# Bayesian inference: Interacting particle approaches

Sebastian Reich

University of Potsdam & SFB 1294 Data Assimilation

June 9, 2022

# Content

1. Computational Bayesian inference
   - Coupling of measures
   - Invariance and ergodicity
2. Interacting particle systems for sampling
   - Overdamped Langevin dynamics
   - Gradient log density estimator
3. Kalman–Wasserstein gradient flow structure
   - Gradient flow structures
   - Numerical implementation

# Computational Bayesian inference

Publications:

SR, **A dynamical systems perspective for intermittent data assimilation**, BIT, 51, 235–359, 2011

SR, **Data assimilation: The Schrödinger perspective**, Acta Numerica, 635–711, 2019

Daniel Huang, Jiaoyang Huang, SR & Andrew Stuart, **Efficient derivative-free Bayesian inference for large scale inverse problems**, arXiv:2204.04386

Edoardo Calvello, SR & Andrew Stuart, **Ensemble Kalman methods: A mean field perspective**, in preparation

**Prior (forecast)**:

$$\Theta_\mathrm{f} \sim \pi_\mathrm{f}$$

**negative log-likelihood**:

**nonl. regression:** $l(y|\theta) = \frac{1}{2}(g(\theta) - y)^\mathrm{T} R^{-1}(g(\theta) - y)$

**logistic regression:** $l(y|\theta) = -y \log \sigma(\theta^\mathrm{T}\phi_x) - (1-y)\log(1 - \sigma(\theta^\mathrm{T}\phi_x))$

$g$ forward map, $R$ error covariance matrix, $y$ the data ($y \in \mathbb{R}/y \in \{0,1\}$), $\phi_x$ feature map, $x \in \mathbb{R}^J$, $\sigma(t) = 1/(1+\exp(-t))$.

**Bayesian posterior (analysis)**:

$$\pi_\mathrm{a}(\theta|y) \propto e^{-l(y|\theta)}\,\pi_\mathrm{f}(\theta).$$

# Sampling the posterior

**Monte Carlo**: Compute **realisations** $\theta_\mathrm{a}^{(i)}$, $i = 1, \ldots, M$, from a random variable (RV)

$$\Theta_\mathrm{a} \sim \pi_\mathrm{a}$$

to approximate posterior expectation values

$$\mathbb{E}_\mathrm{a}[f] \approx \frac{1}{M} \sum_{i=1}^{M} f(\theta_\mathrm{a}^{(i)}) .$$

**Catch**: The random variable $\Theta_\mathrm{a}$ is **not fully specified** by Bayes' theorem:

- approaches based on **coupling of measures**,
- and those based on **invariance and ergodicity**.

# Coupling approaches I

Find a pair of random variables

$$(\Theta_{\rm f}, \Theta_{\rm a}) \sim \pi_{\rm fa}(\theta_{\rm f}, \theta_{\rm a}) = \pi_{\rm a}(\theta_{\rm a}|\theta_f)\,\pi_{\rm f}(\theta_f)$$

such that

$$\Theta_{\rm f} \sim \pi_{\rm f}, \qquad \Theta_{\rm a} \sim \pi_{\rm a}\,. \tag{1}$$

**Catch**: Joint distribution $\pi_{\rm fa}(\theta_{\rm f}, \theta_{\rm a})$ is not uniquely determined by its marginals (1):

- **Examples**: sequential Monte-Carlo, ensemble Kalman filter
- **optimal transportation** (minimise expected distance between $\Theta_{\rm f}$ and $\Theta_{\rm a}$; transport equation)
- **Schrödinger bridges** (minimise the Kullback–Leibler divergence to some reference measure; stochastic optimal control)

# Invariance and ergodicity I

Do the opposite: **successively decouple**.

Define a sequence of random variables (**stochastic process**) $\Theta_\tau$, $\tau \geq 0$, with $\Theta_0 = \Theta_\mathrm{f}$ and

$$\Theta_\mathrm{a} := \lim_{\tau \to \infty} \Theta_\tau \sim \pi_\mathrm{a} \, .$$

**Catch**: Such stochastic processes typically satisfy:

- $\Theta_0 \sim \pi_\mathrm{a}$ implies $\Theta_\tau \sim \pi_\mathrm{a}$ for all $\tau > 0$ (**invariance**) and
- $\Theta_\mathrm{a}$ is **independent** of $\Theta_0 = \Theta_\mathrm{f}$ (**ergodicity**), that is,

$$\pi_\mathrm{fa}(\theta_\mathrm{f}, \theta_\mathrm{a}) = \pi_\mathrm{f}(\theta_\mathrm{f})\, \pi_\mathrm{a}(\theta_\mathrm{a}|y) \, .$$

SFB 1294 Data Assimilation

**Examples**:

- **Langevin dynamics**

$$\mathrm{d}\Theta_\tau = -\nabla_\theta V(\Theta_\tau)\,\mathrm{d}\tau + \sqrt{2}\,\mathrm{d}W_\tau\,. \tag{2}$$

  with

$$V(\theta) = -\log \pi_{\mathrm{a}}(\theta|y)\,.$$

  The SDE (2) is ergodic with unique invariant measure $\pi_{\mathrm{a}}$ under appropriate conditions.

- In discrete time, we got **Markov chain Monte Carlo** (MCMC) methods.

**Discrete-time Langevin**:

$$\text{stochastic process } \{\Theta_n\}_{n\geq 0}, \qquad \lim_{n\to\infty} \Theta_n \sim \pi_{\mathrm{a}}.$$

**Idea**.[1] Let $\Theta_n \sim \tilde{\pi}_n$; for any $\gamma > 0$:

$$\begin{aligned} \text{diffusion:} \quad & \tilde{\pi}_{n+1/2} \propto \tilde{\pi}_n^{1/(1+\gamma)}, \\ \text{Bayes/drift:} \quad & \tilde{\pi}_{n+1} \propto \pi_{\mathrm{a}}^{\gamma/(1+\gamma)} \tilde{\pi}_{n+1/2} \end{aligned}$$

**Theorem**. If $\tilde{\pi}_n = \pi_{\mathrm{a}}$, then $\tilde{\pi}_{n+1} = \pi_{\mathrm{a}}$. Convergence is exponential for all $\gamma > 0$.

[1] Huang et al, arXiv:2204.04386

**Diffusion step**:

$$\Theta_n \sim \tilde{\pi}_n = \mathrm{N}(\mu_n, \Sigma_n)$$

implies

$$\tilde{\pi}_{n+1/2} = \mathrm{N}(\mu_n, (1+\gamma)\Sigma_n).$$

**Update step**:

deterministic: $\Theta_{n+1/2} = \mu_n + (1+\gamma)^{1/2}(\Theta_n - \mu_n)$

stochastic: $\Theta_{n+1/2} = \Theta_n + \gamma^{1/2}\Sigma_n^{1/2}\Xi_n, \quad \Xi_n \sim \mathrm{N}(0, I).$

**Bayes/drift step**:

i) Extended observations

$$\tilde{y}_{\mathrm{obs}} = \begin{pmatrix} y_{\mathrm{obs}} \\ \mu_0 \end{pmatrix}, \quad \tilde{G} = \begin{pmatrix} G \\ I \end{pmatrix} \quad \tilde{R} = \begin{pmatrix} R & 0 \\ 0 & \Sigma_0 \end{pmatrix}.$$

ii) negative log "likelihood" function

$$\tilde{l}(\theta|\tilde{y}_{\mathrm{obs}}) := -\log \pi_{\mathrm{a}}(\theta) = \frac{1}{2}(\tilde{G}\theta - \tilde{y}_{\mathrm{obs}})^{\mathrm{T}}\tilde{R}^{-1}(\tilde{G}\theta - \tilde{y}_{\mathrm{obs}}).$$

iii) Kalman filter step with $\tau = \gamma/(1+\gamma)$, likelihood $\tilde{l}(\theta|\tilde{y}_{\mathrm{obs}})$ and prior $\mathrm{N}(\mu_{n+1/2}, \Sigma_{n+1/2})$.

**Continuous-time limit**:[2] $(\gamma \to 0)$

$$\dot{\Theta}_\tau = -\frac{1}{2}\Sigma_{\Theta_\tau}\left\{G^{\mathrm{T}}R^{-1}(G\Theta_\tau + G\mu_{\Theta_\tau} - 2y_{\mathrm{obs}}) + \Sigma_0^{-1}(\Theta_\tau + \mu_{\Theta_\tau} - 2\mu_0)\right\} + \Sigma_{\Theta_\tau}^{1/2}\dot{W}_\tau.$$

Alternatively:

$$\dot{\Theta}_\tau = -\frac{1}{2}\Sigma_{\Theta_\tau}\left\{G^{\mathrm{T}}R^{-1}(G\Theta_\tau + G\mu_{\Theta_\tau} - 2y_{\mathrm{obs}}) + \Sigma_0^{-1}(\Theta_\tau + \mu_{\Theta_\tau} - 2\mu_0)\right\} + 2(\Theta_\tau - \mu_{\Theta_\tau}).$$

[2] Pidstrigach & SR, FoCM, 2022, Huang et al, arXiv:2204.04386

# Interacting particle systems for sampling

Publications:

Sahani Pathiraja & SR, **Discrete gradients for computational Bayesian inference**, J. Comput. Dyn., 6, 236–251, 2019.

Dimitra Maoutsa, SR & Manfred Opper, **Interacting particle solutions of Fokker–Planck equations through gradient-log-density estimation**, Entropy, 22, 0802, 2020

SFB 1294 Data Assimilation

**Nonlinear SDE**:

$$\mathrm{d}\Theta_\tau = f(\Theta_\tau)\mathrm{d}\tau + \sqrt{2\sigma}\mathrm{d}W_\tau, \qquad \Theta_0 \sim \pi_0,$$

$W_\tau$ standard Brownian motion and e.g. $f(\theta) = \nabla_\theta \log \pi_{\mathrm{a}}(\theta|y)$.

**Fokker–Planck equation**: $\Theta_\tau \sim \pi_\tau$

$$\begin{aligned}\partial_\tau \pi_\tau &= -\nabla \cdot (\pi_\tau f) + \sigma \Delta \pi_\tau, \\ &= -\nabla \cdot (\pi_\tau \{f - \sigma \nabla \log \pi_\tau\})\end{aligned}$$

**Nonlinear SDE**:

$$\mathrm{d}\Theta_\tau = f(\Theta_\tau)\mathrm{d}\tau + \sqrt{2\sigma}\mathrm{d}W_\tau, \qquad \Theta_0 \sim \pi_0,$$

$W_\tau$ standard Brownian motion and e.g. $f(\theta) = \nabla_\theta \log \pi_{\mathrm{a}}(\theta|y)$.

**Fokker–Planck equation**: $\Theta_\tau \sim \pi_\tau$

$$\begin{aligned}\partial_\tau \pi_\tau &= -\nabla \cdot (\pi_\tau f) + \sigma \Delta \pi_\tau, \\ &= -\nabla \cdot (\pi_\tau \{f - \sigma \nabla \log \pi_\tau\})\end{aligned}$$

**Mean-field ODE**

$$\dot{\Theta}_\tau = f(\Theta_\tau) - \sigma \nabla \log \pi_\tau.$$

**Gaussian case**:

$$\Theta_\tau \sim \mathrm{N}(\mu_\tau, \Sigma_\tau) \quad \Longrightarrow \quad -\nabla \log \pi_\tau(\theta) = \Sigma_\tau^{-1}(\theta - \mu_\tau).$$

[3] Carrillo et al, Calc. Var. Part. Diff. Eqs, 2019

**Gaussian case**:

$$\Theta_\tau \sim \mathrm{N}(\mu_\tau, \Sigma_\tau) \quad \Longrightarrow \quad -\nabla \log \pi_\tau(\theta) = \Sigma_\tau^{-1}(\theta - \mu_\tau).$$

**Interacting particle dynamics**: $\Theta_0^{(i)} \sim \pi_0$, $i = 1, \ldots, M$,

$$\dot{\Theta}_\tau^{(i)} = f(\Theta_\tau^{(i)}) - \sigma \nabla \log \tilde{\pi}_\tau(\Theta_\tau^{(i)})$$

with approximative density $\tilde{\pi}_\tau$:

- Gaussian

$$\tilde{\pi}_\tau(\theta) = \mathrm{n}(\theta; \mu_\tau^M, \Sigma_\tau^M)$$

- Gaussian mixture[3]

$$\tilde{\pi}_\tau(\theta) = \frac{1}{M} \sum_{i=1}^{M} \mathrm{n}(\theta; \Theta_\tau^{(i)}, \gamma I).$$

[3] Carrillo et al, Calc. Var. Part. Diff. Eqs, 2019

**Variational formulation**:[4]

$$\partial_\alpha \log \pi := r^{(\alpha)} + \arg\min_\phi \mathcal{L}_\alpha[\phi, \pi]$$

$\partial_\alpha = \partial_{\theta^{(\alpha)}}$, $r^{(\alpha)}$ the $\alpha$th component of an appropriate reference function $r : \mathbb{R}^{N_\theta} \to \mathbb{R}^{N_\theta}$,

$$\begin{aligned}\mathcal{L}_\alpha[\phi, \pi] &:= \int \pi(\theta) \left(\phi^2(\theta) + 2r^{(\alpha)}(\theta)\,\phi(\theta) + 2\partial_\alpha \phi(\theta)\right) \mathrm{d}\theta \\ &= \int \pi(\theta) \left(\phi(\theta) + r^{(\alpha)}(\theta) - \partial_\alpha \log \pi(\theta)\right)^2 + \\ &\qquad \text{terms independent of } \phi\end{aligned}$$

[4] A. Hyvärinen, J. Mach. Learn. Res., 2005

**Estimator:**

$$\mathcal{L}_\alpha[\phi, \pi_\tau] \approx \mathcal{L}_\alpha[\phi, \pi_\tau^M]$$

$$:= \frac{1}{M} \sum_{i=1}^{M} \left( \phi^2(\Theta_\tau^{(i)}) + 2r^{(\alpha)}(\Theta_\tau^{(i)})\, \phi(\Theta_\tau^{(i)}) + 2\partial_\alpha \phi(\Theta_\tau^{(i)}) \right)$$

and

$$\partial_\alpha \log \pi_\tau(\theta) \approx r^{(\alpha)}(\theta) + \arg\min_{\phi \in \mathcal{F}} \mathcal{L}_\alpha[\phi, \pi_\tau^M]\,(\theta).$$

SFB 1294 Data Assimilation

**Estimator:**

$$\mathcal{L}_\alpha[\phi,\pi_\tau] \approx \mathcal{L}_\alpha[\phi,\pi_\tau^M] := \frac{1}{M}\sum_{i=1}^{M}\left(\phi^2(\Theta_\tau^{(i)}) + 2r^{(\alpha)}(\Theta_\tau^{(i)})\,\phi(\Theta_\tau^{(i)}) + 2\partial_\alpha\phi(\Theta_\tau^{(i)})\right)$$

and

$$\partial_\alpha \log \pi_\tau(\theta) \approx r^{(\alpha)}(\theta) + \arg\min_{\phi\in\mathcal{F}} \mathcal{L}_\alpha[\phi,\pi_\tau^M]\,(\theta).$$

**Interacting particle ODE**: $i = 1,\ldots,M$,

$$\dot{\Theta}_\tau^{(i)} = f(\Theta_\tau^{(i)}) - \sigma\left(r^{(\alpha)}(\Theta_\tau^{(i)}) + \phi_\tau^{(\alpha)}(\Theta_\tau^{(i)})\right).$$

with

$$\phi_\tau^{(\alpha)} := \arg\min_{\phi\in\mathcal{F}} \mathcal{L}_\alpha[\phi,\pi_\tau^M]$$

## Remarks

- Approximation space $\mathcal{F}$: (i) $L$-dimensional (random feature) space

$$\phi_\tau(\theta) = \sum_{l=1}^{L} \alpha_\tau^{(l)} \phi_l(\theta)$$

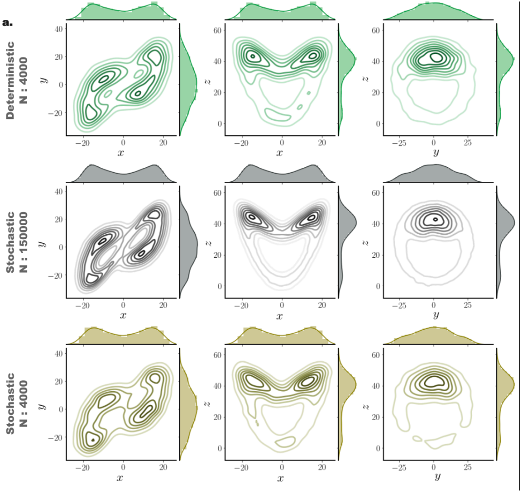
(ii) RKHS with kernel $k(\theta, \theta')$

$$\phi_\tau(\theta) = \sum_{i=1}^{M} \alpha_\tau^{(i)} k(\theta, \Theta_\tau^{(i)})$$

- (ii) with $\gamma = 1$, $f(\theta) = \nabla_\theta \log \pi_{\mathrm{a}}(\theta|y)$, $r = f$, leads equations related to **Stein variational gradient descent**.[5]

[5] Q. Liu & D. Wang, NEURIPS, 2016

# Example: Chaotic attractor Lorenz 63



# Kalman–Wasserstein gradient flow structure

Publications:

SR & Colin Cotter, **Ensemble filter techniques for intermittent data assimilation**, in Radon Series on Computational and Applied Mathematics, Volume 13, 91-134, 2013.

Alfredo Garbuno-Inigo, Nikolas Nüsken & SR, **Affine invariant interacting Langevin dynamics for Bayesian inference**, SIADS, 19, 1633–1658, 2020.

SR & Simon Weissmann, **Fokker–Planck particle systems for Bayesian inference: Computational approaches**, SIAM/ASA JUQ, 9, 446–482, 2021.

Jakiw Pidstrigach & SR, **Affine-invariant ensemble transform methods for logistic regression**, FoCM, 2022.

**Overdamped Langevin dynamics**

$$\mathrm{d}\Theta_\tau = -\nabla_\theta V(\Theta_\tau)\,\mathrm{d}\tau + \sqrt{2}\,\mathrm{d}W_\tau\,.$$

is **not** invariant under **affine transformations**

$$\theta' = A\theta + b;$$

that is

$$\mathrm{d}\Theta'_\tau = -AA^{\mathrm{T}}\nabla_{\theta'} V(\Theta'_\tau)\,\mathrm{d}\tau + \sqrt{2}A\,\mathrm{d}W_\tau\,.$$

## Affine invariant sampling methods[6]

[6]Weare & Goodman, Comm. Appl. Math. Comput. Sci., 2010; Matthews et al, Stats. Comput., 2017

[7]SR, BIT, 2011

[8]Garbuno-Inigo et al, SIADS, 2020a

[9]Garbuno-Inigo et al, SIADS, 2020b

## Affine invariant sampling methods[6]

Inspired by **ensemble Kalman–Bucy filter**:[7]

$$\mathrm{d}\Theta_\tau = -\Sigma_\tau \nabla_\theta G(\Theta_\tau) R^{-1} \left( G(\Theta_\tau)\mathrm{d}\tau + R^{1/2}\mathrm{d}W_\tau - y\mathrm{d}\tau \right)$$

[6] Weare & Goodman, Comm. Appl. Math. Comput. Sci., 2010; Matthews et al, Stats. Comput., 2017

[7] SR, BIT, 2011

[8] Garbuno-Inigo et al, SIADS, 2020a

[9] Garbuno-Inigo et al, SIADS, 2020b

SFB 1294 Data Assimilation

## Affine invariant sampling methods[6]

Inspired by **ensemble Kalman–Bucy filter**:[7]

$$\mathrm{d}\Theta_\tau = -\Sigma_\tau \nabla_\theta G(\Theta_\tau) R^{-1} \left( G(\Theta_\tau)\mathrm{d}\tau + R^{1/2}\mathrm{d}W_\tau - y\mathrm{d}\tau \right)$$

**Ensemble Kalman sampler** (EKS)[8] / **affine invariant Langevin dynamics** (ALDI)[9]

$$\mathrm{d}\Theta_\tau = -\Sigma_\tau \nabla_\theta V(\Theta_\tau)\,\mathrm{d}\tau + \sqrt{2}\Sigma_\tau^{1/2}\,\mathrm{d}W_\tau.$$

[6] Weare & Goodman, Comm. Appl. Math. Comput. Sci., 2010; Matthews et al, Stats. Comput., 2017

[7] SR, BIT, 2011

[8] Garbuno-Inigo et al, SIADS, 2020a

[9] Garbuno-Inigo et al, SIADS, 2020b

**Nonlinear** (affine invariant) **Fokker–Planck equation**[10]

$$\begin{aligned}\partial_\tau \pi_\tau &= -\nabla_\theta \cdot \left(\pi_\tau \Sigma_\tau \left\{\nabla_\theta \log \pi_\tau - \nabla_\theta \log \pi_{\mathrm{a}}\right\}\right) \\ &= -\mathrm{grad}_\pi^{\mathrm{AI}}\, \mathrm{KL}\left(\pi_\tau || \pi_{\mathrm{a}}\right)\end{aligned}$$

$\Sigma_t$ the covariance matrix of $\Theta_\tau \sim \pi_\tau$.

Metric $g_\pi(\rho_1, \rho_2)$ on space of densities induced by **Mahalanobis distance** on $\mathbb{R}^D$:

$$\|a\|^2_{\Sigma_\tau^{-1}} := a^{\mathrm{T}} \Sigma_\tau a$$

[10] Otto, Comm. Part. Diff. Eqs., 2001, SR & Cotter, CUP, 2015; Garbuno-Inigo et al, SIADS, 2020a

**Nonlinear** (affine invariant) **Fokker–Planck equation**[10]

$$\begin{aligned}\partial_\tau \pi_\tau &= -\nabla_\theta \cdot \left(\pi_\tau \Sigma_\tau \left\{\nabla_\theta \log \pi_\tau - \nabla_\theta \log \pi_{\rm a}\right\}\right)\\ &= -\text{grad}_\pi^{\rm AI}\, \text{KL}\left(\pi_\tau || \pi_{\rm a}\right)\end{aligned}$$

$\Sigma_t$ the covariance matrix of $\Theta_\tau \sim \pi_\tau$.

Metric $g_\pi(\rho_1, \rho_2)$ on space of densities induced by **Mahalanobis distance** on $\mathbb{R}^D$:

$$\|a\|^2_{\Sigma_\tau^{-1}} := a^{\rm T} \Sigma_\tau a$$

It holds that

$$\frac{\rm d}{{\rm d}\tau}{\rm KL}(\pi_\tau || \pi_{\rm a}) = -\int_{\mathbb{R}^N} \pi_\tau \left\| \nabla_\theta \frac{\delta {\rm KL}(\pi_\tau||\pi_{\rm a})}{\delta \pi_\tau}\right\|^2_{\Sigma_\tau^{-1}} \leq 0\,.$$

[10] Otto, Comm. Part. Diff. Eqs., 2001, SR & Cotter, CUP, 2015; Garbuno-Inigo et al, SIADS, 2020a

**Implementation** of **a**ffine **i**nvariant **L**angevin **d**ynamics (**ALDI**):

$$\mathrm{d}\Theta_\tau^{(i)} = -\Sigma_\tau^M \nabla_\theta V(\Theta_\tau^{(i)})\,\mathrm{d}\tau + \frac{D+1}{M}(\Theta_\tau^{(i)} - \bar{\theta}_\tau^M) + \sqrt{2}(\Sigma_\tau^M)^{1/2}\,\mathrm{d}W_\tau^{(i)}\,,$$

$i = 1, \ldots, M$, $\Theta_\tau \in \mathbb{R}^D$.

**Implementation** of **a**ffine **i**nvariant **L**angevin **d**ynamics (**ALDI**):

$$\mathrm{d}\Theta_\tau^{(i)} = -\Sigma_\tau^M \nabla_\theta V(\Theta_\tau^{(i)})\,\mathrm{d}\tau + \frac{D+1}{M}(\Theta_\tau^{(i)} - \bar{\theta}_\tau^M) + \sqrt{2}(\Sigma_\tau^M)^{1/2}\,\mathrm{d}W_\tau^{(i)}\,,$$

$i = 1, \ldots, M$, $\Theta_\tau \in \mathbb{R}^D$.

**Remarks.**

- **correction term in orange** is needed for invariance of $\pi_a$ (**multiplicative noise**),
- **invariance** and **ergodicity** holds provided $M \geq D + 1$,
- ALDI is **affine invariant** for any $M \geq 2$,
- **derivative-free formulation** .

Can we avoid the computation of gradients?

**Idea**: Introduce localised covariance matrices

$$\Sigma_\tau(\theta) := \frac{1}{C} \int (\theta' - \bar{\theta}_\tau)(\theta' - \bar{\theta}_\tau)^{\mathrm{T}} \, e^{-\frac{1}{2\gamma}\|\theta' - \theta\|^2_{\Sigma_\tau}} \, \pi_\tau(\theta') \, \mathrm{d}\theta' \,, \qquad (3)$$

$\bar{\theta}_\tau$ localised mean, $\gamma > 0$, $C > 0$ a scaling constant.

**Localised** ALDI dynamics:

$$\mathrm{d}\Theta_\tau = -\Sigma_\tau(\Theta)\nabla_\theta V(\Theta_\tau)\,\mathrm{d}\tau + \nabla_\theta \cdot \Sigma_\tau(\Theta_\tau)\,\mathrm{d}\tau + \sqrt{2}\,\Sigma_\tau(\Theta_\tau)^{1/2}\,\mathrm{d}W_\tau$$

Let $\overline{V}_\tau(\theta)$ denote the expectation of $V(\Theta')$ w.r.t. density defined in (3), that is,

$$\tilde{\pi}_\tau(\theta'|\theta) = \frac{1}{C} e^{-\frac{1}{2\gamma}\|\theta'-\theta\|^2_{\Sigma_\tau}} \pi_\tau(\theta')\,.$$

**Catch:** $\tilde{\pi}_\tau(\theta'|\theta)$ is **close to Gaussian** with mean $\theta$ for $\gamma \ll 1$.

Allows for **derivative-free implementation** of ALDI/EnKBF with controllable errors as $\gamma \ll 1$ and $M \to \infty$:

$$\Sigma_\tau \nabla_\theta V \approx \Sigma_\tau \overline{\nabla_\theta V}_\tau \approx \overline{(\Theta' - \bar{\theta}_\tau)(V(\Theta') - \bar{V}_\tau)}_\tau\,.$$

Nonlinear forward operator

$$g(\theta_1, \theta_2) = (\theta_1 - \theta_2)^2$$

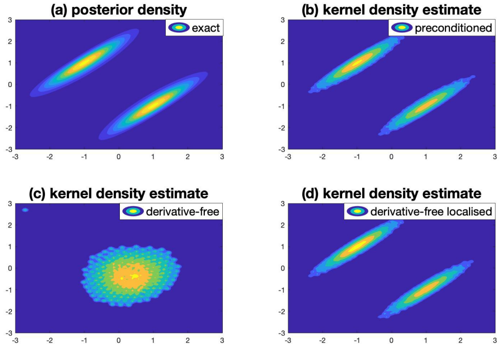



Figure: Bimodal posterior (a) and ALDI results (b)-(d) for two-dimensional multimodal posterior distribution.

# Conclusions

- Increasing interest in interacting particle systems for sampling, inference, and optimisation
- Fruitful exchange between methods based on ergodicyt & invariance and those based on coupling of measures
- Solid comparison is largely missing
- Affine invariance is highly desirable for applications in the natural sciences
- Gradient flow structures in the space of probability measures also appear as desirable; but in which metric and under which cost functional?