Computationally tractable Bayesian inference for interacting particle system models

KIT, 4 March 2026

Aurélien Castre, joint work with Richard Nickl

University of Cambridge (UK)

A weakly interacting particle system

- For $W \in \dot{H}^{\alpha+1}(\mathbb{T}^d)$, $\alpha > 1 + d/2$, consider the system of SDEs

$$\begin{cases} \mathrm{d}X_t^{n,i} = -\frac{1}{n}\sum_{j\neq i} \nabla W(X_t^{n,i} - X_t^{n,j})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^i \\ X_0^{n,i} \sim \phi \end{cases}, \quad i = 1, \dots, n.$$

  with $\phi \in H^\beta(\mathbb{T}^d)$ some positive density.

A weakly interacting particle system

- For $W \in \dot{H}^{\alpha+1}(\mathbb{T}^d)$, $\alpha > 1 + d/2$, consider the system of SDEs

$$\begin{cases} \mathrm{d}X_t^{n,i} = -\frac{1}{n}\sum_{j\neq i} \nabla W(X_t^{n,i} - X_t^{n,j})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^i \\ X_0^{n,i} \sim \phi \end{cases}, \quad i = 1, \dots, n.$$

  with $\phi \in H^\beta(\mathbb{T}^d)$ some positive density.
- With $P_t^{n,k} := \mathrm{Law}(X_t^{n,1}, \dots, X_t^{n,k})$, Lacker (Probab. Math. Phys., 2023) shows 'propagation of chaos', i.e. for any $t > 0$,

$$\|P_t^{n,k} - P_t^{\otimes k}\|_{\mathrm{TV}} = O\left(\frac{k}{n}\right) \text{ as } n \to \infty,$$

A weakly interacting particle system

- For $W \in \dot{H}^{\alpha+1}(\mathbb{T}^d)$, $\alpha > 1 + d/2$, consider the system of SDEs

$$\begin{cases} \mathrm{d}X_t^{n,i} = -\frac{1}{n}\sum_{j\neq i} \nabla W(X_t^{n,i} - X_t^{n,j})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^i \\ X_0^{n,i} \sim \phi \end{cases}, \quad i = 1, \dots, n.$$

  with $\phi \in H^\beta(\mathbb{T}^d)$ some positive density.
- With $P_t^{n,k} := \mathrm{Law}(X_t^{n,1}, \dots, X_t^{n,k})$, Lacker (Probab. Math. Phys., 2023) shows 'propagation of chaos', i.e. for any $t > 0$,

$$\|P_t^{n,k} - P_t^{\otimes k}\|_{\mathrm{TV}} = O\left(\frac{k}{n}\right) \text{ as } n \to \infty,$$

  where P_t has density $\rho_t = \rho(t,x)$ solving the **McKean-Vlasov PDE**

$$\begin{cases} \partial_t \rho = \Delta\rho + \nabla \cdot (\rho \nabla W * \rho) \\ \rho(0) = \phi. \end{cases}$$

A weakly interacting particle system

- For $W \in \dot{H}^{\alpha+1}(\mathbb{T}^d)$, $\alpha > 1 + d/2$, consider the system of SDEs

$$\begin{cases} \mathrm{d}X_t^{n,i} = -\frac{1}{n}\sum_{j\neq i} \nabla W(X_t^{n,i} - X_t^{n,j})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^i \\ X_0^{n,i} \sim \phi \end{cases}, \quad i = 1, \ldots, n.$$

  with $\phi \in H^\beta(\mathbb{T}^d)$ some positive density.
- With $P_t^{n,k} := \mathrm{Law}(X_t^{n,1}, \ldots, X_t^{n,k})$, Lacker (Probab. Math. Phys., 2023) shows 'propagation of chaos', i.e. for any $t > 0$,

$$\|P_t^{n,k} - P_t^{\otimes k}\|_{\mathrm{TV}} = O\left(\frac{k}{n}\right) \text{ as } n \to \infty,$$

  where P_t has density $\rho_t = \rho(t,x)$ solving the **McKean-Vlasov PDE**

$$\begin{cases} \partial_t \rho = \Delta\rho + \nabla \cdot (\rho \nabla W * \rho) \\ \rho(0) = \phi. \end{cases}$$

- **Goal:** infer W from measurements of $\rho = \rho_W$ (nonlinear inverse problem).

Data

Given some 'ground truth' $W_0 \in \dot{H}^{\alpha+1}$, observe regression data $Z_N = (Y_i, t_i, x_i)_{i=1}^N$ of the form

$$Y_i = \rho_{W_0}(t_i, x_i) + \varepsilon_i, \quad i = 1, \dots, N,$$

where $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and (t_i, x_i) are i.i.d. uniform on $[0, T] \times \mathbb{T}^d$.

The Bayesian approach

Follow approach popularised by Stuart (Acta Numerica, 2010) for inverse problems.

The Bayesian approach

Follow approach popularised by Stuart (Acta Numerica, 2010) for inverse problems.

- Put a (rescaled) Gaussian prior $\Pi = \Pi_N$ on W of regularity $\alpha + 1 > d/2 + 2$ such that $\int_{\mathbb{T}^d} W = 0$,

The Bayesian approach

Follow approach popularised by Stuart (Acta Numerica, 2010) for inverse problems.

- Put a (rescaled) Gaussian prior $\Pi = \Pi_N$ on W of regularity $\alpha + 1 > d/2 + 2$ such that $\int_{\mathbb{T}^d} W = 0$,
- This induces the posterior distribution

$$\mathrm{d}\Pi(W \mid Z_N) \propto e^{\ell_N(W)}\, \mathrm{d}\Pi(W),$$

with $\ell_N(W) = -\frac{1}{2} \sum_{i=1}^{N} |Y_i - \rho_W(t_i, x_i)|^2$

The Bayesian approach

Follow approach popularised by Stuart (Acta Numerica, 2010) for inverse problems.

- Put a (rescaled) Gaussian prior $\Pi = \Pi_N$ on W of regularity $\alpha + 1 > d/2 + 2$ such that $\int_{\mathbb{T}^d} W = 0$,
- This induces the posterior distribution

$$\mathrm{d}\Pi(W \mid Z_N) \propto e^{\ell_N(W)} \, \mathrm{d}\Pi(W),$$

  with $\ell_N(W) = -\frac{1}{2} \sum_{i=1}^{N} |Y_i - \rho_W(t_i, x_i)|^2$
- Extract information about W from the posterior distribution, e.g. by computing the posterior mean $\bar{W}_N := \int W \, \mathrm{d}\Pi(W \mid Z_N)$ or credible sets. This last step typically relies on **MCMC algorithms**.

Bayesian computations

Bayesian computations

- Discretise the prior by projecting it onto $\dot{E}_K$, the space of mean-zero trigonometric polynomials of degree at most K, so that W is represented by the vector of dimension $D \simeq K^d \gg 1$ of its Fourier coefficients.

Bayesian computations

- Discretise the prior by projecting it onto $\dot{E}_K$, the space of mean-zero trigonometric polynomials of degree at most K, so that W is represented by the vector of dimension $D \simeq K^d \gg 1$ of its Fourier coefficients. W is now **band-limited**: $\hat{W}_k = 0$ for all $|k| > K$.

Bayesian computations

- Discretise the prior by projecting it onto $\dot{E}_K$, the space of mean-zero trigonometric polynomials of degree at most K, so that W is represented by the vector of dimension $D \simeq K^d \gg 1$ of its Fourier coefficients. W is now **band-limited**: $\hat{W}_k = 0$ for all $|k| > K$.
- Sample from the discretised posterior distribution. In high dimensions ($D \gg 1$), we typically want to exploit **gradient** information.

Bayesian computations

- Discretise the prior by projecting it onto $\dot{E}_K$, the space of mean-zero trigonometric polynomials of degree at most K, so that W is represented by the vector of dimension $D \simeq K^d \gg 1$ of its Fourier coefficients. W is now **band-limited**: $\hat{W}_k = 0$ for all $|k| > K$.
- Sample from the discretised posterior distribution. In high dimensions ($D \gg 1$), we typically want to exploit **gradient** information.

Unadjusted Langevin Algorithm (ULA)

Choose a step size $\gamma > 0$ and an initialiser W_0. For $\xi_k \overset{\text{iid}}{\sim} \mathcal{N}(0, I_D)$, do

$$W_{k+1} = W_k + \gamma \nabla \log(\mathrm{d}\Pi(W_k \mid Z_N)) + \sqrt{2\gamma}\xi_{k+1},$$

where the gradient is taken with respect to W.

Bayesian computations

- Discretise the prior by projecting it onto $\dot{E}_K$, the space of mean-zero trigonometric polynomials of degree at most K, so that W is represented by the vector of dimension $D \simeq K^d \gg 1$ of its Fourier coefficients. W is now **band-limited**: $\hat{W}_k = 0$ for all $|k| > K$.
- Sample from the discretised posterior distribution. In high dimensions ($D \gg 1$), we typically want to exploit **gradient** information.

Unadjusted Langevin Algorithm (ULA)

Choose a step size $\gamma > 0$ and an initialiser W_0. For $\xi_k \stackrel{\text{iid}}{\sim} \mathcal{N}(0, I_D)$, do

$$W_{k+1} = W_k + \gamma \nabla \log(\mathrm{d}\Pi(W_k \mid Z_N)) + \sqrt{2\gamma}\xi_{k+1},$$

where the gradient is taken with respect to W.

- Compute desired quantities, e.g. $\hat{W}_N = M^{-1} \sum_{k=1}^{M} W_k$

Challenges

Challenges

- Computational hurdle: the map $\mathcal{G} : W \mapsto \rho_W$ is non-linear, meaning the posterior could be **non-log-concave** or even **multi-modal**. Because it is also **high-dimensional** ($D \gg 1$), sampling from this Gibbs measures is highly non-trivial.

Challenges

- Computational hurdle: the map $\mathcal{G} : W \mapsto \rho_W$ is non-linear, meaning the posterior could be **non-log-concave** or even **multi-modal**. Because it is also **high-dimensional** ($D \gg 1$), sampling from this Gibbs measures is highly non-trivial.

- Statistical hurdle: even if we can compute the posterior mean $\bar{W}_N$, we have no initial guarantees that $\bar{W}_N$ will land close to the true W_0 (**consistency**).

Injectivity and stability

Injectivity and stability

- It is not clear that $W \mapsto \rho_W$ is even injective: $\phi = 1$ implies $\rho_W = 1$ for all W!

Injectivity and stability

- It is not clear that $W \mapsto \rho_W$ is even injective: $\phi = 1$ implies $\rho_W = 1$ for all W!
- Assume that the initial condition $\phi \in H^\beta$ satisfies $\inf_x \phi(x) > \phi_{\min} > 0$ and is **ζ-deconvolvable**, i.e. its Fourier coefficients $\hat{\phi}_k$ do not decay too fast:

$$|\hat{\phi}_k| \geq c|k|^{-\zeta}, \quad \text{for some } \zeta > \beta + d/2$$

Injectivity and stability

- It is not clear that $W \mapsto \rho_W$ is even injective: $\phi = 1$ implies $\rho_W = 1$ for all W!
- Assume that the initial condition $\phi \in H^\beta$ satisfies $\inf_x \phi(x) > \phi_{\min} > 0$ and is **ζ-deconvolvable**, i.e. its Fourier coefficients $\hat{\phi}_k$ do not decay too fast:

$$|\hat{\phi}_k| \geq c|k|^{-\zeta}, \quad \text{for some } \zeta > \beta + d/2$$

  The deconvolvability condition propagates to $\rho_W(t)$ for t small enough by time continuity.

Injectivity and stability

- It is not clear that $W \mapsto \rho_W$ is even injective: $\phi = 1$ implies $\rho_W = 1$ for all W!
- Assume that the initial condition $\phi \in H^\beta$ satisfies $\inf_x \phi(x) > \phi_{\min} > 0$ and is **ζ-deconvolvable**, i.e. its Fourier coefficients $\hat{\phi}_k$ do not decay too fast:

$$|\hat{\phi}_k| \geq c|k|^{-\zeta}, \quad \text{for some } \zeta > \beta + d/2$$

  The deconvolvability condition propagates to $\rho_W(t)$ for t small enough by time continuity.

Stability estimate

For any $W \in \dot{E}_K$ and $W_0 \in H^{\alpha+1}$, and for ϕ as above

$$\|W - P_{E_K} W_0\|_{L^2} \lesssim K^{3\zeta/2} \left(\|\rho_W - \rho_{W_0}\|_{L^2([0,T];L^2)} + \|\rho_W - \rho_{W_0}\|_{H^1([0,T];H^{-2})} \right)$$

Consistency

Using this stability estimate, we can use the framework of Nickl (EMS Press, 2023) for Gaussian priors to show a posterior contraction rate.

Theorem (Nickl, Pavliotis, Ray, AoS 2025)

For any $W_0 \in H^{\alpha+1}$, under the previous assumptions on ϕ, for some $\theta > 0$

$$\Pi(W \in E_{K_N} : \|W - W_0\|_{L^2} \lesssim N^{-\theta} \mid Z_N) \to 1 \text{ in } P^N_{W_0}\text{-probability}$$

In particular,

$$\|\bar{W}_N - W_0\|_{L^2} = O_{P^N_{W_0}}(N^{-\theta})$$

Consistency

Using this stability estimate, we can use the framework of Nickl (EMS Press, 2023) for Gaussian priors to show a posterior contraction rate.

Theorem (Nickl, Pavliotis, Ray, AoS 2025)

For any $W_0 \in H^{\alpha+1}$, under the previous assumptions on ϕ, for some $\theta > 0$

$$\Pi(W \in E_{K_N} : \|W - W_0\|_{L^2} \lesssim N^{-\theta} \mid Z_N) \to 1 \text{ in } P^N_{W_0}\text{-probability}$$

In particular,

$$\|\bar{W}_N - W_0\|_{L^2} = O_{P^N_{W_0}}(N^{-\theta})$$

Note: this solves the consistency problem! We now need to compute $\bar{W}_N$.

Geometry of the likelihood function

Geometry of the likelihood function

- Recall $\mathrm{d}\Pi(W \mid Z_N) \propto e^{\ell_N(W)}\,\mathrm{d}\Pi(W), \ell_N(W) = -\frac{1}{2}\sum_{i=1}^{N} |Y_i - \rho_W(t_i, x_i)|^2$. We would like to show $-\ell(W) := -\ell_1(W)$ is convex. Assuming for the moment that $W \mapsto \rho_W$ is twice differentiable,

$$-\nabla^2 \ell(W) = \nabla\rho_W(t,x)^\top \nabla\rho_W(t,x) + (\rho_W(t,x) - Y)\nabla^2\rho_W(t,x).$$

Geometry of the likelihood function

- Recall $\mathrm{d}\Pi(W \mid Z_N) \propto e^{\ell_N(W)}\,\mathrm{d}\Pi(W), \ell_N(W) = -\frac{1}{2}\sum_{i=1}^N |Y_i - \rho_W(t_i, x_i)|^2$. We would like to show $-\ell(W) := -\ell_1(W)$ is convex. Assuming for the moment that $W \mapsto \rho_W$ is twice differentiable,

$$-\nabla^2 \ell(W) = \nabla \rho_W(t,x)^\top \nabla \rho_W(t,x) + (\rho_W(t,x) - Y)\nabla^2 \rho_W(t,x).$$

- So the 'average' Hessian satisfies, for any h with $\|h\|_{\mathbb{R}^D} \leq 1$ (for some appropriate norm $\|\cdot\|_*$)

$$\boxed{h^\top E_{W_0}[-\nabla^2 \ell(W)]h = \|h^\top \nabla \rho_W\|^2_{L^2_\lambda([0,T]\times\mathbb{T}^d)} + O(\|\rho_W - \rho_{W_0}\|_*).}$$

Geometry of the likelihood function

- Recall $\mathrm{d}\Pi(W \mid Z_N) \propto e^{\ell_N(W)}\,\mathrm{d}\Pi(W), \ell_N(W) = -\frac{1}{2}\sum_{i=1}^N |Y_i - \rho_W(t_i, x_i)|^2$. We would like to show $-\ell(W) := -\ell_1(W)$ is convex. Assuming for the moment that $W \mapsto \rho_W$ is twice differentiable,

$$-\nabla^2 \ell(W) = \nabla \rho_W(t,x)^\top \nabla \rho_W(t,x) + (\rho_W(t,x) - Y)\nabla^2 \rho_W(t,x).$$

- So the 'average' Hessian satisfies, for any h with $\|h\|_{\mathbb{R}^D} \leq 1$ (for some appropriate norm $\|\cdot\|_*$)

$$\boxed{h^\top E_{W_0}[-\nabla^2 \ell(W)]h = \|h^\top \nabla \rho_W\|^2_{L^2_\lambda([0,T]\times\mathbb{T}^d)} + O(\|\rho_W - \rho_{W_0}\|_*).}$$

- Nickl and Wang (JEMS, 2023) introduce **gradient stability**: $\|h^\top \nabla \rho_W\|^2_{L^2_T L^2} \gtrsim \|h\|^2_{\mathbb{R}^d}$, but no formula is available for $\nabla \rho_W$ since ρ_W is the solution to a non-linear PDE!

Linearisation of $W \mapsto \rho_W$

Linearisation of $W \mapsto \rho_W$

- Viewing $W \mapsto \rho_W$ as a mapping between Banach spaces, it suffices to compute its Fréchet derivative, since then $D\rho_W[h] = h^\top \nabla \rho_W$.

Linearisation of $W \mapsto \rho_W$

- Viewing $W \mapsto \rho_W$ as a mapping between Banach spaces, it suffices to compute its Fréchet derivative, since then $D\rho_W[h] = h^\top \nabla \rho_W$.
- Note that the pairs (W, ρ_W) are exactly the zeros of the explicitly defined map

$$f(W, \rho) = (\partial_t \rho - \Delta \rho - \nabla \cdot (\rho \nabla W * \rho), \rho(0) - \phi)$$

Linearisation of $W \mapsto \rho_W$

- Viewing $W \mapsto \rho_W$ as a mapping between Banach spaces, it suffices to compute its Fréchet derivative, since then $D\rho_W[h] = h^\top \nabla \rho_W$.
- Note that the pairs (W, ρ_W) are exactly the zeros of the explicitly defined map

$$f(W, \rho) = (\partial_t \rho - \Delta \rho - \nabla \cdot (\rho \nabla W * \rho), \rho(0) - \phi)$$

- We can formally differentiate $f(W, \rho_W) = 0$ with respect to W to get a good candidate for $D\rho_W$:

$$\boxed{D\rho_W[H] \stackrel{?}{=} -(D_2 f(W, \rho_W))^{-1} D_1 f(W, \rho_W)[H]}$$

Linearisation of $W \mapsto \rho_W$

- Viewing $W \mapsto \rho_W$ as a mapping between Banach spaces, it suffices to compute its Fréchet derivative, since then $D\rho_W[h] = h^\top \nabla \rho_W$.
- Note that the pairs (W, ρ_W) are exactly the zeros of the explicitly defined map

$$f(W, \rho) = (\partial_t \rho - \Delta\rho - \nabla \cdot (\rho \nabla W * \rho), \rho(0) - \phi)$$

- We can formally differentiate $f(W, \rho_W) = 0$ with respect to W to get a good candidate for $D\rho_W$:

$$\boxed{D\rho_W[H] \stackrel{?}{=} -(D_2 f(W, \rho_W))^{-1} D_1 f(W, \rho_W)[H]}$$

- Note that $D_2 f(W, \rho_W)$ is a **linear** parabolic partial differential operator, so it is automatically invertible by linear PDE theory and the right-hand side is well-defined.

Linearisation of $W \mapsto \rho_W$

- Viewing $W \mapsto \rho_W$ as a mapping between Banach spaces, it suffices to compute its Fréchet derivative, since then $D\rho_W[h] = h^\top \nabla \rho_W$.
- Note that the pairs (W, ρ_W) are exactly the zeros of the explicitly defined map

$$f(W, \rho) = (\partial_t \rho - \Delta \rho - \nabla \cdot (\rho \nabla W * \rho), \rho(0) - \phi)$$

- We can formally differentiate $f(W, \rho_W) = 0$ with respect to W to get a good candidate for $D\rho_W$:

$$\boxed{D\rho_W[H] \stackrel{?}{=} -(D_2 f(W, \rho_W))^{-1} D_1 f(W, \rho_W)[H]}$$

- Note that $D_2 f(W, \rho_W)$ is a **linear** parabolic partial differential operator, so it is automatically invertible by linear PDE theory and the right-hand side is well-defined.
- This formal derivation is made rigorous using the implicit function theorem in Banach spaces.

Linearisation of $W \mapsto \rho_W$

- For given initial condition $\phi \in H^\beta$, regard $W \mapsto \rho_W$ as a map
$$\dot{W}^{2,\infty} \to L^2([0,T];H^{\beta+1}) \cap H^1([0,T];H^{\beta-1})$$

Linearisation of $W \mapsto \rho_W$

- For given initial condition $\phi \in H^\beta$, regard $W \mapsto \rho_W$ as a map

$$\dot{W}^{2,\infty} \to L^2([0,T]; H^{\beta+1}) \cap H^1([0,T]; H^{\beta-1})$$

- This map is well-defined due to existence, uniqueness and regularity theory for McKean-Vlasov PDEs.

Linearisation of $W \mapsto \rho_W$

- For given initial condition $\phi \in H^\beta$, regard $W \mapsto \rho_W$ as a map
$$\dot{W}^{2,\infty} \to L^2([0,T]; H^{\beta+1}) \cap H^1([0,T]; H^{\beta-1})$$
- This map is well-defined due to existence, uniqueness and regularity theory for McKean-Vlasov PDEs.

Theorem (C. and Nickl, 2026)

For any $W, H \in \dot{W}^{2,\infty}$, the Fréchet derivative $D\rho_W[H]$ is the unique $\mathbf{u} \in L^2([0,T]; H^{\beta+1}) \cap H^1([0,T]; H^{\beta-1})$ solving the linear parabolic PDE

$$\begin{cases} \partial_t \mathbf{u} - \Delta \mathbf{u} - \nabla \cdot (\mathbf{u} \nabla W * \rho_W) - \nabla \cdot (\rho_W \nabla W * \mathbf{u}) = \nabla \cdot (\rho_W \nabla H * \rho_W) \\ \mathbf{u}(0) = 0 \end{cases}$$

Moreover, the map $W \mapsto \rho_W$ is C^∞ in the Fréchet sense and higher order derivatives can be computed similarly.

The posterior is locally log-concave

Using that $D\rho_W[H]$ solves a PDE, we can now prove that gradient stability holds. For any $W, H \in \dot{E}_K$, provided ϕ is ζ-deconvolvable,

$$\|H^\top \nabla \rho_W\|^2_{L^2([0,T];L^2)} = \|D\rho_W[H]\|^2_{L^2([0,T];L^2)} \gtrsim K^{-6\zeta} \|H\|^2_{L^2}.$$

The posterior is locally log-concave

Using that $D\rho_W[H]$ solves a PDE, we can now prove that gradient stability holds. For any $W, H \in \dot{E}_K$, provided ϕ is ζ-deconvolvable,

$$\|H^\top \nabla \rho_W\|^2_{L^2([0,T];L^2)} = \|D\rho_W[H]\|^2_{L^2([0,T];L^2)} \gtrsim K^{-6\zeta} \|H\|^2_{L^2}.$$

Theorem (C. and Nickl, 2026)

For any $W_0 \in H^{\alpha+1}$ and any ζ-deconvolvable initial condition ϕ,

$$\inf_{W \in \mathcal{B}} \lambda_{\min}(E_{W_0}[-\nabla^2 \ell(W)]) \gtrsim K^{-6\zeta} > 0$$

on a neighbourhood $\mathcal{B}$ of $P_{E_K} W_0$ in $\dot{E}_K$ of radius D^{-w} for appropriate $w > 0$.

The posterior is locally log-concave

Using that $D\rho_W[H]$ solves a PDE, we can now prove that gradient stability holds. For any $W, H \in \dot{E}_K$, provided ϕ is ζ-deconvolvable,

$$\|H^\top \nabla \rho_W\|^2_{L^2([0,T];L^2)} = \|D\rho_W[H]\|^2_{L^2([0,T];L^2)} \gtrsim K^{-6\zeta}\|H\|^2_{L^2}.$$

Theorem (C. and Nickl, 2026)

For any $W_0 \in H^{\alpha+1}$ and any ζ-deconvolvable initial condition ϕ,

$$\inf_{W \in \mathcal{B}} \lambda_{\min}(E_{W_0}[-\nabla^2 \ell(W)]) \gtrsim K^{-6\zeta} > 0$$

on a neighbourhood $\mathcal{B}$ of $P_{E_K} W_0$ in $\dot{E}_K$ of radius D^{-w} for appropriate $w > 0$.

Combining the above with the consistency result, we can construct a **strongly globally log-concave** proxy posterior $\tilde{\Pi}(\,\cdot \mid Z_N)$ such that

$$\mathcal{W}_2^2(\tilde{\Pi}(\,\cdot \mid Z_N), \Pi(\,\cdot \mid Z_N)) = O_{P^N_{W_0}}(e^{-N^p}), \quad \text{for some } p > 0$$

Polynomial-time mixing of ULA

Polynomial-time mixing of ULA

- Even for unimodal, locally log-concave posteriors, ‘cold-start’ ULA can take an exponential number of steps to mix in high-dimensions (Bandeira, Maillard, Nickl, Wang, Phil. Trans. R. Soc. A 2023).

Polynomial-time mixing of ULA

- Even for unimodal, locally log-concave posteriors, ‘cold-start’ ULA can take an exponential number of steps to mix in high-dimensions (Bandeira, Maillard, Nickl, Wang, Phil. Trans. R. Soc. A 2023).
- We study ‘warm-start’ ULA and assume we are given a feasible initialiser W_{init} into $\mathcal{B}$. Ideas for initialisation can be found in Wang (2026).

Polynomial-time mixing of ULA

- Even for unimodal, locally log-concave posteriors, ‘cold-start’ ULA can take an exponential number of steps to mix in high-dimensions (Bandeira, Maillard, Nickl, Wang, Phil. Trans. R. Soc. A 2023).
- We study ‘warm-start’ ULA and assume we are given a feasible initialiser W_{init} into $\mathcal{B}$. Ideas for initialisation can be found in Wang (2026).
- We start from W_{init} and run ULA on the proxy $\tilde{\Pi}(\,\cdot \mid Z_N)$.

Polynomial-time mixing of ULA

- Even for unimodal, locally log-concave posteriors, 'cold-start' ULA can take an exponential number of steps to mix in high-dimensions (Bandeira, Maillard, Nickl, Wang, Phil. Trans. R. Soc. A 2023).
- We study 'warm-start' ULA and assume we are given a feasible initialiser W_{init} into $\mathcal{B}$. Ideas for initialisation can be found in Wang (2026).
- We start from W_{init} and run ULA on the proxy $\tilde{\Pi}(\cdot \mid Z_N)$.

Theorem (C. and Nickl, 2026)

Under regularity conditions and ζ-deconvolvability of ϕ, for any precision level $\varepsilon \geq N^{-P}, P > 0$, the warm-start ULA requires iterations growing as

$$O(N^{b_1} D^{b_2} \varepsilon^{-b_3}), b_1, b_2, b_3 > 0$$

to produce an output $\hat{W}_\varepsilon$ that satisfies, with high $P^N_{W_0}$-probability

$$\|\hat{W}_\varepsilon - \bar{W}_N\|_{L^2} \leq \varepsilon, \text{ as well as } \|\hat{W}_\varepsilon - W_0\|_{L^2} \leq \varepsilon$$

Thank you!

References

A. Castre, R. Nickl, On gradient stability in nonlinear PDE models and inference in interacting particle systems, *arXiv 2026*

S. Wang, Global polynomial-time estimation in statistical nonlinear inverse problems via generalized stability, *arXiv 2026*

R. Nickl, G. A. Pavliotis, K. Ray, Bayesian non-parametric inference in McKean-Vlasov models, *Annals of Statistics, 2025*

R. Nickl, S. Wang, On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms, *J. Eur. Math. Soc., 2024*

R. Nickl, *Bayesian non-linear statistical inverse problems*, Zürich Lectures in Advanced Mathematics, EMS press, 2023

Ideas for initialisation

Ideas for initialisation

Notice that the operator $\mathcal{L}_W\rho \coloneqq \partial_t\rho - \Delta\rho - \nabla\cdot(\rho\nabla W * \rho)$ is **linear** in W.

Ideas for initialisation

Notice that the operator $\mathcal{L}_W\rho \coloneqq \partial_t\rho - \Delta\rho - \nabla\cdot(\rho\nabla W * \rho)$ is **linear** in W. Following the plug-in M-estimator idea of Wang (2026), we can construct an initialiser W_{init} by solving two **quadratic** optimisation problems:

Ideas for initialisation

Notice that the operator $\mathcal{L}_W \rho := \partial_t \rho - \Delta \rho - \nabla \cdot (\rho \nabla W * \rho)$ is **linear** in W. Following the plug-in M-estimator idea of Wang (2026), we can construct an initialiser W_{init} by solving two **quadratic** optimisation problems:

- A regression step, e.g.

$$\hat{\rho} \in \arg\min_\rho \sum_{i=1}^{N} |Y_i - \rho(t_i, x_i)|^2 + \mathrm{pen}(\rho)$$

Ideas for initialisation

Notice that the operator $\mathcal{L}_W \rho := \partial_t \rho - \Delta\rho - \nabla \cdot (\rho \nabla W * \rho)$ is **linear** in W. Following the plug-in M-estimator idea of Wang (2026), we can construct an initialiser W_{init} by solving two **quadratic** optimisation problems:

- A regression step, e.g.

$$\hat{\rho} \in \arg\min_\rho \sum_{i=1}^{N} |Y_i - \rho(t_i, x_i)|^2 + \mathrm{pen}(\rho)$$

- A deterministic step, reminiscent of the **PINN** methodology,

$$\hat{W} \in \arg\min_W \|\mathcal{L}_W \hat{\rho}\|^2_{L^2([0,T];L^2)} + \mathrm{pen}(W)$$