# Mean-field limits for Consensus-Based and Ensemble Kalman Sampling

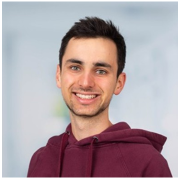
Journées McKean–Vlasov à Jussieu

Urbain Vaes
urbain.vaes@inria.fr

MATHERIALS – Inria Paris & CERMICS – École des Ponts ParisTech

October 2024

# Collaborators and reference



Nicolai Gerber

UNIVERSITÄT BONN

Hausdorff Center for Mathematics

Franca Hoffmann

Caltech

Caltech

**References:**

- N. J. Gerber, F. Hoffmann, and UV. Arxiv preprint, 2023
  Mean-field limits for Consensus-Based Optimization and Sampling
- UV. Arxiv preprint, 2024
  Sharp propagation of chaos for the Ensemble Langevin Sampler

# Outline


Motivation

The classical synchronous coupling approach

Extending the synchronous coupling approach for CBO/S

Extending the synchronous coupling approach for EKS

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

**Global optimization problem**:

$$\text{Find } x \in \arg\min_{x \in \mathbf{R}^d} \mathcal{F} \qquad (\mathcal{F} \colon \mathbf{R}^d \to \mathbf{R})$$

## CBO interacting particle system

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

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

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

[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**:

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

## CBS interacting particle system

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

- $\beta$ is "inverse temperature" parameter.
- $\mu_t^J$ 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:

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

[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

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

## CBS mean field limit

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

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

# Notation: Wasserstein distances[1]

## Wasserstein distance in $\mathbf{R}^d$ (here $|\cdot|$ is always the Euclidean norm)

$$\text{For } \mu,\nu \in \mathcal{P}(\mathbf{R}^d), \qquad W_p(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \left(\mathbf{E}_{(X,Y)\sim\gamma}|X-Y|^p\right)^{\frac{1}{p}}$$

## Wasserstein distance in $\mathbf{R}^{dJ}$

$$\text{For } f^J, g^J \in \mathcal{P}(\mathbf{R}^{dJ}), \qquad W_p(f^J,g^J) = \inf_{\gamma\in\Gamma\left(f^J,g^J\right)} \left(\mathbf{E}_{(\mathbf{X},\mathbf{Y})\sim\gamma}\frac{1}{J}\sum_{j=1}^{J}|X^j-Y^j|^p\right)^{\frac{1}{p}}$$

- With this normalization, $W_p(\mu^{\otimes J},\nu^{\otimes J}) \leqslant W_p(\mu,\nu)$.
- For associated empirical measures, $\mathbf{E}\left[W_p(\mu_f^J,\mu_g^J)^p\right] \leqslant W_p(f^J,g^J)^p$.

Below $f^J, \overline{f}^J \in \mathcal{P}(\mathbf{R}^{dJ})$ are joint laws, while $\mu^J, \overline{\mu}^J \in \mathcal{P}(\mathcal{P}(\mathbf{R}^d))$ are empirical measures:

$$\mu^J = \frac{1}{J}\sum_{j=1}^{J}\delta_{X^j}, \qquad \overline{\mu}^J = \frac{1}{J}\sum_{j=1}^{J}\delta_{\overline{X}^j}.$$

[1] L.-P. Chaintron and A. Diez. Kinet. Relat. Models, 2022.

# 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 $\lambda$ such that

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

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

## Convergence of mean field CBS[3]

If $\pi \propto \mathrm{e}^{-\mathcal{F}}$ 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. SIAM J. Optim., 2024.
[3] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and UV. Stud. Appl. Math., 2022.

# Convergence for the interacting particle systems

Let $f_t^J = \mathrm{Law}\big(X_t^1, \dots, X_t^J\big)$. By the triangle inequality,

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

## 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

$$\mu_t^J \xrightarrow[J \to \infty]{\mathrm{Law}} \overline{\rho}_t, \qquad \text{(no rate)}, \qquad \mu_t^J := \frac{1}{J} \sum_{j=1}^{J} \delta_{X_t^j}.$$

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

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

**Our goal**: obtain an estimate of the form $W_p\Big(f_t^J, \overline{\rho}_t^{\otimes J}\Big) \leqslant C J^{-\frac{1}{2}}$.

[1]H. Huang and J. Qiu. Math. Methods Appl. Sci., 2022.
[2]M. Fornasier, T. Klock, and K. Riedl. SIAM J. Optim., 2024.

# Outline

Motivation

The classical synchronous coupling approach

Extending the synchronous coupling approach for CBO/S

Extending the synchronous coupling approach for EKS

# Introduction of synchronous coupling

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

**Interacting particle system**:

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

**Mean field limit**:

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

## Synchronous coupling

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

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

with same initial condition and driving Browian motions.

# Using the synchronously coupled to prove propagation of chaos

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

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

**Key fact:** mean field processes are i.i.d. with law $\overline{X}_t^j \sim \overline{\rho}_t$, so

$$
W_2\left(f_t^J, \overline{\rho}_t^{\otimes J}\right) = W_2\left(f_t^J, \overline{f}_t^J\right), \qquad \overline{f}_t^J = \mathrm{Law}\left(\overline{X}_t^1, \ldots, \overline{X}_t^J\right).
$$

By definition of Wasserstein distance and exchangeability,

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

## 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.

# Side-remark: from infinite-dimensional chaos to empirical chaos

A triangle inequality gives

$$\begin{aligned}\left(\mathbf{E} W_2\left(\mu_t^J, \overline{\rho}_t\right)^2\right)^{\frac{1}{2}} &\leqslant \left(\mathbf{E} W_2\left(\mu_t^J, \overline{\mu}_t^J\right)^2\right)^{\frac{1}{2}} + \left(\mathbf{E} W_2\left(\overline{\mu}_t^J, \overline{\rho}_t\right)^2\right)^{\frac{1}{2}} \\ &\leqslant \left(\mathbf{E}\, \frac{1}{J} \sum_{j=1}^{J} \left| X_t^j - \overline{X}_t^j \right|^2\right)^{\frac{1}{2}} + \left(\mathbf{E} W_2\left(\overline{\mu}_t^J, \overline{\rho}_t\right)^2\right)^{\frac{1}{2}} \\ &\lesssim J^{-\frac{1}{2}} + J^{-\alpha},\end{aligned}$$

for $\alpha > 0$ depending on dimension[1].

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

# Why the classical Sznitman approach fails for CBO/CBS

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

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

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

**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

Motivation

The classical synchronous coupling approach

Extending the synchronous coupling approach for CBO/S

Extending the synchronous coupling approach for EKS

# Main result: quantitative mean field limits

## Assumption (focusing on the unbounded $\mathcal{F}$ setting for simplicity here)

- **Local Lischitz continuity.** $\mathcal{F}$ is bounded from below by $\mathcal{F}_\star = \inf \mathcal{F}$ and satisfies

$$\forall x, y \in \mathbf{R}^d, \qquad |\mathcal{F}(x) - \mathcal{F}(y)| \leq L_f \bigl(1 + |x| + |y|\bigr)^s |x - y|, \qquad s \geqslant 0.$$

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

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

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

If $\mathcal{F}$ 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.

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

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

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

# 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

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

- First term can be shown to scale as $CJ^{-\frac{p}{2}}$ using classical approach;
- Second term 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.

# Details of the proof: first term (1/2)

- Starting point: the following is an upper bound for $\left|X_t^j - \overline{X}_t^j\right|^p \mathbf{1}_{\{\theta_J > T\}}$.

$$\left|X_{t\wedge\theta_J}^j - \overline{X}_{t\wedge\theta_J}^j\right|^p \leqslant \left|\int_0^{t\wedge\theta_J} b\left(X_s^j, \mu_s^J\right) - b\left(\overline{X}_s^j, \overline{\rho}_s\right)\, \mathrm{d}s\right|^p + \left|\int_0^{t\wedge\theta_J} \sigma\left(X_s^j, \mu_s^J\right) - \sigma\left(\overline{X}_s^j, \overline{\rho}_s\right)\, \mathrm{d}W_s\right|^p.$$

- By Doob's optional stopping and Burkholder–Davis–Gundy,

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|X_{s\wedge\theta_J}^j - \overline{X}_{s\wedge\theta_J}^j\right|^p\right] \leqslant (2T)^{p-1}\mathbf{E}\int_0^{t\wedge\theta_J}\left|b\left(X_s^j, \mu_s^J\right) - b\left(\overline{X}_s^j, \overline{\rho}_s\right)\right|^p \mathrm{d}s + C_{\mathrm{BDG}}2^{p-1}T^{\frac{p}{2}-1}\mathbf{E}\int_0^{t\wedge\theta_J}\left\|\sigma\left(X_s^j, \mu_s^J\right) - \sigma\left(\overline{X}_s^j, \overline{\rho}_s\right)\right\|_{\mathrm{F}}^p \mathrm{d}s =: A_t + B_t.$$

- Both terms handled similarly. For the drift, by the triangle inequality,

$$A_t \lesssim \int_0^t \mathbf{E}\left|b\left(X_{s\wedge\theta_J}^j, \mu_{s\wedge\theta_J}^J\right) - b\left(\overline{X}_{s\wedge\theta_J}^j, \overline{\mu}_{s\wedge\theta_J}^J\right)\right|^p \mathrm{d}s + \int_0^t \mathbf{E}\left|b\left(\overline{X}_s^j, \overline{\mu}_s^J\right) - b\left(\overline{X}_s^j, \overline{\rho}_s\right)\right|^p \mathrm{d}s =: A_t^{(1)} + A_t^{(2)}.$$

# Details of the proof: first term (2/2)

- In order to bound $A_t^{(1)}$, recall that $b(x,\mu) = -x + \mathcal{M}_\beta(\mu)$, so

$$\mathbf{E}\left|b\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right) - b\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right|^p \lesssim \mathbf{E}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^p + \mathbf{E}\left|\mathcal{M}_\beta(\mu^J_{s\wedge\theta_J}) - \mathcal{M}_\beta(\overline{\mu}^J_{s\wedge\theta_J})\right|^p$$

By local $W_p$ Lipschitz continuity of $\mathcal{M}_\beta$ and definition of $\theta_J$,

$$\mathbf{E}\left|\mathcal{M}_\beta(\mu^J_{s\wedge\theta_J}) - \mathcal{M}_\beta(\overline{\mu}^J_{s\wedge\theta_J})\right|^p \lesssim C(R)\ \mathbf{E}\left|W_p\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right|^p \leqslant C(R)\ \mathbf{E}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^p.$$

- In order to bound $A_t^{(2)}$, we use known results[1,2]

$$\mathbf{E}\left|b\left(\overline{X}^j_s,\overline{\mu}^J_s\right) - b\left(\overline{X}^j_s,\overline{\rho}_s\right)\right|^p \propto \mathbf{E}\left|\mathcal{M}_\beta\left(\overline{\mu}^J_s\right) - \mathcal{M}_\beta\left(\overline{\rho}_s\right)\right|^p \lesssim J^{-\frac{p}{2}}.$$

Putting everything together and using Grönwall,

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^p\right] \lesssim J^{-\frac{p}{2}}.$$

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

# Details of the proof: second term

The more difficult part is to bound

$$\begin{aligned}\mathbf{P}[\theta_J(R) \leqslant T] &= \mathbf{P}\left[\sup_{t\in[0,T]} \frac{1}{J}\sum_{j=1}^{J}\left|\overline{X}_t^j\right|^p \geqslant R\right] \\ &\leqslant \mathbf{P}\left[\frac{1}{J}\sum_{j=1}^{J} Z_j \geqslant R\right], \qquad Z_j := \sup_{t\in[0,T]}\left|\overline{X}_t^j\right|^p.\end{aligned}$$

Let $X = \frac{1}{J}\sum_{j=1}^{J} Z_j$. By the Marcinkiewicz–Zygmund inequality, it holds for $r \geqslant 2$ that

$$\mathbf{E}\left|X - \mathbf{E}[X]\right|^r \lesssim J^{-r}\mathbf{E}\left[\left(\sum_{j=1}^{J}\left|Z_j - \mathbf{E}[Z_j]\right|^2\right)^{\frac{r}{2}}\right] \leqslant J^{-\frac{r}{2}}\mathbf{E}\left[\left|Z_1 - \mathbf{E}[Z_1]\right|^r\right],$$

where we used Jensen's inequality and exchangeability. If $R > \mathbf{E}[X]$, then

$$\mathbf{P}\left[X \geq R\right] \leqslant \mathbf{P}\left[\left|X - \mathbf{E}[X]\right|^r \geq \left(R - \mathbf{E}[X]\right)^r\right] \leqslant \mathbf{E}\left[\frac{\left|X - \mathbf{E}[X]\right|^r}{\left(R - \mathbf{E}[X]\right)^r}\right] \leqslant \frac{CJ^{-\frac{r}{2}}}{\left(R - \mathbf{E}[X]\right)^r},$$

where we used Markov's inequality.

# Outline

Motivation

The classical synchronous coupling approach

Extending the synchronous coupling approach for CBO/S

Extending the synchronous coupling approach for EKS

# Extension to Ensemble Kalman Sampler

## Ensemble Kalman Sampler (EKS) to sample from $\pi \propto \mathrm{e}^{-f}$

$$\mathrm{d}X_t^j = -\mathcal{C}(\mu_t^J)\nabla\mathcal{F}(X_t^j)\,\mathrm{d}t + \sqrt{2\mathcal{C}(\mu_t^J)}\,\mathrm{d}W_t^j, \qquad j = 1, \dots, J.$$

**Formal mean field limit:**

$$\mathrm{d}\overline{X}_t = -\mathcal{C}(\overline{\rho}_t)\nabla\mathcal{F}(\overline{X}_t)\,\mathrm{d}t + \sqrt{2\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}\overline{W}_t, \qquad \overline{\rho}_t = \mathrm{Law}(\overline{X}_t).$$

**Additional technical difficulties**:

- Covariance is a quadratic nonlinearity,
- "One-sided" local Lipschitz continuity does not hold.

## Local Lipschitz continuity of $\mathcal{C}\colon \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}^{d\times d}$

$$\forall(\mu,\nu) \in \mathcal{P}_2(\mathbf{R}^d) \times \mathcal{P}_2(\mathbf{R}^d), \qquad \left\|\mathcal{C}(\mu) - \mathcal{C}(\nu)\right\|_{\mathrm{F}} \leqslant 2\Big(W_2(\mu,\delta_0) + W_2(\nu,\delta_0)\Big)W_2(\mu,\nu).$$

# Sharp propagation of chaos for EKS

## Synchronous coupling for EKS

$$
\begin{aligned}
\mathrm{d}X_t^j &= -\mathcal{C}(\mu_t^J)\nabla\mathcal{F}(X_t^j)\,\mathrm{d}t + \sqrt{2\mathcal{C}(\mu_t^J)}\,\mathrm{d}W^{(j)}, \qquad X_0^j = x_0^j, \qquad j = 1,\dots,J, \\
\mathrm{d}\overline{X}_t^j &= -\mathcal{C}(\overline{\rho}_t)\nabla\mathcal{F}(\overline{X}_t^j\,\mathrm{d}t + \sqrt{2\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W^{(j)}, \qquad X_0^j = x_0^j, \qquad j = 1,\dots,J.
\end{aligned}
$$

First almost optimal propagation of chaos result proved by Ding and Li[1,2]:

$$
\forall \varepsilon > 0, \qquad \exists C_\varepsilon > 0, \qquad \mathbf{E}\left[\left|X_T^j - \overline{X}_T^j\right|^2\right] \leqslant C_\varepsilon J^{-1+\varepsilon}.
$$

## Theorem: sharp propagation of chaos[3]

If $\mathcal{F}$ is strongly convex with quadratic growth and $\overline{\rho}_0$ has infinitely many moments, then

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

[1] Z. Ding and Q. Li. Stat. Comput., 2021.
[2] Z. Ding and Q. Li. SIAM J. Math. Anal., 2021.
[3] UV. Arxiv preprint, 2024.

# Mean field limit for ensemble Kalman sampler: idea of the proof

**Key idea:** covariance function $\mathcal{C}\colon \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}^{d\times d}$ is Lipschitz continuous on

$$P_R := \Big\{\nu \in \mathcal{P}(\mathbf{R}^d) : W_2(\nu, \delta_0) \leqslant R\Big\}.$$

- Motivates letting $\theta_J(R) = \tau_J(R) \wedge \overline{\tau}_J(R)$ with

$$\tau_J(R) = \inf\Big\{t \geqslant 0 : W_2(\mu_t^J, \delta_0) \geqslant R\Big\}, \qquad \mu_t^J := \tfrac{1}{J}\textstyle\sum_{j=1}^J \delta_{X^j},$$
$$\overline{\tau}_J(R) = \inf\Big\{t \geqslant 0 : W_2(\overline{\mu}_t^J, \delta_0) \geqslant R\Big\}, \qquad \overline{\mu}_t^J := \tfrac{1}{J}\textstyle\sum_{j=1}^J \delta_{\overline{X}^j}.$$

- Then decompose

$$\mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2\right] = \mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2 \mathbf{1}_{\{\theta_J > T\}}\right] + \mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2 \mathbf{1}_{\{\theta_J \leqslant T\}}\right]$$

- First term can be shown to scale as $C_R J^{-1}$ using classical approach;
- Second term requires to bound

$$\mathbf{P}\left[\theta_J \leqslant T\right] \leqslant \underbrace{\mathbf{P}\left[\overline{\tau}_J \leqslant T\right]}_{\lesssim J^{-a} \quad \forall a>0} + \underbrace{\mathbf{P}\left[\tau_J \leqslant T \leqslant \overline{\tau}_J\right]}_{\lesssim J^{-???}}.$$

# Bounding $\mathbf{P}[\tau_J \leqslant T]$

$$\mathbf{P}\left[\tau_J \leqslant T < \overline{\tau}_J\right] \leqslant \mathbf{P}\left[\sup_{t\in[0,T]} W_2\left(\mu^J_{t\wedge\theta_J}, \delta_0\right) = R\right]$$
$$= \mathbf{P}\left[\sup_{t\in[0,T]} W_2\left(\mu^J_{t\wedge\theta_J}, \overline{\mu}^J_{t\wedge\theta_J}\right) + \sup_{t\in[0,T]} W_2\left(\overline{\mu}^J_{t\wedge\theta_J}, \delta_0\right) \geqslant R\right]$$
$$\leqslant \mathbf{P}\left[\sup_{t\in[0,T]} W_2\left(\mu^J_{t\wedge\theta_J}, \overline{\mu}^J_{t\wedge\theta_J}\right) \geqslant \frac{R}{2}\right] + \mathbf{P}\left[\sup_{t\in[0,T]} W_2\left(\overline{\mu}^J_{t\wedge\theta_J}, \delta_0\right) \geqslant \frac{R}{2}\right],$$

- First term decreases as $J^{-1}$ by Markov and estimate for stopped particle systems
- Second term decreases as $J^{-\alpha}$ for all $\alpha \geqslant 0$, as before.

In the end, this leads to a suboptimal estimate: for all $q > 2$,

$$\mathbf{E}\left[\left|X^j_t - \overline{X}^j_t\right|^2 \mathbf{1}_{\{\theta_J \leqslant T\}}\right] \leqslant \mathbf{E}\left[\left|X^j_t - \overline{X}^j_t\right|^q\right]^{\frac{2}{q}} \mathbf{P}\left[\theta_J \leqslant T\right]^{\frac{q-2}{q}} \lesssim J^{-\frac{q-2}{q}}.$$

# Details: making the estimate optimal

- Take $R$ sufficiently large and let $\theta_{J,r} = \tau_{J,r} \wedge \overline{\tau}_{J,r}$ with $r \geqslant 2$ and

$$\tau_{J,r}(R) = \inf\Big\{t \geqslant 0 : W_r(\mu_t^J, \delta_0) \geqslant R\Big\}, \qquad \mu_t^J := \tfrac{1}{J} \textstyle\sum_{j=1}^J \delta_{X^j},$$
$$\overline{\tau}_{J,r}(R) = \inf\Big\{t \geqslant 0 : W_r(\overline{\mu}_t^J, \delta_0) \geqslant R\Big\}. \qquad \overline{\mu}_t^J := \tfrac{1}{J} \textstyle\sum_{j=1}^J \delta_{\overline{X}^j}.$$

- Prove propagation of chaos in $L^r$ for the stopped particle system:

$$\mathbf{E}\left[\sup_{t\in[0,T]} \left|X^j_{t\wedge\theta_{J,r}} - \overline{X}^j_{t\wedge\theta_{J,r}}\right|^r\right] \leq C J^{-\frac{r}{2}}.$$

- Reasoning as before, prove that $\mathbf{P}\left[\tau_J \leqslant T < \overline{\tau}_J\right] \lesssim J^{-\frac{r}{2}}$.
- Conclude as before

$$\mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2\right] = \mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2 \mathbf{1}_{\{\theta_J > T\}}\right] + \mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2 \mathbf{1}_{\{\theta_J \leqslant T\}}\right]$$
$$\lesssim J^{-1} + \mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^q\right]^{\frac{2}{q}} \left(J^{-\frac{r}{2}}\right)^{\frac{q-2}{q}} \lesssim J^{-1}.$$

# 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!