Mean-field limits for Consensus-Based Optimization and Sampling

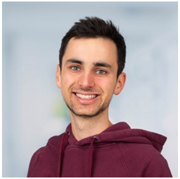
ECCOMAS – Minisymposium *Novel Kinetic Approaches*

Urbain Vaes
urbain.vaes@inria.fr

MATHERIALS – Inria Paris & CERMICS – École des Ponts ParisTech

June 4, 2024

Collaborators and reference



Nicolai Gerber

Hausdorff Center for Mathematics

Franca Hoffmann

Caltech

Reference:

- N. J. Gerber, F. Hoffmann, and UV. Arxiv preprint, 2023
  Mean-field limits for Consensus-Based Optimization and Sampling

Outline

Consensus-based optimization (CBO)[1,2]

Global optimization problem:

$$\text{Find } x \in \underset{x \in \mathbf{R}^d}{\arg\min} f \qquad (f : \mathbf{R}^d \to \mathbf{R})$$

CBO interacting particle system

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

- β is "inverse temperature" parameter.
- μ_t^J is empirical measure $\mu_t^J = \frac{1}{J}\sum_{j=1}^J \delta_{X_t^j}$.
- $\mathcal{M}_\beta : \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^d$ is weighted mean operator:

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

.

[1] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. Math. Models Methods Appl. Sci., 2017.
[2] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. Mathematical Models and Methods in Applied Sciences, 2018.

Consensus-based sampling (CBS)[1]

Sampling problem:

$$\text{Generate samples from distribution } \pi \propto \mathrm{e}^{-f} \qquad (f\colon \mathbf{R}^d \to \mathbf{R})$$

CBS interacting particle system

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

- β is "inverse temperature" parameter.
- μ_t^J is empirical measure $\mu_t^J = \frac{1}{J}\sum_{j=1}^J \delta_{X_t^j}$,
- $\mathcal{C}_\beta\colon \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^{d\times d}$ is weighted covariance operator:

$$\mathcal{C}_\beta(\mu) = \frac{\int (x \otimes x)\,\mathrm{e}^{-\beta f(x)}\,\mu(\mathrm{d}x)}{\int \mathrm{e}^{-\beta f(x)}\,\mu(\mathrm{d}x)} - \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu),$$

.

[1] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and UV. Stud. Appl. Math., 2022.

Mean field limits

Taking formally $J \to \infty$ in the interacting particle systems leads to

CBO mean field limit

$$\begin{cases} \mathrm{d}\overline{X}_t = -\left(\overline{X}_t - \mathcal{M}_\beta(\overline{\rho}_t)\right)\mathrm{d}t + \sqrt{2}\sigma\left|\overline{X}_t - \mathcal{M}_\beta(\overline{\rho}_t)\right|\mathrm{d}\overline{W}_t, \\ \overline{\rho}_t = \mathrm{Law}(\overline{X}_t). \end{cases}$$

CBS mean field limit

$$\begin{cases} \mathrm{d}\overline{X}_t = -\left(\overline{X}_t - \mathcal{M}_\beta(\overline{\rho})\right)\mathrm{d}t + \sqrt{2(1+\beta)\mathcal{C}_\beta(\overline{\rho}_t)}\,\mathrm{d}\overline{W}_t, \\ \overline{\rho}_t = \mathrm{Law}(\overline{X}_t). \end{cases}$$

- Nonlinear Markov processes in $\mathbf{R}^d$: future depends on $\overline{X}_t$ and its distribution
- Associated Fokker–Planck equations are nonlinear and nonlocal.

Convergence results in mean field law for CBO and CBS

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

Convergence of mean field CBO[1,2]

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

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

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

Convergence of mean field CBS[3]

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

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

[1] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. Mathematical Models and Methods in Applied Sciences, 2018.
[2] M. Fornasier, T. Klock, and K. Riedl. Arxiv preprint, 2021.
[3] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and UV. Stud. Appl. Math., 2022.

By the triangle inequality,

$$\mathbf{E}\Big[W_2(\mu_t^J,\nu)\Big] \leqslant \underbrace{\mathbf{E}\Big[W_2(\mu_t^J,\overline{\rho}_t)\Big]}_{\to 0 \text{ as } J\to\infty ???} + \underbrace{W_2(\overline{\rho}_t,\nu)}_{\leqslant C\,\mathrm{e}^{-\lambda t}}, \qquad \nu = \begin{cases} \delta_{x_*} & \text{for CBO,} \\ \mathrm{e}^{-f} & \text{for CBS.} \end{cases}$$

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

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

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

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

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

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

[1]H. Huang and J. Qiu. Math. Methods Appl. Sci., 2022.
[2]M. Fornasier, T. Klock, and K. Riedl. Arxiv preprint, 2021.

Outline

Introduction of synchronous coupling

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

Interacting particle system:

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

Mean field limit:

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

Synchronous coupling

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

$$\begin{aligned} \mathrm{d}X_t^j &= -\Big(X_t^j - \mathcal{M}\big(\mu_t^J\big)\Big)\,\mathrm{d}t + \mathrm{d}W_t^j, \qquad X_0^j = x_0^j, \qquad j = 1, \ldots, J, \\ \mathrm{d}\overline{X}_t^j &= -\Big(\overline{X}_t^j - \mathcal{M}(\overline{\rho}_t)\Big)\,\mathrm{d}t + \mathrm{d}W_t^j, \qquad \overline{X}_0^j = x_0^j, \qquad j = 1, \ldots, J, \end{aligned}$$

with same initial condition and driving Browian motions.

Using the synchronously coupled system as a pivot

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

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

Key triangle inequality

$$
\mathbf{E}\Big[W_2(\mu_t^J, \overline{\rho}_t)\Big] \leqslant \underbrace{\mathbf{E}\Big[W_2(\mu_t^J, \overline{\mu}_t^J)\Big]}_{\leqslant C J^{-???}} + \underbrace{\mathbf{E}\Big[W_2(\overline{\mu}_t^J, \overline{\rho}_t)\Big]}_{\leqslant C J^{-\alpha}}, \qquad \overline{\mu}_t^J = \frac{1}{J}\sum_{j=1}^{J} \delta_{\overline{X}_t^j}.
$$

- Second term controlled[1] independently of particle system.
- First term satisfies, by definition of the Wasserstein distance and by exchangeability

$$
\mathbf{E}\left[W_2(\mu_t^J, \overline{\mu}_t^J)^2\right] \leqslant \mathbf{E}\left[\frac{1}{J}\sum_{i=1}^{n}\left|X_t^j - \overline{X}_t^j\right|^2\right] \leqslant \mathbf{E}\left[\left|X_t^1 - \overline{X}_t^1\right|^2\right].
$$

$\rightsquigarrow$ It only remains to control $\mathbf{E}\left[\left|X_t^1 - \overline{X}_t^1\right|^2\right]$.

[1] N. Fournier and A. Guillin. Probab. Theory Related Fields, 2015.

Bounding the remaining term (using Sznitman's approach[1])

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

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

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

$$
\forall (\mu, \nu) \in \mathcal{P}_1(\mathbf{R}^d) \times \mathcal{P}_1(\mathbf{R}^d), \qquad \left|\mathcal{M}(\mu) - \mathcal{M}(\nu)\right| \leqslant W_2(\mu, \nu).
$$

$$
\begin{aligned}
\mathbf{E}\left[\left|X_t^1 - \overline{X}_t^1\right|^2\right] &\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 + \mathbf{E}\left|\mathcal{M}\left(\mu_s^J\right) - \mathcal{M}\left(\overline{\rho}_s\right)\right|^2 \,\mathrm{d}s \\
&\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 + \mathbf{E}\left|\mathcal{M}\left(\mu_s^J\right) - \mathcal{M}\left(\overline{\mu}_s^J\right)\right|^2 + \mathbf{E}\left|\mathcal{M}\left(\overline{\mu}_s^J\right) - \mathcal{M}\left(\overline{\rho}_s\right)\right|^2 \,\mathrm{d}s \\
&\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 + \mathbf{E}\left[W_2\left(\mu_s^J, \overline{\mu}_s^J\right)^2\right] \mathrm{d}s + C_{\mathrm{MC}} J^{-1} \\
&\lesssim \int_0^t \mathbf{E}\left|X_s^1 - \overline{X}_s^1\right|^2 \mathrm{d}s + C_{\mathrm{MC}} J^{-1} \qquad \overset{\text{Grönwall}}{\rightsquigarrow} \qquad \mathbf{E}\left[|X_t^1 - \overline{X}_t^1|^2\right] \leqslant C(t) J^{-1}.
\end{aligned}
$$

[1] A.-S. Sznitman. In École d'Été de Probabilités de Saint-Flour XIX—1989. Springer, Berlin, 1991.

Why the classical Sznitman approach fails for CBO/CBS

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

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

Technical difficulties:

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

$$\mathbf{E}\left|\mathcal{M}_\beta\left(\overline{\mu}_s^J\right) - \mathcal{M}_\beta(\overline{\rho}_s)\right|^2,$$

but estimates are given in the literature[1,2].

[1] P. Doukhan and G. Lang. Bernoulli, 2009.
[2] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. Statist. Sci., 2017.

Outline

Main result: 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

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

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

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

Main theorem[1], holds for both CBO and CBS

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

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

[1] N. J. Gerber, F. Hoffmann, and UV. Arxiv preprint, 2023.

Main ingredients of the proof

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

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

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

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

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

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

Sketch of the proof: stopping time approach[1]

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

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

- Then decompose

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

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

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

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

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

[1] D. J. Higham, X. Mao, and A. M. Stuart. SIAM J. Numer. Anal., 2002.

Conclusion and perspectives

- We presented optimal mean field estimates for CBO/S.
- These estimates are valid over a finite time horizon.
- **Desirable improvement**: prove uniform-in-time estimates:

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

Thank you for your attention!