:

- For $\beta \to \infty$ the measure $\mathcal{R}_\beta \rho$ concentrates on δ_{θ_*} (Laplace's method).
- Two step method to prove convergence on the mean-field level for a large class of objective functions f.

CBS:

- Laplace approximation: rescaling of the covariance by $(1-\alpha^2)(1+\beta)$ enables recovery of a Gaussian approximation of the desired target measure $\exp\big(-f(\bullet)\big)$.
- Fixing the scale at $(1-\alpha^2)$ allows the covariance to remain small when optimization of $f(\bullet)$ is the desired goal.
- Laplace method allows us to provide convergence guarantees beyond the Gaussian setting.

CBO & CBS

- algorithm for optimization and sampling;
- is based on a stochastic interacting particle system:
  - can be parallelized easily;
  - can be studied from a mean field viewpoint.
- is derivative-free, so well suited for PDE inverse problems;
- converges exponentially fast at the mean-field level;
- is affine-invariant, so convergence rate is independent of target in Gaussian setting.

Perspectives for CBX:

- Can we obtain uniform-in-time mean-field limit results ?
- Can we make a rigorous connection to the discrete-time algorithms?
- Can we study the method with adaptive β?
- Can we prove convergence at the particle level?
- Overarching framework!
- Comparison to other methods and software implementations!

Thank you for your attention!

Konstantin ALTHAUS, Iason PAPAIOANNOU, and Elisabeth ULLMANN. Consensus-based rare event estimation. arXiv preprint, 2304.09077, 2023. eprint: arXiv:2304.09077.

Hyeong-Ohk BAE et al. A constrained consensus based optimization algorithm and its application to finance. Appl. Math. Comput., 416:Paper No. 126726, 10, 2022.

Rafael BAILO et al. Cbx: python and julia packages for consensus-based interacting particle methods. 2024. eprint: arXiv:2403.14470.

Erhan BAYRAKTAR, Ibrahim EKREN, and Hongyi ZHOU. Uniform-in-time weak propagation of chaos for consensus-based optimization. arXiv preprint, 2502.00582, 2025.

Giacomo BORGHI, Sara GRASSI, and Lorenzo PARESCHI. Consensus based optimization with memory effects: random selection and applications. Chaos Solitons Fractals, 174:Paper No. 113859, 17, 2023.

Giacomo BORGHI, Michael HERTY, and Lorenzo PARESCHI. An adaptive consensus based method for multi-objective optimization with uniform Pareto front approximation. Appl. Math. Optim., 88(2):Paper No. 58, 43, 2023.

References II

Leon Bungert, Tim Roith, and Philipp Wacker. Polarized consensus-based dynamics for optimization and sampling. arXiv preprint, 2211.05238, 2022. arXiv: 2211.05238 [math.OC].

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 UV. Consensus Based Sampling. Studies In Applied Mathematics, 2106.02519, October 2021.

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

J.A. Carrillo, C. Totzeck, and U. Vaes. Consensus-based optimization and ensemble kalman inversion for global optimization problems with constraints. arXiv:2111.02970, 2021.

José A. Carrillo, Shi Jin, Lei Li, and Yuhua Zhu. A consensus-based global optimization method for high dimensional machine learning problems. ESAIM Control Optim. Calc. Var., 27:Paper No. S5, 22, 2021.

References III

José A. Carrillo, Shi Jin, Haoyu Zhang, and Yuhua Zhu. An interacting particle consensus method for constrained global optimization. 2024. eprint: arXiv:2405.00891.

Jose A. Carrillo, Nicolas Garcia Trillos, Sixu Li, and Yuhua Zhu. FedCBO: reaching group consensus in clustered federated learning through consensus-based optimization. 2023. eprint: arXiv:2305.02894.

Jingrun Chen, Shi Jin, and Liyao Lyu. A consensus-based global optimization method with adaptive momentum estimation. Commun. Comput. Phys., 31(4):1296–1316, 2022.

M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In Handbook of uncertainty quantification. Vol. 1, 2, 3, pages 311–428. Springer, Cham, 2017.

P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. J. R. Stat. Soc. Ser. B Stat. Methodol., 68(3):411–436, 2006.

Massimo Fornasier, Hui-Lin Huang, Lorenzo Pareschi, and Philippe Sunnen. Consensus-based optimization on hypersurfaces: well-posedness and mean-field limit. Mathematical Models and Methods in Applied Sciences, 2020.

References IV

- Massimo Fornasier, Timo Klock, and Konstantin Riedl. Consensus-based optimization methods converge globally. Arxiv preprint, 2103.15130, 2021.
- Massimo Fornasier, Peter Richtárik, Konstantin Riedl, and Lukang Sun. Consensus-based optimization with truncated noise. 2023. eprint: arXiv:2310.16610.
- Massimo Fornasier and Lukang Sun. A pde framework of consensus-based optimization for objectives with multiple global minimizers. 2024. eprint: arXiv:2403.06662.
- A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. SIAM J. Appl. Dyn. Syst., 19(1):412–441, 2020.
- Nicolai Gerber, Franca Hoffmann, Dohyeon Kim, and Urbain Vaes. Uniform-in-time propagation of chaos for consensus-based optimization. 2026. arXiv: 2505.08669 [math.PR].
- Nicolai J. Gerber, Franca Hoffmann, Dohyeon Kim, and Urbain Vaes. Uniform-in-time propagation of chaos for consensus-based optimization. arXiv:2505.08669, 2025. eprint: arXiv:2505.08669.

Nicolai Jurek Gerber, Franca Hoffmann, and Urbain Vaes. Mean-field limits for consensus-based optimization and sampling. 2023. eprint: arXiv:2312.07373.

Nicolai Jurek Gerber, Franca Hoffmann, and Urbain Vaes. Mean-field limits for consensus-based optimization and sampling. 2024. arXiv: 2312.07373 [math.PR].

J. Goodman and J. Weare. Ensemble samplers with affine invariance. Commun. Appl. Math. Comput. Sci., 5(1):65–80, 2010.

Sara Grassi and Lorenzo Pareschi. From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit. Math. Models Methods Appl. Sci., 31(8):1625–1657, 2021.

Seung-Yeal Ha, Myeongju Kang, Dohyun Kim, Jeongho Kim, and Insoon Yang. Stochastic consensus dynamics for nonconvex optimization on the Stiefel manifold: mean-field limit and convergence. Math. Models Methods Appl. Sci., 32(3):533–617, 2022.

Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM J. Numer. Anal., 40(3):1041–1063, 2002.

Hui HUANG and Hicham KOUHKOUH. Uniform-in-time mean-field limit estimate for the consensus-based optimization. arXiv preprint, 2411.03986, 2024.

Hui HUANG and Jinniao QIU. On the mean-field limit for the consensus-based optimization. Math. Methods Appl. Sci., 45(12):7814–7831, 2022.

Hui HUANG and Jinniao QIU. On the mean-field limit for the consensus-based optimization. Math. Methods Appl. Sci., 45(12):7814–7831, 2022.

Hui HUANG, Jinniao QIU, and Konstantin RIEDL. On the global convergence of particle swarm optimization methods. Appl. Math. Optim., 88(2):Paper No. 30, 44, 2023.

M. A. IGLESIAS, K. J. H. LAW, and A. M. STUART. Ensemble Kalman methods for inverse problems. Inverse Problems, 29(4):045001, 20, 2013.

James KENNEDY and Russell EBERHART. Particle swarm optimization. In Proceedings of ICNN'95-international conference on neural networks, volume 4, pages 1942–1948. IEEE, 1995.

Kathrin KLAMROTH, Michael STIGLMAYR, and Claudia TOTZECK. Consensus-based optimization for multi-objective problems: a multi-swarm approach. arXiv preprint, 2211.15737, 2022. arXiv: 2211.15737 [math.OC].

- Dongnam Ko, Seung-Yeal Ha, Shi Jin, and Doheon Kim. Convergence analysis of the discrete consensus-based optimization algorithm with random batch interactions and heterogeneous noises. Math. Models Methods Appl. Sci., 32(6):1071–1107, 2022.
- Q. Liu and D. Wang. Stein variational gradient descent: a general purpose Bayesian inference algorithm. In Advances In Neural Information Processing Systems, pages 2378–2386, 2016.
- Jingcheng Lu, Eitan Tadmor, and Anil Zenginoglu. Swarm-based gradient descent method for non-convex optimization. 2022. eprint: arXiv:2211.17157.
- P. D. Miller. Applied asymptotic analysis. Volume 75 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2006, pages xvi+467.
- 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.
- A. M. Stuart. Inverse problems: a Bayesian perspective. Acta Numer., 19:451–559, 2010.

References VIII

Claudia Totzeck and Marie-Therese Wolfram. Consensus-based global optimization with personal best. Math. Biosci. Eng., 17(5):6026–6044, 2020.