



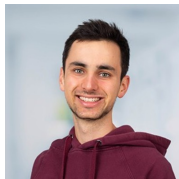
# Mean-field limits for Consensus-Based Optimization and Sampling

CMX seminar

Urbain Vaes  
[urbain.vaes@inria.fr](mailto:urbain.vaes@inria.fr)

MATERIALS – Inria Paris & CERMICS – École des Ponts ParisTech

February 2025



Nicolai Gerber  
universität  
**uulm**

Hausdorff Center for  
Mathematics



Franca Hoffmann

**Caltech**

Caltech



Dohyeon Kim

**Caltech**

Caltech

## References:

- ▶ N. J. Gerber, F. Hoffmann, and UV. [Arxiv preprint, 2023](#)  
Mean-field limits for Consensus-Based Optimization and Sampling
- ▶ N. J. Gerber, F. Hoffmann, D. Kim, and UV. [In preparation, 2025](#)  
Uniform-in-time propagation of chaos for Consensus-Based Optimization

Motivation

The classical synchronous coupling approach

Extending the synchronous coupling approach for CBO/S

Towards uniform-in-time propagation of chaos

## Paradigmatic inverse problem

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;
- ▶  $\eta$  is observational noise.

Two difficulties<sup>1</sup> associated with this problem are the following:

- ▶ Because of the noise, it might be that  $y \notin \text{Ran}(\mathcal{G})$ ;
- ▶ The problem might be underdetermined.

Additionally, in many PDE applications,

- ▶  $\mathcal{G}$  is expensive to evaluate;
- ▶ The derivatives of  $\mathcal{G}$  are difficult to calculate;
- ▶  $\theta$  is a function  $\rightarrow$  infinite dimension.

<sup>1</sup>M. Dashti and A. M. Stuart. In Handbook of uncertainty quantification. Vol. 1, 2, 3. Springer, Cham, 2017.



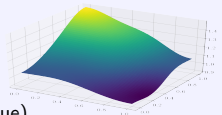
# Example: inference of the thermal conductivity in a plate

**Mathematical model:**

$$\begin{aligned} -\nabla \cdot (\theta(x) \nabla T(x)) &= f(x), & x \in \Omega, \\ T(x) &= 0, & x \in \partial\Omega. \end{aligned}$$

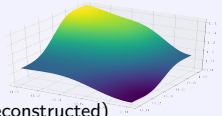
**Unknown parameter:**

Thermal conductivity  $\theta(x)$



(true)

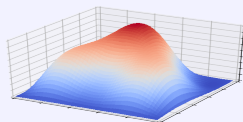
**MAP estimator:**



(reconstructed)

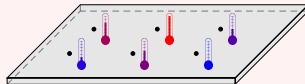
Forward problem

**Solution:**



Temperature field  $T(x)$

**Data:**



Noisy temperature measurements:

$$y = (T(x_1), \dots, T(x_m)) + \eta.$$

Inverse problem

# Probabilistic approach for solving “ $y = \mathcal{G}(\theta) + \eta$ ”<sup>1</sup>

## Bayesian approach to inverse problems

Modeling step:

- Probability distribution on parameter:  $\theta \sim \pi$ , encoding our **prior knowledge**;
- Probability distribution for noise:  $\eta \sim \nu$ .

An application of **Bayes' theorem** gives the **posterior distribution**:

$$\rho^y(\theta) \propto \pi(\theta) \nu(y - \mathcal{G}(\theta)) = \text{prior} \times \text{likelihood}.$$

(In infinite dimension, use Radon–Nikodym derivative.)

In the Gaussian case where  $\pi = \mathcal{N}(m, \Sigma)$  and  $\nu = \mathcal{N}(0, \Gamma)$ ,

$$\rho^y(\theta) \propto \exp \left( - \left( \frac{1}{2} |y - \mathcal{G}(\theta)|_{\Gamma}^2 + \frac{1}{2} |\theta - m|_{\Sigma}^2 \right) \right) =: \exp(-\mathcal{F}(\theta)).$$

where  $|x|_A := \sqrt{x^T A^{-1} x}$ .

**Two approaches for extracting information:**

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

<sup>1</sup>A. M. Stuart. *Acta Numer.*, 2010.

# Brief review of the recent literature on interacting particle methods

- ▶ 1994: Ensemble Kalman filter<sup>1</sup> (6,611 citations);
- ▶ 1995: Particle swarm optimization<sup>2</sup> (**90,668** citations);
- ▶ 2006: Sequential Monte Carlo samplers<sup>3</sup> (2,255 citations);
- ▶ 2010: Affine-invariant many-particle MCMC<sup>4</sup> (3,505 citations);
- ▶ 2013: Ensemble Kalman inversion<sup>5</sup> (473 citations);
- ▶ 2016: Stein variational gradient descent<sup>6</sup> (1,285 citations);
- ▶ 2017: Consensus-based optimization<sup>7</sup> (185 citations);
- ▶ 2020: Ensemble Kalman sampling<sup>8</sup> (233 citations);

Often **parallelizable**, and some can be studied through **mean-field equations**.

<sup>1</sup>G. Evensen. *Journal of Geophysical Research: Oceans*, 1994.

<sup>2</sup>J. Kennedy and R. Eberhart. In *Proceedings of ICNN'95-international conference on neural networks*. iee, 1995.

<sup>3</sup>P. Del Moral, A. Doucet, and A. Jasra. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2006.

<sup>4</sup>J. Goodman and J. Weare. *Commun. Appl. Math. Comput. Sci.*, 2010.

<sup>5</sup>M. A. Iglesias, K. J. H. Law, and A. M. Stuart. *Inverse Problems*, 2013.

<sup>6</sup>Q. Liu and D. Wang. In *Advances In Neural Information Processing Systems*, 2016.

<sup>7</sup>R. Pinnau, C. Totzeck, O. Tse, and S. Martin. *Math. Models Methods Appl. Sci.*, 2017.

<sup>8</sup>A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart. *SIAM J. Appl. Dyn. Syst.*, 2020.

## Global optimization problem:

$$\text{Find } x \in \arg \min_{x \in \mathbf{R}^d} \mathcal{F} \quad (\mathcal{F}: \mathbf{R}^d \rightarrow \mathbf{R})$$

## CBO interacting particle system

$$dX_t^j = -\left(X_t^j - \mathcal{M}_\beta(\mu_t^J)\right) dt + \sqrt{2}\sigma \left|X_t^j - \mathcal{M}_\beta(\mu_t^J)\right| dW_t^j, \quad j = 1, \dots, J,$$

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

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

<sup>1</sup>R. Pinnau, C. Totzeck, O. Tse, and S. Martin. [Math. Models Methods Appl. Sci.](#), 2017.

<sup>2</sup>J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. [Mathematical Models and Methods in Applied Sciences](#), 2018.

# Typical evolution of CBO dynamics

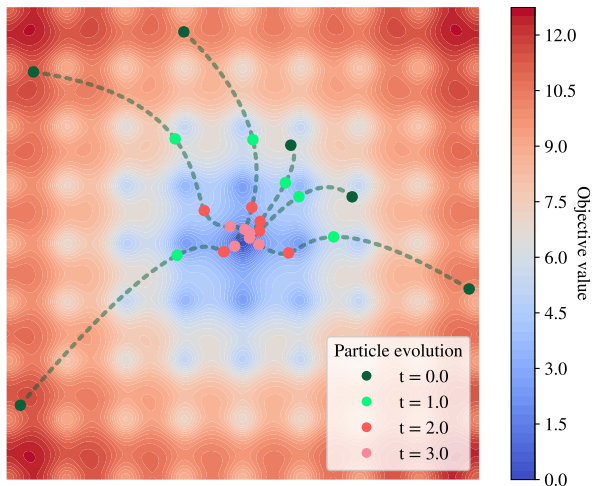
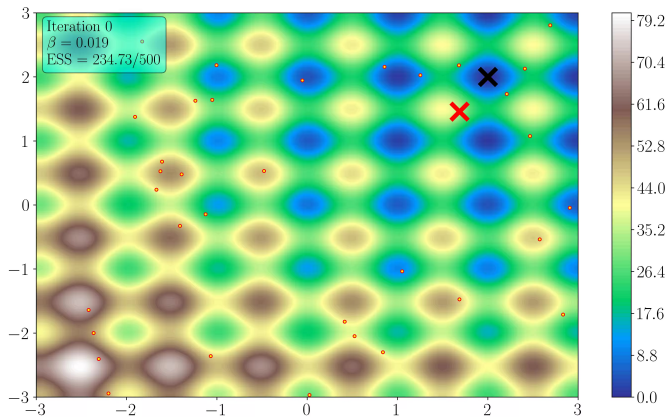


Figure: Ensemble evolution for the Ackley function

# Video illustration





Software library in Python (lead T. Roith) and Julia (lead R. Bailo):

- ▶ Offers high-performance implementation of the method;
- ▶ Implements a number of extensions (different noises, mini-batching, sampling, ...)
- ▶ Provides general interface that can accommodate extensions.



CBX: Python and Julia Packages for Consensus-Based Interacting Particle Methods

Rafael Bailo<sup>1</sup>, Alethea Barbaro<sup>2</sup>, Susana N. Gomes<sup>3</sup>, Konstantin Riedl<sup>4,5</sup>, Tim Roith<sup>6</sup>, Claudia Totzeck<sup>7</sup>, and Urbain Vaes<sup>8,9</sup>

<sup>1</sup> Mathematical Institute, University of Oxford, United Kingdom <sup>2</sup> Delft University of Technology, The Netherlands <sup>3</sup> Mathematics Institute, University of Warwick, United Kingdom <sup>4</sup> Technical University of Munich, Germany <sup>5</sup> Munich Center for Machine Learning, Germany <sup>6</sup> Helmholtz Imaging, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany <sup>7</sup> University of Wuppertal, Germany <sup>8</sup> MATHERIALS team, Inria Paris, France <sup>9</sup> École des Ponts ParisTech, Marne-la-Vallée, France

<sup>1</sup>R. Bailo, A. Barbaro, S. N. Gomes, K. Riedl, T. Roith, C. Totzeck, and U. Vaes. [Journal of Open Source Software](#), 2024.

## Sampling problem:

Generate samples from distribution  $\pi \propto e^{-\mathcal{F}}$  ( $\mathcal{F}: \mathbf{R}^d \rightarrow \mathbf{R}$ )

## CBS interacting particle system

$$dX_t^j = -\left(X_t^j - \mathcal{M}_\beta(\mu_t^J)\right) dt + \sqrt{2(1 + \beta) \mathcal{C}_\beta(\mu_t^J)} dW_t^j, \quad j = 1, \dots, J,$$

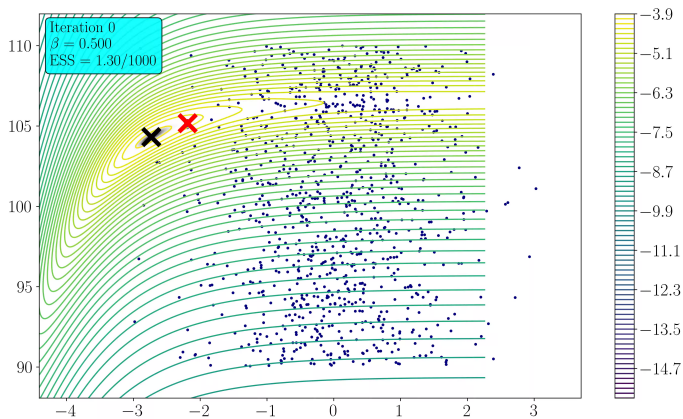
- ▶  $\beta$  is “inverse temperature” parameter.
- ▶  $\mu_t^J$  is empirical measure  $\mu_t^J = \frac{1}{J} \sum_{j=1}^J \delta_{X_t^j}$ .
- ▶  $\mathcal{C}_\beta: \mathcal{P}(\mathbf{R}^d) \rightarrow \mathbf{R}^{d \times d}$  is weighted covariance operator:

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

<sup>1</sup>J. A. Carrillo, F. Hoffmann, A. M. Stuart, and UV. *Stud. Appl. Math.*, 2022.



# Illustration



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

## CBO mean field limit

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

## CBS mean field limit

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

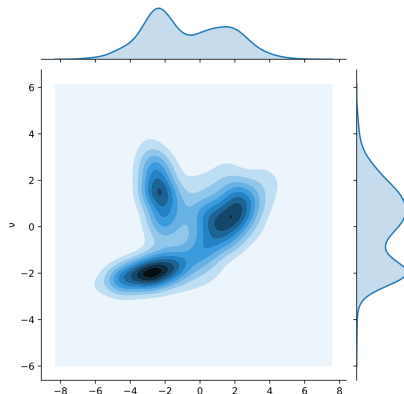
- ▶ Nonlinear Markov processes in  $\mathbf{R}^d$ : future depends on  $\bar{X}_t$  and its distribution;
- ▶ Associated Fokker–Planck equations are nonlinear and nonlocal.

# Notation: Wasserstein distances<sup>1</sup>

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

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

Here  $\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^d) : \text{proj}_{\#}^x \gamma = \mu, \text{proj}_{\#}^y \gamma = \nu\}$ .



# Convergence results in mean field law for CBO and CBS

Recall  $\mathcal{W}_2 : \mathcal{P}_2(\mathbf{R}^d) \times \mathcal{P}_2(\mathbf{R}^d) \rightarrow \mathbf{R}$  denotes the Wasserstein-2 metric.

## Convergence of mean field CBO<sup>1,2</sup>

Under mild conditions including existence of a unique minimizer, there is  $\lambda$  such that

$$\forall t \in [0, T_\beta], \quad \mathcal{W}_2(\bar{\rho}_t, \delta_{x_*}) \leq \mathcal{W}_2(\bar{\rho}_0, \delta_{x_*}) e^{-\lambda t}, \quad x_* = \arg \min_{x \in \mathbf{R}^d} \mathcal{F}.$$

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

## Convergence of mean field CBS<sup>3</sup>

If  $\pi \propto e^{-\mathcal{F}}$  is Gaussian and  $\bar{\rho}_0$  is Gaussian, then

$$\forall t \geq 0, \quad \mathcal{W}_2(\bar{\rho}_t, \pi) \leq C e^{-\left(\frac{\beta}{1+\beta}\right)t}.$$

<sup>1</sup>J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse. [Mathematical Models and Methods in Applied Sciences](#), 2018.

<sup>2</sup>M. Fornasier, T. Klock, and K. Riedl. [SIAM J. Optim.](#), 2024.

<sup>3</sup>J. A. Carrillo, F. Hoffmann, A. M. Stuart, and UV. [Stud. Appl. Math.](#), 2022.

# Key characteristics of consensus based optimization

- ▶ Derivative-free, making them versatile and widely applicable:

```
from cbx.dynamics import CB0
f = lambda x: x[0]**2 + x[1]**2
x = CB0(f, d=2).optimize()
```

- ▶ Can be easily implemented in parallel;
- ▶ (For the sampling variant) Affine invariant: convergence rate independent of target;
- ▶ Theoretical guarantees for the mean field equations.

**Question:** how to obtain convergence guarantees in the finite-size setting?

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

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

- ▶ With this normalization,  $\mathcal{W}_p(\mu^{\otimes J}, \nu^{\otimes J}) \leq \mathcal{W}_p(\mu, \nu)$ .
- ▶ For associated empirical measures,  $\mathbf{E} [\mathcal{W}_p(\mu_f^J, \mu_g^J)^p] \leq \mathcal{W}_p(f^J, g^J)^p$ .

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

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

<sup>1</sup>L.-P. Chaintron and A. Diez. *Kinet. Relat. Models*, 2022.

# Convergence for the interacting particle systems

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

$$\mathcal{W}_2\left(f_t^J, \nu^{\otimes J}\right) \leq \underbrace{\mathcal{W}_2\left(f_t^J, \bar{\rho}_t^{\otimes J}\right)}_{\rightarrow 0 \text{ as } J \rightarrow \infty ???} + \underbrace{\mathcal{W}_2(\bar{\rho}_t, \nu)}_{\leq C e^{-\lambda t}}, \quad \nu = \begin{cases} \delta_{x_*} & \text{for CBO,} \\ e^{-\mathcal{F}} & \text{for CBS.} \end{cases}$$

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

► <sup>1</sup>Based on a compactness argument, it was shown that

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

► <sup>2</sup>For all  $\varepsilon > 0$ , there is  $C_\varepsilon > 0$  such that for all  $J$  there is  $\Omega_\varepsilon \subset \Omega$  satisfying

$$\mathbf{P}[\Omega \setminus \Omega_\varepsilon] \leq \varepsilon \quad \text{and} \quad \mathbf{E}\left[\mathcal{W}_2\left(f_t^J, \bar{\rho}_t^{\otimes J}\right) \mid \Omega_\varepsilon\right] \leq C_\varepsilon J^{-\frac{1}{2}}, \quad C_\varepsilon \xrightarrow[\varepsilon \rightarrow 0]{} \infty.$$

**Our goal:** obtain an estimate of the form  $\sup_{t \geq 0} \mathcal{W}_p\left(f_t^J, \bar{\rho}_t^{\otimes J}\right) \leq C J^{-\frac{1}{2}}$ .

<sup>1</sup>H. Huang and J. Qiu. *Math. Methods Appl. Sci.*, 2022.

<sup>2</sup>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

Towards uniform-in-time propagation of chaos



# Introduction of synchronous coupling

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

**Interacting particle system:**

$$dX_t^j = -\left(X_t^j - \mathcal{M}(\mu_t^J)\right) dt + e^{-t} dW_t^j, \quad X_0^j = x_0^j \stackrel{\text{i.i.d.}}{\sim} \bar{\rho}_0 \quad j = 1, \dots, J.$$

**Mean field limit:**

$$\begin{cases} d\bar{X}_t = -\left(\bar{X}_t - \mathcal{M}(\bar{\rho}_t)\right) dt + e^{-t} d\bar{W}_t, \\ \bar{\rho}_t = \text{Law}(\bar{X}_t). \end{cases}$$

## Synchronous coupling

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

$$\begin{aligned} dX_t^j &= -\left(X_t^j - \mathcal{M}(\mu_t^J)\right) dt + e^{-t} dW_t^j, & X_0^j &= x_0^j, & j &= 1, \dots, J, \\ d\bar{X}_t^j &= -\left(\bar{X}_t^j - \mathcal{M}(\bar{\rho}_t)\right) dt + e^{-t} dW_t^j, & \bar{X}_0^j &= x_0^j, & j &= 1, \dots, J, \end{aligned}$$

with same initial condition and driving Brownian motions.

# Using the synchronously coupled to prove propagation of chaos

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

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

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

$$\mathcal{W}_2\left(f_t^J, \bar{\rho}_t^{\otimes J}\right) = \mathcal{W}_2\left(f_t^J, \bar{f}_t^J\right), \quad \bar{f}_t^J = \text{Law}\left(\bar{X}_t^1, \dots, \bar{X}_t^J\right).$$

By definition of Wasserstein distance and exchangeability,

$$\mathcal{W}_2\left(f_t^J, \bar{f}_t^J\right)^2 \leq \mathbf{E}\left[\frac{1}{J} \sum_{j=1}^J \left|X_t^j - \bar{X}_t^j\right|^2\right] = \mathbf{E}\left[\left|X_t^1 - \bar{X}_t^1\right|^2\right].$$

# Bounding the remaining term (using Sznitman's approach<sup>1</sup>)

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

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

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

$$\forall(\mu, \nu) \in \mathcal{P}_1(\mathbf{R}^d) \times \mathcal{P}_1(\mathbf{R}^d), \quad \left| \mathcal{M}(\mu) - \mathcal{M}(\nu) \right| \leq \mathcal{W}_2(\mu, \nu).$$

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

<sup>1</sup>A.-S. Sznitman. In *École d'Été de Probabilités de Saint-Flour XIX—1989*. Springer, Berlin, 1991.

Sznitman's approach shows that

$$\mathcal{W}_p(f_t^J, \bar{\rho}_t^{\otimes J}) = \mathcal{O}\left(\frac{1}{\sqrt{J}}\right) \quad \text{as } J \rightarrow \infty.$$

**Question:** Can we say anything about the convergence of the **empirical measure**  $\mu_t^J$ ?

$$\begin{aligned} \left(\mathbf{E} \mathcal{W}_p(\mu_t^J, \bar{\rho}_t)^p\right)^{\frac{1}{p}} &\leq \left(\mathbf{E} \mathcal{W}_p(\mu_t^J, \bar{\mu}_t^J)^p\right)^{\frac{1}{p}} + \left(\mathbf{E} \mathcal{W}_p(\bar{\mu}_t^J, \bar{\rho}_t)^p\right)^{\frac{1}{p}} \\ &\leq \left(\mathbf{E} \frac{1}{J} \sum_{j=1}^J \left|X_t^j - \bar{X}_t^j\right|^p\right)^{\frac{1}{p}} + \left(\mathbf{E} \mathcal{W}_2(\bar{\mu}_t^J, \bar{\rho}_t)^p\right)^{\frac{1}{p}} \\ &\lesssim J^{-\frac{1}{2}} + J^{-\alpha}, \end{aligned}$$

for  $\alpha > 0$  depending on dimension<sup>1</sup>.

<sup>1</sup>N. Fournier and A. Guillin. [Probab. Theory Related Fields](#), 2015.

# Why the classical Sznitman approach fails for CBO/CBS

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

$$\begin{aligned}dX_t^j &= -\left(X_t^j - \mathcal{M}_\beta(\mu_t^J)\right) dt + \sqrt{2}\sigma \left|X_t^j - \mathcal{M}_\beta(\mu_t^J)\right| dW_t^j, & X_0^j &= x_0^j. \\d\bar{X}_t^j &= -\left(\bar{X}_t^j - \mathcal{M}_\beta(\bar{\rho}_t)\right) dt + \sqrt{2}\sigma \left|\bar{X}_t^j - \mathcal{M}_\beta(\bar{\rho}_t)\right| dW_t^j, & \bar{X}_0^j &= x_0^j.\end{aligned}$$

### Technical difficulties:

- ▶  $\mathcal{M}_\beta: \mathcal{P}_1(\mathbf{R}^d) \rightarrow \mathbf{R}^d$  is **not globally Lipschitz** continuous in general.
- ▶ Presence of multiplicative noise that depends on  $\mu_t^J$ .
- ▶ Usual Monte Carlo estimates do not enable to bound

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

but estimates are given in the literature<sup>1,2</sup>.

<sup>1</sup>P. Doukhan and G. Lang. [Bernoulli](#), 2009.

<sup>2</sup>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

Towards uniform-in-time propagation of chaos

# 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

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

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

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

Main theorem<sup>1</sup>, holds for both CBO and CBS

If  $\mathcal{F}$  satisfies the above assumption and  $\bar{\rho}_0$  has infinitely many moments, then

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

<sup>1</sup>N. J. Gerber, F. Hoffmann, and UV. [Arxiv preprint](#), 2023.

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

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

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

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

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

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



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

$$\theta_J = \inf \left\{ t \geq 0 : \mathcal{W}_2(\bar{\mu}_t^J, \delta_0) \geq R \right\}, \quad \bar{\mu}_t^J := \frac{1}{J} \sum_{j=1}^J \delta_{\bar{X}_t^j}.$$

- ▶ Then decompose

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

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

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

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

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

<sup>1</sup>D. J. Higham, X. Mao, and A. M. Stuart. [SIAM J. Numer. Anal.](#), 2002.

# Outline

Motivation

The classical synchronous coupling approach

Extending the synchronous coupling approach for CBO/S

Towards uniform-in-time propagation of chaos

## Revisiting the toy example

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

**Interacting particle system:**

$$dX_t^j = -\left(X_t^j - \mathcal{M}(\mu_t^J)\right) dt + e^{-t} dW_t^j, \quad X_0^j = x_0^j \stackrel{\text{i.i.d.}}{\sim} \bar{\rho}_0 \quad j = 1, \dots, J.$$

**Mean field limit:**

$$\begin{cases} d\bar{X}_t = -\left(\bar{X}_t - \mathcal{M}(\bar{\rho}_t)\right) dt + e^{-t} d\bar{W}_t, \\ \bar{\rho}_t = \text{Law}(\bar{X}_t). \end{cases}$$

**Moment decay estimate:** by Itô's formula,

$$\begin{aligned} \frac{d}{dt} \mathbf{E} \left| X_t^j - \mathcal{M}(\mu_t^J) \right|^2 &\leq -2 \mathbf{E} \left| X_t^j - \mathcal{M}(\mu_t^J) \right|^2 + d e^{-2t} \\ &\stackrel{\text{Grönwall}}{\leq} \left( \mathbf{E} \left| X_0^j - \mathcal{M}(\mu_0^J) \right|^2 + \frac{d}{2} \right) e^{-2t}. \end{aligned}$$

Similarly

$$\frac{d}{dt} \mathbf{E} \left| \bar{X}_t - \mathbf{E} \bar{X}_t \right|^2 \leq \left( \mathbf{E} \left| \bar{X}_0 - \mathbf{E} \bar{X}_0 \right|^2 + \frac{d}{2} \right) e^{-2t}.$$

## Synchronous coupling for toy example

$$\begin{aligned} dX_t^j &= -\left(X_t^j - \mathcal{M}(\mu_t^J)\right) dt + e^{-t} dW_t^j, & X_0^j &= x_0^j, & j &= 1, \dots, J, \\ d\bar{X}_t^j &= -\left(\bar{X}_t^j - \mathcal{M}(\bar{\rho}_t)\right) dt + e^{-t} dW_t^j, & \bar{X}_0^j &= x_0^j, & j &= 1, \dots, J. \end{aligned}$$

$$\begin{aligned} \frac{d}{dt} \frac{1}{2J} \sum_{j=1}^J \left| X_t^j - \bar{X}_t^j \right|^2 &= -\frac{1}{J} \sum_{j=1}^J \left\langle X_t^j - \bar{X}_t^j, X_t^j - \mathcal{M}(\mu_t^J) - \bar{X}_t^j + \mathcal{M}(\bar{\mu}_t^J) \right\rangle \\ &\quad + \frac{1}{J} \sum_{j=1}^J \left\langle X_t^j - \bar{X}_t^j, \mathcal{M}(\bar{\mu}_t^J) - \mathcal{M}(\bar{\rho}_t) \right\rangle. \end{aligned}$$

The first term is **nonpositive**. By the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \mathbf{E} \left| X_t^1 - \bar{X}_t^1 \right|^2 &\leq \sqrt{\mathbf{E} \left| X_t^1 - \bar{X}_t^1 \right|^2} \sqrt{\mathbf{E} \left| \mathcal{M}(\bar{\mu}_t^J) - \mathcal{M}(\bar{\rho}_t) \right|^2}. \\ \rightsquigarrow \frac{d}{dt} \sqrt{\mathbf{E} \left| X_t^1 - \bar{X}_t^1 \right|^2} &\leq \sqrt{\mathbf{E} \left| \mathcal{M}(\bar{\mu}_t^J) - \mathcal{M}(\bar{\rho}_t) \right|^2} \lesssim \frac{e^{-t}}{\sqrt{J}}. \end{aligned}$$

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

$$\begin{aligned} dX_t^j &= -\left(X_t^j - \mathcal{M}_\beta(\mu_t^J)\right) dt + \sqrt{2}\sigma \left|X_t^j - \mathcal{M}_\beta(\mu_t^J)\right| dW_t^j, & X_0^j &= x_0^j, \\ d\bar{X}_t^j &= -\left(\bar{X}_t^j - \mathcal{M}_\beta(\bar{\rho}_t)\right) dt + \sqrt{2}\sigma \left|\bar{X}_t^j - \mathcal{M}_\beta(\bar{\rho}_t)\right| dW_t^j, & \bar{X}_0^j &= x_0^j. \end{aligned}$$

$$\begin{aligned} \frac{d}{dt} \frac{1}{2J} \sum_{j=1}^J \left|X_t^j - \bar{X}_t^j\right|^2 &= -\frac{1}{J} \sum_{j=1}^J \left\langle X_t^j - \bar{X}_t^j, X_t^j - \mathcal{M}(\mu_t^J) - \bar{X}_t^j + \mathcal{M}(\bar{\mu}_t^J) \right\rangle \\ &\quad - \frac{1}{J} \sum_{j=1}^J \left\langle X_t^j - \bar{X}_t^j, \underbrace{\mathcal{M}(\mu_t^J) - \mathcal{M}_\beta(\mu_t^J) - \mathcal{M}(\bar{\mu}_t^J) + \mathcal{M}_\beta(\bar{\mu}_t^J)}_{\text{Small ???}} \right\rangle \\ &\quad + \frac{1}{J} \sum_{j=1}^J \left\langle X_t^j - \bar{X}_t^j, \underbrace{\mathcal{M}_\beta(\bar{\mu}_t^J) - \mathcal{M}_\beta(\bar{\rho}_t)}_{\text{Small when } J \gg 1} \right\rangle + \sigma \dots \end{aligned}$$

**Assumption:** for simplicity here, from now on we assume

- ▶ no noise ( $\sigma = 0$ );
- ▶  $f$  bounded and globally Lipschitz;
- ▶ initialization in a compact set:  $x_0^j \sim \bar{\rho}_0$  with  $\bar{\rho}_0$  compactly supported.

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

$$\mathfrak{M}_p(\mu) := \int |x - \mathcal{M}(\mu)|^p \mu(dx).$$

## Key ingredient: Stability of weighted mean

Then there exists  $C_{\mathcal{M}} > 0$  such that for all  $\mu, \nu \in \mathcal{P}_2(\mathbf{R}^d)$ ,

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

## Key ingredient: Moment estimate for particle system

For all  $p > 0$ , there exists  $\lambda_p > 0$  such that

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

$$\rightsquigarrow \quad \left| \mathcal{M}(\mu_t^J) - \mathcal{M}_{\beta}(\mu_t^J) - \mathcal{M}(\bar{\mu}_t^J) + \mathcal{M}_{\beta}(\bar{\mu}_t^J) \right| \lesssim e^{-\frac{\lambda_2}{2} t} \mathcal{W}_2\left(\mu_t^J, \bar{\mu}_t^J\right)$$

## Main theorem

Suppose that

- ▶ function  $f$  is bounded and globally Lipschitz;
- ▶ probability distribution  $\bar{\rho}_0$  has finite moments of all orders;
- ▶ noise coefficient  $\sigma$  is sufficiently small.

Then there exists  $C_{\text{MFL}}(\alpha, \underline{f}, \bar{f}, L_f, \sigma, d)$  such that

$$\sup_{t \geq 0} \mathbf{E} \left[ |X_t^1 - \bar{X}_t^1|^2 \right] \leq \frac{C_{\text{MFL}}}{J}.$$

**Future directions:**

- ▶ Uniform-in-time estimate for Consensus-Based Sampling;
- ▶ Discrete-time estimates;

Thank you for your attention!

## Main theorem

Suppose that

- ▶ function  $f$  is bounded and globally Lipschitz;
- ▶ probability distribution  $\bar{\rho}_0$  has finite moments of all orders;
- ▶ noise coefficient  $\sigma$  is sufficiently small.

Then there exists  $C_{\text{MFL}}(\alpha, \underline{f}, \bar{f}, L_f, \sigma, d)$  such that

$$\sup_{t \geq 0} \mathbf{E} \left[ |X_t^1 - \bar{X}_t^1|^2 \right] \leq \frac{C_{\text{MFL}}}{J}.$$

**Future directions:**

- ▶ Uniform-in-time estimate for Consensus-Based Sampling;
- ▶ Discrete-time estimates;

**Thank you for your attention!**



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

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

$$\begin{aligned} \left|X_{t \wedge \theta_J}^j - \bar{X}_{t \wedge \theta_J}^j\right|^p &\leq \left|\int_0^{t \wedge \theta_J} b\left(X_s^j, \mu_s^J\right) - b\left(\bar{X}_s^j, \bar{\rho}_s\right) ds\right|^p \\ &\quad + \left|\int_0^{t \wedge \theta_J} \sigma\left(X_s^j, \mu_s^J\right) - \sigma\left(\bar{X}_s^j, \bar{\rho}_s\right) dW_s\right|^p. \end{aligned}$$

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

$$\begin{aligned} \mathbf{E} \left[ \sup_{s \in [0, t]} \left|X_{s \wedge \theta_J}^j - \bar{X}_{s \wedge \theta_J}^j\right|^p \right] &\leq (2T)^{p-1} \mathbf{E} \int_0^{t \wedge \theta_J} \left|b\left(X_s^j, \mu_s^J\right) - b\left(\bar{X}_s^j, \bar{\rho}_s\right)\right|^p ds \\ &\quad + C_{\text{BDG}} 2^{p-1} T^{\frac{p}{2}-1} \mathbf{E} \int_0^{t \wedge \theta_J} \left\| \sigma\left(X_s^j, \mu_s^J\right) - \sigma\left(\bar{X}_s^j, \bar{\rho}_s\right) \right\|_{\mathbf{F}}^p ds =: A_t + B_t. \end{aligned}$$

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

$$\begin{aligned} A_t &\lesssim \int_0^t \mathbf{E} \left| b\left(X_{s \wedge \theta_J}^j, \mu_{s \wedge \theta_J}^J\right) - b\left(\bar{X}_{s \wedge \theta_J}^j, \bar{\mu}_{s \wedge \theta_J}^J\right) \right|^p ds \\ &\quad + \int_0^t \mathbf{E} \left| b\left(\bar{X}_s^j, \bar{\mu}_s^J\right) - b\left(\bar{X}_s^j, \bar{\rho}_s\right) \right|^p ds =: A_t^{(1)} + A_t^{(2)}. \end{aligned}$$

## 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

$$\begin{aligned} \mathbf{E} \left| b \left( X_{s \wedge \theta_J}^j, \mu_{s \wedge \theta_J}^J \right) - b \left( \bar{X}_{s \wedge \theta_J}^j, \bar{\mu}_{s \wedge \theta_J}^J \right) \right|^p &\lesssim \mathbf{E} \left| X_{s \wedge \theta_J}^j - \bar{X}_{s \wedge \theta_J}^j \right|^p \\ &\quad + \mathbf{E} \left| \mathcal{M}_\beta(\mu_{s \wedge \theta_J}^J) - \mathcal{M}_\beta(\bar{\mu}_{s \wedge \theta_J}^J) \right|^p \end{aligned}$$

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

$$\begin{aligned} \mathbf{E} \left| \mathcal{M}_\beta(\mu_{s \wedge \theta_J}^J) - \mathcal{M}_\beta(\bar{\mu}_{s \wedge \theta_J}^J) \right|^p &\lesssim C(R) \mathbf{E} \left| \mathcal{W}_p \left( \mu_{s \wedge \theta_J}^J, \bar{\mu}_{s \wedge \theta_J}^J \right) \right|^p \\ &\leq C(R) \mathbf{E} \left| X_{s \wedge \theta_J}^j - \bar{X}_{s \wedge \theta_J}^j \right|^p. \end{aligned}$$

- In order to bound  $A_t^{(2)}$ , we use known results<sup>1,2</sup>

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

Putting everything together and using Grönwall,

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

<sup>1</sup>P. Doukhan and G. Lang. [Bernoulli](#), 2009.

<sup>2</sup>S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. [Statist. Sci.](#), 2017.

The more difficult part is to bound

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

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

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

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

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

where we used Markov's inequality.