Statistics and Computing (2025) 35:64
https://doi.org/10.1007/s11222-025-10595-w

ORIGINAL PAPER

Metropolis-adjusted interacting particle sampling

Björn Sprungk[1] · **Simon Weissmann**[2] · **Jakob Zech**[3]

Received: 21 December 2023 / Accepted: 20 February 2025

Abstract
In recent years, various interacting particle samplers have been developed to sample from complex target distributions, such as those found in Bayesian inverse problems. These samplers are motivated by the mean-field limit perspective and implemented as ensembles of particles that move in the product state space according to coupled stochastic differential equations. The ensemble approximation and numerical time stepping used to simulate these systems can introduce bias and affect the invariance of the particle system with respect to the target distribution. To correct for this, we investigate the use of a Metropolization step, similar to the Metropolis-adjusted Langevin algorithm. We examine Metropolization of either the whole ensemble or smaller subsets of the ensemble, and prove basic convergence of the resulting ensemble Markov chain to the target distribution. Our numerical results demonstrate the benefits of this correction in numerical examples for popular interacting particle samplers such as ALDI, CBS, and stochastic SVGD.

Keywords Metropolis–Hastings · Interacting particle systems · Bayesian inference

1 Introduction

Generating samples or computing expectations with respect to a given target distribution π in $\mathbb{R}^d$ is a ubiquitous task in applied mathematics, computational physics, statistics, and data science. Applications are broad and include for example Bayesian inference, generative modeling, and hypothesis testing and model fitting.

Various methods tackling this problem have been proposed and analyzed in the literature. A classical and nowadays standard method is Markov-Chain Monte Carlo (MCMC) (Brooks et al. 2011) and, in particular, the popular Metropolis–Hastings (MH) algorithm (Metropolis et al. 1953; Hastings 1970). Recently, novel approaches that couple the target distribution with a reference distribution π_0 through a deterministic "transport map" have emerged, such as polynomial transports (Marzouk et al. 2016; Jaini et al. 2019), tensor-train transports (Dolgov et al. 2020), normalizing flows (Rezende and Mohamed 2015), and neural ODEs (Chen et al. 2018). The resulting sampling methods aim to transform initial iid samples following the reference distribution to samples (approximately) following the target distribution by applying the transport map samplewise.

Another way to achieve such a transformation of an initial ensemble of particles or samples following π_0 is by applying suitable stochastic dynamics to the ensemble which for time $t \to \infty$ yield particles approximately distributed according to the target π. The resulting ensemble dynamics are often interacting, i.e., the drift or diffusion term for each particle depends on the whole ensemble. Such stochastic interacting particle systems emerge from various ideas and approaches: (i) as ensemble approximations of π-invariant stochastic differential equations of Langevin or McKean–Vlasov type (Garbuno-Inigo et al. 2020a, b; Grenander and Miller 1994), (ii) by adapting methods from particle swarm optimization to construct samplers (Carrillo et al. 2022), or (iii) from gradient flows to minimize some objective quantifying the difference

Björn Sprungk, Simon Weissmann and Jakob Zech have contributed equally to this work.

✉ Simon Weissmann
simon.weissmann@uni-mannheim.de

Björn Sprungk
bjoern.sprungk@math.tu-freiberg.de

Jakob Zech
jakob.zech@uni-heidelberg.de

[1] Faculty of Mathematics and Computer Science, Technische Universität Bergakademie Freiberg, 09599 Freiberg, Germany

[2] Institute of Mathematics, Universität Mannheim, 68138 Mannheim, Germany

[3] Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg, 69120 Heidelberg, Germany

Published online: 11 March 2025

between the target π and a current approximation π_t (Liu and Wang 2016), (Nüsken and Renger 2023; Gallego and Insua 2018). For each of these approaches we consider a particular example in this work: (i) an affine invariant interacting Langevin sampler (ALDI) (Garbuno-Inigo et al. 2020a, b), (ii) a consensus-based sampler (CBS) (Carrillo et al. 2022), and (iii) a stochastic version of Stein variational gradient descent (SVGD) (Gallego and Insua 2018).

In practice, simulating the resulting stochastic dynamical system requires a time discretization and a suitable numerical integration scheme. For example, approximating the Langevin dynamics via the Euler–Maruyama scheme leads to the so-called unadjusted Langevin algorithm (ULA) (Roberts and Tweedie 1996). For ULA it can be shown that the time-stepping scheme causes a bias, so that the limiting distribution has an error of the size of the time discretization step, see e.g. (Vempala and Wibisono (2019), Theorem 2). The Metropolis-adjusted Langevin algorithm (MALA) circumvents this problem by introducing an MH acceptance/rejection-step after each Langevin update (Besag 1994; Roberts and Tweedie 1996).

In this paper, we propose a similar approach for recent *interacting* particle systems with $M \in \mathbb{N}$ particles. That is, we view the time-discrete interactive update of the particles as a proposal within an MH scheme. This results in a Markov chain in the M-fold product space, $\mathbb{R}^d \times \cdots \times \mathbb{R}^d = \mathbb{R}^{Md}$, that corrects for the bias introduced by the time-discretization of the ensemble dynamics. Compared to non-interacting Markov chains, we see several potential advantages. The use of an entire ensemble of particles enables the simultaneous exploration of a broader area, helping to avoid getting stuck in a single mode. Additionally, the information obtained from the whole particle system can be used to infer important properties of the target distribution, such as its local curvature, and provide proposals that are better adapted to it—similarly to adaptive MCMC, see e.g., Radu and Yang (2009). Another significant advantage is that particle methods are inherently parallelizable, as each particle can be updated independently before interaction steps, allowing for efficient implementation. Consequently, they provide a natural approach to parallelize Markov chain Monte Carlo sampling.

The concept of ensemble MCMC was introduced in Christen and Fox (2010), Goodman and Weare (2010), where particles, referred to as "walkers", are updated individually using so-called walk or stretch moves that involve only two of the M particles. Although, earlier instances of algorithms which propagate ensembles of particles where the moves are constructed using the current ensemble can be found in Gilks et al. (1994), Braak (2006). Following Goodman and Weare (2010) further ensemble MCMC algorithms have been proposed in recent years (Clarté et al. 2022; Coullon and Webber 2021; Dunlop and Stadler 2022; Leimkuhler et al. 2018; Zhang and Sutton 2011; Zhu 2019). In the works (Coullon and Webber 2021; Dunlop and Stadler 2022; Leimkuhler et al. 2018; Zhang and Sutton 2011; Zhu 2019) the ensemble is used to estimate the target covariance empirically. The ensemble covariance is then applied as a preconditioner or covariance for proposing new states, e.g., based on a Gauss-Newton update or the (generalized) preconditioned Crank-Nicolson proposal (Cotter et al. 2013; Rudolf and Sprungk 2018). Similar to Goodman and Weare (2010), the authors of Coullon and Webber (2021), Dunlop and Stadler (2022), Leimkuhler et al. (2018), Zhang and Sutton (2011) use a sequential particle-wise update and, hence, Metropolization. The updating schemes in Clarté et al. (2022), Zhu (2019) are slightly different, since there the particle to be updated is chosen at random according to suitable algorithm-dependent probabilities, but still one particle at a time is evolved.

Outline The remainder of this paper is organized as follows: In this section, we explain interacting particle systems and present our main ideas, describe our contributions, and introduce notation. In Sect. 2, we present several Metropolization strategies for interacting particle systems and provide convergence results for each of them. Section 3 discusses common examples of interacting particle systems that are based on various underlying stochastic dynamics, and we explain how they align with our Metropolization schemes. Finally, in Sect. 4 we report on numerical results for all presented interacting particle methods. Moreover, for the convenience of the reader, in Sect. A we review the basic methodology of the Metropolis–Hastings algorithm, and also discuss classic results related to its convergence which are required in the main text.

Notation and conventions Throughout we consider an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{R}^d$ to be equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$, and we assume the target probability distribution π on $\mathbb{R}^d$ to be absolutely continuous with respect to Lebesgue measure. By abuse of notation, we use the same symbols to denote the Lebesgue densities and the corresponding distributions they represent. Moreover, we denote by $\mathcal{P}(\mathbb{R}^d)$ the set of probability densities on $\mathbb{R}^d$. As usual, $\mathrm{N}(\mu, \Sigma)$ stands for a normal distribution with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, and $\mathrm{U}([0,1])$ denotes a uniform distribution on $[0,1]$.

For a transition kernel $P : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$, and a measure μ on $\mathbb{R}^d$ we use the usual notation μP for the measure

$$\mu P(A) := \int_{\mathbb{R}^d} P(z, A)\mu(dz) \qquad \forall A \in \mathcal{B}(\mathbb{R}^d) \tag{1}$$

and inductively $\mu P^k := (\mu P^{k-1})P$ for all $k \geq 2$. Moreover, we denote by $L^1_\mu(\mathbb{R})$ the Lebesgue space of μ-integrable functions $F : \mathbb{R}^d \to \mathbb{R}$. Similarly $L^2_\mu(\mathbb{R})$ stands for the square integrable functions w.r.t. the measure μ. For $F \in L^1_\mu(\mathbb{R})$ we write $\mathbb{E}_\mu[F] := \int_{\mathbb{R}^d} F(x)\mu(\mathrm{d}x)$ and additionally $\mathbb{V}_\mu[F] := \mathbb{E}_\mu[(F - \mathbb{E}_\mu[F])^2]$ in case $F \in L^2_\mu(\mathbb{R})$.

We write $\mathcal{C}(\mathbb{R}^d) \subset \mathbb{R}^{d\times d}$ for the set of symmetric positive semidefinite matrices of size $d \times d$. The set of all symmetric positive definite matrices is denoted by $\mathcal{C}^+(\mathbb{R}^d)$. For $C \in \mathcal{C}(\mathbb{R}^d)$, we write $\sqrt{C}$ to denote the unique square root of C in $\mathcal{C}(\mathbb{R}^d)$. The n-dimensional identity matrix is denoted by $\mathrm{Id}_n \in \mathbb{R}^{n\times n}$.

We use boldface notation $\mathbf{x}$ to denote vectors in $\mathbb{R}^{Md}$. They are always interpreted as an ensemble of M vectors in $\mathbb{R}^d$ which in turn are denoted by $x^{(1)}, \ldots, x^{(M)}$. The ensemble excluding the ith particle will be denoted by $\mathbf{x}^{-(i)}$. More precisely

$$\mathbf{x} := \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(M)} \end{pmatrix} \in \mathbb{R}^{Md} \quad \text{and}$$

$$\mathbf{x}^{-(i)} := \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(i-1)} \\ x^{(i+1)} \\ \vdots \\ x^{(M)} \end{pmatrix} \in \mathbb{R}^{M(d-1)}. \tag{2}$$

For random variables we use upper case notation, for example $\boldsymbol{X} = ((X^{(1)})^\top, \ldots, (X^{(M)})^\top)^\top$. The notation $\boldsymbol{X} = \mathbf{x}$ signifies that a draw of this random variable yielded the value $\mathbf{x}$ and we use $X^{(i)} \in \mathbb{R}^d$ or $\boldsymbol{X} \in \mathbb{R}^{Md}$ as shorthand notation for $X^{(i)}$ being an $\mathbb{R}^d$-valued and $\boldsymbol{X}$ being an $\mathbb{R}^{Md}$-valued random variable, respectively.

1.1 Interacting particle systems

The starting point of our method are dynamical systems that transform a single particle $X_0 \sim \pi_0$ at time $t = 0$ into a particle following the target distribution π as $t \to \infty$. The dynamics of the particle are described by a stochastic differential equation (SDE) of McKean–Vlasov type (Pathiraja et al. 2021):

$$\mathrm{d}X_t = \phi(X_t, \pi_t)\mathrm{d}t + \sqrt{\sigma(X_t, \pi_t)}\mathrm{d}B_t, \tag{3}$$

where $\pi_t : \mathbb{R}^d \to [0, \infty)$ is the probability density of $X_t \in \mathbb{R}^d$ at time t, $B_t \in \mathbb{R}^d$ is a (standard) Brownian motion, $\phi : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ is referred to as the *drift*, and $\sigma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathcal{C}(\mathbb{R}^d)$ is the *diffusion*. Moreover, we assume existence of a unique strong solution X_t of (3) throughout the paper.

Example 1.1 (Langevin dynamics). One classical example of (3) is

$$\mathrm{d}X_t = C\nabla \log \pi(X_t)\mathrm{d}t + \sqrt{2C}\mathrm{d}B_t \tag{4}$$

for a fixed covariance matrix $C \in \mathcal{C}^+(\mathbb{R}^d)$. The density π_t of X_t satisfies the corresponding Fokker–Planck equation

$$\partial_t \pi_t = \nabla \cdot (\pi_t C \nabla \log(\pi)) + \mathrm{Tr}(C\nabla^2 \pi_t),$$

which describes the gradient flow in the space of probability measures w.r.t. the Wasserstein metric (Jordan et al. 1998). Under suitable assumptions (π satisfies a Poincaré inequality) one can show exponential convergence of π_t to π as $t \to \infty$, e.g., Markowich and Villani (2000).

Note that the drift and diffusion in (4) are independent of π_t. This is in contrast to the closely related "Kalman-Wasserstein dynamics" (Garbuno-Inigo et al. 2020a):

Example 1.2 (Kalman-Wasserstein dynamics). Choosing C in (4) as

$$C(\pi_t) := \int_{\mathbb{R}^d} (x - m(\pi_t))(x - m(\pi_t))^\top \pi_t(x)\,\mathrm{d}x \in \mathcal{C}(\mathbb{R}^d),$$
$$m(\pi_t) := \int_{\mathbb{R}^d} x\pi_t(x)\,\mathrm{d}x \in \mathbb{R}^d \tag{5a}$$

yields a McKean–Vlasov Langevin dynamic of the form (3) with

$$\phi(x, \rho) = C(\rho)\nabla \log \pi(x), \qquad \sigma(x, \rho) = 2C(\rho), \tag{5b}$$

for $x \in \mathbb{R}^d$ and $\rho \in \mathcal{P}(\mathbb{R}^d)$.

For strongly log-concave target measures π and under the additional assumption that $C(\pi_t)$ does not degenerate, the resulting Markov process $(X_t)_{t\geq 0}$ is ergodic with unique invariant distribution π and there is exponential convergence of π_t to π in the Kullback–Leibler divergence $\mathrm{KL}(\pi_t \,\|\, \pi)$ as $t \to \infty$ (Garbuno-Inigo et al. (2020a), Proposition 2). Potential advantages of replacing C by $C(\pi_t)$ are (i) faster convergence of $\pi_t \to \pi$ due to the preconditioning and (ii) affine-invariance of the resulting dynamics, see Garbuno-Inigo et al. (2020b), Leimkuhler et al. (2018).

Ensemble discretization Solving (3) by numerical methods requires discretization. In terms of the distribution π_t, this is achieved by replacing π_t with the empirical distribution of

an ensemble of $M \in \mathbb{N}$ particles $X_t^{(i)} \in \mathbb{R}^d$, $i = 1, \dots, M$, initialized iid as $X_0^{(i)} \sim \pi_0, i = 1, \dots, M$, at time $t = 0$. The particles are collected into a vector $\boldsymbol{X}_t \in \mathbb{R}^{Md}$ representing the state of the whole ensemble (cf. (2)). Equation (3) then formally becomes a coupled system of SDEs

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{\Phi}(\boldsymbol{X}_t)\,\mathrm{d}t + \sqrt{\boldsymbol{\Sigma}(\boldsymbol{X}_t)}\,\mathrm{d}\boldsymbol{B}_t, \tag{6}$$

for some suitable drift, diffusion, and standard Brownian motion

$$\boldsymbol{\Phi} : \mathbb{R}^{Md} \to \mathbb{R}^{Md}, \qquad \boldsymbol{\Sigma} : \mathbb{R}^{Md} \to \mathcal{C}(\mathbb{R}^{Md}), \qquad \boldsymbol{B}_t \in \mathbb{R}^{Md}.$$

Again, we assume well-definedness of the solution $\boldsymbol{X}_t$ of (6) throughout.

Example 1.3 (Interacting Langevin Dynamics). We continue the example of McKean–Vlasov Langevin dynamics from Example 1.1

$$\mathrm{d}X_t = C(\pi_t)\nabla \log \pi(X_t)\mathrm{d}t + \sqrt{2C(\pi_t)}\mathrm{d}B_t. \tag{7}$$

The computation of $C(\pi_t)$ in (5a) requires to approximate an integral w.r.t. π_t. Using Monte Carlo integration based on an ensemble of M particles following these dynamics we obtain the ensemble version (Garbuno-Inigo et al. 2020a) of the mean-field dynamics (7)

$$\begin{aligned}\mathrm{d}X_t^{(i)} &= C(\boldsymbol{X}_t)\nabla_x \log \pi(X_t^{(i)})\,\mathrm{d}t + \sqrt{2C(\boldsymbol{X}_t)}\mathrm{d}B_t^{(i)} \\ &\quad i \in \{1, \dots, M\},\end{aligned} \tag{8}$$

where

$$\begin{aligned}C(\boldsymbol{X}_t) &:= \frac{1}{M}\sum_{i=1}^{M}\left(X_t^{(i)} - m(\boldsymbol{X}_t)\right)\left(X_t^{(i)} - m(\boldsymbol{X}_t)\right)^\top \in \mathcal{C}(\mathbb{R}^d), \\ m(\boldsymbol{X}_t) &:= \frac{1}{M}\sum_{i=1}^{M} X_t^{(i)} \in \mathbb{R}^d\end{aligned} \tag{9}$$

denote the empirical covariance and mean of the ensemble $\boldsymbol{X}_t$. Note that the system of SDEs (8) is completely coupled since individual particles interact via $C(\boldsymbol{X}_t)$.

Considering the associated Fokker–Planck equation of (8), it can be shown that the invariant distribution of the particles in (8) in general may have a bias, i.e. does not equal π. To address this issue, the authors in Nüsken and Reich (2019), Garbuno-Inigo et al. (2020b) propose the following modification called affine invariant Langevin dynamics (ALDI):

$$\begin{aligned}\mathrm{d}X_t^{(i)} &= C(\boldsymbol{X}_t)\nabla \log \pi(X_t^{(i)})\,\mathrm{d}t + \frac{d+1}{M}(X_t^{(i)} - m(\boldsymbol{X}_t))\,\mathrm{d}t \\ &\quad + \sqrt{2C(\boldsymbol{X}_t)}\mathrm{d}B_t^{(i)}, \qquad i \in \{1, \dots, M\}.\end{aligned} \tag{10}$$

Denote by $\boldsymbol{\pi}_t$ the density of $\boldsymbol{X}_t$. For strongly log-concave target measures π and under additional assumptions such as $M > d + 1$, one can guarantee $\boldsymbol{\pi}_t \to \boldsymbol{\pi}$ in total variation as $t \to \infty$, where $\boldsymbol{\pi}(\mathbf{x}) := \prod_{i=1}^{M} \pi(x^{(i)})$ (Garbuno-Inigo et al. (2020b), Proposition 4.5).

Time discretization Besides discretizing the distribution, a numerical time-stepping scheme to approximately simulate (6) is required. For a fixed time step size $h > 0$, let $\boldsymbol{\zeta}_k \sim \mathrm{N}(0, \mathrm{Id}_{Md}), k \in \mathbb{N}$. Then the Euler–Maruyama discretization of (6) reads

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + h\boldsymbol{\Phi}(\boldsymbol{X}_k) + \sqrt{h\boldsymbol{\Sigma}(\boldsymbol{X}_k)}\boldsymbol{\zeta}_{k+1} \tag{11}$$

with $\boldsymbol{X}_k = (X_k^{(1)}, \dots, X_k^{(M)}) \in \mathbb{R}^{Md}$ and initialized as $X_0^{(i)} \sim \pi_0$ iid for $i = 1, \dots, M$.

Example 1.4 (Unadjusted Langevin algorithm). The Langevin dynamics (4) do not require an ensemble approximation. However, we may consider M particles $X_t^{(i)}, i = 1, \dots, M$, each individually following (4) without interaction. This can be viewed as (non-interacting) ensemble dynamics of the form (6). The corresponding time-discretized system is

$$\begin{aligned}X_{k+1}^{(i)} &= X_k^{(i)} + hC\nabla \log \pi(X_k^{(i)}) \\ &\quad + \sqrt{2hC}\zeta_{k+1}^{(i)}, \qquad i \in \{1, \dots, M\},\end{aligned} \tag{12}$$

with $\zeta_{k+1}^{(i)} \sim \mathrm{N}(0, \mathrm{Id}_d)$ iid. For $C = \mathrm{Id}_d$, this is known as the (parallel) ULA (Grenander and Miller 1994; Roberts and Tweedie 1996).

Example 1.5 (Unadjusted interacting Langevin dynamics). For the interacting Langevin dynamics (10), we obtain the time-discrete interacting particle system

$$\begin{aligned}X_{k+1}^{(i)} &= X_k^{(i)} + hC(\boldsymbol{X}_k)\nabla \log \pi(X_k^{(i)}) \\ &\quad + h\frac{d+1}{M}(X_k^{(i)} - m(\boldsymbol{X}_k)) \\ &\quad + \sqrt{2hC(\boldsymbol{X}_k)}\zeta_{k+1}^{(i)}, \qquad i \in \{1, \dots, M\},\end{aligned} \tag{13}$$

with $\zeta_{k+1}^{(i)} \sim \mathrm{N}(0, \mathrm{Id}_d)$ and $C(\boldsymbol{X}_k)$ as in (9).

It is well-known that the introduction of the time discretization may lead to a bias, i.e. in general it does *not* hold that $X_k^{(i)}$ converges in distribution to π as $k \to \infty$. This applies for example to ULA (Vempala and Wibisono 2019).

1.2 Main idea and contributions

Besag (1994) suggested in 1994 to correct the unadjusted Langevin algorithm with a Metropolization to obtain a π-invariant one-particle Markov chain $(X_k)_{k\in\mathbb{N}}$, which lead

to MALA, whereas a similar Metropolization correction already appeared in Rossky et al. (1978) for the Monte Carlo simulation of Langevin-type dynamics from physics. We adopt this idea to correct for the bias in general time-discrete *interacting particle systems* (11). To this end, we view (11) as the proposal mechanism for an ensemble Markov chain $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ in the product state space $\mathbb{R}^{Md}$. We propose to apply Metropolization in three ways: (i) ensemble-wise, i.e., accept or reject the whole ensemble of all proposed particles, (ii) particle-wise, i.e., accept or reject each proposed particle individually in a sequential manner, and (iii) block-wise, i.e., accept or reject each block of particles individually in a sequential manner. Here a *block* is understood as a fixed subset of particles and identified with $\mathbf{b} \subset \{1, \dots, M\}$. Methods (i) and (ii) can be seen as a special case of (iii) with either just one batch representing the whole ensemble, or batches consisting of only one particle. The presentation in Sect. 2 is therefore focused on the more general variant (iii). A high-level version of the novel block-wise Metropolization is summarized in Algorithm 1.

Algorithm 1 Block-wise Metropolization

fix a partition $\bigcup_{j=1}^{L} \mathbf{b}_j = \{1, \dots, M\}$
draw $\mathbf{x}_0 \in \mathbb{R}^{Md}$ according to $\otimes_{i=1}^{M} \pi_0$ and set initial ensemble state $\boldsymbol{X}_0 = \mathbf{x}_0$
for $k = 0, \dots, N$ **do**
  given $\boldsymbol{X}_k = \mathbf{x}_k$ initialize $\mathbf{x} = (x^{(1)}, \dots, x^{(M)})$ by $\mathbf{x} = \mathbf{x}_k$
  for $j = 1, \dots, L$ **do**
    draw block proposal $(y^{(i)})_{i\in\mathbf{b}_j} \in \mathbb{R}^{|\mathbf{b}_j|d}$ according particle proposals such as (12) or (13) based on current ensemble $\mathbf{x}$
    update $(x^{(i)})_{i\in\mathbf{b}_j} = (y^{(i)})_{i\in\mathbf{b}_j}$ only with block acceptance probability
  end for
  set $\boldsymbol{X}_{k+1} = \mathbf{x}$
end for
return ensemble chain $(\boldsymbol{X}_k)_{k=0}^{N+1}$

Besides the sequential updating of (blocks of) particles, one might also consider a *simultaneous version* where the proposal and acceptance/rejection for each particle takes place in parallel and independently of the other particles. Such a method is computationally appealing, as it enables full parallelization over the whole ensemble. However, as it turns out, a simultaneous Metropolization yields in general a biased algorithm, i.e., the ensemble Markov chain does not have the correct invariant measure. We discuss this in detail in Sect. B. Nonetheless, under suitable assumptions on interaction structure of the particle system one can allow for a simultaneous acceptance/rejection of the particles within each block in the sequential block updating. This *simultaneous within block* Metropolization potentially yields a lower autocorrelation in the resulting ensemble Markov chain than accepting/rejection the whole block and is discussed in Sect. 2.2. We emphasize the following potential advantages of combining the Metropolis–Hastings mechanism with interacting particle systems:

- From an interacting particle sampling perspective we *correct for the immanent bias* of the particle dynamics due to numerical time-stepping schemes and finite ensemble approximations of $\phi(\cdot, \pi_t)$ and $\sigma(\cdot, \pi_t)$ in (3). In particular, this allows in principle to take large time steps h in (11) without provoking an instability of the time-discrete dynamical system.
- From an MCMC sampling perspective the interacting particle dynamics provide in each iteration not just one new state $x_k \in \mathbb{R}^d$ but M new states which can be computed *in parallel*. This yields a computational advantage. Moreover, in comparison to simply performing, e.g., parallel MALA, the interaction of the particles may yield more *efficient proposal kernels* due to estimating, e.g., the target covariance empirically by the ensemble, and, thus, lead to more efficient MCMC sampling. In particular, we obtain *affine-invariant* MCMC methods if the underlying interacting particle dynamics are affine-invariant themselves such as those proposed in Garbuno-Inigo et al. (2020a, b), Carrillo et al. (2022). The benefit of affine-invariance is an improved mixing, particularly, for highly anisotropic target distributions since the proposal is then adapted to the anisotropy, see e.g., Christen and Fox (2010), Goodman and Weare (2010) for numerical illustrations and Rudolf and Sprungk (2022) for a theoretical analysis of affine-invariant MCMC including the independence of spectral gaps for arbitrarily anisotropic Gaussian targets.

Finally, we want to highlight that the purpose of the presented framework is to introduce the Metropolis–Hastings mechanism to a broad class of interacting particle systems. We do not compare different interacting particle methods with each other, nor do we compare the proposed framework to other adaptive MCMC methods.

Contributions We summarize the main contributions of this paper:

1. We propose a new MCMC method by combining a Metropolization step with interacting particle dynamics in the M-fold product state space.
2. We discuss several variants of this method for differing sizes of the blocks which are Metropolized. Basic convergence results as well as numerical indications suggesting that suitable choices of block sizes lead to more efficient sampling algorithms are presented.
3. We give several concrete examples based on different particle dynamics, including the recently introduced

stochastic SVGD (Nüsken and Renger 2023; Gallego and Insua 2018), ALDI (Garbuno-Inigo et al. 2020a, b), and CBS (Carrillo et al. 2022).
4. We present numerical experiments to compare the different variants of our algorithm, demonstrating in particular improved robustness and higher efficiency due to interaction and (block-wise) Metropolization.
5. In the appendix, we provide counterexamples to show that a *simultaneous* particle-wise Metropolization strategy does in general not yield an unbiased algorithm.

2 Metropolis-adjusted interacting particle sampling

Consider again an interacting particle system (6) resulting from a finite ensemble approximation of suitable π-invariant McKean–Vlasov dynamics as in (3), and its time discretization (11) with step size $h > 0$, i.e.

$$X_{k+1} = X_k + h\Phi(X_k) + \sqrt{h\Sigma(X_k)}\zeta_{k+1}. \tag{14}$$

As motivated above time- and ensemble discretization introduces a bias in the ensemble dynamics (14), i.e., the particle-wise marginals of the invariant measure of (14) are in general not equal to π. This can be corrected my Metropolization viewing (14) rather as a proposal mechanism in a Metropolis–Hastings scheme. However, several ways how to Metropolize are possible:

- Propose and accept/reject the ensemble as a whole,
- propose and accept/reject the individual particles,
- or propose and accept/reject blocks of particles as a whole.

The first variant, referred to as *ensemble wise Metropolization*, is maybe the most natural and offers the possibility of parallelization. However, it may require rather small steps sizes h in order to ensure that all proposed particle moves are accepted. The second variant, referred to as *particle-wise Metropolization*, on the other hand will allow in general for larger step sizes for the same average acceptance rate, but cannot be parallelized. The third variant, which we call *block-wise Metropolization*, serves as the intermediate level between the previous two extremes. This variant allows for larger step sizes h than ensemble-wise Metropolization and for parallelization over the blocks. To this end, we partition the ensemble of M particles into blocks of particles of size $B \leq M$. Thus, choosing $B = M$ recovers the first Metropolization variant and $B = 1$ the second one. We therefore focus in our subsequent presentation on the case of block-wise Metropolization.

Besides the sequential updating of blocks, cf. Gibbs sampling, one could also consider a *simultaneous* Metropolization, i.e., we propose and accept/reject the blocks in parallel. This turns out to be biased again in general, which is demonstrated in detail in Sect. B. However, under certain assumptions we can allow for a modification of the block-wise Metropolization[1] where rather than the block as a whole each particle within the block is accepted or rejected individually. This is called **simultaneous within block-wise Metropolization** and may yield a better mixing, i.e., reduce the autocorrelation in the ensemble Markov chain, while still maintaining the advantage of parallel computation. Previous ensemble MCMC algorithms introduced in Goodman and Weare (2010), Leimkuhler et al. (2018), Coullon and Webber (2021), Dunlop and Stadler (2022) are only following a particle-wise Metropolization, thereby not allowing for a direct parallelization.

Throughout we assume the following:

Assumption 2.1 *For each* $\mathbf{x} \in \mathbb{R}^{Md}$ *the matrix* $\boldsymbol{\Sigma}(\mathbf{x}) \in \mathbb{R}^{Md\times Md}$ *in* (14) *is invertible.*

This assumptions ensures the existence of Lebesgue densities of the corresponding ensemble proposal kernel on $\mathbb{R}^{Md}$ derived from Equation (14). Another important consequence of the invertibility of $\boldsymbol{\Sigma}(\mathbf{x})$ is related to the so-called *subspace property* of the ensemble dynamics, see e.g. Iglesias et al. (2013). This means that for certain $\boldsymbol{\Phi}$ each particle $X_k^{(i)}$ of the ensemble Markov chain $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ would remain in the span of the initial ensemble $V_0 := \operatorname{span}(X_0^{(1)}, \dots, X_0^{(M)})$, and, thus, the ensemble Markov chain would not explore the entire space $\mathbb{R}^{Md}$, but rather stay within $V_0 \times \cdots \times V_0 = V_0^M$. However, if Assumption 2.1 is satisfied, the subspace property does not hold (except for $V_0 = \mathbb{R}^d$).

Remark 2.2 Assumption 2.1 is not necessarily satisfied in common particle dynamics, in particular, if these dynamics involve the empirical covariance matrix $C(\boldsymbol{X}_k)$ of the ensemble, e.g.,

$$\boldsymbol{\Sigma}(\mathbf{x}) = \begin{pmatrix} C(\mathbf{x}) & & \\ & \ddots & \\ & & C(\mathbf{x}) \end{pmatrix}.$$

In these cases, $\boldsymbol{\Sigma}(\mathbf{x})$ may satisfy invertibility only for almost every $\mathbf{x} \in \mathbb{R}^{Md}$. This can be addressed (both theoretically and practically) by replacing $\boldsymbol{\Sigma}(\mathbf{x})$ with an almost everywhere equal $\tilde{\boldsymbol{\Sigma}}(\mathbf{x})$ which satisfies invertibility for all $\mathbf{x} \in \mathbb{R}^{Md}$.

Remark 2.3 Equation (14) generates an ensemble Markov chain $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ in the product state space $\mathbb{R}^{Md}$. We emphasize that the dynamics of each individual particle $(X_k^{(i)})_{k\in\mathbb{N}}$,

[1] We thank the anonymous reviewer for suggesting this variant of the algorithm.

$i = 1, \dots, M$, does not necessarily satisfy the Markov property with respect to the filtration $\mathcal{F}_t^{(i)} = \sigma(X_s^{(i)}, s \le t)$, since the particle-wise drift and diffusion term may depend on the entire ensemble X_k.

2.1 Block-wise metropolization

In the following, let $\mathbf{\Phi} : \mathbb{R}^{M\,d} \to \mathbb{R}^d$ be given by

$$\mathbf{\Phi}(\mathbf{x}) = \left(\Phi_i\left(x^{(i)},\ \mathbf{x}^{-(i)}\right)\right)_{i=1,\dots,M},$$

with $\Phi_i : \mathbb{R}^d \times \mathbb{R}^{(M-1)d} \to \mathbb{R}^d$ and $\mathbf{\Sigma} : \mathbb{R}^{M\,d} \to \mathcal{C}(\mathbb{R}^d)$ be given by

$$\mathbf{\Sigma}(\mathbf{x}) = \left(\Sigma_{i,j}\left(x^{(i)},\ x^{(j)},\ \mathbf{x}^{-(i,j)}\right)\right)_{i,j=1,\dots,M}$$

where $\Sigma_{i,j} : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{(M-2)d} \to \mathbb{R}^{d\times d}$. Given a "block" $\mathbf{b} \subseteq \{1, \dots, M\}$ of indices, we denote for $\mathbf{x} \in \mathbb{R}^{Md}$

$$\mathbf{x}^{(\mathbf{b})} := \left(x^{(j)}\right)_{j\in\mathbf{b}} \in \mathbb{R}^{|\mathbf{b}|d} \quad \text{and}$$
$$\mathbf{x}^{-(\mathbf{b})} := \left(x^{(j)}\right)_{j\in\{1,\dots,M\}\setminus\mathbf{b}} \in \mathbb{R}^{(M-|\mathbf{b}|)d}.$$

Moreover, we introduce for any ensemble state $\mathbf{x} \in \mathbb{R}^{Md}$ and block $\mathbf{b} \subseteq \{1, \dots, M\}$ the following functions

$$\mathbf{\Phi}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{x}^{(\mathbf{b})}) := \left(\Phi_i\left(x^{(i)},\ \mathbf{x}^{-(i)}\right)\right)_{i\in\mathbf{b}} \in \mathbb{R}^{|\mathbf{b}|d},$$
$$\mathbf{\Sigma}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{x}^{(\mathbf{b})}) := \left(\Sigma_{i,j}\left(x^{(i)},\ x^{(j)},\ \mathbf{x}^{-(i,j)}\right)\right)_{i,j\in\mathbf{b}} \in \mathbb{R}^{|\mathbf{b}|d\times|\mathbf{b}|d}.$$

2.1.1 Algorithm

Block proposal and transition kernel Given the current ensemble state $\mathbf{x} \in \mathbb{R}^{Md}$ we define a proposal kernel $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}} : \mathbb{R}^{|\mathbf{b}|d} \times \mathcal{B}(\mathbb{R}^{|\mathbf{b}|d}) \to [0, 1]$ for the particles within a block $\mathbf{b} \subseteq \{1, \dots, M\}$ by

$$\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \cdot) := \mathrm{N}\left(\mathbf{z} + h\,\mathbf{\Phi}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}),\, h\,\mathbf{\Sigma}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z})\right). \tag{15}$$

Under Assumption 2.1, for every $\mathbf{x}^{-(\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$, the proposal kernel $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}$ possesses a Lebesgue density $q_{\mathbf{x}^{-(\mathbf{b})}} : \mathbb{R}^{|\mathbf{b}|d} \times \mathbb{R}^{|\mathbf{b}|d} \to (0, \infty)$

$$q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y}) = \frac{1}{\det\left(2\pi\, h\, \mathbf{\Sigma}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z})\right)^{1/2}} \exp\left(-\frac{1}{2h}\left\|\mathbf{\Sigma}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z})^{-1/2}\left(\mathbf{y} - \mathbf{z} - h\mathbf{\Phi}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z})\right)\right\|^2\right).$$

This allows to define block-wise acceptance probabilities of the form

$$\boldsymbol{\alpha}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y}) := \begin{cases} \min\left(1, \frac{\boldsymbol{\pi}(\mathbf{y})\, q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{y},\mathbf{z})}{\boldsymbol{\pi}(\mathbf{z})\, q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z},\mathbf{y})}\right) & \text{if } \boldsymbol{\pi}(\mathbf{z})\, q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y}) > 0 \\ 1 & \text{else} \end{cases} \tag{16}$$

with $\mathbf{y}, \mathbf{z} \in \mathbb{R}^{|\mathbf{b}|d}$, where $\boldsymbol{\pi}(\mathbf{z}) := \prod_{i=1}^{|\mathbf{b}|} \pi(z^{(i)})$ for $\mathbf{z} \in \mathbb{R}^{|\mathbf{b}|d}$.

Remark 2.4 For many relevant interacting particle methods we have

$$\mathbf{\Phi}(\mathbf{x}) = \begin{pmatrix} \Phi(x^{(1)}, \mathbf{x}^{-(1)}) \\ \vdots \\ \Phi(x^{(M)}, \mathbf{x}^{-(M)}) \end{pmatrix},$$
$$\mathbf{\Sigma}(\mathbf{x}) = \begin{pmatrix} \Sigma(x^{(1)}, \mathbf{x}^{-(1)}) & & \\ & \ddots & \\ & & \Sigma(x^{(M)}, \mathbf{x}^{-(M)}) \end{pmatrix}, \tag{17}$$

with $\Phi : \mathbb{R}^d \times \mathbb{R}^{(M-1)d} \to \mathbb{R}^d$ and $\Sigma : \mathbb{R}^d \times \mathbb{R}^{(M-1)d} \to \mathcal{C}(\mathbb{R}^d)$. In that case, for any block $\mathbf{b} = \{i_1, \dots, i_{|\mathbf{b}|}\}$ we have for $\mathbf{z}, \mathbf{y} \in \mathbb{R}^{|\mathbf{b}|d}$

$$q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y}) = \prod_{j=1}^{|\mathbf{b}|} q_{\mathbf{x}^{-(i_j)}}\left(z^{(j)}, y^{(j)}\right)$$

where

$$q_{\mathbf{x}^{-(i)}}(z, y) = \frac{\exp\left(-\frac{1}{2h}\left\|\Sigma(z, \mathbf{x}^{-(i)})^{-1/2}\left(y - z - h\Phi(z, \mathbf{x}^{-(i)})\right)\right\|^2\right)}{\det\left(2\pi\, h\, \Sigma(z, \mathbf{x}^{-(i)})\right)^{1/2}}.$$

In other words, for any choice of block $\mathbf{b} = \{i_1, \dots, i_{|\mathbf{b}|}\}$ the proposal kernel takes a product form

$$\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \cdot) = \bigotimes_{j=1}^{|\mathbf{b}|} Q_{\mathbf{x}^{-(i_j)}}(z^{(j)}, \cdot),$$
$$Q_{\mathbf{x}^{-(i)}}(z, \cdot) := \mathrm{N}\left(z + h\,\Phi(z, \mathbf{x}^{-(i)}),\, h\,\Sigma(z, \mathbf{x}^{-(i)})\right). \tag{18}$$

Finally, we introduce the block MH transition kernel $\mathbf{P}_{\mathbf{x}^{-(\mathbf{b})}} : \mathbb{R}^{|\mathbf{b}|d} \times \mathcal{B}(\mathbb{R}^{|\mathbf{b}|}d) \to [0, 1]$, $\mathbf{x}^{-(\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$ describing the update of the $|\mathbf{b}|$ particles within a block $\mathbf{b}$ by

$$\mathbf{P}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathrm{d}\mathbf{y}) := \boldsymbol{\alpha}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y})\, \mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathrm{d}\mathbf{y}) + \mathbf{r}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z})\delta_{\mathbf{z}}(\mathrm{d}\mathbf{y}),$$

where $\mathbf{r}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}) := 1 - \int_{\mathbb{R}^{|\mathbf{b}|d}} \boldsymbol{\alpha}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y})\, \mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathrm{d}\mathbf{y})$. The transition kernel $\mathbf{P}_{\mathbf{x}^{-(\mathbf{b})}}$ is reversible w.r.t. $\boldsymbol{\pi} = \Pi_{i=1}^{|\mathbf{b}|}\pi$ by construction for any $\mathbf{x}^{-(\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$.

Algorithm 2 Block-wise Metropolized interacting particle sampling

Input:

- target density π on $\mathbb{R}^d$
- a partition $\dot{\bigcup}_{j=1}^{L} \mathbf{b}_j = \{1, \dots, M\}$
- ensemble dependent block proposal kernels $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b}_j)}}$ with density $\mathbf{q}_{\mathbf{x}^{-(\mathbf{b}_j)}} : \mathbb{R}^{|\mathbf{b}_j|d} \times \mathbb{R}^{|\mathbf{b}_j|d} \to (0, \infty)$ for arbitrary $\mathbf{x}^{-(\mathbf{b}_j)} \in \mathbb{R}^{(M-|\mathbf{b}_j|)d}$ and $j \in \{1, \dots, L\}$
- initial probability distribution π_0 on $\mathbb{R}^d$

Output: ensemble Markov chain $(\boldsymbol{X}_k)_{k\in\{1,\dots,N\}}$ in state space $\mathbb{R}^{Md}$

draw $\mathbf{x}_0 \sim \otimes_{i=1}^{M} \pi_0$ and set initial state $\boldsymbol{X}_0 = \mathbf{x}_0 \in \mathbb{R}^{Md}$
for $k = 0, \dots, N$ **do**
  given $\boldsymbol{X}_k = \mathbf{x}_k$ initialize $\mathbf{x} = (x_k^{(1)}, \dots, x_k^{(M)})$
  for $j = 1, \dots, L$ **do**
    draw proposal $\mathbf{y} \sim \mathbf{Q}_{\mathbf{x}^{-(\mathbf{b}_j)}}(\mathbf{x}^{(\mathbf{b}_j)}, \cdot)$
    compute block acceptance probability $\boldsymbol{\alpha}_{\mathbf{x}^{-(\mathbf{b}_j)}}(\mathbf{x}^{(\mathbf{b}_j)}, \mathbf{y}) \in [0, 1]$ in (16)
    draw $u_j \sim \mathrm{U}([0, 1])$ and set

$$\mathbf{x}^{(\mathbf{b}_j)} = \begin{cases} \mathbf{y} & \text{if } u_j \le \boldsymbol{\alpha}_{\mathbf{x}^{-(\mathbf{b}_j)}}(\mathbf{x}^{(\mathbf{b}_j)}, \mathbf{y}) \\ \mathbf{x}^{(\mathbf{b}_j)} & \text{else} \end{cases}$$

  end for
  set $\boldsymbol{X}_{k+1} = \mathbf{x}$
end for

Ensemble Update Consider now a partition of the ensemble of M particles via

$$\{1, \dots, M\} = \dot{\bigcup}_{i=1}^{L} \mathbf{b}_i. \tag{19}$$

We define transition kernels $\mathbf{P}^{(\mathbf{b}_j)} \colon \mathbb{R}^{Md} \times \mathcal{B}(\mathbb{R}^{Md}) \to [0, 1]$, which only update particles corresponding to the block $\mathbf{b}_j$, via

$$\mathbf{P}^{(\mathbf{b}_j)}(\mathbf{x}, \mathrm{d}\mathbf{y}) = \bigotimes_{k \notin \mathbf{b}_j} \delta_{x^{(k)}}(\mathrm{d}y^{(k)}) \otimes \mathbf{P}_{\mathbf{x}^{-(\mathbf{b}_j)}}(\mathbf{x}^{(\mathbf{b}_j)}, \mathrm{d}\mathbf{y}^{(\mathbf{b}_j)}).$$

By construction each $\mathbf{P}^{(\mathbf{b}_j)}$, $j = 1, \dots, L$, is invariant w.r.t. product target $\boldsymbol{\pi} = \Pi_{i=1}^{M} \pi$.

The sequential block-wise Metropolization of the interacting particle system (14) is defined as the sequential application of the transition kernels $\mathbf{P}^{(\mathbf{b}_j)}$, i.e.,

$$\mathbf{P}(\mathbf{x}, \mathrm{d}\mathbf{y}) := \mathbf{P}^{(\mathbf{b}_1)} \cdots \mathbf{P}^{(\mathbf{b}_L)}(\mathbf{x}, \mathrm{d}\mathbf{y}). \tag{20}$$

An algorithmic description of $\mathbf{P}$ in (20) is given in Algorithm 2. Moreover, in Fig. 1, we provide a graphical illustration of the proposed method. We point out that $\mathbf{P}$ is $\boldsymbol{\pi}$-invariant since all $\mathbf{P}^{(\mathbf{b}_j)}$ are $\boldsymbol{\pi}$-invariant, but in general $\mathbf{P}$ is not $\boldsymbol{\pi}$-reversible (cf. deterministic scan Gibbs sampling).

Remark 2.5 The definition of $\mathbf{P}$ can be modified to a palindromic scan through the blocks, i.e.,

$$\mathbf{P}^{(\mathbf{b}_1)} \cdots \mathbf{P}^{(\mathbf{b}_L)} \mathbf{P}^{(\mathbf{b}_L)} \cdots \mathbf{P}^{(\mathbf{b}_1)}(\mathbf{x}, \mathrm{d}\mathbf{y})$$

which then yields $\boldsymbol{\pi}$-reversibility. Further common approaches to obtain a reversible modification of $\mathbf{P}$ are selecting randomly which of the (fixed) blocks $\mathbf{b}_1, \dots, \mathbf{b}_L$ is to be updated in each step (cf. random scan Gibbs sampling)

$$\sum_{j=1}^{L} \frac{1}{L} \mathbf{P}^{(\mathbf{b}_j)}(\mathbf{x}, \mathrm{d}\mathbf{y})$$

or updating all blocks in a random order (cf. random sweep Gibbs sampling)

$$\sum_{\psi \in \Psi_M} \frac{1}{M!} \mathbf{P}^{(\mathbf{b}_{\psi(1)})} \cdots \mathbf{P}^{(\mathbf{b}_{\psi(M)})}(\mathbf{x}, \mathrm{d}\mathbf{y})$$

where Ψ_M denotes the set of all $M!$ permutations $\psi : \{1, \dots, M\} \to \{1, \dots, M\}$.

Moreover, one could also think about generalizing the construction by allowing for random blocks $\mathbf{b}$, i.e., updating each time a random selection of $|\mathbf{b}|$ particles, where also the number $|\mathbf{b}| \in \{1, \dots, M\}$ could be drawn randomly in each step. However, we focus in the following on deterministic blocks and the deterministic scan block updating.

2.1.2 Convergence and ergodicity

Algorithm 2 and the associated transition kernel $\mathbf{P}$ correspond to a Metropolis-within-Gibbs algorithm where the blocks of particles $\mathbf{x}^{(\mathbf{b})}$ are the "coordinates". In order to show convergence we can leverage on a classical result of Roberts and Rosenthal (2006) about the ergodicity of Metropolis-within-Gibbs algorithms:

Theorem 2.6 (Cf. Roberts and Rosenthal (2006)). *Let the $\boldsymbol{\pi}$-invariant Markov chain $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ be generated by Algorithm 2. Assume that for any $\mathbf{b} \subseteq \{1, \dots, M\}$ the Lebesgue density of $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}$ satisfies $q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y}) > 0$ for all $\mathbf{z}, \mathbf{y} \in \mathbb{R}^{|\mathbf{b}|d}$ and all $\mathbf{x}^{-(\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$. If*

$$\lim_{n\to\infty} \mathbb{P}\left(\forall i = 1, \dots, M \; \exists k \in \{1, \dots, n\} \text{ s.t. } X_k^{(i)} \neq X_0^{(i)} \,\middle|\, \boldsymbol{X}_0 = \mathbf{x}\right) = 1 \tag{21}$$

for all $\mathbf{x} \in \mathbb{R}^{Md}$, then $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ is ergodic and satisfies a law of large numbers

$$\mathbf{S}_N^M(F) := \frac{1}{NM} \sum_{k=1}^{N} \sum_{i=1}^{M} F\left(X_k^{(i)}\right)$$

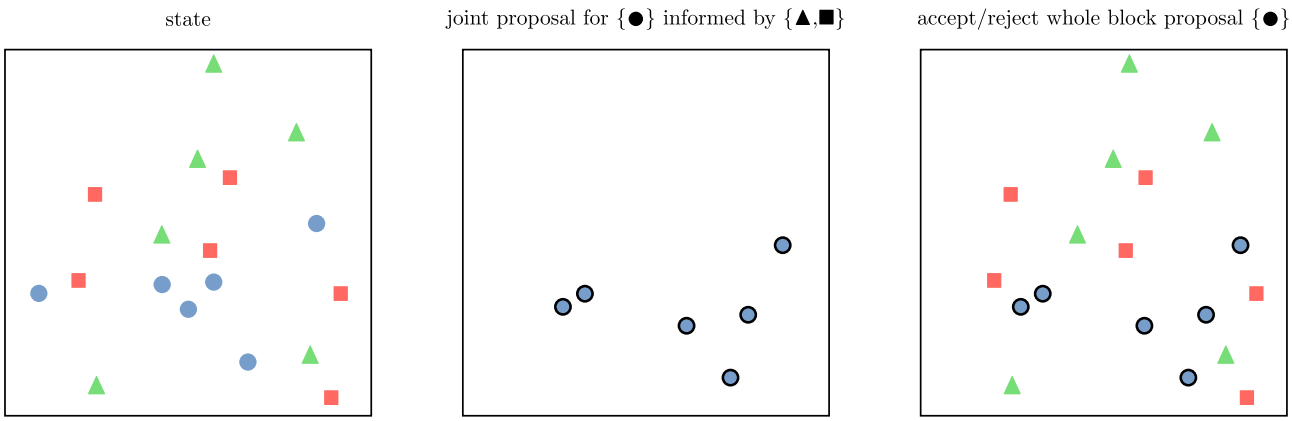


Fig. 1 Illustration of Algorithm 2: For the block of particles {●}, a joint proposal informed by the remaining blocks {▲,■} is computed. The joint proposal is then either accepted or rejected as a whole. This is sequentially repeated for each block

$$\xrightarrow[N\to\infty]{a.\,s.} \mathbb{E}_\pi[F] \qquad \forall F \in L^1_\pi(\mathbb{R}). \tag{22}$$

The additional assumption (21) for ergodicity in Theorem 2.6 is rather mild. It states that for any initial ensemble $\boldsymbol{X}_0 = \mathbf{x}$, the probability that each particle is updated at least once during n steps tends to 1 as $n \to \infty$. In practice, this should be satisfied in any reasonable setting.

Remark 2.7 (Ensemble-wise Metropolization). For the special case $B = |\mathbf{b}| = M$, i.e., proposing and accepting/rejecting the ensemble as a whole the resulting transition kernel $\mathbf{P} = \mathbf{P}^{(\mathbf{b})}$ is a standard Metropolis–Hastings kernel. Therefore, we can apply standard results and obtain the statement of Theorem 2.6 without the technical assumption (21).

We note that the particles $(X_k^{(i)})_{i=1}^M$ within the ensemble $\boldsymbol{X}_k$ become iid π distributed as $k \to \infty$, since the limit distribution $\boldsymbol{\pi}$ is of product form. However, $X_k^{(i)}$ and $X_{k+1}^{(j)}$ do in general not become uncorrelated for $i \neq j$ as $k \to \infty$, which is due to the chains interacting:

Example 2.8 Let $\pi = \mathrm{N}(0, 1)$, $M = 2$ and consider the $\boldsymbol{\pi} = \pi \otimes \pi$-invariant transition kernel

$$\mathbf{P}(\mathbf{x}, \cdot) = \mathrm{N}\left(A\mathbf{x}, \frac{1}{2} I_2\right), \qquad A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Then for the Markov chain $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ generated with transition kernel $\mathbf{P}$ holds for all $k \in \mathbb{N}$ that $\boldsymbol{X}_k \sim \boldsymbol{\pi}$ implies $\boldsymbol{X}_{k+1} \sim \boldsymbol{\pi}$, but $\mathrm{Corr}(X_k^{(1)}, X_{k+1}^{(2)}) = \mathrm{Corr}(X_k^{(2)}, X_{k+1}^{(1)}) = \frac{1}{2}$.

2.2 Simultaneous within block metropolization

We now discuss another sequential block Metropolization, for which all particles within the block are accepted or rejected individually. This allows for a full parallelization within the blocks and might improve the mixing of the ensemble Markov chain. However, this variant requires a structural assumptions on the ensemble dynamics:

Definition 1 The particle dynamics (14) satisfy **conditional within-block-independence (CWBI)** with respect to a partition of the M particles into L non-empty, disjoint blocks $\mathbf{b}_1, \ldots, \mathbf{b}_L$ if for any ensemble state $\mathbf{x} \in \mathbb{R}^{Md}$ and any chosen block $\mathbf{b} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_L\}$ we have for any $i \in \mathbf{b}$ with a slight abuse of notation that

$$\Phi_i\left(x^{(i)}, \mathbf{x}^{-(i)}\right) = \Phi_i\left(x^{(i)}, \mathbf{x}^{-(\mathbf{b})}\right),$$

and

$$\Sigma_{i,i}\left(x^{(i)}, x^{(i)}, \mathbf{x}^{-(i)}\right) = \Sigma_{i,i}\left(x^{(i)}, \mathbf{x}^{-(\mathbf{b})}\right)$$

as well as for any $j \neq i, i, j \in \mathbf{b}$ that

$$\Sigma_{i,j}\left(x^{(i)}, x^{(j)}, \mathbf{x}^{-(i,j)}\right) = 0.$$

Thus, conditioned on the particles $\mathbf{x}^{-(\mathbf{b})}$ there is no interaction between the particles $x^{(i)}$, $i \in \mathbf{b}$, within the same block $\mathbf{b}$.

Remark 2.9 For many common interacting particle systems the CWBI condition is not satisfied per se. However, one can often modify the interacting particle dynamics for a given block partition $\{\mathbf{b}_1, \ldots, \mathbf{b}_L\}$ accordingly to satisfy CWBI. We comment on these kinds of modifications in the next chapter for the three popular interacting particle systems discussed there.

2.2.1 Algorithm

Block proposal and transition kernel Fix again a block $\mathbf{b} = \{i_1, \dots, i_{|\mathbf{b}|}\} \subseteq \{1, \dots, M\}$. Under the CWBI condition the proposal kernel $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}$ takes the special product form (cf. Remark 2.4)

$$\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathrm{d}\mathbf{y}) = Q_{\mathbf{x}^{-(\mathbf{b})},i_1}(z^{(1)}, \mathrm{d}y^{(1)}) \otimes \cdots \otimes Q_{\mathbf{x}^{-(\mathbf{b})},i_{|\mathbf{b}|}}(z^{(|\mathbf{b}|)}, \mathrm{d}y^{(|\mathbf{b}|)}), \tag{23}$$

where for $i \in \mathbf{b}$, $z \in \mathbb{R}^d$, and $\mathbf{x}^{(-\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$

$$Q_{\mathbf{x}^{-b},i}(z, \cdot) := \mathrm{N}\left(z + h\,\Phi_i(z, \mathbf{x}^{(-\mathbf{b})}),\, h\,\Sigma_{i,i}(z, \mathbf{x}^{(-\mathbf{b})})\right). \tag{24}$$

This product structure allows to define particle-wise acceptance probabilities for $i \in \mathbf{b}$

$$\alpha_{\mathbf{x}^{-(\mathbf{b})},i}(z, y) := \begin{cases} \min\left(1, \frac{\pi(y)\, q_{\mathbf{x}^{-(\mathbf{b})},i}(y,z)}{\pi(z)\, q_{\mathbf{x}^{-(\mathbf{b})},i}(z,y)}\right) & \text{if } \pi(z)\, q_{\mathbf{x}^{-(\mathbf{b})},i}(z, y) > 0 \\ 1 & \text{else} \end{cases} \tag{25}$$

where

$$q_{\mathbf{x}^{-(\mathbf{b})},i}(z, y) = \frac{\exp\left(-\frac{1}{2h}\left\|\boldsymbol{\Sigma}_{i,i}(z, \mathbf{x}^{-(\mathbf{b})})^{-1/2}\left(y - z - h\Phi_i(z, \mathbf{x}^{-(\mathbf{b})})\right)\right\|^2\right)}{\det\left(2\pi\, h\, \boldsymbol{\Sigma}_{i,i}(z, \mathbf{x}^{-(\mathbf{b})})\right)^{1/2}}.$$

denotes the Lebesgue density of the kernel $Q_{\mathbf{x}^{-b},i}$ in (24) which is well-defined under Assumption 2.1.

We then define particle-wise transition kernels $P_{\mathbf{x}^{-(\mathbf{b})},i} : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$, for the particles in block $\mathbf{b}$

$$P_{\mathbf{x}^{-(\mathbf{b})},i}(z, \mathrm{d}y) := \alpha_{\mathbf{x}^{-(\mathbf{b})},i}(z, y)\, Q_{\mathbf{x}^{-(\mathbf{b})},i}(z, \mathrm{d}y) + r_{\mathbf{x}^{-(\mathbf{b})},i}(z)\delta_z(\mathrm{d}y),$$

with $r_{\mathbf{x}^{-(\mathbf{b})},i}(z) := 1 - \int_{\mathbb{R}^{|\mathbf{b}|d}} \alpha_{\mathbf{x}^{-(\mathbf{b})},i}(z, y)\, Q_{\mathbf{x}^{-(\mathbf{b})},i}(z, \mathrm{d}y)$ and use them to define the following simultaneous-within-block transition kernel $\widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})}} : \mathbb{R}^{|\mathbf{b}|d} \times \mathcal{B}(\mathbb{R}^{|\mathbf{b}|}d) \to [0, 1]$, $\mathbf{x}^{-(\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$

$$\widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathrm{d}\mathbf{y}) := P_{\mathbf{x}^{-(\mathbf{b})},i_1}(z^{(1)}, \mathrm{d}y^{(1)}) \otimes \cdots \otimes P_{\mathbf{x}^{-(\mathbf{b})},i_{|\mathbf{b}|}}(z^{(|\mathbf{b}|)}, \mathrm{d}y^{(|b|)}).$$

The transition kernel $\widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})}}$ is in fact reversible w.r.t. $\boldsymbol{\pi} = \Pi_{i=1}^{|\mathbf{b}|}\pi$ since for any $A_1, \dots, A_{|\mathbf{b}|}, B_1, \dots, B_{|\mathbf{b}|} \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\begin{aligned}
&\int_{A_1 \times \cdots \times A_{|\mathbf{b}|}} \widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, B_1 \times \cdots \times B_{|\mathbf{b}|})\, \boldsymbol{\pi}(\mathrm{d}\mathbf{z}) \\
&= \prod_{j=1}^{|\mathbf{b}|} \int_{A_j} \widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})},i_j}(z^{(j)}, B_j)\, \boldsymbol{\pi}(\mathrm{d}z^{(j)}) \\
&= \prod_{j=1}^{|\mathbf{b}|} \int_{B_j} \widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})},i_j}(z^{(j)}, A_j)\, \boldsymbol{\pi}(\mathrm{d}z^{(j)}) \\
&= \int_{B_1 \times \cdots \times B_{|\mathbf{b}|}} \widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, A_1 \times \cdots \times A_{|\mathbf{b}|})\, \boldsymbol{\pi}(\mathrm{d}\mathbf{z}).
\end{aligned}$$

Ensemble update As for the previous block-wise Metropolization we now apply the block-wise transition kernels sequentially over all blocks and obtain the simultaneous-within-block transition kernel

$$\widetilde{\mathbf{P}}(\mathbf{x}, \mathrm{d}\mathbf{y}) := \widetilde{\mathbf{P}}^{(\mathbf{b}_1)} \cdots \widetilde{\mathbf{P}}^{(\mathbf{b}_L)}(\mathbf{x}, \mathrm{d}\mathbf{y}) \tag{26}$$

where analogously to above

$$\widetilde{\mathbf{P}}^{(\mathbf{b}_j)}(\mathbf{x}, \mathrm{d}\mathbf{y}) = \bigotimes_{k \notin \mathbf{b}_j} \delta_{x^{(k)}}(\mathrm{d}y^{(k)}) \otimes \widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b}_j)}}(\mathbf{x}^{(\mathbf{b}_j)}, \mathrm{d}\mathbf{y}^{(\mathbf{b}_j)}).$$

Again, $\widetilde{\mathbf{P}}$ is $\boldsymbol{\pi}$-invariant but in general not $\boldsymbol{\pi}$-reversible. The same approaches of random scan and random sweep can be applied to derive reversible modifications of $\widetilde{\mathbf{P}}$. An algorithmic description of $\widetilde{\mathbf{P}}$ is given in Algorithm 3, and a graphical illustration is provided in Fig. 2. We emphasize that $\widetilde{\mathbf{P}}$ is block-wise sequential but within the blocks it performs an *simultaneous particle-wise Metropolization*. We comment on the possibility of a *simultaneous block-wise Metropolization* in Remark 2.11 and Sect. B.

2.2.2 Convergence and ergodicity

Since the transition kernels $\widetilde{\mathbf{P}}_{\mathbf{x}^{-(\mathbf{b})}}$ are not of MH form but products of particle-wise MH kernels we can not directly apply Theorem 2.6. However, the main idea for the ergodicity proof of Metropolis-within-Gibbs algorithms in Roberts and Rosenthal (2006) can be carried over to $\widetilde{\mathbf{P}}$.

Theorem 2.10 *Let the $\boldsymbol{\pi}$-invariant Markov chain $(X_k)_{k\in\mathbb{N}}$ be generated by Algorithm 3. Assume that for any $\mathbf{b} \subseteq \{1, \dots, M\}$ the Lebesgue density of $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}$ satisfies $q_{\mathbf{x}^{-(\mathbf{b})}}(\mathbf{z}, \mathbf{y}) > 0$ for all $\mathbf{z}, \mathbf{y} \in \mathbb{R}^{|\mathbf{b}|d}$ and all $\mathbf{x}^{-(\mathbf{b})} \in \mathbb{R}^{(M-|\mathbf{b}|)d}$. If*

$$\lim_{n\to\infty} \mathbb{P}\left(\exists k \in \{1, \dots, n\}\ s.t.\ \forall i = 1, \dots, M\ holds\ X_k^{(i)} \neq X_{k-1}^{(i)} \,\middle|\, X_0 = \mathbf{x}\right) = 1 \tag{27}$$

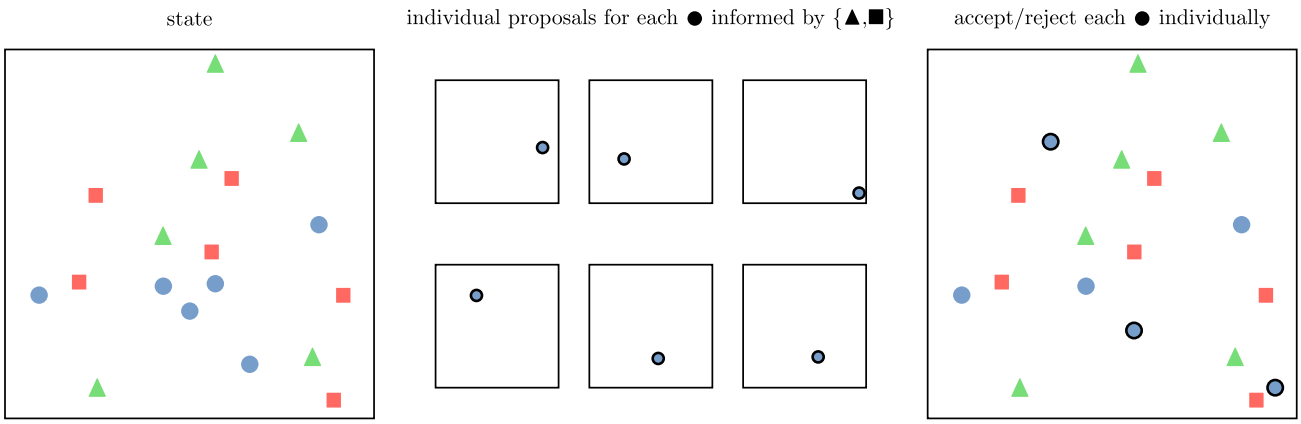


Fig. 2 Illustration of Algorithm 3: For each particle ● in the block {●}, an individual proposal informed by the remaining blocks {▲,■} is computed. The proposals are independently accepted or rejected in parallel. This is sequentially repeated for each block

Algorithm 3 Simultaneous-within-block Metropolized interactive particle sampling

Input:

- target density π on $\mathbb{R}^d$
- a partition $\dot{\bigcup}_{j=1}^{L} \mathbf{b}_j = \{1, \dots, M\}$
- ensemble dependent block proposal kernels $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b}_j)}}$ of form (23) with density $\mathbf{q}_{\mathbf{x}^{-(\mathbf{b}_j)}} : \mathbb{R}^{|\mathbf{b}_j|d} \times \mathbb{R}^{|\mathbf{b}_j|d} \to (0, \infty)$ for arbitrary $\mathbf{x}^{-(\mathbf{b}_j)} \in \mathbb{R}^{(M-|\mathbf{b}_j|)d}$ and $j \in \{1, \dots, L\}$
- initial probability distribution π_0 on $\mathbb{R}^d$

Output: ensemble Markov chain $(\boldsymbol{X}_k)_{k\in\{1,\dots,N\}}$ in state space $\mathbb{R}^{Md}$

draw $\mathbf{x}_0 \sim \bigotimes_{i=1}^{M} \pi_0$ and set initial state $\boldsymbol{X}_0 = \mathbf{x}_0 \in \mathbb{R}^{Md}$
for $k = 0, \dots, N$ **do**
  given $\boldsymbol{X}_k = \mathbf{x}_k$ initialize $\mathbf{x} = (x_k^{(1)}, \dots, x_k^{(M)})$
  for $j = 1, \dots, L$ **do**
    $\forall i \in \mathbf{b}_j$: draw proposal $y^{(i)} \sim Q_{\mathbf{x}^{-(\mathbf{b}_j)},i}(x^{(i)}, \cdot)$ independently
    $\forall i \in \mathbf{b}_j$: compute particle acceptance probability $\alpha_{\mathbf{x}^{-(\mathbf{b}_j)},i}(x^{(i)}, y^{(i)}) \in [0, 1]$
    $\forall i \in \mathbf{b}_j$: draw $u_i \sim \mathrm{U}([0, 1])$ independently and set

$$x^{(i)} = \begin{cases} y^{(i)} & \text{if } u_i \le \alpha_{\mathbf{x}^{-(\mathbf{b}_j)},i}(x^{(i)}, y^{(i)}) \\ x^{(i)} & \text{else} \end{cases}$$

  end for
  set $\boldsymbol{X}_{k+1} = \mathbf{x}$
end for

for all $\mathbf{x} \in \mathbb{R}^{Md}$*, then* $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ *is ergodic and satisfies a law of large numbers* (22).

Proof Ergodicity and a strong law or large numbers hold for any inital distribution if the Markov chain $(\boldsymbol{X}_k)_{k\in\mathbb{N}}$ is aperiodic, irreducible and Harris recurrent, see Robert and Casella (2004), Section 6 for statements and definitions. The first two are easy to verify for $\widetilde{\mathbf{P}}$ and the third is, given irreducibility, equivalent to

$$\mathbb{P}(\boldsymbol{X}_n \in A \text{ for all } n \in \mathbb{N} \mid \boldsymbol{X}_0 = \mathbf{x}) = 0 \quad \forall \mathbf{x} \in \mathbb{R}^{Md}\ \forall A \in \mathcal{B}(\mathbb{R}^{Md}) \colon \boldsymbol{\pi}(A) = 0,$$

see Roberts and Rosenthal (2006), Theorem 6. Since π and thus $\boldsymbol{\pi}$ are absolutely continuous w.r.t. the Lebesgue measure and $q_{\mathbf{x}^{-(\mathbf{b})}} > 0$ we have for any $\mathbf{x} \in \mathbb{R}^{Md}$ and $A \in \mathcal{B}(\mathbb{R}^{Md})$ with $\boldsymbol{\pi}(A) = 0$ that

$$\mathbb{P}\left(\boldsymbol{X}_k \in A \mid \boldsymbol{X}_k^{(i)} \neq \boldsymbol{X}_{k-1}^{(i)}\ \forall i,\ \boldsymbol{X}_0 = \mathbf{x}\right) = 0.$$

Thus, by introducing the event $D_n := \{\forall k \in \{1, \dots, n\}\ \exists i \in \{1, \dots, M\} \colon \boldsymbol{X}_k^{(i)} = \boldsymbol{X}_{k-1}^{(i)}\}$, i.e., in each iteration $k = 1, \dots, n$ there is at least on particle which does not move, we have by law of total probability

$$\mathbb{P}(\boldsymbol{X}_n \in A \mid \boldsymbol{X}_0 = \mathbf{x}) \le \mathbb{P}(D_n \mid \boldsymbol{X}_0 = \mathbf{x}).$$

Thus, since

$$\begin{aligned} \mathbb{P}(\boldsymbol{X}_n \in A\ \forall n \in \mathbb{N} \mid \boldsymbol{X}_0 = \mathbf{x}) &\le \lim_{n\to\infty} \mathbb{P}(\boldsymbol{X}_n \in A \mid \boldsymbol{X}_0 = \mathbf{x}) \\ &\le \lim_{n\to\infty} \mathbb{P}(D_n \mid \boldsymbol{X}_0 = \mathbf{x}), \end{aligned}$$

we have that (27) ensures Harris recurrence. □

Again, the additional assumption (27) for ergodicity can be considered rather mild. Here it states that for any initial ensemble $\boldsymbol{X}_0 = \mathbf{x}$, the probability that all particles are updated during one step with $\widetilde{\mathbf{P}}$ tends to 1 as $n \to \infty$. Again, this should be satisfied in any reasonable setting. However, note that (27) is a stronger assumption than (21) for the ergodicity of block-wise Metropolization in Theorem 2.6.

Remark 2.11 (Simultaneous block-wise Metropolization). For interacting particle systems described by (14) it seems unintuitive to update the blocks in the ensemble sequentially, i.e., to use a different drift and diffusion term for each of the blocks due to interaction and the sequentially updated ensemble. Moreover, from a computational perspective, a *simultaneous* (rather than sequential) block-wise

Metropolization would be most desirable as it allows for the parallel processing of all particles within the ensemble: this method, in each step of the algorithm, decides for each block in the ensemble independently, whether the proposal for this block is accepted or rejected. Formally, this is described by the ensemble transition operator

$$\mathbf{P}_{\text{sim}}(\mathbf{x}, \mathrm{d}\mathbf{y}) = \mathbf{P}_{\mathbf{x}^{-(\mathbf{b}_1)}}(\mathbf{x}^{(\mathbf{b}_1)}, \mathrm{d}\mathbf{y}^{(\mathbf{b}_1)}) \otimes \cdots \otimes \mathbf{P}_{\mathbf{x}^{-(\mathbf{b}_L)}}(\mathbf{x}^{(\mathbf{b}_L)}, \mathrm{d}\mathbf{y}^{(\mathbf{b}_L)}).$$

Also $\widetilde{\mathbf{P}}_{\text{sim}}$ can be defined analogously. However, it turns out that this approach in general fails to ensure ergodicity and even $\boldsymbol{\pi}$-invariance. A detailed discussion including counterexamples for the case where all blocks have size one is provided in Sect. B.

2.3 Asymptotic variance

2.3.1 General case

Let $(\tilde{\boldsymbol{X}}_k)_{k\in\mathbb{N}}$ be a $\boldsymbol{\pi}$-invariant ensemble Markov chain starting in stationarity $\tilde{\boldsymbol{X}}_0 \sim \boldsymbol{\pi}$, and let $F \in L^2_\pi(\mathbf{R})$. Then the asymptotic variance of the path-average estimator of an ensemble Markov $\mathbf{S}^M_N(F)$, see (22), is defined by

$$\boldsymbol{\sigma}^2_F := \lim_{N\to\infty} N\,\mathbb{V}\left[\mathbf{S}^M_N(F)\right],$$

and, in case of existence, also given by

$$\boldsymbol{\sigma}^2_F = \frac{\mathbb{V}_\pi[F]}{M}\left[1 + 2\sum_{k=1}^{\infty}\mathrm{Corr}\left(\frac{1}{M}\sum_{i=1}^{M}F\left(\tilde{X}^{(i)}_0\right), \frac{1}{M}\sum_{j=1}^{M}F\left(\tilde{X}^{(j)}_k\right)\right)\right]. \tag{28}$$

This follows from $\mathbf{S}^M_N(F) = \frac{1}{N}\sum_{k=1}^N G(\tilde{\boldsymbol{X}}_k)$ where $G(\mathbf{x}) := \frac{1}{M}\sum_{j=1}^M F(x^{(i)})$ with $\mathbb{V}_{\boldsymbol{\pi}}[G] = \frac{1}{M}M\mathbb{V}_\pi[F]$. By standard MCMC arguments, we can further obtain a central limit theorem for $\mathbf{S}^M_N(F)$. For the case of reversible ensemble transition kernels $\mathbf{P}$, i.e., for ensemble-wise Metropolization or the variants explained in Remark 2.5, we have given the assumptions of Theorem 2.6 and that $\boldsymbol{\sigma}^2_F$ in (28) is finite that

$$\sqrt{N}\left(\mathbf{S}^M_N(F) - \mathbb{E}_\pi[F]\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathrm{N}\left(0, \boldsymbol{\sigma}^2_F\right). \tag{29}$$

The same holds for reversible $\widetilde{\mathbf{P}}$ under the assumptions of Theorem 2.10 and $\boldsymbol{\sigma}^2_F < \infty$.

Moreover, also for the non-reversible transition kernels $\mathbf{P}_{\text{seq}}$ and $\widetilde{\mathbf{P}}_{\text{seq}}$ a central limit theorem (29) with $\boldsymbol{\sigma}^2_F$ as in (28) can be derived given additional sufficient conditions. For the latter we refer again to (Roberts and Rosenthal (2004), Section 5) for more details.

2.3.2 Comparison to non-interacting particles

Let us consider the special case of non-interacting particles, i.e., we assume for the moment that $(\tilde{X}^{(i)}_k)_{k\in\mathbb{N}}$ are mutually independent π-reversible Markov chains for all $i = 1, \ldots, M$ starting in stationarity $\tilde{X}^{(i)}_0 \sim \pi$ iid. Then the estimator

$$\mathbf{S}^M_N(F) = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{N}\sum_{k=1}^{N}F(X^{(i)}_k)\right)$$

is simply an average of M iid path average estimators $\frac{1}{N}\sum_{k=1}^N F(X^{(i)}_k)$, $i = 1, \ldots, M$, of the independent particle Markov chains $(X^{(i)}_k)_{k\in\mathbb{N}}$ and, thus, its asymptotic variance $\boldsymbol{\sigma}^2_F$ can be written as

$$\boldsymbol{\sigma}^2_F = \frac{\mathbb{V}_\pi[F]}{M}\left[1 + 2\sum_{k=1}^{\infty}\mathrm{Corr}\left(F\left(X^{(i)}_0\right), F\left(X^{(i)}_k\right)\right)\right]$$

with arbitrary $i \in \{1, \ldots, M\}$. Thus, for non-interacting particles and sufficiently large N we have

$$\mathbb{V}(S_{MN}(F)) \approx \mathbb{V}(\mathbf{S}^M_N(F))$$

where $S_{MN}(F)$ is the estimator based on MN iterates of a *one-particle* Markov chain $(\tilde{X}^{(i)}_k)_{k\in\mathbb{N}}$, as already mentioned in Goodman and Weare (2010). Thus, sampling based on non-interacting particle systems has the *only* advantage of parallel computation.

In our numerical experiments, see Sect. 4, we observe that the interaction of the M particles can give improvements in the form

$$\mathbb{V}(S_{MN}(F)) > \mathbb{V}(\mathbf{S}^M_N(F))$$

due to allowing each particle chain $(X^{(i)}_k)_{k\in\mathbb{N}}$ to take larger steps resulting from exploiting approximate information on π provided by the ensemble in the proposal kernels $\mathbf{Q}$. However, a rigorous proof of this statement is beyond the scope of this work.

3 Algorithmic examples

In this section we discuss several interacting particle methods which can be used to build proposals for the Metropolization schemes presented in Sect. 2. They are inspired from continuous-time systems (3), which we translate into a time-discrete interacting particle system through ensemble and

time discretization. For each example we check if Assumption 2.1 is satisfied and comment on modifications for the CWBI condition.

3.1 Parallel Metropolis-adjusted Langevin algorithm (pMALA)

For comparison we start with a *non-interacting* system, namely parallel MALA. This will serve to illustrate the benefits of interaction of particles in our numerical examples.

Ensemble update and proposal density Recall the particle-wise update scheme from Example 1.4 arising from discretization of the Langevin dynamics introduced in Example 1.1 and Example 1.4, i.e.

$$X_{k+1}^{(i)} = X_k^{(i)} + h\nabla \log \pi(X_k^{(i)}) + \sqrt{2h}\zeta_{k+1}^{(i)}, \quad k \in \mathbb{N},\ j = 1, \dots, M, \tag{30}$$

with iid $\zeta_{k+1}^{(i)} \sim \mathrm{N}(0, \mathrm{Id}_d)$. This system satisfies CWBI for any block partition due to the non-interaction of the particles and for any block $\mathbf{b}$ the corresponding proposal kernel $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}$ has the product form (23) with ($\mathbf{x}^{-(\mathbf{b})}$-independent) Lebesgue density

$$q_{\mathbf{x}^{-(\mathbf{b})},i}(x, y) = q(x, y) = \frac{1}{(4h\pi)^{d/2}} \exp\left(-\frac{1}{4h}\| (y - (x + h\nabla \log \pi(x))) \|^2\right), \qquad x, y \in \mathbb{R}^d,$$

i.e., $\Phi_i(x, \mathbf{x}^{-(i)}) = \nabla \log \pi(x)$ and $\Sigma_{i,i}(x, x, \mathbf{x}^{-(i)}) = 2\mathrm{Id}_d$ in (17). Hence, besides CWBI also Assumption 2.1 is satisfied.

3.2 Metropolis-adjusted ALDI (MA-ALDI)

We continue with the ALDI discretization introduced in Example 1.3 for the continuous Wasserstein dynamics described by (5).

Ensemble update and proposal density In order to avoid the subspace property, see Iglesias et al. (2013), we introduce a sightly modified version of the update scheme (10):

$$\begin{aligned} X_{k+1}^{(i)} = X_k^{(i)} &+ h(\gamma \mathrm{Id}_d + (1-\gamma)C(\boldsymbol{X}_k))\nabla \log \pi(X_k^{(i)}) \\ &+ h(1-\gamma)\frac{d+1}{M}(X_k^{(i)} - m(\boldsymbol{X}_k)) \\ &+ \sqrt{2h(\gamma \mathrm{Id}_d + (1-\gamma)C(\boldsymbol{X}_k))}\zeta_{k+1}^{(i)}. \end{aligned} \tag{31}$$

Here, $\zeta_{k+1}^{(i)} \sim \mathrm{N}(0, \mathrm{Id}_d)$ iid for all $i = 1, \dots, M$ and $k \geq 0$, and $\gamma \in [0, 1]$ is a fixed parameter that determines the contribution of the covariance matrix $C(\boldsymbol{X}_k)$ in the update.

When $\gamma = 0$, the update scheme (31) reduces to the ALDI method (10), while $\gamma = 1$ corresponds to the pMALA method discussed in Example 1.4 or the previous subsection. By choosing $\gamma \in [0, 1]$, we enable a smooth transition from pMALA to MA-ALDI. Additionally, the term $\gamma\mathrm{Id}$ in $\Sigma_\gamma(x^{(i)}, \mathbf{x}^{-(i)}) := 2(\gamma\mathrm{Id} + (1-\gamma)C(\mathbf{x}))$ serves as regularization and guarantees the positive-definiteness of $\Sigma_\gamma(x^{(i)}, \mathbf{x}^{-(i)}) \in \mathcal{C}^+(\mathbb{R}^d)$ for all $\mathbf{x} \in \mathbb{R}^{Md}$ when $\gamma > 0$, which breaks the well-known subspace property (Iglesias et al. 2013). In other words, this regularization acts as a stabilization of possible degeneration of the particle system also known as the ensemble collapse. Moreover, this interacting particle system satisfies the assumptions in Remark 2.4 with $\Sigma = \Sigma_\gamma$ from above and $\Phi = \Phi_\gamma$ where

$$\begin{aligned} \Phi_\gamma(x^{(i)}, \mathbf{x}^{-(i)}) = (\gamma\mathrm{Id} &+ (1-\gamma)C(\mathbf{x}))\nabla \log \pi(x^{(i)}) \\ &+ h(1-\gamma)\frac{d+1}{M}(x^{(i)} - m(\mathbf{x})). \end{aligned}$$

I.e., the block proposal kernel takes the a product form (18) where the proposal density $q_{\mathbf{x}^{-(i)}}$ of $Q_{\mathbf{x}^{-(i)}}$ in (18) is given by

$$\begin{aligned} &q_{\mathbf{x}^{-(i)}}(z, y) \\ &= \frac{1}{\det(\sqrt{2\pi\, h\, \Sigma_\gamma(z, \mathbf{x}^{-(i)})})} \\ &\exp\left(-\frac{1}{4h}\| \Sigma_\gamma(z, \mathbf{x}^{-(i)})^{-1/2}\left(y - (z + h\Phi_\gamma(z, \mathbf{x}^{-(i)}))\right)\|^2\right). \end{aligned}$$

The regularization of the covariance is not necessary, i.e., we can allow for $\gamma = 0$, if $C(\boldsymbol{X}_k) \in \mathcal{C}^+(\mathbb{R}^d)$ for all $k \in \mathbb{N}$. The latter necessarily requires having $M \geq d + 1$ particles $X_k^{(i)}$, since $\mathrm{rank}(C(\boldsymbol{X}_k)) \leq M - 1$. Note that $C(\boldsymbol{X}_k) \in \mathcal{C}^+(\mathbb{R}^d)$ implies that $C(\boldsymbol{X}_{k+1}) \in \mathcal{C}^+(\mathbb{R}^d)$ almost surely for (31) with $\gamma = 0$ within Algorithm 3. This holds since the probability of proposing blocks of particles which lie in (a finite union) of strict subspaces of $\mathbb{R}^{|\mathbf{b}|}$ is zero. Thus, we introduce:

Assumption 3.1 *At least one of the following conditions is met: (i) $\gamma > 0$ or (ii) $M > d$ and for the initial ensemble $\mathbf{x}_0 \in \mathbb{R}^{Md}$ it holds that $C(\mathbf{x}_0) \in \mathcal{C}^+(\mathbb{R}^d)$.*

Given Assumption 3.1 the crucial Assumption 2.1 is satisfied for ALDI. Finally, it is worth mentioning that the algorithm can be further adapted to eliminate the need for computing the gradients of $\log \pi$. Additional details on this adjustment are provided in Appendix C.

Simultaneous-within-block implementation In the following, we want to present a heuristic modification of the interacting particle dynamic (31) allowing for a simultaneous-within-block implementation satisfying the product form (18). We consider again a partition $\{1,\dots,M\} = \cup_{j=1}^{L} \mathbf{b}_j$ with $|b_j| \geq 2$. In order to guarantee CWBI condition, i.e., that there is no interaction between the particles within a block, we will exclude these particles from the empirical approximation of the mean and covariances. That is, we define the block-wise empirical means and covariances as

$$C_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k) = \frac{1}{M-|\mathbf{b}_i|}\sum_{i=1,\ i\notin\mathbf{b}_j}^{M}\left(X_k^{(i)} - m_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k)\right)\left(X_k^{(i)} - m_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k)\right)^{\top},$$

$$m_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k) = \frac{1}{M-|\mathbf{b}_j|}\sum_{i=1,\ i\notin\mathbf{b}_j}^{M} X_k^{(i)}.$$

Finally, we replace the update scheme (31) by the block-wise update

$$\begin{aligned} X_{k+1}^{(i)} = X_k^{(i)} &+ h(\gamma \mathrm{Id}_d + (1-\gamma)C_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k))\nabla \log \pi(X_k^{(i)}) \\ &+ h(1-\gamma)\frac{d+1}{M}(X_k^{(i)} - m_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k)) \\ &+ \sqrt{2h(\gamma \mathrm{Id}_d + (1-\gamma)C_{\boldsymbol{X}_k^{-(\mathbf{b}_j)}}(\boldsymbol{X}_k))}\zeta_{k+1}^{(i)}, \\ &i \in \mathbf{b}_j,\ j = 1,\dots,L. \end{aligned} \tag{32}$$

3.3 Metropolis-adjusted consensus based sampling (MA-CBS)

Motivated by the consensus based optimization (CBO) scheme (Pinnau et al. 2017), the authors of Carrillo et al. (2022) propose a modification leading to the so-called consensus based sampling (CBS) method. While CBO aims to find a global minimizer of some objective function $\mathcal{V}: \mathbb{R}^d \to \mathbb{R}_+$, CBS aims to generate approximate samples from measures of the form $\pi(x) \propto \exp(-\mathcal{V}(x))$, $x \in \mathcal{X} = \mathbb{R}^d$. It is worth noting that the algorithm generates samples from a Gaussian approximation of the target measure, i.e. in general π is not an invariant measure of the corresponding SDE (Carrillo et al. 2022).

The theoretical study of CBS in Carrillo et al. (2022) was based on its continuous-time formulation in the mean-field limit represented by the McKean–Vlasov SDE

$$\mathrm{d}X_t = -(X_t - m_\pi(\pi_t))\,\mathrm{d}t + \sqrt{2\lambda^{-1}C_\pi(\pi_t)}\,\mathrm{d}B_t, \quad X_0 \sim \pi_0, \tag{33}$$

i.e., $\phi(x,\rho) = -(x - m_\pi(\rho))$ and $\sigma(x,\rho) = 2\lambda^{-1}C_\pi(\rho)$ in (3). Here B_t is a d-dimensional Brownian motion, π_t denotes the probability density function of the state X_t, $t \geq 0$, and $m_\pi(\rho)$, $C_\pi(\rho)$ denote the weighted mean and covariance of a probability distribution ρ defined as

$$m_\pi(\rho) = \frac{1}{\int_{\mathbb{R}^d}\rho(x)\pi(x)\,\mathrm{d}x}\int_{\mathbb{R}^d} x\,\pi(x)\,\rho(x)\,\mathrm{d}x,$$

$$C_\pi(\rho) = \frac{1}{\int_{\mathbb{R}^d}\rho(x)\pi(x)\,\mathrm{d}x}\int_{\mathbb{R}^d}\left((x - m_\pi(\rho))(x - m_\pi(\rho))^{\top}\right)\pi(x)\,\rho(x)\,\mathrm{d}x.$$

To give a better intuition of the dynamical system one may derive the evolution of the (unweighted) mean and covariance defined in (5a) of the solution to (33), which are given by

$$\partial_t(m(\pi_t)) = -m(\pi_t) + m_\pi(\pi_t)$$

$$\partial_t(C(\pi_t)) = -2C(\pi_t) + 2\frac{1}{\lambda}C_\pi(\pi_t).$$

One important property of CBS is the property that once initialized with a Gaussian distribution, the law π_t remains Gaussian. In this case, finding a stationary distribution of (33) corresponds to finding a Gaussian approximation of the target measure π perserving the mean and covariance. Under certain conditions, such as the convexity of the potential, the existence of a Gaussian steady state can be established (Carrillo et al. (2022), Theorem 2). In the one-dimensional setting, this steady state is arbitrarily close to the Laplace approximation of π_* (Carrillo et al. (2022), Theorem 3).

Ensemble update and proposal density Discretizing (33) by the Euler–Maruyama scheme in time and using an ensemble as empirical approximation for π_t we obtain the following update

$$X_{k+1}^{(i)} = X_k^{(i)} - h(X_k^{(i)} - m_\pi(\boldsymbol{X}_k)) + \sqrt{4hC_\pi(\boldsymbol{X}_k)}\zeta_{k+1}^{(i)},\quad i = 1,\dots,M,\ k \in \mathbb{N},$$

where $\zeta_{k+1}^{(i)} \sim \mathrm{N}(0, \mathrm{Id}_d)$ iid and

$$m_\pi(\mathbf{x}) = \frac{1}{\sum_{i=1}^M \pi(x^{(i)})}\sum_{i=1}^{M}\pi(x^{(i)})x^{(i)},$$

$$C_\pi(\mathbf{x}) = \frac{1}{\sum_{i=1}^M \pi(x^{(i)})}\sum_{i=1}^{M}\pi(x^{(i)})\left((x^{(i)} - m_\pi(\mathbf{x}))(x^{(i)} - m_\pi(\mathbf{x}))^{\top}\right)$$

denote the weighted empirical mean and covariance of the ensemble $\mathbf{x} \in \mathbb{R}^{Md}$. We mention that in the original work (Carrillo et al. (2022)) the authors introduced a rescaled time, so that our system slightly differs from the one in Carrillo et al. (2022). The proposed Metropolization is applicable to both discrete time systems.

Similar to the ALDI method the weighted empirical covariance $C_\pi(\boldsymbol{X}_k)$ is not necessarily positive definite. We can again either use sufficiently many particles $M > d$ or introduce a regularization of the covariance which yields the update

$$X_{k+1}^{(i)} = X_k^{(i)} - h(X_k^{(i)} - m_\pi(\boldsymbol{X}_k)) + \sqrt{4h(\gamma \mathrm{Id} + (1-\gamma)C_\pi(\boldsymbol{X}_k))}\zeta_{k+1}^{(i)}, \tag{34}$$

for $i = 1, \dots, M$, $k \in \mathbb{N}$, with $\gamma \in [0, 1]$ fixed. Also (34) satisfies the assumptions of Remark 2.4 with $\Phi = \Phi_\gamma$ and $\Sigma = \Sigma_\gamma$ where

$$\Phi_\gamma(x^{(i)}, \mathbf{x}^{-(i)}) = x^{(i)} - h(x^{(i)} - m_\pi(\mathbf{x}))$$

and

$$\Sigma_\gamma(x^{(i)}, \mathbf{x}^{-(i)}) = 2(\gamma \mathrm{Id}_d + (1-\gamma)C_\pi(\mathbf{x})),$$

i.e., the particle-wise proposal densities in the product block proposal kernel (18) are

$$q_{\mathbf{x}^{-(i)}}(z, y) = \frac{1}{\det(\sqrt{2\pi\, h\, \Sigma_\gamma(z, \mathbf{x}^{-(i)})})} \exp\left(-\frac{1}{4h}\| \Sigma_\gamma(z, \mathbf{x}^{-(i)})^{-1/2}\left(y - (z + h\Phi_\gamma(z, \mathbf{x}^{-(i)}))\right)\|^2\right).$$

3.4 Metropolis-adjusted Stein variational gradient descent (MA-SVGD)

Stein variational gradient descent (SVGD) is a non-parametric particle-based Bayesian inference method that aims to approximate a target distribution (Liu and Wang (2016)). The algorithm can formally be viewed as a particle and time-discretization of the mean field equation

$$\mathrm{d}X_t = \Phi(X_t, \pi_t)\mathrm{d}t, \quad X_0 \sim \pi_0,$$
$$\Phi(x, \rho) = \int_{\mathbb{R}^d} \left[K(y, x)\nabla \log \pi(y) + \nabla_y K(y, x)\right] \pi_k(\mathrm{d}y) \tag{35}$$

where π_t denotes the distribution of X_t and K is a fixed kernel function (Korba et al. 2020; Liu 2017).

The *deterministic* interacting particle system termed SVGD and originally proposed in Liu and Wang (2016) reads

$$X_{k+1}^{(i)} = X_k^{(i)} + h_{k+1}\Phi(X_k^{(i)}, \boldsymbol{X}_k),$$
$$\Phi(x, \mathbf{x}) = \frac{1}{M}\sum_{j=1}^{M} K(x^{(j)}, x)\nabla \log \pi(x^{(j)}) + \nabla_{x^{(j)}} K(x^{(j)}, x)$$

where $h_k > 0$ denote the step size in the kth step. While this method has shown promising results in certain applications, in practice it requires a careful tuning of the kernel function and convergence in high dimensions tends to be slow. Moreover, convergence results have only been established for the long time behavior of the mean field limit in continuous time, where exponential convergence of the Kullback–Leibler divergence was shown under strong assumptions on the underlying kernel K and the target π (Korba et al. 2020; Duncan et al. 2023). In fact, these results suggest that for many standard kernels K exponential convergence to equilibrium is not possible, or convergence is not guaranteed (also see (Gorham and Mackey (2017), Theorem 6)). This makes Metropolization of a stochastic variant a compelling option to ensure convergence. We also mention that, in the finite-particle regime, the convergence of the kernelized Stein discrepancy (KSD) has recently been established under relaxed assumptions on the kernel (Shi and Mackey 2023). However, the obtained convergence rate remains notably slow.

Ensemble update and proposal density Motivated by the Langevin dynamics, a *stochastic* version of SVGD has recently been proposed in Gallego and Insua (2018). This approach is driven by a $\boldsymbol{\pi}$-invariant coupled system of SDEs (Nüsken and Renger 2023; Gallego and Insua 2018) that describes the time-continuous dynamics of the ensemble $\boldsymbol{X}_t$. Notably, the mean field limit for $M \to \infty$ of this system of SDEs is the same as in (35). We shortly recall the dynamics from Gallego and Insua (2018) and refer to this paper for further explanations and details.

The stochastic particle dynamics reads

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + h\boldsymbol{\Phi}(\boldsymbol{X}_k) + \sqrt{h\boldsymbol{\Sigma}(\boldsymbol{X}_k)}\zeta_{k+1}, \tag{36}$$

where $\zeta_k \sim \mathrm{N}(0, \mathrm{Id}_{Md})$,

$$\boldsymbol{\Phi}(\mathbf{x}) := (\Phi(x^{(1)}, \mathbf{x}^{(-1)}), \dots, \Phi(x^{(M)}, \mathbf{x}^{(-M)}))^\top,$$
$$\Phi(x^{(j)}, \mathbf{x}^{-(j)}) := \frac{1}{M}\sum_{i=1}^{M} K(x^{(j)}, x^{(i)})\nabla \log \pi(x^{(i)}) + \nabla_{x^{(i)}} K(x^{(j)}, x^{(i)}),$$

and

$$\boldsymbol{\Sigma}(\mathbf{x}) = \frac{2}{M}\boldsymbol{S}^\top \operatorname{diag}_d(\mathcal{K}(\mathbf{x}), \dots, \mathcal{K}(\mathbf{x}))\boldsymbol{S}.$$

Here the positive definite matrix $\mathcal{K}(\mathbf{x}) \in \mathcal{C}^+(\mathbb{R}^d)$ is given by

$$\mathcal{K}(\mathbf{x})_{i,j} = K(x^{(i)}, x^{(j)}) \qquad \forall\, i, j = 1, \dots, M, \tag{37}$$

and $\boldsymbol{S} \in \mathbb{R}^{Md\times Md}$ denotes a permutation matrix defined by

$$\boldsymbol{S} = \begin{pmatrix} \mathbf{1}_{1,1} & \dots & \mathbf{1}_{M,1} \\ \vdots & \ddots & \vdots \\ \mathbf{1}_{1,d} & \dots & \mathbf{1}_{M,d} \end{pmatrix}, \qquad \mathbf{1}_{k,\ell} \in \mathbb{R}^{M\times d} \quad \text{with}$$

$$(\mathbf{1}_{k,\ell})_{n,m} = \begin{cases} 1 & \text{if } k = n,\ \ell = m \\ 0 & \text{else.} \end{cases}$$

Note that $\boldsymbol{S}$ is an orthogonal matrix such that $\boldsymbol{S}^\top\boldsymbol{S} = \boldsymbol{S}\boldsymbol{S}^\top = \mathrm{Id}_{Md}$, and it holds that

$$\sqrt{h\boldsymbol{\Sigma}(\mathbf{x})} = \sqrt{2h/M}\boldsymbol{S}^\top \operatorname{diag}_d(\sqrt{\mathcal{K}(\mathbf{x})}, \dots, \sqrt{\mathcal{K}(\mathbf{x})})\boldsymbol{S}.$$

Moreover, letting $\mathbf{K}_{ij} := K(x^{(i)}, x^{(j)})\,\mathrm{Id}_d \in \mathbb{R}^{d\times d}$, we can write

$$\boldsymbol{\Sigma}(\mathbf{x}) = \frac{2}{M}\begin{pmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{M1} & \dots & \mathbf{K}_{MM} \end{pmatrix}.$$

This definition of the diffusion ensures that the associated Fokker–Planck equation of (36), as $h \to 0$, admits $\boldsymbol{\pi}$ as a stationary solution even for the finite particle setting (Gallego and Insua (2018), Proposition 2). This is complementary to SVGD viewed as finite particle approximation of (35) (Gallego and Insua (2018), Proposition 3). It is worthwhile to mention that (36) does not satisfy the assumptions of Remark 2.4. However, the associated block proposal kernels $\mathbf{Q}_{\mathbf{x}^{-(\mathbf{b})}}$ possess positive Lebesgue densities, since $\mathcal{K}(\mathbf{x})$ and, therefore, $\boldsymbol{\Sigma}(\mathbf{x})$ is positive definite for distinct particles $x^{(i)} \neq x^{(j)}$, $i \neq j$. Hence, Assumption 2.1 is given as long as are particles in the ensemble are distinct.

Remark 3.2 Similar to CBS and ALDI, we can improve the stability of the system by introducing regularization in order to avoid the kernel matrix $\mathcal{K}(\mathbf{x})$ from becoming close to singular in case the particles collapse. To do so, one can replace $\boldsymbol{\Sigma}(\mathbf{x})$ by $\boldsymbol{\Sigma}_\gamma(\mathbf{x}) := \gamma\mathrm{Id}_{Md} + (1-\gamma)\boldsymbol{\Sigma}(\mathbf{x})$, $\gamma \in (0, 1)$. However, in our numerical examples such a regularization was not necessary and, hence, not applied.

4 Numerical experiments

To evaluate the performance of our proposed Metropolis-adjusted interactive particle sampling (MA-IPS) methods, we conduct three numerical experiments in which we implement (MA-)ALDI, (MA-)SVGD, and (MA-)CBS. The implementation was done in `MATLAB` and we utilized the discrete dynamics (31), (34), and (36) for the unadjusted algorithms, along with the corresponding Metropolis-adjusted methods based on Algorithms 2 and 3. We adhere to the following naming conventions for our algorithms:

- *ALDI* corresponds to the chain generated by (31), *CBS* corresponds to the chain generated by (34), and *SVGD* corresponds to the chain generated by (36),
- *pMALA* corresponds to parallel MALA, i.e. to M independent chains each generated by Algorithm 4 with proposal (30),
- we use the prefix *MA*, to indicate that it's the Metropolized version of an algorithm,
- we use the suffix *bw* to indicate block-wise Metropolization as in Algorithm 2, e.g. MA-ALDI-bw corresponds to Algorithm 2 with proposal (31).
- we use the suffix *ew* to indicate the special case of ensemble-wise Metropolization as in Algorithm 2 with $L = 1$, e.g. MA-ALDI-ew corresponds to Algorithm 2 with $L = 1$ and proposal (31),
- we use the suffix *pw* to indicate the special case of particle-wise sequential Metropolization as in Algorithm 2 with $L = 1$, e.g. MA-ALDI-pw corresponds to Algorithm 2 with $|\mathbf{b}_j| = B = 1$, $j = 1, \dots, L = M$, and proposal (31),
- we use the suffix *sim-bw* to indicate the simultaneous within blockMetropolization as in Algorithm 3, e.g. MA-ALDI-sim-bw corresponds to Algorithm 3 with proposal (32).

Let us briefly describe the three experiments below:

1. The first experiment illustrates the bias of the unadjusted interactive particle samplers for a one-dimensional non-Gaussian target distribution, thus, emphasizing the need for Metropolization.
2. The second experiment is a more detailed study of MA-IPS using a four-dimensional multivariate Gaussian target distribution. It demonstrates the benefits of interaction (compared to independent parallel Markov chains) and compares the performance of the various Metropolization schemes, in particular, under consideration of parallelization. Moreover, also the optimal tuning of the considered MA-IPS methods is studied empirically by the relation of the average acceptance rate and the obtained mean squared error of $\mathbf{S}_N^M(F)$ for a chosen F.

3. The final experiment applies the MA-IPS methods to a multivariate non-Gaussian target distribution and confirms our observations in the Gaussian setting.

In the second and third experiments, we place detailed focus on the comparison of MA-ALDI and pMALA due to their natural relationship. For CBS and SVGD, we compare only the Metropolis-adjusted and unadjusted versions. The MA-CBS algorithm utilizes only zeroth-order information, such that making a fair comparison would correspond to a (parallel) random walk MH algorithm. The algorithm MA-SVGD requires additional hyperparameter tuning, particularly in selecting an underlying kernel K, complicating direct comparisons with pMALA and MA-ALDI.

Computational complexity In many applications, such as PDE driven inverse problems, the main computational cost stems from the evaluations of the target distribution π and it's gradient. All other computations can be considered as relatively minor overhead. In all variants of our algorithm, the evaluations of π and its gradient can be performed in parallel within each block for each sweep through the ensemble.

In terms of the ensemble size M, the number of blocks L, and the spatial dimension d, in the worst case the computation of the square root of all empirical covariances in one step of the algorithm behaves like $O(LMd^3)$, i.e. linear in the ensemble size M and the number of blocks L, and cubic in d.

4.1 One-dimensional example: bimodal distribution

We study the effect of Metropolization on interacting particle systems for ALDI, CBS, and SVGD, by comparing the performance of each method with and without ensemble Metropolization.

As a target we consider the (unnormalized) probability density

$$\pi(x) \propto \exp\left(-\frac{1}{2\sigma^2}|x^2-1|^2\right)\cdot\exp\left(-\frac{1}{2}|x-m_0|^2\right) \qquad x\in\mathbb{R},$$

with $\sigma = \sqrt{0.5}$ and $m_0 = 0.8$. This target density corresponds to the posterior density in a Bayesian inverse problem with Gaussian prior distribution $\mathrm{N}(m_0, 1)$, forward model $G(x) = x^2$ and additive Gaussian noise with zero mean and variance σ^2.

For each particle sampling method we have simulated $N_{\text{burn}} = 10^4$ iterations of burn-in followed by another $N = 10^5$ iterations. Both, the Metropolis-adjusted and unadjusted variant, were implemented with $M = 10$ particles, same realization of the iid initialization $X_0^{(j)} \sim \mathrm{N}(m_0, 1)$, $j = 1, \ldots, M$ and same step size $h > 0$ in the Euler–Maruyama scheme. For both ALDI versions we used step size $h = 0.0725$ and for both CBS versions the step size $h = 0.05$. For both SVGD versions we used step size $h = 0.0001$ and the Gaussian kernel

$$k_s(x_1, x_2) = \exp\left(-\frac{1}{2s^2}|x_1 - x_2|^2\right)$$

with $s = 0.01$.

Table 1 Comparison of the MSE for adjusted and unadjusted ALDI for various step sizes

h	MA-ALDI-ew		(Unadjusted) ALDI
	Acceptance rate	MSE	MSE
0.01	0.93	0.0067	0.0054
0.04	0.82	0.005	0.018
0.0725	0.7	0.006	–
0.1	0.61	0.0045	–
0.125	0.5	0.0038	–

In Figs. 3, 4, 5 we compare the histograms of the particles obtained by the Metropolis-adjusted and unadjusted algorithms accumulated over all N iterations. For all three interactive particle methods, the invariant measures of the unadjusted implementations exhibit a visible bias relative to the target π, whereas this bias is not present for their Metropolized counterparts. It is worth mentioning again that the unadjusted CBS algorithm only produces a Gaussian approximation.

Finally, in Table 1 we compare the results of both ALDI versions for different choices of step sizes h. Here, we averaged over 10 seeds and computed both the averaged acceptance rate and the mean squared error (MSE) for estimating $\mathbb{E}_\pi[X^2] = \int_{\mathbb{R}} x^2\pi(x)\,\mathrm{d}x$ using an average over the entire chain. Note that unadjusted ALDI lacks stability when the step size h is too large, causing various simulated paths to diverge to infinity. This is the case when $h \geq 0.0725$.

4.2 Multivariate Gaussian distribution

We now consider a 4-dimensional multivariate Gaussian target distribution $\mathrm{N}(0, C)$ with target density

$$\pi(x) \propto \exp\left(-\frac{1}{2}\|C^{-1/2}x\|^2\right),$$
$$C = \mathrm{diag}(1, 0.1, 0.01, 0.001).$$

As the coordinates are weighted differently, we expect a positive effect from the ensemble preconditioner. In this section we illustrate the benefits of interaction, compare ensemble-wise to particle-wise Metropolization (for MA-ALDI), and

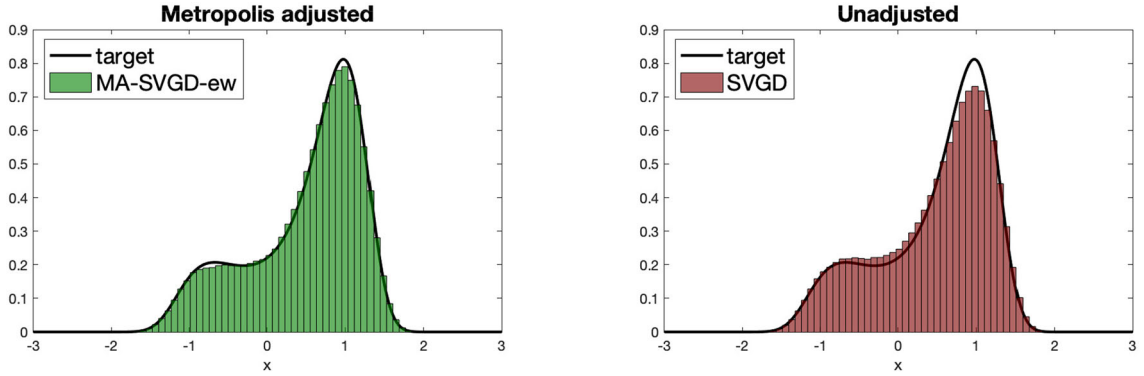


Fig. 3 Comparison of MA-SVGD-ew and SVGD with $M = 10$ particles and step size $h = 0.0001$. MA-SVGD-ew achieved an acceptance rate of 53%

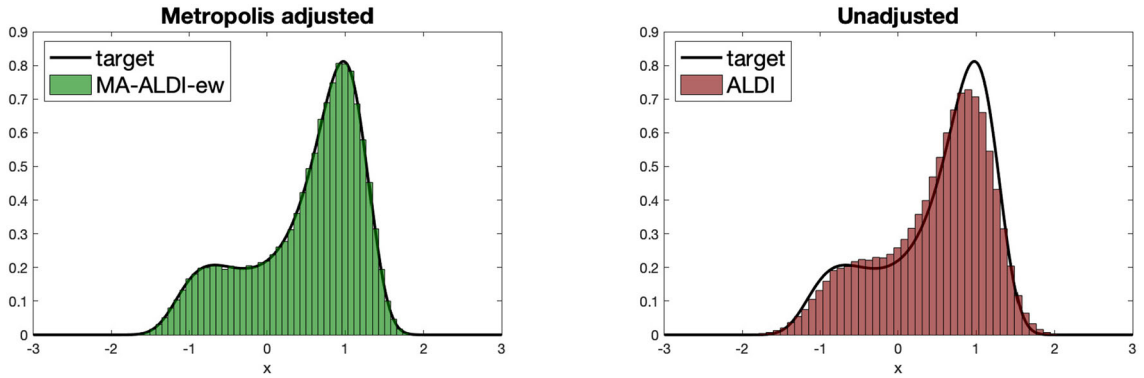


Fig. 4 Comparison of MA-ALDI-ew and ALDI with $M = 10$ particles and step size $h = 0.0725$. MA-ALDI-ew achieved an acceptance rate of 70%

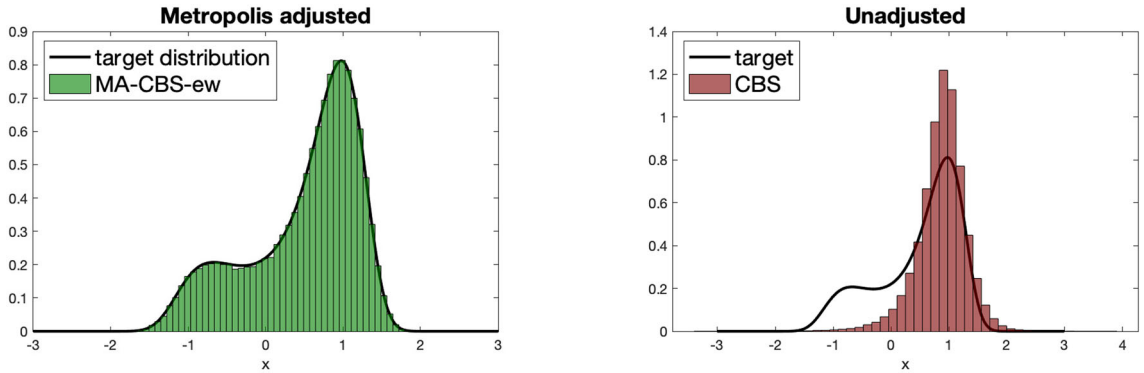


Fig. 5 Comparison of MA-CBS-ew and CBS with $M = 10$ particles and step size $h = 0.05$. MA-CBS-ew achieved an acceptance rate of 52%

consider optimal tuning of MA-ALDI, MA-CBS, and MA-SVGD by controlling the acceptance rate.

To this end, we consider the quantity of interest $f(x) = x^\top C^{-1} x$ where for $X \sim \pi$ we have $f(X) \sim \chi^2_4$. Our goal is then to estimate

$$P_{\text{ref}} := \mathbb{P}(f(X) \leq q_{0.5}) = \mathbb{E}_\pi \left[\mathbb{1}_{(-\infty, q_{0.5}]}(f(X)) \right] = \frac{1}{2} \tag{38}$$

where $q_{0.5}$ denotes the 0.5-quantile of the χ^2_4 distribution and $\mathbb{1}_A$ the indicator function of a set A. Let us mention that the presented results below are very similar when we directly estimate $\mathbb{E}_\pi[f(X)]$ rather than P_{ref}. The corresponding estimators based on ensemble Markov chains $(\boldsymbol{X}_k)_{k \in \mathbb{N}}$ are then

$$\widehat{P}_N := \frac{1}{N} \sum_{k=1}^{N} \mathbf{F}(\boldsymbol{X}_k), \quad \mathbf{F}(\mathbf{x}) := \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{(-\infty, q_{0.5}]}(f(x^{(i)})) \in \mathbb{R}.$$

For evaluating the efficicency of these estimators we also compute the associated autocorrelations $\mathrm{Corr}(\mathbf{F}(\boldsymbol{X}_1), \mathbf{F}(\boldsymbol{X}_{1+k})) \approx \rho_k := c_k/c_0$ where

$$c_k := \frac{1}{N-k} \sum_{i=1}^{N-k} \left(\mathbf{F}(\boldsymbol{X}_i) - \widehat{P}_N\right)\left(\mathbf{F}(\boldsymbol{X}_{i+k}) - \widehat{P}_N\right)$$

since the ρ_k yield information about the asymptotic variance (47) in the central limit theorem.

Superiority of interacting particles In Fig. 6, we compare MA-ALDI-ew for different values of the covariance regularization parameter $\gamma \in \{0.001, 0.1, 1\}$ to pMALA, in order to show the improved performance for proposals generated by *interacting* particle systems. All variants are initialized by iid samples of the initial distribution $\mathrm{N}(0, 5 \cdot \mathrm{Id})$. Note that for $\gamma = 1$ the proposal of MA-ALDI-ew and pMALA are in fact the same, but MA-ALDI-ew uses ensemble-wise Metropolization while pMALA uses particle-wise Metropolization.

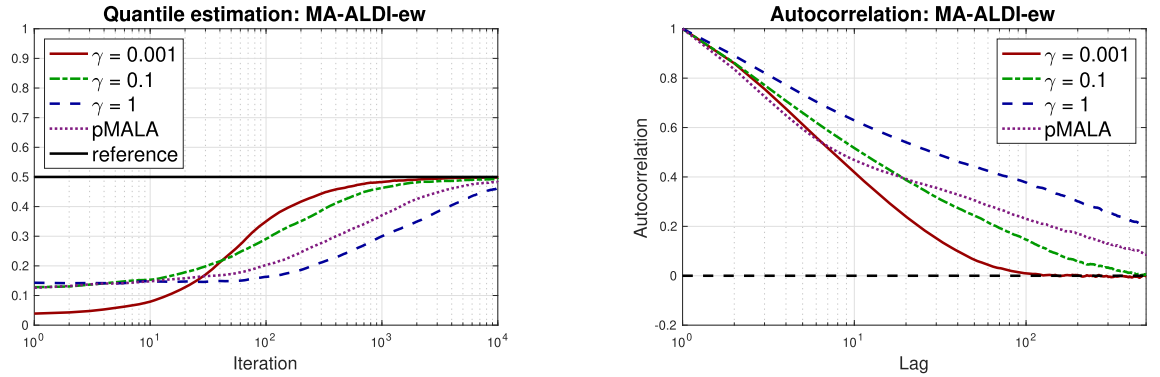
Figure 6a shows the evolution of $\widehat{P}_N$ averaged over 100 independent runs for each method. For each method, we tuned the step size h such that the average acceptance rate was around 50%. To be more precise, we have applied $h \in \{0.07, 0.008, 0.0008\}$ for the corresponding $\gamma \in \{0.001, 0.1, 1\}$ and $h = 0.0023$ for pMALA. Each method is implemented using $M = 10$ particles. For $\gamma = 0.001$ and $\gamma = 0.1$ the MA-ALDI-ew based estimators converge much faster than pMALA. This observation is confirmed by the significantly faster decaying autocorrelation depicted on the right plot in Fig. 6. On the other hand for $\gamma = 1$ it is observed that MA-ALDI-ew performs worse than pMALA, suggesting that for $\gamma = 1$ the particle-wise Metropolization is more effective than ensemble-wise Metropolization.

Superiority of particle- and block-wise Metropolization In Figs. 7 and 8 we compare the ensemble-wise Metropolization MA-ALDI-ew with the particle- and block-wise Metropolizations MA-ALDI-pw and MA-ALDI-bw for $\gamma = 0.001$. Moreover, we compare the heuristic simultaneous within block-wise variant of MA-ALDI based on the proposal (32) satisfying the product form (18). The results are obtained by averaging over 10 independent runs of the corresponding methods, where each variant is initialized by iid samples of the initial distribution $\mathrm{N}(0, 0.01 \cdot \mathrm{Id})$. We have again tuned the step size h for all three MA-ALDI versions such that an acceptance rate of approximately 50% is achieved. Note that we have increased the ensemble size to $M = 100$ in this experiment. Similar as for $\gamma = 1$, the particle-wise Metropolization outperforms the ensemble-wise Metropolization also for $\gamma = 0.001$ and $\gamma = 0.1$. We observe a clear benefit of applying particle- and block-wise Metropolization in Fig. 7a, where we do not take into account possibilities for parallel computing. The result changes, when allowing access to multiple cores, such that the computation of the ensemble- and block-wise Metropolization can be done in parallel. Here, we assume that the primary computational cost arises from evaluating π and its gradient, making the assumption of full parallelism reasonable. In Fig. 8, we observe a clear advantage of block-wise Metropolization the more cores we incorporate. In addition, Table 2 shows the estimated integrated autocorrelation scaled by the associated cost of the applied algorithm, i.e. the quantity

$$c(\mathrm{int\_ac}, B, \mathrm{cores}) = \mathrm{int\_ac} \cdot N \cdot M \cdot \frac{1}{\min\{\mathrm{cores}, B\}}. \quad (39)$$

Here we assume all blocks in (19) to be of equal size $|\mathbf{b}_j| = B \in \mathbb{N}$ for all $j = 1, \ldots, L$ (and thus $M = LB$ is a multiple of B). Moreover, $\mathrm{cores} \in \mathbb{N}$ refers to the number of processors available for parallelization, N is the number of iterations, and int_ac refers to the estimated integrated autocorrelation. Note that ensemble-wise Metropolization and parallel MALA correspond to $B = M$ and particle-wise Metropolization corresponds to $B = 1$. Notably the MA-ALDI-sim-bw benefits from both the particle-wise acceptance schedule and parallelization within the blocks resulting in the overall best performance.

Optimal tuning Finally, We study the dependence of the mean squared error (MSE) of $\mathbf{S}_N^M(F)$ on the average acceptance rate for the considered ensemble-wise MA-IPS in Fig. 9. Here, we control the acceptance rate through the step size h of the Euler–Maruyama scheme. The MSE was estimated over 100 independent chains of length $5 \cdot 10^4$ for each method and step size. In this experiment, we initialize each variant by an iid sample of the initial distribution $\mathrm{N}(0, 5 \cdot \mathrm{Id})$ with ensemble size $M = 10$. We observe slightly different optimal average acceptance rates for MA-SVGD-ew, MA-ALDI-ew ($\gamma = 0.001$), and MA-CBS-ew ($\gamma = 0$), with MA-ALDI-ew notably performing most sensitive w.r.t. the acceptance rate, but also achieving the smallest MSE—by an order of magnitude smaller compared to MA-SVGD-ew and two orders of magnitude compared to MA-CBS-ew. Note that we have applied $\gamma > 0$ for MA-ALDI-ew to avoid numerical instabilities due to the small number of particles M. In Fig. 10, we show the expected MSE in dependence of the applied step size h, where we compare the Metropolis-adjusted and unadjusted variants (UA). We again observe that Metropolization significantly improves the performance. Finally, for MA-SVGD-ew we can additionally control the acceptance rate through the kernel function: We use the Gaus-

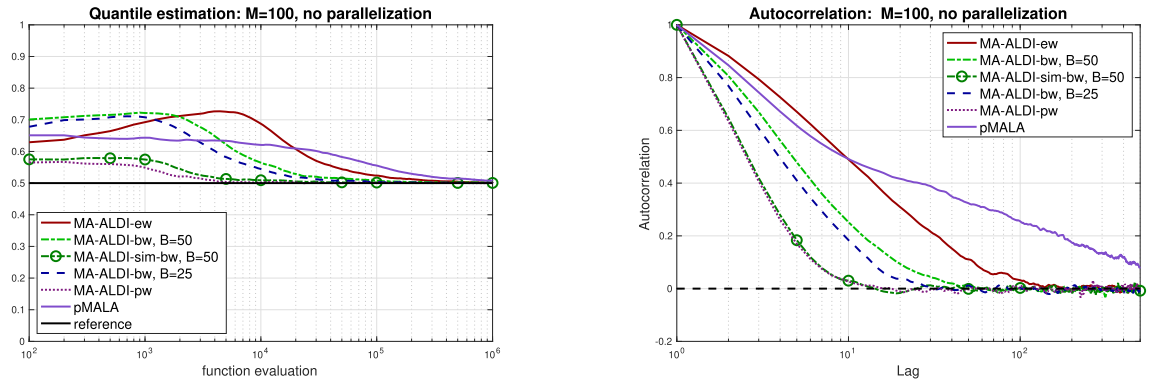


(a) quantile estimation (b) autocorrelation

Fig. 6 Estimation of P_{ref} in (38) and the corresponding estimated autocorrelation using either pMALA or MA-ALDI-ew for different choices of γ. For each method the step size h was tuned to obtain an acceptance rate of approximately 50%

Table 2 Comparison of the quantity in (39) measuring the computational efficiency of the method for different numbers of cores available and different variants of Metropolization ($\gamma = 0.001$)

Cores	ew	bw, B = 50	sim-bw, B = 50	bw, B = 25	pw	pMALA
1	2077.8	631.92	169.89	409.56	257.39	9529.58
25	83.11	25.28	6.8	16.38	257.39	381.18
50	41.56	12.64	3.4	16.4	257.39	190.59
100	20.78	12.64	3.4	16.4	257.39	95.3
h	0.06	0.15	0.8	0.225	0.8	0.0023



(a) no parallelization (b) no parallelization

Fig. 7 Estimation of P_{ref} in (38) and the corresponding estimated autocorrelation for different versions of MA-ALDI with $\gamma = 0.001$, and pMALA. For all variants the step size h was tuned to obtain an acceptance rate of approximately 50%

sian kernel

$$k_s(x_1, x_2) \propto \exp\left(-\frac{1}{2s^2}\|x_1 - x_2\|^2\right),$$

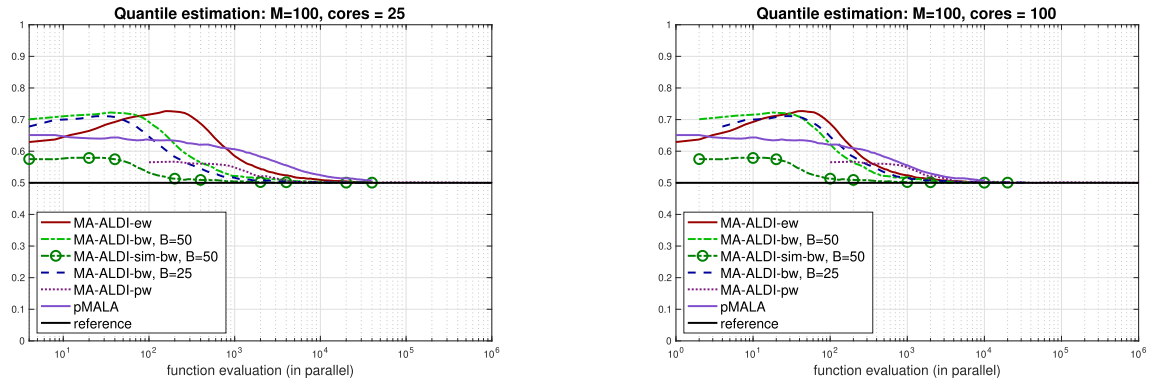
and can also steer the performance of (MA-)SVGD by the variance parameter s^2. The results are shown in Fig. 11. The optimal average acceptance rate seems to be almost the same as for tuning h (right). The dependence of the MSE and average acceptance rate is explicitly shown in the middle and right plot of Fig. 11.

4.3 ODE-based nonlinear inverse problem

We consider the one-dimensional elliptic equation

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}s}\left(e^{\theta(s)}\frac{\mathrm{d}}{\mathrm{d}s}p(s)\right) = 0 \quad s \in D := (0,1) \\ p(0) = 0, \;\; p(1) = 2 \end{cases} \tag{40}$$

and the inverse problem of recovering the unknown coefficient $\theta(\cdot) \in L^\infty(D)$ from noisy observations $y = F(\theta) + \xi \in \mathbb{R}^K$, where $\xi \in \mathbb{R}^K$ denotes observational noise. Denote by $H^1(D)$ the standard Sobolev space of $L^2(D)$-integrable



(a) 25 cores (b) 100 cores

Fig. 8 Same estimation of P_{ref} in (38) as in Fig. 7 but with different assumptions on accessible cores

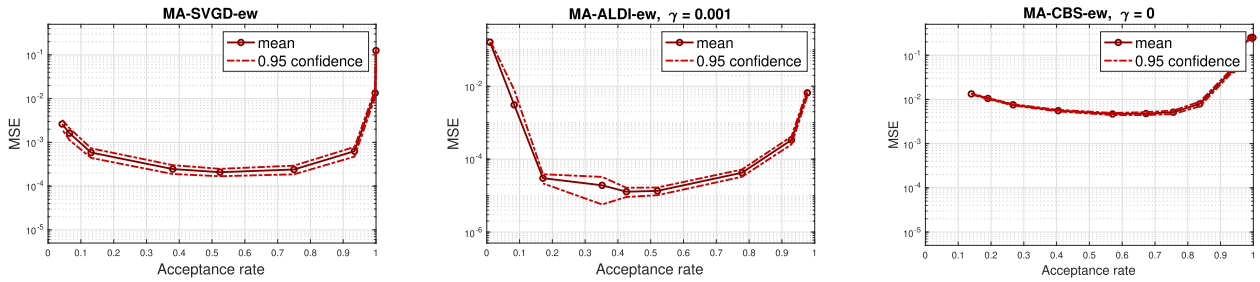


Fig. 9 Comparison of the expected MSE depending on the chosen step size h for MA-SVGD-ew (left), MA-ALDI-ew with $\gamma = 0.001$ (middle), and MA-CBS-ew with $\gamma = 0$ (right)

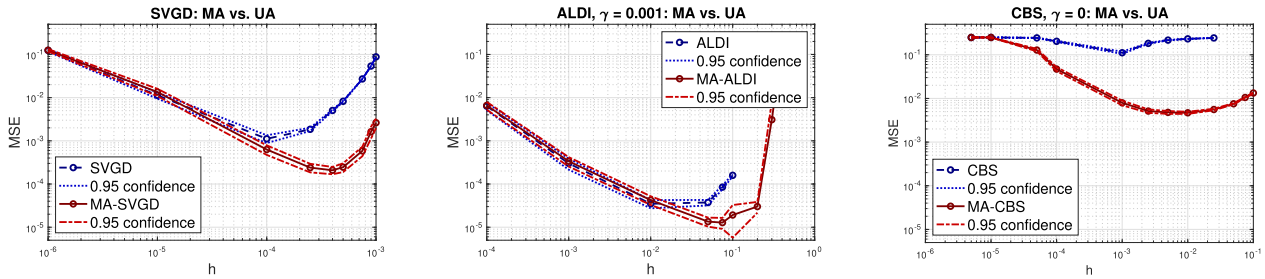


Fig. 10 Comparison of the expected MSE depending on the chosen step size h rate for SVGD (left), ALDI with $\gamma = 0.001$ (middle), and CBS with $\gamma = 0$ (right)

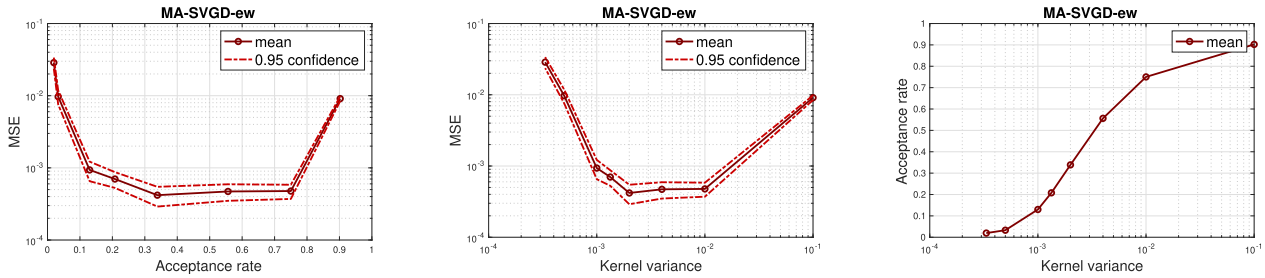


Fig. 11 Comparison of the expected MSE depending on the expected acceptance rate and the choice of kernel variance for MA-SVGD-ew

functions whose derivatives are also $L^2(D)$-integrable (see, e.g., Adams and Fournier (2003)). The forward operator $F : L^\infty(D) \to \mathbb{R}^K$ is defined by $F = \mathcal{O} \circ G$ - the solution operator $G : L^\infty(D) \to H^1(D)$ of (40) composed with the linear observation operator $\mathcal{O} : H^1(D) \to \mathbb{R}^K$ providing function values of K equidistant observation points $s_k = \frac{k}{K+1}$, $k = 1, \dots, K$, such that for $p \in H^1(D)$ we have $\mathcal{O}p(\cdot) = (p(s_1), \dots, p(s_K))^\top$. The solution of (40) is given in closed form by

$$p(s) = G(\theta)(s) = 2I_s(e^{-\theta(\cdot)})/I_1(e^{-\theta(\cdot)}) \tag{41}$$

where $I_s(f(\cdot)) := \int_0^s f(s)\,\mathrm{d}s$.

We consider a Gaussian process prior for θ given by

$$\theta(\cdot, x) = B_\phi x := \sum_{i=1}^{d} x_i \phi_i(\cdot),$$

where $\phi_i(s) = \frac{\sqrt{2}}{\pi}\sin(i\pi s)$ and $x_i \sim \mathrm{N}(0, \lambda_i)$ independently with $\lambda_i = i^{-2}$. Thus, the resulting inverse problem is to recover the coefficients $x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ with prior information $\mathrm{N}(0, \Gamma_0)$, where $\Gamma_0 = \mathrm{diag}(\lambda_1, \dots, \lambda_d)$. Assuming additive Gaussian noise $\xi \sim \mathrm{N}(0, \Gamma)$, $\Gamma = 0.01^2 \cdot \mathrm{Id}$, the resulting (unnormalized) posterior density is

$$\pi(x) \propto \exp\left(-\frac{1}{2}\|\Gamma^{-1/2}(y - F(B_\phi x))\|^2 - \frac{1}{2}\|\Gamma_0^{-1/2}x\|^2\right).$$

For the numerical implementation we replace G by a numerical solution operator for (40) on the grid $D_\delta \subset D$ with mesh size $\delta = 2^{-10}$ and restriction of the unknown parameter $\theta(\cdot, x)$ to D_δ. More specifically, we apply a trapezoidal rule to approximate the integrals in (41) and compute the exact gradient of this numerical formula with respect to $x \in \mathbb{R}^d$. We consider a partially observed system with $K = 4$ and $d = 5$ terms in the Gaussian process model.

We apply different versions of MA-ALDI with different choices of $\gamma \in \{0.001, 0.1, 1\}$ and compare them to pMALA. All variants have been implemented with an ensemble-size $M = 10$. For the comparison of the different schemes, we consider the quantity of interest

$$f(x) := \int_0^1 e^{\theta(s,x)}\,\mathrm{d}s = I_1(e^{\theta(\cdot,x)}) \tag{42}$$

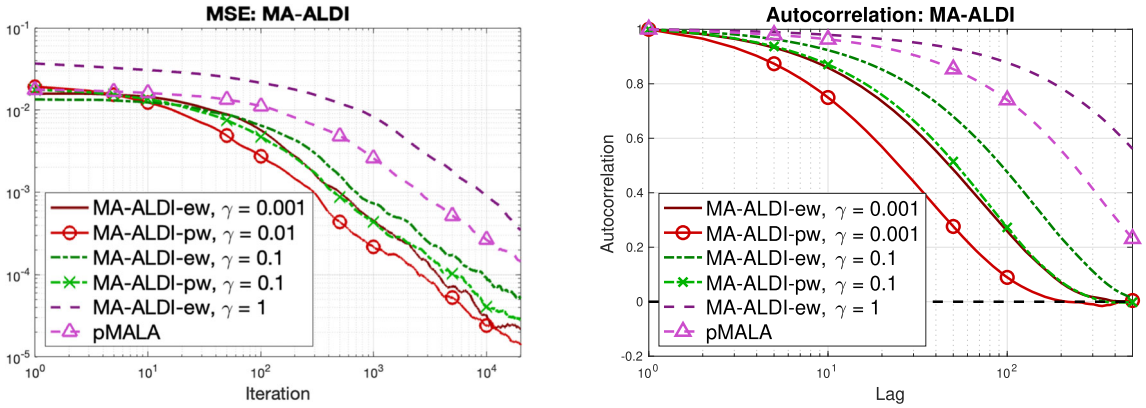
where the integral $I_1(e^{\theta(\cdot,x)})$, $x \in \mathbb{R}^d$, is again approximated by a trapezoidal rule, and use $\mathbf{S}_N^M(f)$ as estimator for the posterior expectation $\pi(f)$. The corresponding autocorrelation is estimated as in the previous section. Moreover, the MSE of $\mathbf{S}_N^M(f)$ computed by averaging over 100 seeds, where we use a reference value for $\pi(f)$ based on running the (single chain) preconditioned Crank-Nicolson Metropolis algorithm (Cotter et al. 2013) for $N_{\mathrm{ref}} = 10^7$ iterations. The resulting estimated MSE and autocorrelation are plotted in Fig. 12. Again, we observe that MA-ALDI outperforms pMALA. In particular, the particle-wise MA-ALDI with highest interaction ($\gamma = 0.001$) performs best among all considered algorithms.

Moreover, we show the resulting posterior approximation averaged over the realized Markov chains pushed forward through the truncated KL-expansion, i.e., $\widehat{\theta}_k = B_\phi X_k$. In Fig. 13 we plot the pointwise mean $\frac{1}{N}\sum_{k=N_{\mathrm{burn}}}^{N_{\mathrm{burn}}+N} \widehat{\theta}_k$ plus/minus the pointwise empirical standard deviation. Here, we used again a burn-in of $N_{\mathrm{burn}} = 10^3$ iterations but then only take $N = 200$ iterations. We observe a smaller deviation to the posterior (approximated by the preconditioned Crank-Nicolson Metropolis) for MA-ALDI-ew ($\gamma = 0.001$) than for pMALA which illustrates again the faster convergence of the Metropolis-adjusted *interacting* particle system.

5 Conclusion

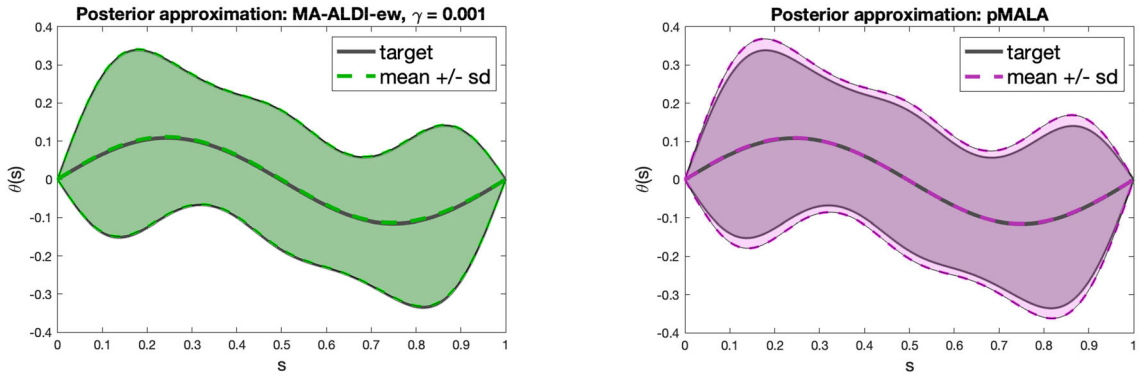
The success of MCMC methods, particularly in anisotropic and nonlinear problems, heavily relies on the quality of the proposal distribution. Ideally, additional information on the target distribution, such as its covariance, should be used. While this is not available in practice, it can for example be estimated along the *path* of the chain which has led to the developement of adaptive MCMC methods. In the present work, we consider an alternative approach that evolves an *ensemble* of $M \in \mathbb{N}$ interacting particles and leverages the information gained by the entire ensemble to generate a proposal for the next update. One key advantage is that this method provides an effective and natural means of parallelization which takes full advantage of the additional information provided by the ensemble. This can be crucial in the treatment of real-world problems. For instance, in engineering and science, solving a (Bayesian) inverse problems often involves simulating a complex physical process at each step of the chain. Each of these simulations can take minutes or even hours to complete, which renders any sequential algorithm and any MCMC approach that mixes only slowly infeasible.

The present study investigated fundamental variants of Metropolizing interacting particle systems that evolve $M \in \mathbb{N}$ particles based on some SDE of McKean–Vlasov type. To this end, we partition the ensembles into blocks of equal size and sequentially propose and accept or reject all particles in each block as a whole. This includes as special cases the ensemble- and particle-wise Metropolization where either the entire ensemble is considered as one block and each individual particle is taken as a block. While sequen-



(a) quantile estimation (b) autocorrelation

Fig. 12 MSE for estimating $\pi(f)$ with f as in (42) and the corresponding estimated autocorrelation for different versions of MA-ALDI and γ, and pMALA. For MA-ALDI-ew and pMALA the step size h was tuned to obtain an acceptance rate of approximately 50%. For MA-ALDI-pw we use the same h as for MA-ALDI-ew



(a) MA-ALDI-ew (b) pMALA

Fig. 13 Posterior approximation (averaged over 200 iterations after burn-in) for MA-ALDI-ew with $\gamma = 0.001$ and pMALA. The step size h was chosen such that the acceptance rate is approximately 50%

tial particle-wise updating has been proposed and discussed previously, e.g., in Coullon and Webber (2021), Dunlop and Stadler (2022), Goodman and Weare (2010), Leimkuhler et al. (2018), the ensemble- and sequential block-wise Metropolziation are novel to the best of our knowledge. These two variants also allow for the parallel computation of the likelihood for each particle in the block. For suitable interactive particle systems which satisfy a conditional independence of particles within each block we also proposed a *simultaneous Metropolization* within the block, i.e., each particle is updated and accepted/rejected independently and in parallel to the other particles of the block. This allows for a further exploitation of parallelization and also can improve the mixing of the ensemble Markov chain. In general, all discussed variants allow for the construction of affine-invariant MCMC methods through affine-invariant particle dynamics. Furthermore, we provide a theoretical analysis of the Metropolized interacting particle sampling methods, establishing basic ergodicity under mild and common assumptions. Finally, in the appendix we also discuss why a simultaneous updating of the blocks (instead of sequential updating) does in general not yield the correct invariant distribution.

We presented a detailed empirical study comparing these methods for several common particle dynamics. Our findings show that the interaction of the particles can significantly improve mixing compared to trivially running M independent MCMC chains (in parallel). Moreover, depending on the situation, we observed that the particle- and block-wise Metropolization seem to outperform the ensemble-wise variant. Overall, our study suggests that proposals based on interacting particle systems can provide significant improvements over traditional MCMC methods. However, comparing the proposed methods to established adaptive MCMC methods,

such as preconditioned MALA with online covariance estimation, is left for future work.

Other possible directions for future work include additionally using the history of the Markov chain, e.g., estimating the target covariance also along the path of the ensemble chain which may reduce the estimation error for the covariance. Also the application of localization techniques as discussed e.g., in Huang et al. (2022), Reich and Weissmann (2021) within the MA-IPS approach seems beneficial. Finally, while we provide basic convergence results, a solid theoretical analysis of the superiority of interacting ensembles over independent, parallel Markov chains remains an open and interesting avenue for future work.

A Preliminaries on Markov chain Monte Carlo

We recall the basic terminology and ideas of MCMC sampling required in the following.

Throughout let π be a given absolutely continuous (target) probability distribution on $\mathbb{R}^d$ with the Lebesgue density also denoted by $\pi : \mathbb{R}^d \to [0, \infty)$. To approximately sample from π, we construct a Markov chain $(X_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^d$ that converges to π in distribution as $k \to \infty$. We denote the associated transition kernel by $P : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$, i.e., the chain is characterized by $X_k = x$ implying $X_{k+1} \sim P(x, \cdot)$. Note that if $X_0 \sim \pi_0$ for some initial distribution π_0 then $X_k \sim \pi_0 P^k$, (cf. (1)).

We say that

- π is an *invariant measure* of P iff

$$\pi = \pi P,$$

in which case we call P and $(X_k)_{k\in\mathbb{N}}$ π*-invariant*,
- the Markov chain is π*-reversible* iff it satisfies the *detailed balance condition*

$$\int_A P(x, B)\pi(\mathrm{d}x) = \int_B P(z, A)\pi(\mathrm{d}z) \qquad \forall A, B \in \mathcal{B}(\mathbb{R}^d), \tag{43}$$

- the Markov chain is *ergodic* iff it holds

$$\lim_{k\to\infty} d_{\mathrm{TV}}(\pi_0 P^k, \pi) = 0, \tag{44}$$

where π_0 is the initial distribution and d_{TV} stands for the total variation distance,

- the Markov chain *satisfies a strong law of large numbers* iff

$$S_N(F) := \frac{1}{N}\sum_{k=1}^{N} F(X_k) \xrightarrow[N\to\infty]{\text{a. s.}} \mathbb{E}_\pi[F] \qquad \forall F \in L^1_\pi(\mathbb{R}), \tag{45}$$

- the Markov chain *satisfies a central limit theorem (CLT) for* $F \in L^2_\pi(\mathbb{R})$ iff the *asymptotic variance* of $S_N(F)$ is finite

$$\sigma_F^2 = \lim_{N\to\infty} N\,\mathbb{V}[S_N(F)] \in (0, \infty)$$

and

$$\sqrt{N}\,(S_N(F) - \mathbb{E}_\pi[F]) \xrightarrow[N\to\infty]{\mathcal{L}} \mathrm{N}\left(0, \sigma_F^2\right). \tag{46}$$

Due to $X_k \sim \pi_0 P^k$, the Markov chain $(X_k)_{k\in\mathbb{N}}$ can be interpreted as the realization of a fixed point iteration under the mapping P. Hence π being an invariant measure of P is a necessary condition to obtain ergodicity. Additionally, let us mention that π-reversibility is sufficient to ensure π-invariance the latter following from (43) with $B = \mathbb{R}^d$.

While a strong law of large numbers holds under mild conditions, the central limit theorem is more nuanced. First note that, the asymptotic variance of $S_N(F)$ can be expressed as

$$\sigma_F^2 = \mathbb{V}_\pi[F]\left[1 + 2\sum_{k=1}^{\infty} \mathrm{Corr}(F(\tilde{X}_0), F(\tilde{X}_{0+k})\right]. \tag{47}$$

where $(\tilde{X}_k)_{k\in\mathbb{N}}$ denotes the Markov chain with the same π-invariant transition kernel P but starting in stationarity $\tilde{X}_0 \sim \pi$. However, in general neither the finiteness of σ_F^2 nor a CLT (46) is a given. A classical result (Kipnis and Varadhan (1986)) states that for reversible, irreducible and aperiodic Markov chains the finiteness of σ_F^2 also yields a CLT (46). Moreover, several other non-trivial conditions for (46) can be given. We refer to Roberts and Rosenthal (2006), Section 5 for more details.

A.1 Metropolis–Hastings algorithm

The key question is how to obtain transition kernels that ensure ergodicity and a strong law of large numbers. The MH algorithm (Metropolis et al. 1953; Hastings 1970) is a standard method achieving this under rather mild assumptions.

The algorithm is based on a *proposal kernel* $Q : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$, that assigns a probability measure $Q(x, \cdot)$

Algorithm 4 Metropolis-Hastings

Input:

- target density π on $\mathbb{R}^d$
- proposal kernel Q with density $q : \mathbb{R}^d \times \mathbb{R}^d \to (0, \infty)$
- initial probability distribution π_0 on $\mathbb{R}^d$

Output: Markov chain $(X_k)_{k\in\{1,\dots,N\}}$ in state space $\mathbb{R}^d$

draw $x_0 \sim \pi_0$ and set initial state $X_0 = x_0$
for $k = 0, \dots, N$ **do**
  given $X_k = x_k$ draw proposal $y_{k+1} \sim Q(x_k, \cdot)$
  compute acceptance probability $\alpha(x_k, y_{k+1})$ in (48)
  draw $u \sim \mathrm{U}([0, 1])$ and set

$$X_{k+1} = \begin{cases} y_{k+1} & \text{if } u \le \alpha(x_k, y_{k+1}) \\ x_k & \text{else} \end{cases}$$

end for

on $\mathbb{R}^d$ to every $x \in \mathbb{R}^d$, in combination with an acceptance-rejection step. Throughout, we assume that $Q(x, \cdot)$ possesses a Lebesgue density for each $x \in \mathbb{R}^d$, i.e. there exists $q : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ such that

$$Q(x, A) = \int_A q(x, y)\, \mathrm{d}y \qquad \forall A \in \mathcal{B}(\mathbb{R}^d).$$

In the kth step, if $X_k = x_k$ and y_{k+1} is a proposed value drawn from $Q(x_k, \cdot)$, then X_{k+1} is set to y_{k+1} with probability $\alpha(x_k, y_{k+1})$ defined as follows:

$$\alpha(x_k, y_{k+1}) : = \begin{cases} \min\left(1, \frac{\pi(y_{k+1})q(y_{k+1}, x_k)}{\pi(x_k)q(x_k, y_{k+1})}\right) & \text{if } \pi(x_k)q(x_k, y_{k+1}) > 0 \\ 1 & \text{otherwise.} \end{cases} \tag{48}$$

If X_{k+1} is not set to y_{k+1}, it is set to x_k. The resulting transition kernel is

$$P(x, \mathrm{d}y) = \alpha(x, y)q(x, y)\mathrm{d}y + r(x)\delta_x(\mathrm{d}y), \quad r(x) := 1 - \int_{\mathbb{R}^d} \alpha(x, y)q(x, y)\mathrm{d}y. \tag{49}$$

It can easily be checked that P is in fact π-reversible. We present the full algorithm in Algorithm 4 and refer to (Robert and Casella (2004), Section 7.3) for more details.

It is left to choose a suitable proposal kernel Q. As Theorem A.1 shows ergodicity is already ensured if $Q(x, \cdot)$ has a positive Lebesgue density q, i.e. $q(x, y) > 0$ for all $x, y \in \mathbb{R}^d$. Nonetheless, in practice the efficiency of the algorithm crucially depends on the choice of Q. A standard (albeit crude) proposal satisfying the positivity condition is $Q(x, \cdot) = \mathrm{N}(x, h\mathrm{Id}_d)$, $h > 0$, also known as Random Walk-MH algorithm.

Theorem A.1 Robert and Casella (2004), Section 6.7.2 and 7.3.2. *Let π be absolutely continuous with respect to the Lebesgue measure and let Q possess a positive Lebesgue density $q : \mathbb{R}^d \times \mathbb{R}^d \to (0, \infty)$, and let π_0 be any initial probability distribution. Then, the Markov chain $(X_k)_{k\in\mathbb{N}}$ generated by Algorithm 4 with $X_0 \sim \pi_0$*

1. *Is ergodic (44) and satisfies a strong law of large numbers (45),*
2. *Satisfies the central limit theorem (46) for any $F \in L^2_\pi(\mathbb{R})$ for which σ_F^2 in (47) is nonzero and finite.*

A.2 Metropolis-adjusted Langevin algorithm

A popular proposal kernel Q is obtained through the Euler-Maruyama discretization

$$X_{k+1} = X_k + h\nabla \log \pi(X_k) + \sqrt{2h}\xi_{k+1}, \quad \xi_{k+1} \sim \mathrm{N}(0, \mathrm{Id}_d) \tag{50}$$

of the Langevin dynamics (4) (with $C = \mathrm{Id}_d$) introduced in Example 1.1. Here $h > 0$ is a fixed step size, and generating a Markov chain $(X_k)_{k\in\mathbb{N}}$ through (50) is also known as the *unadjusted Langevin algorithm (ULA)*. While the continuous dynamics (4) has π as an invariant distribution, see e.g., Pavliotis (2014), it is known that the Markov chain (50) has a bias that scales linearly in h (Vempala and Wibisono (2019), Theorem 2).

Nevertheless, the continuous-time result suggests using (50) as the proposal mechanism, yielding a proposal kernel Q with positive Lebesgue density

$$q(x, y) = \frac{1}{(4\pi h)^{d/2}} \exp\left(-\frac{1}{4h}\|y - x - h\nabla \log \pi(x)\|^2\right).$$

Algorithm 4 with this choice of proposal kernel is known as MALA. According to Theorem A.1, and contrary to the Markov chain (50), the Metropolised Markov chain generated by Algorithm 4 necessarily does have π as its invariant distribution. Moreover, it satisfies ergodicity and a law of large numbers. Furthermore, a common rule of thumb for choosing the step size h is to achieve an average acceptance rate $\bar{\alpha} = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \alpha(x, y)q(x, y)\mathrm{d}y\, \pi(x)\mathrm{d}x$ of roughly 57.4%. Given suitable assumptions this tuning of h can indeed be shown to be optimal in terms of a small asymptotic variance σ_F^2 in (46) for sufficiently large d, see Roberts and Rosenthal (2001).

B On simultaneous block-wise metropolization

From a computational viewpoint, it would be advantegous to decide for each block independently and in parallel whether to accept or reject it, as this facilitates the embarassingly parallel processing of all M particles in the ensemble in each step of the algorithm. However, as we illustrate in the following, the corresponding "simultaneous" transition kernel is in general not invariant with respect to the product target measure $\boldsymbol{\pi}$ or even an M-coupling of π.

To simplify the discussion, we focus on the special case where all blocks are of size one, i.e. the particle-wise variant. Moreover, we assume the setting of Remark 2.4, i.e., the ensemble-wise proposal kernel would take a product form

$$\mathbf{Q}(\mathbf{x}, \mathrm{d}\mathbf{y}) = Q_{\mathbf{x}^{-(1)}}(x^{(1)}, \mathrm{d}y^{(1)}) \otimes \cdots \otimes Q_{\mathbf{x}^{-(M)}}(x^{(M)}, \mathrm{d}y^{(M)}),$$

where

$$Q_{\mathbf{x}^-}(z, \cdot) := \mathrm{N}(z + h\,\Phi(z, \mathbf{x}^-), h\,\Sigma(z, \mathbf{x}^-)) \quad \forall z \in \mathbb{R}^d,\ \mathbf{x}^- \in \mathbb{R}^{(M-1)d}.$$

Under Assumption 2.1, i.e., $\Sigma(z, \mathbf{x}^-)$ is regular for any $z \in \mathbb{R}^d$ and $\mathbf{x}^- \in \mathbb{R}^{(M-1)d}$, $Q_{\mathbf{x}^-}(z, \cdot)$ possesses the Lebesgue density

$$q_{\mathbf{x}^-}(z, y) = \frac{1}{\det\left(2\pi\, h\, \Sigma(z, \mathbf{x}^-)\right)^{1/2}} \exp\left(-\frac{1}{2h}\left\|\Sigma(z, \mathbf{x}^-)^{-1/2}\left(y - z - h\Phi(z, \mathbf{x}^-)\right)\right\|^2\right) > 0. \tag{51}$$

We then introduce the acceptance probability for the ith particle as

$$\alpha_{\mathbf{x}^{-(i)}}(x^{(i)}, y) := \begin{cases} \min\left(1, \frac{\pi(y)\, q_{\mathbf{x}^{-(i)}}(y, x^{(i)})}{\pi(x^{(i)})\, q_{\mathbf{x}^{-(i)}}(x^{(i)}, y)}\right) & \text{if } \pi(x^{(i)})\, q_{\mathbf{x}^{-(i)}}(x^{(i)}, y) > 0 \\ 1 & \text{else.} \end{cases} \tag{52}$$

Finally, we consider particle-wise transition kernels

$$P_{\mathbf{x}^{-(i)}}(x^{(i)}, \mathrm{d}y) := \alpha_{\mathbf{x}^{-(i)}}(x^{(i)}, y)\, Q_{\mathbf{x}^{-(i)}}(x^{(i)}, \mathrm{d}y) + r_{\mathbf{x}^{-(i)}}(x^{(i)})\delta_x(\mathrm{d}y). \tag{53}$$

Applying these to the ith particle for each $i \in \{1, \ldots, M\}$ simultaneously instead of sequentially yields the transition kernel $\mathbf{P}_{\mathrm{sim}} \colon \mathbb{R}^{Md} \times \mathcal{B}(\mathbb{R}^{Md}) \to [0, 1]$

Algorithm 5 Simultaneous particle-wise Metropolized interactive particle sampling

Input:

- target density π on $\mathbb{R}^d$
- ensemble dependent proposal kernel $Q_{\mathbf{x}^-}$ with density $q_{\mathbf{x}^-} : \mathbb{R}^d \times \mathbb{R}^d \to (0, \infty)$ in (51)
- initial probability distribution π_0 on $\mathbb{R}^d$

Output: ensemble Markov chain $(\boldsymbol{X}_k)_{k\in\{1,\ldots,N\}}$ in state space $\mathbb{R}^{Md}$

1: draw $\mathbf{x}_0 \sim \otimes_{i=1}^M \pi_0$ and set initial state $\boldsymbol{X}_0 = \mathbf{x}_0 \in \mathbb{R}^{Md}$
2: **for** $k = 0, \ldots, N$ **do**
3: $\quad \forall i$: given $\boldsymbol{X}_k = \mathbf{x}_k$ draw proposal $y^{(i)} \sim Q_{\mathbf{x}_k^{-(i)}}(x_k^{(i)}, \cdot)$ independently
4: $\quad \forall i$: compute particle acceptance probability $\alpha_{\mathbf{x}_k^{-(i)}}(x_k^{(i)}, y^{(i)}) \in [0, 1]$ in (52)
5: $\quad \forall i$: draw $u_i \sim \mathrm{U}([0, 1])$ independently and set

$$X_{k+1}^{(i)} = \begin{cases} y^{(i)} & \text{if } u_i \le \alpha_{\mathbf{x}_k^{-(i)}}(x_k^{(i)}, y^{(i)}) \\ x_k^{(i)} & \text{else} \end{cases}$$

6: **end for**

$$\mathbf{P}_{\mathrm{sim}}(\mathbf{x}, \mathrm{d}\mathbf{y}) = P_{\mathbf{x}^{-(1)}}(x^{(1)}, \mathrm{d}y^{(1)}) \otimes \cdots \otimes P_{\mathbf{x}^{-(M)}}(x^{(M)}, \mathrm{d}y^{(M)}). \tag{54}$$

The associated algorithmic description is given in Algorithm 5.

As for the sequential updates discussed in Sect. 2.1, $\boldsymbol{\pi}$-reversibility does not hold, since in general

$$\prod_{i=1}^M \left(\alpha_{\mathbf{x}^{-(i)}}(x^{(i)}, y^{(i)})\, q_{\mathbf{x}^{-(i)}}(x^{(i)}, y^{(i)})\, \pi(x^{(i)})\right) \neq \prod_{i=1}^M \left(\alpha_{\mathbf{y}^{-(i)}}(y^{(i)}, x^{(i)})\, q_{\mathbf{y}^{-(i)}}(y^{(i)}, x^{(i)})\, \pi(y^{(i)})\right)$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{Md}$ with $x^{(i)} \neq y^{(j)}$ for all $i, j = 1, \ldots, M$—except for the case of noninteraction, i.e., $q_{\mathbf{x}^-} = q$ and, thus, $\alpha_{\mathbf{x}^-} = \alpha$ does not depend on the other particles in the ensemble.

Regarding the $\boldsymbol{\pi}$ invariance of $\mathbf{P}_{\mathrm{sim}}$, it is worth noting that each $P_{\mathbf{x}^{-(i)}}$ is π-invariant. However, due to the interaction, this does not directly imply $\boldsymbol{\pi}$ invariance of $\mathbf{P}_{\mathrm{sim}}$. Although, the particle-wise marginals of $\boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{P}_{\mathrm{sim}}$ coincide as shown below in Proposition B.1 the product transition kernel $\mathbf{P}_{\mathrm{sim}}$ is in general not $\boldsymbol{\pi}$ invariant as illustrated by several counterexamples below.

Proposition B.1 *For the transition kernel* $\mathbf{P}_{\mathrm{sim}} \colon \mathbb{R}^{Md} \times \mathcal{B}(\mathbb{R}^{Md}) \to [0, 1]$ *given in* (54) *associated to Algorithm* 5 *we have that* $\boldsymbol{\pi}\mathbf{P}_{\mathrm{sim}}$ *is an* M-coupling *of* π, *i.e. the particle-wise marginals of* $\boldsymbol{X} \sim \boldsymbol{\pi}\mathbf{P}_{\mathrm{sim}}$ *are* $X^{(i)} \sim \pi$ *for all* $i = 1, \ldots, M$.

Proof For any $i = 1, \ldots, M$ and any $A_i \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\boldsymbol{\pi}\mathbf{P}_{\mathrm{sim}}(\mathbb{R}^d \times \cdots \times \mathbb{R}^d \times A_i \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^{Md}} \mathbf{P}_{\text{sim}}(\mathbf{x}, \mathbb{R}^d \times \cdots \\
&\times \mathbb{R}^d \times A_i \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d)\, \boldsymbol{\pi}(\mathrm{d}\mathbf{x}) \\
&= \int_{\mathbb{R}^{Md}} P_{\mathbf{x}^{-(i)}}(x^{(i)}, A_i)\ \boldsymbol{\pi}(\mathrm{d}\mathbf{x}) \\
&= \int_{\mathbb{R}^{(M-1)d}} \left(\int_{\mathbb{R}^d} P_{\mathbf{x}^{-(i)}}(x^{(i)}, A_i)\, \pi(\mathrm{d}x^{(i)}) \right) \bigotimes_{j \neq i} \pi(\mathrm{d}x^{(j)}) \\
&= \int_{\mathbb{R}^{(M-1)d}} \pi(A_i) \bigotimes_{j \neq i} \pi(\mathrm{d}x^{(j)}) = \pi(A_i)
\end{aligned}
$$

due to the π-invariance of $P_{\mathbf{x}^-}$ for any $\mathbf{x}^- \in \mathbb{R}^{(M-1)d}$. □

Example B.2 Let us consider a discrete state space $\mathcal{X} = \{x_1, x_2\}$ of two elements $x_1 \neq x_2$ and the uniform distribution $\pi(x_1) = \pi(x_2) = \frac{1}{2}$ on $\mathcal{X}$ as the target measure. Consider parametrized proposal kernels $Q_z \colon \mathcal{X} \times \mathcal{X} \to [0, 1]$ with parameter $z \in \mathcal{X}$ which we write as right stochastic matrices $\mathrm{Q}_z \in [0, 1]^{|\mathcal{X}|}$ where the element q_{ij} of Q_z in the ith row and jth column denotes the probability $Q_z(x_i, x_j)$:

$$
\mathrm{Q}_{x_1} = \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix}, \qquad \mathrm{Q}_{x_2} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.
$$

The resulting parametrized acceptance probabilities $\alpha_z \colon \mathcal{X} \times \mathcal{X} \to [0, 1]$ are given by (cf. (48))

$$
\alpha_{x_1}(x, y) = \begin{cases} 1/2, & \text{if } x = x_2, y = x_1 \\ 1, & \text{else} \end{cases}, \qquad \alpha_{x_2}(x, y) \equiv 1,
$$

and the rejection probabilies by (cf. (49))

$$
r_{x_1}(x) = \begin{cases} 1/3, & x = x_1 \\ 0, & x = x_2 \end{cases}, \qquad r_{x_2}(x) \equiv 0.
$$

This yields as π-invariant MH transition kernel $P_z \colon \mathcal{X} \times \mathcal{X} \to [0, 1]$ again written as right stochastic matrices P_z (cf. (49))

$$
\mathrm{P}_{x_1} = \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix}, \qquad \mathrm{P}_{x_2} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.
$$

The resulting product transition kernel

$$
\mathbf{P}_{\text{sim}}(\mathbf{x}, \mathbf{y}) = P_{x^{(2)}}(x^{(1)}, y^{(1)}) \cdot P_{x^{(1)}}(x^{(2)}, y^{(2)})
$$

is then, written as well as a (4×4)-right stochastic matrix where the rows and columns in the matrix correspond to the lexicographically ordered states (x_1, x_1), (x_1, x_2), (x_2, x_1), (x_2, x_2) in $\mathcal{X}^2$,

$$
\mathbf{P}_{\text{sim}} \simeq \begin{pmatrix} 4/9 & 2/9 & 2/9 & 1/9 \\ 1/6 & 2/6 & 1/6 & 2/6 \\ 1/6 & 2/6 & 1/6 & 2/6 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.
$$

The associated invariant measure $\boldsymbol{\nu} \colon \mathcal{X}^2 \to [0, 1]$ is given by

$$
\begin{aligned}
&\boldsymbol{\nu}(x_1, x_1) = \frac{45}{173}, \quad \boldsymbol{\nu}(x_1, x_2) = \frac{49}{173}, \quad \boldsymbol{\nu}(x_2, x_1) = \frac{35}{173}, \\
&\boldsymbol{\nu}(x_2, x_2) = \frac{44}{173}
\end{aligned}
$$

which does not correspond to $\boldsymbol{\pi} \equiv \frac{1}{4}$. Moreover, also the particlewise marginals $\nu^{(i)}$ of $\boldsymbol{\nu}$ do not coincide with π:

$$
\begin{aligned}
&\nu^{(1)}(x_1) = \frac{94}{173}, \quad \nu^{(1)}(x_2) = \frac{79}{173}, \qquad \nu^{(2)}(x_1) = \frac{80}{173}, \\
&\nu^{(2)}(x_2) = \frac{93}{173}.
\end{aligned}
$$

Example B.3 Let us consider the continuous state space $\mathcal{X} = \mathbb{R}$ equipped with the triangular target distribution π given by the Lebesgue density

$$
\pi(x) = \begin{cases} \frac{1}{2} + 2\min\{x, 1 - x\}, & \text{if } x \in [0, 1] \\ 0, & \text{else.} \end{cases}
$$

Again, we consider an ensemble Markov chain of $M = 2$ interacting particles $\boldsymbol{X}_k = (X_k^{(1)}, X_k^{(2)})^\top \in \mathbb{R}^2$. As ensemble proposal kernel we choose the following

$$
\begin{aligned}
&\mathbf{Q}(\mathbf{x}, \mathrm{d}\mathbf{y}) = Q_{x^{(2)}}(x^{(1)}, \mathrm{d}y^{(1)}) \otimes Q_{x^{(1)}}(x^{(2)}, \mathrm{d}y^{(2)}), \\
&Q_z(x, \mathrm{d}y) = \mathrm{N}\left(\frac{x + z}{2}, \frac{1}{4} \right),
\end{aligned}
$$

which corresponds particle-wise to $\mathrm{N}(m(\boldsymbol{X}_k), h)$ for $h = \frac{1}{4}$. Consider the transition kernel $\mathbf{P}_{\text{sim}}$ resulting from particle-wise Metropolization. Note that for $M = 2$ we can decompose $\mathbf{P}_{\text{sim}}$ as follows

$$
\begin{aligned}
&\mathbf{P}_{\text{sim}}((x_1, x_2), \mathrm{d}y_1, \mathrm{d}y_2) \\
&\quad = \alpha_{x_2}(x_1, y_1)\alpha_{x_1}(x_2, y_2) q_{x_2}(x_1, y_1) q_{x_1}(x_2, y_2) \mathrm{d}y_1 \mathrm{d}y_2 \\
&\qquad + \alpha_{x_2}(x_1, y_1) r_{x_1}(x_2) q_{x_2}(x_1, y_1) \mathrm{d}y_1 \delta_{x_2}(\mathrm{d}y_2) \\
&\qquad + r_{x_2}(x_1) \alpha_{x_1}(x_2, y_2) q_{x_1}(x_2, y_2) \mathrm{d}y_2 \delta_{x_1}(\mathrm{d}y_1) \\
&\qquad + r_{x_2}(x_1) r_{x_1}(x_2) \delta_{x_1}(\mathrm{d}y_1) \delta_{x_2}(\mathrm{d}y_2)
\end{aligned}
$$

where q_z denotes the Lebesgue density of Q_z. We then discretize the state space to obtain a transition matrix $\mathbf{P}_{\text{sim}} \in \mathbb{R}^{n \times n}$ and compute its invariant measure as approximation

to the true invariant measure of the operator $\mathbf{P}_{\text{sim}}$. Since $\alpha_z(x, y) = 0$ for $y \notin [0, 1]$ it suffices to discretize $[0, 1]^2$. Here we use a uniform grid with grid size $\Delta x = 0.01$ in each dimension. Thus, $n = 101^2$.The invariant measure of the matrix $\mathbf{P}_{\text{sim}} \in \mathbb{R}^{n\times n}$ is then computed numerically and rearranged to yield $\boldsymbol{\pi}_{\text{inv}} \in [0, 1]^{101\times 101}$. It is displayed in comparison to an analogously discretized version of the true product target $\boldsymbol{\pi} = \pi \otimes \pi$ in Fig. 14. We do notice a bias, although a small one of relative size 10^{-3} to 10^{-2}. Since the crucial object for sampling purposes is not necesarilly the invariant measure in the ensemble space but the particle-wise marginals of it we also compare these in Fig. 15. However, the results are similar here. A small bias is observable, again the relative size compared to the true target are of order 10^{-3}.

Example B.4 We provide another numerical example similar to the previous one. Here again, $\mathcal{X} = \mathbb{R}$ but now $\pi = \mathbb{U}[0, 1]$ is the uniform distribution on $[0, 1]$. We consider $M = 2$ interacting particles $\boldsymbol{X}_k = (X_k^{(1)}, X_k^{(2)})^\top \in \mathbb{R}^2$ based on the following ensemble proposal kernel

$$\mathbf{Q}(\mathbf{x}, \mathrm{d}\mathbf{y}) = Q_{x^{(2)}}(x^{(1)}, \mathrm{d}y^{(1)}) \otimes Q_{x^{(1)}}(x^{(2)}, \mathrm{d}y^{(2)}),$$
$$Q_z(x, \mathrm{d}y) = \mathrm{N}\left(x, \frac{1}{4}\left(0.01 + \frac{0.99}{2}(z - x)^2\right)\right)$$

which corresponds particle-wise to $\mathrm{N}(x, h(\gamma + (1 - \gamma)C(\boldsymbol{X}_k)))$ with $\gamma = 0.01$ and $h = \frac{1}{4}$. Analogously, we discretize the state space or $[0, 1]^2$, respectively, using a uniform grid with grid size $\Delta x = 0.01$ and compute numerically the invariant measure of the resulting transition matrix $\mathbf{P}_{\text{sim}} \in \mathbb{R}^{101\times 101}$. The results are shown in Figs. 16 and 17. Also for this example we do notice a bias which is even larger than in the previous example, i.e., we observe a relative error of order 10^{-2} to 10^{-1} for the joint target and 10^{-2} for the particle-wise marginals.

We suspect that the bias of simultaneous particle-wise Metropolization is larger for smaller ensemble sizes than for bigger ones. In particular, the bias may vanish as $M \to \infty$ for suitable interacting particle systems, i.e., if the dynamics of each particle converge to their only time-discretized mean field limit as $M \to \infty$, then the corresponding proposal distributions $Q_{x^{-(i)}}(x^{(i)}, \cdot)$ should also converge to a limit proposal distribution $Q_\infty(x^{(i)}, \cdot)$ which does not depend on the other particles anymore, e.g., $Q_\infty(x^{(i)}, \cdot) = \mathrm{N}(x^{(i)} + hC(\pi_t)\nabla \log \pi(x^i), 2hC(\pi_t))$ in case of ALDI. However, for independent proposal kernels $Q_{x^{-(i)}}(x^{(i)}, \cdot) = Q(x^{(i)}, \cdot)$ the transition kernel of simultaneous particle-wise Metropolization $\mathbf{P}_{\text{sim}}$ is in fact $\boldsymbol{\pi}$-invariant. Therefore, we suspect that the bias of $\mathbf{P}_{\text{sim}}$ is largest for the $M = 2$ particle case considered in the numerical counterexamples above.

C Derivative-free implementation of MA-ALDI

From an inverse problem perspective, the interacting Langevin system (8) can also be viewed as modification of the ensemble Kalman inversion (EKI) (Iglesias et al. (2013)) and allows for derivative-free implementations avoiding the computation of $\nabla \log \pi$. Let π be of the form

$$\pi(x) \propto \exp(-\frac{1}{2}\|\Gamma^{-1/2}(G(x) - z)\|^2),$$

where $G : \mathbb{R}^d \to \mathbb{R}^{d_z}$ is a possibly nonlinear differentiable mapping, $z \in \mathbb{R}^{d_z}$, and $\Gamma \in \mathcal{C}^+(\mathbb{R}^{d_z})$. Then $\nabla \log \pi$ computes as

$$\nabla \log \pi(x) \propto -\nabla G(x)\Gamma^{-1}(G(x) - z),$$

for $x \in \mathbb{R}^d$. The key idea of the derivative-free implementation is to apply the approximation

$$C(\mathbf{x})\nabla G(x)\Gamma^{-1}(G(x) - z) \approx C^{xG}(\mathbf{x})\Gamma^{-1}(G(x) - z)$$

within the particle system, where

$$C^{xG}(\mathbf{x}) := \frac{1}{M}\sum_{i=1}^{M}(G(\mathbf{x}^{(i)}) - m(G(\mathbf{x}))(\mathbf{x}^{(i)} - m(\mathbf{x}))^\top,$$
$$m(G(\mathbf{x})) := \frac{1}{M}\sum_{i=1}^{M} G(x^{(i)}).$$

Using a second order Taylor expansion, one can show that the approximation error scales with the spread of the particle system (Weissmann (2022), Lemma 4.5). The derivative-free modification of (8)

$$\mathrm{d}X_t^{(i)} = -C^{xG}(\boldsymbol{X}_t)\Gamma^{-1}(G(X_t^{(i)}) - z) + \sqrt{2C(\boldsymbol{X}_t)}\mathrm{d}B_t^{(i)},$$
$$i \in \{1, \ldots, M\}, \tag{55}$$

is often referred to the Ensemble Kalman sampler (EKS) (Garbuno-Inigo et al. (2020a)). Note that for linear maps G both (8) and (55) coincide. For nonlinear maps G and hence, non-Gaussian distributions π, localisation of the empirical covariance helps to improve the approximation of π through the EKS (Reich and Weissmann (2021)). Finally, one can similarly apply our proposed Metropolis-adjusted scheme to EKS in order to reduce the resulting bias through avoiding computation of derivatives.

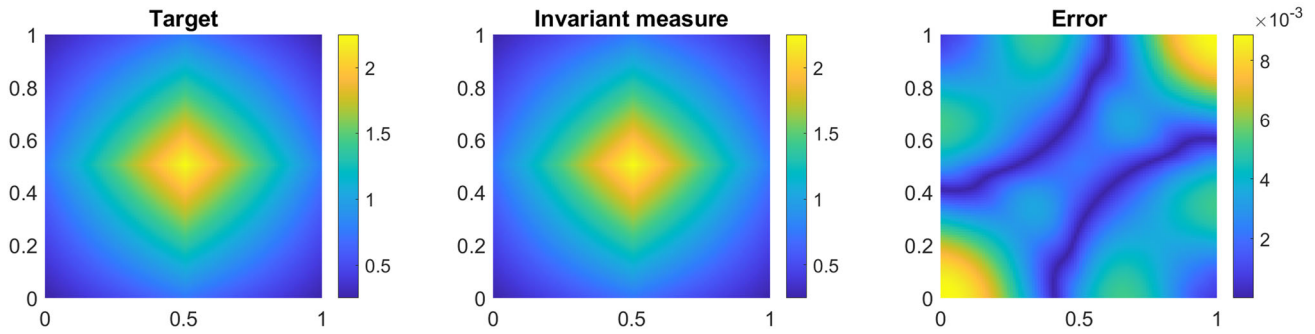


Fig. 14 Comparison true product target and numerically computed invariant measure of simultaneous particle-wise Metropoliziation $\mathbf{P}_{\text{sim}}$ for $M = 2$ particles in Example B.3

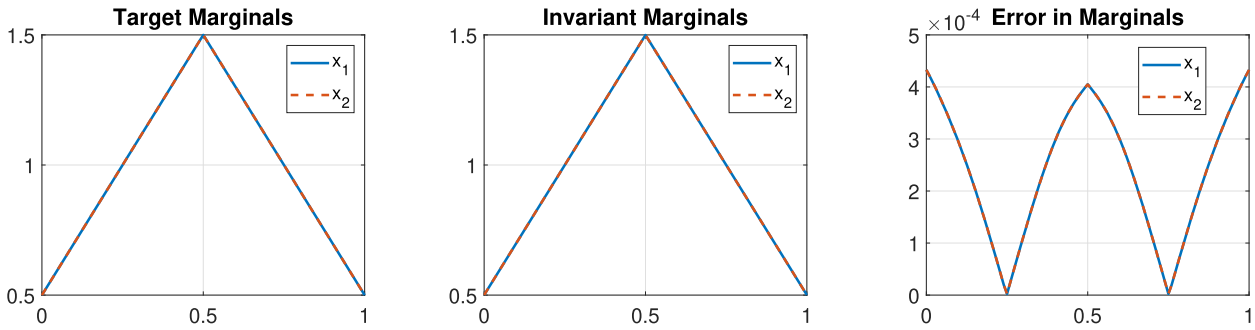


Fig. 15 Comparison true product target and numerically computed invariant measure of simultaneous particle-wise Metropoliziation $\mathbf{P}_{\text{sim}}$ for $M = 2$ particles in Example B.3

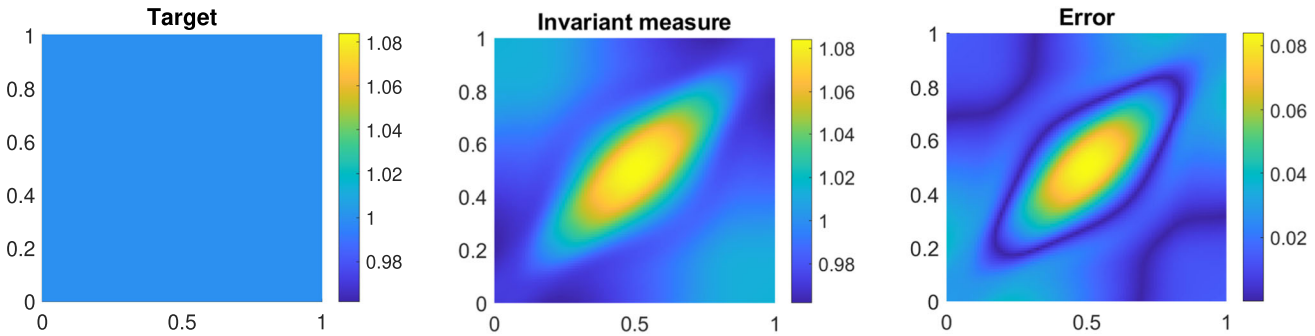


Fig. 16 Comparison true product target and numerically computed invariant measure of simultaneous particle-wise Metropoliziation $\mathbf{P}_{\text{sim}}$ for $M = 2$ particles in Example B.4

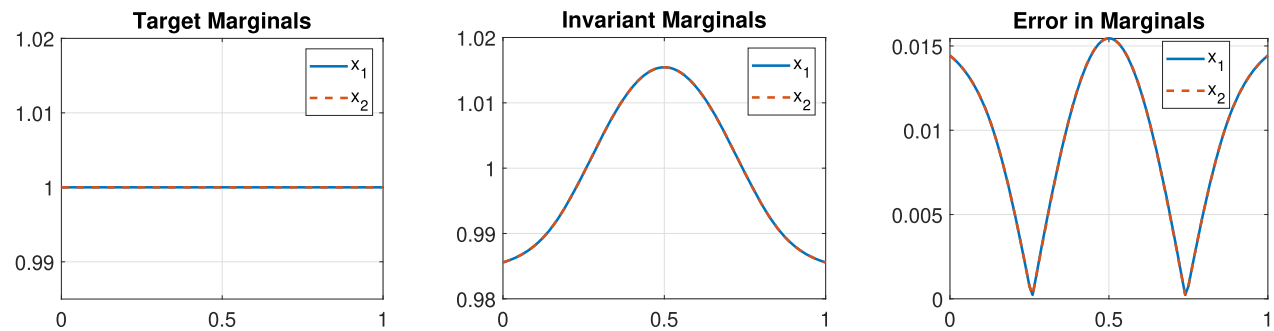


Fig. 17 Comparison true product target and numerically computed invariant measure of simultaneous particle-wise Metropoliziation $\mathbf{P}_{\text{sim}}$ for $M = 2$ particles in Example B.4

Acknowledgements We thank Daniel Rudolf for very helpful and enduring discussions. We thank the anonymous reviewer who suggested to parallelize within each block as discussed in Sect. 2.1.

Author Contributions All authors contributed equally to this work.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Pure and Applied Mathematics. Academic Press, (2003). https://books.google.de/books?id=R5A65Koh-EoC

Besag, J.: Discussion of "Representations of knowledge in complex systems". J. Roy. Stat. Soc. Ser. B **56**(4), 591–592 (1994)

Braak, C.J.F.T.: A Markov chain monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. Stat. Comput. **16**, 239–249 (2006). https://doi.org/10.1007/s11222-006-8769-1

Brooks, S., Gelman, A., Jones, G., Meng (Eds.), X.-L.: Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, New York, NY (2011). https://doi.org/10.1201/b10905

Carrillo, J.A., Hoffmann, F., Stuart, A.M., Vaes, U.: Consensus-based sampling. Stud. Appl. Math. **148**(3), 1069–1140 (2022). https://doi.org/10.1111/sapm.12470

Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.: Neural ordinary differential equations. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS'18, pp. 6572–6583. Curran Associates Inc., Red Hook, NY, USA (2018)

Christen, J.A., Fox, C.: A general purpose sampling algorithm for continuous distributions (the t-walk). Bayesian Anal. **5**(2), 263–281 (2010). https://doi.org/10.1214/10-BA603

Clarté, G., Diez, A., Feydy, J.: Collective proposal distributions for nonlinear MCMC samplers: mean-field theory and fast implementation. Electr. J. Stat. **16**(2), 6395–6460 (2022). https://doi.org/10.1214/22-EJS2091

Cotter, S., Roberts, G., Stuart, A., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. Stat. Sci. **28**(3), 283–464 (2013). https://doi.org/10.1214/13-STS421

Coullon, J., Webber, R.J.: Ensemble sampler for infinite-dimensional inverse problems. Stat. Comput. **31**, 28 (2021). https://doi.org/10.1007/s11222-021-10004-y

Craiu, Radu V., J.R., Yang, C,: Learn from thy neighbor: parallel-chain and regional adaptive MCMC. J. Am. Stat. Assoc. **104**(488), 1454–1466 (2009). https://doi.org/10.1198/jasa.2009.tm08393

Dolgov, S., Anaya-Izquierdo, K., Fox, C., Scheichl, R.: Approximation and sampling of multivariate probability distributions in the tensor train decomposition. Stat. Comput. **30**(3), 603–625 (2020). https://doi.org/10.1007/s11222-019-09910-z

Duncan, A., Nüsken, N., Szpruch, L.: On the geometry of Stein variational gradient descent. J. Mach. Learn. Res. **24**(56), 1–39 (2023)

Dunlop, M.M., Stadler, G.: A gradient-free subspace-adjusting ensemble sampler for infinite-dimensional Bayesian inverse problems. arXiv:2202.11088 (2022) https://doi.org/10.48550/ARXIV.2202.11088

Gallego, V., Insua, D.R.: Stochastic gradient MCMC with repulsive forces. ArXiv arXiv:1812.00071 (2018)

Garbuno-Inigo, A., Hoffmann, F., Li, W., Stuart, A.M.: Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. SIAM J. Appl. Dyn. Syst. **19**(1), 412–441 (2020a). https://doi.org/10.1137/19M1251655

Garbuno-Inigo, A., Nüsken, N., Reich, S.: Affine invariant interacting Langevin dynamics for Bayesian inference. SIAM J. Appl. Dyn. Syst. **19**(3), 1633–1658 (2020b). https://doi.org/10.1137/19M1304891

Gilks, W.R., Roberts, G.O., George, E.I.: Adaptive direction sampling. J. R. Stat. Soc. Ser. D **43**(1), 179–189 (1994). (**Accessed 2024-06-25**)

Goodman, J., Weare, J.: Ensemble samplers with affine invariance. Comm. App. Math. and Comp. Sci. (2010). https://doi.org/10.2140/camcos.2010.5.65

Gorham, J., Mackey, L.: Measuring sample quality with kernels. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1292–1301. PMLR, (2017). https://proceedings.mlr.press/v70/gorham17a.html

Grenander, U., Miller, M.I.: Representations of knowledge in complex systems. J. Roy. Statist. Soc. Ser. B **56**(4), 549–603 (1994). https://doi.org/10.1111/j.2517-6161.1994.tb02000.x

Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1), 97–109 (1970). https://doi.org/10.2307/2334940

Huang, D.Z., Huang, J., Reich, S., Stuart, A.M.: Efficient derivative-free Bayesian inference for large-scale inverse problems. Inverse Prob. **38**(12), 125006 (2022). https://doi.org/10.1088/1361-6420/ac99fa

Iglesias, M.A., Law, K.J.H., Stuart, A.M.: Ensemble Kalman methods for inverse problems. Inverse Prob. **29**(4), 045001 (2013). https://doi.org/10.1088/0266-5611/29/4/045001

Jaini, P., Selby, K.A., Yu, Y.: Sum-of-squares polynomial flow. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 3009–3018. PMLR, (2019). https://proceedings.mlr.press/v97/jaini19a.html

Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**, 1–17 (1998). https://doi.org/10.1137/S0036141096303359

Kipnis, C., Varadhan, S.R.S.: Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. Comm. Math. Phys. **104**(1), 1–19 (1986)

Korba, A., Salim, A., Arbel, M., Luise, G., Gretton, A.: A non-asymptotic analysis for Stein variational gradient descent. In: Advances in Neural Information Processing Systems, vol. 33, pp. 4672–4682. Curran Associates,

Inc., (2020). https://proceedings.neurips.cc/paper/2020/file/3202111cf90e7c816a472aaceb72b0df-Paper.pdf
Leimkuhler, B., Matthews, C., Weare, J.: Ensemble preconditioning for Markov chain Monte Carlo simulation. Stat. Comput. **28**(2), 277–290 (2018). https://doi.org/10.1007/s11222-017-9730-1
Liu, Q., Wang, D.: Stein variational gradient descent: A general purpose Bayesian inference algorithm. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16, pp. 2378–2386. Curran Associates Inc., Red Hook, NY, USA (2016)
Liu, Q.: Stein variational gradient descent as gradient flow. In: Advances in Neural Information Processing Systems 30, pp. 3115–3123. Curran Associates, Inc., (2017). http://papers.nips.cc/paper/6904-stein-variational-gradient-descent-as-gradient-flow.pdf
Markowich, P.A., Villani, C.: On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. vol. 19, pp. 1–29 (2000). VI Workshop on Partial Differential Equations, Part II (Rio de Janeiro, 1999)
Marzouk, Y., Moselhy, T., Parno, M., Spantini, A.: In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) Sampling via Measure Transport: An Introduction, pp. 1–41. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-11259-6_23-1
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. **21**(6), 1087–1092 (1953). https://doi.org/10.1063/1.1699114
Nüsken, N., Reich, S.: Note on interacting Langevin diffusion: Gradient structure and ensemble Kalman sampler. Technical Report arXiv:1908.10890v1, University of Potsdam (2019)
Nüsken, N., Renger, D.R.M.: Stein variational gradient descent: Many-particle and long-time asymptotics. Found. Data Sci. (2023). https://doi.org/10.3934/fods.2022023
Pathiraja, S., Reich, S., Stannat, W.: Mckean-Vlasov SDEs in nonlinear filtering. SIAM J. Control. Optim. **59**(6), 4188–4215 (2021). https://doi.org/10.1137/20M1355197
Pavliotis, G.A.: Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin Equations. Texts Appl. Math. (2014). https://doi.org/10.1007/978-1-4939-1323-7
Pinnau, R., Totzeck, C., Tse, O., Martin, S.: A consensus-based model for global optimization and its mean-field limit. Math. Models Methods Appl. Sci. **27**(1), 183–204 (2017). https://doi.org/10.1142/S0218202517400061
Reich, S., Weissmann, S.: Fokker-Planck particle systems for Bayesian inference: computational approaches. SIAM/ASA J. Uncertain. Quantif. **9**(2), 446–482 (2021). https://doi.org/10.1137/19M1303162
Rezende, D., Mohamed, S.: Variational inference with normalizing flows. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 1530–1538. PMLR, Lille, France (2015). http://proceedings.mlr.press/v37/rezende15.html
Robert, C.P., Casella, G.: Monte Carlo statistical methods. Texts Stat. (2004). https://doi.org/10.1007/978-1-4757-4145-2
Roberts, G.O., Rosenthal, J.S.: Optimal scaling for various Metropolis-Hastings algorithms. Stat. Sci. **16**(4), 351–367 (2001). https://doi.org/10.1214/ss/1015346320
Roberts, G.O., Rosenthal, J.S.: General state space Markov chains and MCMC algorithms. Probab. Surv. **1**, 20–71 (2004). https://doi.org/10.1214/154957804100000024
Roberts, G.O., Rosenthal, J.S.: Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. Ann. Appl. Probab. **16**(4), 2123–2139 (2006). https://doi.org/10.1214/105051606000000510
Roberts, G.O., Tweedie, R.L.: Exponential convergence of Langevin distributions and their discrete approximations. Bernoulli **2**(4), 341–363 (1996). https://doi.org/10.2307/3318418
Rossky, P.J., Doll, J.D., Friedman, H.L.: Brownian dynamics as smart Monte Carlo simulation. J. Chem. Phys. **69**(10), 4628–4633 (1978). https://doi.org/10.1063/1.436415
Rudolf, D., Sprungk, B.: Robust random walk-like Metropolis–Hastings algorithms for concentrating posteriors. arXiv:2202.12127 (2022)
Rudolf, D., Sprungk, B.: On a generalization of the preconditioned Crank-Nicolson Metropolis algorithm. Found. Comput. Math. **18**, 309–343 (2018). https://doi.org/10.1007/s10208-016-9340-x
Shi, J., Mackey, L.: A finite-particle convergence rate for stein variational gradient descent. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems, vol. 36, pp. 26831–26844. Curran Associates, Inc., (2023). https://proceedings.neurips.cc/paper_files/paper/2023/file/54e5d7af6250ccab796ad7fe75663ba5-Paper-Conference.pdf
Vempala, S., Wibisono, A.: Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates, Inc., (2019). https://proceedings.neurips.cc/paper/2019/file/65a99bb7a3115fdede20da98b08a370f-Paper.pdf
Weissmann, S.: Gradient flow structure and convergence analysis of the ensemble Kalman inversion for nonlinear forward models. Inverse Prob. **38**(10), 105011 (2022). https://doi.org/10.1088/1361-6420/ac8bed
Zhang, Y., Sutton, C.: Quasi-newton methods for markov chain monte carlo. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 24. Curran Associates, Inc., (2011). https://proceedings.neurips.cc/paper_files/paper/2011/file/e702e51da2c0f5be4dd354bb3e295d37-Paper.pdf
Zhu, M.: Sample adaptive mcmc. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates, Inc., (2019). https://proceedings.neurips.cc/paper_files/paper/2019/file/2cfa8f9e50e0f510ede9d12338a5f564-Paper.pdf