

## Parallel MCMC algorithms: theoretical foundations, algorithm design, case studies

NATHAN E. GLATT-HOLTZ

*Department of Statistics, Indiana University, Bloomington, IN 47405, USA*

ANDREW J. HOLBROOK\*

*Department of Biostatistics, University of California, Los Angeles, CA, 90095, USA*

\*Corresponding author: [aholbroo@g.ucla.edu](mailto:aholbroo@g.ucla.edu)

JUSTIN A. KROMETIS

*National Security Institute, Virginia Tech, Blacksburg, VA 24060, USA*

*Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA*

AND

CECILIA F. MONDAINI

*Department of Mathematics, Drexel University, Philadelphia, PA 19104, USA*

[Received on 3 June 2023; revised on 20 March 2024; accepted on 15 July 2024]

Parallel Markov Chain Monte Carlo (pMCMC) algorithms generate clouds of proposals at each step to efficiently resolve a target probability distribution  $\mu$ . We build a rigorous foundational framework for pMCMC algorithms that situates these methods within a unified ‘extended phase space’ measure-theoretic formalism. Drawing on our recent work that provides a comprehensive theory for reversible single-proposal methods, we herein derive general criteria for multiproposal acceptance mechanisms that yield ergodic chains on general state spaces. Our formulation encompasses a variety of methodologies, including proposal cloud resampling and Hamiltonian methods, while providing a basis for the derivation of novel algorithms. In particular, we obtain a top-down picture for a class of methods arising from ‘conditionally independent’ proposal structures. As an immediate application of this formalism, we identify several new algorithms including a multiproposal version of the popular preconditioned Crank–Nicolson (pCN) sampler suitable for high- and infinite-dimensional target measures that are absolutely continuous with respect to a Gaussian base measure. To supplement the aforementioned theoretical results, we carry out a selection of numerical case studies that evaluate the efficacy of these novel algorithms. First, noting that the true potential of pMCMC algorithms arises from their natural parallelizability and the ease with which they map to modern high-performance computing architectures, we provide a limited parallelization study using TENSORFLOW and a graphics processing unit to scale pMCMC algorithms that leverage as many as 100k proposals at each step. Second, we use our multiproposal pCN algorithm (mpCN) to resolve a selection of problems in Bayesian statistical inversion for partial differential equations motivated by fluid measurement. These examples provide preliminary evidence of the efficacy of mpCN for high-dimensional target distributions featuring complex geometries and multimodal structures.

**Keywords:** Parallel (Multiproposal), Markov Chain Monte Carlo (pMCMC), preconditioned Crank–Nicolson (pCN), Hamiltonian Monte Carlo (HMC), simplicial sampler, Bayesian statistical inversion, graphics processing units (GPU), high-performance computing, Metropolis–Hastings kernels.

## 1. Introduction

The efficient generation of random samples is a central task within many of the quantitative sciences. The workhorse of Bayesian statistics and statistical physics, Markov chain Monte Carlo (MCMC) comprises a large class of algorithms for sampling from arbitrarily complex or high-dimensional probability distributions. The Metropolis–Hastings method (MH) (Hastings, 1970; Metropolis *et al.*, 1953) stands as the seminal MCMC algorithm, and its basic operation underlies most of the MCMC techniques developed to this day. At each iteration, these algorithms choose the next Markov chain state by (1) randomly generating a proposal state according to an auxiliary proposal distribution and (2) accepting or rejecting the proposed state according to a carefully derived threshold. Powerful, modern algorithms such as Hamiltonian (hybrid) Monte Carlo (HMC) (Duane *et al.*, 1987; Neal, 2011) share the same two-step approach but use additional deterministic machinery to guide the random proposal and obtain, e.g., lower autocorrelation between samples. Recent independent works (Andrieu *et al.*, 2020; Glatt–Holtz *et al.*, 2023; Neklyudov *et al.*, 2020) develop an all-encompassing mathematical framework that describes essentially any (reversible) single proposal MH algorithm using three ingredients: a random proposal, an involution on an extended phase space and an accept–reject step. This unified framework illuminates under-appreciated relationships between a variety of known algorithms while providing a means for deriving new methods.

Unfortunately, the overall structure of these MH extensions may fail to fully exploit contemporary parallel computing resources such as multi-core central processing units (CPUs) and many-core graphics processing units (GPUs). While model-specific algorithmic subroutines such as log-likelihood and log-likelihood gradient evaluations sometimes admit parallelization (Holbrook *et al.*, 2022a,b, 2021a,b), the algorithms’ generally sequential nature can lead to under-utilization of increasingly widespread hardware (Brockwell, 2006). Parallel MCMC algorithms (pMCMC) offer a top-down approach to exploiting such infrastructures through their use of multiple proposals at each step. The multiproposal structure makes pMCMC algorithms amenable to conventional parallelization resources such as GPUs (see Section 5.1), and Holbrook (2023b) even demonstrates advantages for implementations that leverage quantum computing. While general, efficient and user-friendly parallel implementations of pMCMC algorithms remain an engineering challenge, there is reason to believe that the pMCMC paradigm may prove useful when other, more prevalent MCMC paradigms—such as, e.g., gradient-based sampling or parallelization using multiple chains (Gelman & Rubin, 1992)—encounter their own challenges.

Whereas gradient-based MCMC methods such as HMC or Metropolis-Adjusted Langevin Algorithm (MALA) can scale inference to extremely high-dimensional settings, they are not always appropriate. Sometimes the gradient is either not available, is computationally prohibitive or involves difficult derivations (see Section 5.2.3). More generally, these algorithms struggle to tackle distributions with challenging geometries, even though there are sometimes solutions. On the one hand, if a target is roughly Gaussian—albeit with an ill-conditioned covariance matrix—then one may effectively implement HMC using an adaptive mass matrix (Holbrook *et al.*, 2022b). On the other hand, if a target is both ill-conditioned and strongly non-Gaussian, then there are few gradient-based options beside Riemannian HMC (Girolami & Calderhead, 2011), a user-intensive method that scales poorly to high dimensions,. Finally, even HMC can lose its competitive advantage over vanilla MH in multimodal settings (Mangoubi *et al.*, 2018).

Parallel implementation of multiple MCMC chains is often a beneficial strategy that adequately leverages hardware resources such as multiple cluster nodes or CPU cores, but the paradigm is not without its own difficulties. First, a lack of communication between chains means that one chain’s progress cannot aid that of another. This is a problem because there is usually little benefit to combining

multiple poorly-mixing chains, where one may measure the quality of the combined Monte Carlo estimator using, say, the potential scale reduction factor (R-hat) of Gelman & Rubin (1992). Thus, the parallel chain strategy may fail when a target displays difficult geometry that confounds individual chains. When such targets are also high-dimensional, memory limitations may inhibit inference:  $C$  chains of length  $S$  defined on a  $D$ -dimensional state space require  $\mathcal{O}(SDC)$  storage. Finally, the multi-chain strategy may also fail when the target distribution is multimodal, as Neal (1996) notes:

*Unfortunately, multiple independent runs will not, in general, produce a sample in which each mode is fairly represented, because the probability of a run reaching a mode will depend more on the mode's 'basin of attraction' than on the total probability in the vicinity of the mode.*

Accordingly, we cast pMCMC as a paradigm that both complements other MCMC paradigms and provides further opportunities for the use of high-performance computing resources in MCMC. Notwithstanding a significant and growing recent literature on pMCMC and related methods, cf. (Calderhead, 2014; Delmas & Jourdain, 2009; Frenkel, 2004; Holbrook, 2023a; Liu *et al.*, 2000; Luo & Tjelmeland, 2019; Neal, 2003; Schwedes & Calderhead, 2021; Tjelmeland, 2004), the subject remains underdeveloped leaving a broad scope for the development and analysis of novel pMCMC methods.

This work provides a unified theoretical foundation for algorithm design in this important direction in the sampling literature, encompassing algorithms that use single (Section 2) or multiple (Section 3) jumps within each proposal set. Our approach then allows us to identify new algorithms while placing existing methods in a broader context and rigorously justifying their reversibility (Section 4). Finally, we provide a series of numerical case studies (Section 5) that demonstrate the efficacy of our new methods and explore the tuning of algorithmic parameters. We continue this introduction with a comprehensive summary of our contributions (Section 1.1) before concluding with an overview of the existing pMCMC literature (Section 1.2).

### 1.1 Contribution overview

Our first contribution is to show that the unified framework developed recently in Glatt-Holtz *et al.* (2023) can be fruitfully and nontrivially extended to encompass a broad class of pMCMC methods. We develop an involutive theory of pMCMC that (1) firmly places these multiproposal algorithms in a broad measure-theoretic context and (2) makes clear when a pMCMC algorithm is unbiased or reversible with respect to a given target distribution  $\mu$ . Note that our formalism includes extensions that allow for proposal cloud resampling.

In Section 2, we begin by introducing a broad class of Markovian kernels of the form

$$P^{\alpha, S, \mathcal{V}}(\mathbf{q}, d\tilde{\mathbf{q}}) = \sum_{j=0}^p \int_Y \alpha_j(\mathbf{q}, \mathbf{v}) \delta_{\Pi_1 \circ S_j(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}), \quad (1.1)$$

where  $\mathbf{q}$  is the current state, the  $\alpha_j$ 's determine the acceptance probabilities, the  $S_j$ 's are involutive operators acting on an extended abstract phase space  $X \times Y$ ,  $\Pi_1$  is the projection onto the first component, namely  $\Pi_1(\mathbf{q}, \mathbf{v}) = \mathbf{q}$ , and  $\mathcal{V}$  is the multiproposal mechanism. Algorithmically, the kernel  $P^{\alpha, S, \mathcal{V}}$  first draws a sample  $\mathbf{v}$  from  $\mathcal{V}(\mathbf{q}, d\mathbf{v})$ , then produces a cloud of proposals

$$(\Pi_1 \circ S_0(\mathbf{q}, \mathbf{v}), \dots, \Pi_1 \circ S_p(\mathbf{q}, \mathbf{v})),$$

and finally draws the next state from this cloud with probabilities

$$(\alpha_0(\mathbf{q}, \mathbf{v}), \dots, \alpha_p(\mathbf{q}, \mathbf{v})).$$

As we illustrate below, (1.1) encompasses a very broad class of pMCMC samplers inclusive of both random walk and Hamiltonian-type methods.

In a series of results (Theorem 2.2, Corollary 2.4, Corollary 2.6, Theorem 2.10 and Theorem 2.11), we identify various conditions on  $\alpha_j$ ,  $S_j$  and  $\mathcal{V}$  such that, for a given target measure  $\mu$ , we have

$$P^{\alpha, S, \mathcal{V}}(\mathbf{q}, d\tilde{\mathbf{q}})\mu(d\mathbf{q}) = P^{\alpha, S, \mathcal{V}}(\tilde{\mathbf{q}}, d\mathbf{q})\mu(d\tilde{\mathbf{q}}), \quad (1.2)$$

i.e.,  $P^{\alpha, S, \mathcal{V}}$  is reversible and hence invariant with respect to a given target  $\mu$ . Notably, Corollary 2.6 includes as a special case any single proposal MH-type algorithm that is reversible with respect to a given target. As such, the theory we develop here may be seen as a strict generalization of Glatt–Holtz *et al.* (2023).

On the other hand, it is worth highlighting some novel insights and challenges to developing the multiproposal theory in contrast to (Andrieu *et al.*, 2020; Glatt–Holtz *et al.*, 2023; Neklyudov *et al.*, 2020). The single proposal setting rightly emphasizes the traditional MH acceptance mechanism, a choice that has shown to be optimal (Tierney, 1998, Section 3). This mechanism frames Corollary 2.6 herein. By contrast, a Barker-type acceptance mechanism (Barker, 1965) reveals its salience in the pMCMC setting, and our results Corollary 2.4, Theorem 2.10 and Theorem 2.11 lead to Algorithms 4 and 5 later in Section 4.

To make this more explicit, we consider pMCMC algorithms that involve acceptance probabilities for the  $j$ th proposed state taking the form

$$\alpha_j(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) = \frac{\pi(\mathbf{q}_j)}{\sum_{k=0}^p \pi(\mathbf{q}_k)}, \quad (1.3)$$

where  $\mathbf{q}_0$  is the current state of the chain,  $(\mathbf{q}_1, \dots, \mathbf{q}_p)$  is the cloud of proposed states made according to some kernel  $\mathcal{V}(\mathbf{q}_0, d\mathbf{q}_1, \dots, d\mathbf{q}_p)$ , and the involutions  $S_j$  are the ‘flip’ operators exchanging the  $j$ th and zeroth elements (see (2.13) below). Here the function  $\pi$  may be the density of the target measure  $\mu$  or taken to be proportional to the likelihood in Bayesian settings. The salience of (1.3) is immediately intuitive—it yields high acceptance probabilities for points in the proposal cloud that inhabit high probability regions of the target distribution.

Of course, we do not expect  $\mathcal{V}$  in conjunction with  $(\alpha_0, \dots, \alpha_p)$  defined according to (1.3) to yield an unbiased algorithm for a given target  $\mu$  in general. But, in Section 2.3, we are able to draw the following wide ranging conclusions by restricting attention to a certain class of ‘conditionally independent’ proposal kernels of the form

$$\mathcal{V}(\mathbf{q}_0, d\mathbf{q}_1, \dots, d\mathbf{q}_p) = \int \prod_{j=1}^p Q(\tilde{\mathbf{q}}_0, d\mathbf{q}_j) \bar{Q}(\mathbf{q}_0, d\tilde{\mathbf{q}}_0), \quad (1.4)$$

where  $Q, \bar{Q}$  are single element Markov kernels on the parameter space on which  $\mu$  sits. Note that, for such kernels  $\mathcal{V}$ , one obtains samples from the current state  $\mathbf{q}_0$  by first producing  $\tilde{\mathbf{q}}_0$  from  $\bar{Q}(\mathbf{q}_0, \cdot)$  and then drawing  $(\mathbf{q}_1, \dots, \mathbf{q}_p)$  independently from  $Q(\tilde{\mathbf{q}}_0, \cdot)$ . For such proposal structures  $\mathcal{V}$ , we are able to write down a Barker-like acceptance probability such that the resulting Metropolized system is unbiased with respect to essentially any target measure  $\mu$ . See Theorem 2.10, for our precise formulation.

Interestingly, the naive choice  $\bar{Q}(\mathbf{q}_0, d\tilde{\mathbf{q}}_0) = \delta_{\mathbf{q}_0}(d\tilde{\mathbf{q}}_0)$  that reduces  $\mathcal{V}$  to  $\prod_{j=1}^p Q(\mathbf{q}_0, d\mathbf{q}_j)$ , i.e.,  $p$  independent proposals around the current state  $\mathbf{q}_0$ , leads to a complicated and potentially computationally onerous acceptance probability  $\alpha_j$ . See, for example, (4.8) in Section 4. On the other hand, inspired by

Tjelmeland (2004), if one carefully chooses  $\bar{Q}$  in relation to  $Q$  we show that one obtains an acceptance probability of the form (1.3). See Theorem 2.11 for precise formulations. Note that in many cases  $\bar{Q}$  is readily identifiable from  $Q$ . Indeed, as observed previously in Tjelmeland (2004) in the specific case of Gaussian proposals, we may take  $Q = \bar{Q}$  under a minimal symmetry condition on  $Q$ . Taken further this case leads us to novel ‘infinite dimensional’ algorithms as we describe presently (see Algorithm 5). Furthermore, in the finite dimensional setting we show that by considering *any* probability density  $r : \mathbb{R}^N \rightarrow \mathbb{R}^+$  we obtain suitable  $Q, \bar{Q}$  by taking

$$Q(\mathbf{q}, d\tilde{\mathbf{q}}) = r(\mathbf{q} - \tilde{\mathbf{q}})d\tilde{\mathbf{q}}, \quad \bar{Q}(\mathbf{q}, d\tilde{\mathbf{q}}) = r(\tilde{\mathbf{q}} - \mathbf{q})d\tilde{\mathbf{q}}, \quad (1.5)$$

thus accommodating non-symmetric proposal structures (see Algorithm 4).

Another natural consideration in developing pMCMC methods is to consider the possibility of performing multiple finite state space jumps between the elements of a proposal cloud. After all, when target evaluations at proposal points represents the main computational burden, there may be relatively little computational overhead associated with additional inter-proposal jumps. In Section 3, we extend the formalism of Section 2 to support multiple jumps within the proposal set. Here our framework provide a rigorous justification for methods advanced in Calderhead (2014).

In summary, the formalisms we develop in Section 2, Section 3 provide a unified basis for concrete algorithm design. In Section 4, we begin to explore this large space of possible methods and to develop a selection of applications within our paradigm. Our most general results Theorem 2.2, Theorem 3.3 and our observations around kernels of the form (1.4), Theorem 2.10, Theorem 2.11, Corollary 2.12, yield the novel and immediately applicable algorithms in Algorithm 4 and in Algorithm 5, Algorithm 6. Meanwhile in Section 4.3, we consider HMC variants that use multiple integration times. This study includes a rigorous and more systematic framing of a method suggested previously in Calderhead (2014). Section 4.4 revisits the simplicial sampler recently introduced in Holbrook (2023a).

Algorithms 5 and 6 provide multiproposal generalizations of the so called preconditioned Crank–Nicolson (pCN) algorithm (Beskos *et al.*, 2008; Cotter *et al.*, 2013). These methods address a class of target measures of the form

$$\mu(d\mathbf{q}) = Z^{-1} \exp(-\phi(\mathbf{q}))\mu_0(d\mathbf{q}), \quad (1.6)$$

where  $\mu_0$  is a centered Gaussian base measure. Such measures  $\mu$  appear naturally in a variety of high- and infinite-dimensional settings (Dashti & Stuart, 2017; Stuart, 2010). Note that while practical implementation of such schemes requires truncation to finite dimensions, deriving algorithms tailored to infinite-dimensional target measures can partially beat the curse of dimensionality in the sense that mixing rates for truncations of the full problem do not depend on the order of the truncation; see Hairer *et al.* (2014), Glatt-Holtz & Mondaini (2021) for some rigorous results in this direction.

The pCN algorithm—so named because it arises from a Crank–Nicolson temporal discretization of an Ornstein–Uhlenbeck process preconditioned by the covariance  $\mathcal{C}$  of  $\mu_0$ —exploits a proposal kernel of the form

$$Q_{pCN}(\mathbf{q}, d\tilde{\mathbf{q}}) = \mathbb{P}(\rho\mathbf{q} + \sqrt{1 - \rho^2}\mathbf{v} \in d\tilde{\mathbf{q}}), \quad \text{where } \mathbf{v} \sim \mu_0. \quad (1.7)$$

Here,  $\rho \in [0, 1]$  is an algorithmic parameter dictating the degree of aggressiveness of the proposal. In the context of (1.4), we show that combining  $Q = \bar{Q} := Q_{pCN}$  and an acceptance probability of the form (1.3) with  $\pi(\mathbf{q}) = \exp(-\phi(\mathbf{q}))$  yields a pMCMC algorithm that is reversible (and hence invariant) with respect to measures of the form (1.6). We dub this method the multiproposal-pCN (mpCN) algorithm. Notably, mpCN stands as a gradient-free methodology that is easy to implement in comparison to other ‘Hilbert space’ methods used for certain parameter estimation problems for partial differential equations

(PDEs) considered in statistical inversion. Here, the existing alternatives to vanilla pCN algorithm,  $\infty$  MALA and  $\infty$  HMC (Beskos *et al.*, 2011; Cotter *et al.*, 2013), typically require adjoint methods and may be complicated or practically impossible to implement. Following the discussion at the beginning of this introduction, even properly implemented versions of these algorithms can suffer under difficult target geometries.

Armed with Algorithms 4 and 5, we conclude this contribution by providing a series of numerical case studies in Section 5. These studies comprise preliminary investigations into the efficacy and the tuning of algorithmic parameters for these pMCMC algorithms. We select these studies with an eye towards effectively leveraging modern computer architectures while identifying significant domains of immediate applicability. In Section 5.1, we demonstrate the natural parallelizability of pMCMC with a limited high-performance computing study. Here, we use a GPU to power Algorithm 4 with up to 100,000 proposals at each iteration. With a simple TENSORFLOW implementation, wall times per iteration for a GPU are shown to be orders of magnitude smaller than those for a comparable sequential implementation. Furthermore, we show that these speedups increase with target dimensionality. Under a fixed computational budget, GPU based parallelization also confers greater sampling efficiency for massively multimodal target distributions, with speedups again increasing with mixture component counts. But the naturally parallel structure of pMCMC makes it amenable to other parallel computing techniques as well. In a companion paper (Holbrook, 2023b), we combine quantum optimization with the Gumbel-max trick and show that quantum algorithms deliver significant theoretical speedups for pMCMC algorithms similar to Algorithm 4.

To test efficacy of mpCN, namely Algorithm 5, in Section 5.2 we consider three stylized nonlinear statistical inverse problems (Dashti & Stuart, 2017; Kaipio & Somersalo, 2006; Stuart, 2010) motivated by fluid measurement, drawing on our concurrent and recent work (Borggaard *et al.*, 2020, 2023). For all of these problems our aim is to provide a principled estimate for an unknown parameter  $\mathbf{q}$  sitting in a Hilbert space  $X$  from the observation of data  $\mathcal{Y}$  described by  $\mathcal{Y} = \mathcal{G}(\mathbf{q}) + \eta$ . Here  $\mathcal{G}$  is a forward map determined from a physical model defined up to the unknown parameter  $\mathbf{q}$ . The additive  $\eta$  is a probabilistic quantity representing measurement error. We in effect invert  $\mathcal{G}$  by placing a (Gaussian) prior  $\mu_0$  on  $\mathbf{q}$  and then invoking Bayes theorem to obtain a posterior of the form (1.6) defined in terms of  $\mathcal{G}$  and the distribution of  $\eta$ . Note that, in each of the problems we present,  $\mathcal{G}$  is a nonlinear map from higher dimensional unknown parameter space to lower dimensional collections of observations so that we are addressing a severely ill-posed inverse problem resulting in a non-Gaussian posterior target measure  $\mu$ ; see Figures 4 and C18 below. Furthermore, the latter two problems presented in Section 5.2.2 and Section 5.2.3 involve a naturally infinite-dimensional parameter space and feature forward maps involving the numerically expensive resolution of a partial differential equation.

As our first example we consider a problem that mimics the mathematical form of PDE inverse problems of interest at a smaller scale. Our aim here is to provide ‘table-top lab’ amenable to computations that can be easily carried out on a personal computer in a matter of minutes or hours for moderate dimensional problems and thus allowing for more comprehensive parameter studies. This problem estimates the coefficients of a  $d \times d$  antisymmetric matrix  $A_{\mathbf{q}}$ —parameterized by its  $d(d - 1)/2$  independent components  $\mathbf{q}$ —from the partial, noisy observations of solutions  $\hat{\mathbf{x}}(\mathbf{q}) = \mathbf{x}(A_{\mathbf{q}})$  of the corresponding problem  $(A_{\mathbf{q}} + \kappa I)\mathbf{x} = \mathbf{g}$ , where here  $\kappa > 0$  and  $\mathbf{g} \in \mathbb{R}^d$  are additional parameters assumed to be known a priori. We find that this class of inverse problems exhibits intricate statistical structures reminiscent of far more complex PDE constrained settings they mimic. See Figure 4 below and compare with, e.g., Figure C18. We therefore think this simple toy model holds independent interest outside of our immediate considerations here.

Our other two inverse problems involve infinite-dimensional unknown parameter spaces and a forward map  $\mathcal{G}$  that involves resolution of a partial differential equation or system of partial differential equations. One infinite-dimensional problem, considered previously in Borggaard *et al.* (2020), involves the estimation of a divergence free fluid flow field  $\mathbf{v}$  through the sparse, noisy observation of a solute that is passively advected and diffuses in the fluid medium. Physically, we can think of estimating the motion of water filling a glass by observing the concentration of a dye that has been introduced. Mathematically, this may be described as observing the solution  $\theta = \theta(\mathbf{q})$  of an advection diffusion equation  $\partial_t \theta + \mathbf{q} \cdot \nabla \theta = \kappa \Delta \theta$ , which we consider here on a periodic two-dimensional domain starting from a given (known) initial condition  $\theta_0$ . Here, as a second infinite-dimensional benchmark for mpCN, we revisit a particular stylized example identified in Borggaard *et al.* (2020), which exploits natural symmetries in this model to produce complex high-dimensional statistical correlation structures that are challenging to sample from in an efficient manner.

Our final statistical inversion problem is a fluid domain shape estimation problem that we develop in concurrent work (Borggaard *et al.*, 2023). For this problem the unknown parameter  $\mathbf{q}$  specifies the shape of a domain upon which a system of PDEs governing a time stationary Stokes flow  $\mathbf{u}$  and an associated solute concentration  $\theta$  are defined. Our aim is therefore to estimate the boundary shape from sparse or volumetrically averaged observations of  $\mathbf{u}$  and  $\theta$ . The forward map  $\mathcal{G}$  therefore requires solving a system of PDEs on an irregular domain, making the derivation and implementation of an appropriate adjoint method a difficult task. Moreover, the typical implementation will involve a (typically third-party, black box) meshing algorithm to re-mesh the domain at each iteration, ruling out use of automatic differentiation algorithms. This problem is therefore a natural fit for exploration of gradient-free algorithms.

## 1.2 Literature review

The pMCMC literature presents a patchwork of independent and sometimes overlapping frameworks for using multiple proposals within a generalized MH algorithm. Although these works come from different disciplines (e.g., statistics (Neal, 2003), physics (Frenkel, 2004), probability (Delmas & Jourdain, 2009), and machine learning (Schwedes & Calderhead, 2021)), they all seek to answer a natural question: “Why not use more than one proposal within Metropolis-Hastings?” Perhaps due to the multi-disciplinary context, the different answers to this question have often come with their own terminology: *waste-recycling* (Frenkel 2004), *parallel MCMC* (Schwedes & Calderhead, 2021), *multiproposal methods* (Holbrook, 2023a), and *generalized* (Calderhead, 2014) or *multiple-try* (Luo & Tjelmeland, 2019) Metropolis-Hastings (MH) are a few. In this paper, we explicitly do not consider the well-known multiple-try MH algorithm of Liu *et al.* (2000), which randomly selects from a cloud of proposals *before* an additional MH step. Here, we are interested in methods that subsume proposal selection into a single step. Those who are interested in a thorough involutive treatment of multiple-try MH should consult (Andrieu *et al.*, 2020), who successfully fit that particular algorithm into a single-proposal framework. Whereas the general multiple-proposal framework we construct in Section 2 admits the single-proposal framework of Glatt–Holtz *et al.* (2023) as a special case, it is difficult to say definitively that there is no theoretical mechanism by which our multiple-proposal extension may be viewed as a special case of the single-proposal theory. We reserve this question for future work.

Neal (2003), Tjelmeland (2004) and Frenkel (2004) provide early contributions to the pMCMC literature by investigating MCMC algorithms that generate multiple proposals at each iteration. Neal (2003) considers the special setting of non-linear state space models and develops an algorithm for sampling from the distribution over hidden states conditioned on an observed sequence. At each iteration, the method generates a ‘pool’ of candidate sequences, an element of which is the current state, in an

iterative manner and selects from among these sequences with probability proportional to the target distribution divided by a ‘pool’ distribution that characterizes the probability of obtaining the individual proposal conditioned on the other proposals. [Neal \(2003\)](#) shows that this strategy with  $P = 10$  proposals greatly outperforms simple Metropolis steps on an iteration by iteration basis and that computing the target probabilities with the help of the linear-time forward-backward algorithm ([Scott, 2002](#)) boosts implementation speeds. [Tjelmeland \(2004\)](#) works in a more general setting and develops multiple proposal generation and multiple transition strategies. Of the latter, the paper’s ‘Transition alternative 1’ relates closely to that of [Neal \(2003\)](#), randomly accepting a proposed state with probability proportional to the target density times the probability of generating the proposal conditioned on the other proposed states. In addition to simply randomly accepting one of the many proposals at each step, [Tjelmeland \(2004\)](#) shows that a carefully weighted average of all proposals provides an unbiased estimator for an arbitrary estimand, assuming that the Markov chain has converged to the target distribution. Among other results, [Tjelmeland \(2004\)](#) shows that empirically optimal proposal scalings grow, and empirical variances of the weighted estimator decrease, with the number of proposals. Section 2, Section 3 establishes a measure-theoretic framework that reduces to the methods of [Tjelmeland \(2004\)](#) as a special case.

Working in the setting of statistical physics, [Frenkel \(2004\)](#) proposes a method called ‘waste-recycling’ that generates multiple proposals at each iteration and constructs a running, weighted estimator that is a special case of that of [Tjelmeland \(2004\)](#), with weights that are products of target probabilities and the probabilities of transitioning from all other proposed states. [Frenkel \(2004\)](#) argues for the correctness of this approach by viewing each iteration’s contribution to the running estimator as a ‘coarsening’ of  $P + 1$  states that satisfies ‘superdetailed balance’ despite not satisfying detailed balance. Finally, [Frenkel \(2004\)](#) applies waste-recycling to estimating the probability distribution over total spin for a 2D, discrete spin Ising model on a  $32^2$  spin lattice and demonstrates an over  $10^{10}$ -fold reduction in statistical error compared to Swendsen-Wang. [Delmas & Jourdain \(2009\)](#) prove consistency and asymptotic normality of the waste-recycling estimator and show that the estimator achieves a variance reduction over MH for certain selection probabilities (alternatively, weights). [Delmas & Jourdain \(2009\)](#) calls these the Boltzmann or Barker selection kernel, and they are equivalent to those of [Frenkel \(2004\)](#); [Neal \(2003\)](#); [Tjelmeland \(2004\)](#). We note that there is no fundamental need to combine the use of multiple proposals with the weighted average approach (as is done in waste-recycling): recent Markov chain importance sampling approaches demonstrate benefits of incorporating rejected proposals via a weighted estimator within the single proposal framework ([Rudolf & Sprungk, 202](#)); [Schuster & Klebanov, 2020](#)).

Next, [Calderhead \(2014\)](#) develops a distinct approach that (1) generates multiple proposals and (2) simulates a *finite* state Markov chain, the states of which are the union of the current state and proposed states, for a fixed number of iterations at each step. [Calderhead \(2014\)](#) describes criteria necessary for the finite state Markov chain transition probabilities to satisfy detailed balance and argues that traditional MH is a special case of this procedure that combines a single proposal with a single transition. [Calderhead \(2014\)](#) does not point out the fact that the specific form of these transition probabilities admits as a special case the form of the transition probabilities and weighting schemes shared by [Frenkel \(2004\)](#); [Neal \(2003\)](#); [Tjelmeland \(2004\)](#). Unlike these previous works, [Calderhead \(2014\)](#) incorporates the partially deterministic proposal schemes of the Metropolis-adjusted Langevin algorithm (MALA) ([Roberts & Stramer, 2002](#)) and HMC into his pMCMC algorithm. Applying these schemes to logistic regression and Bayesian inversion of an ODE, [Calderhead \(2014\)](#) shows, e.g., that pMCMC using MALA or adaptive transitions ([Haario \*et al.\*, 2001](#)) and 1,000 proposals enjoys 5-fold speedups over the algorithm using a single proposal and using multiple intermediate leapfrog steps from HMC transitions

as proposals leads to a roughly 60% decrease in Monte Carlo error. Unfortunately, Calderhead (2014) stops short of proving the correctness of such pMCMC algorithms that leverage extended phase space strategies, and the development of such a theoretical framework is one contribution of the present work (Section 3).

More recently, Yang *et al.* (2018) extend the waste-recycling method of Frenkel (2004) to include HMC transitions and prove that the waste-recycling estimator is unbiased and, as a Rao-Blackwellization, leads to variance reductions over the estimator of Calderhead (2014). Yang *et al.* (2018) accomplish this for two general weighting schemes, one of which is Calderhead (2014)'s generalization of, and one of which is, that of Frenkel (2004); Neal (2003); Tjelmeland (2004). Schwedes & Calderhead (2021) prove similar results to Yang *et al.* (2018) for essentially the same algorithm with the addition of an adaptive proposal mechanism extending (Haario *et al.*, 2001). Schwedes & Calderhead (2021) conduct an empirical study with important consequences for the present work and show that, within the finite state space framework of Calderhead (2014), jumping between proposed states for many iterations leads to significantly lower Monte Carlo error. Remarkably, this study uses  $P$  finite state space iterations for  $P$  proposals as baseline and increases the number of finite state space iterations to as many as  $16P$ . Extrapolating these results, one might expect that a pMCMC algorithm that uses only a single finite state space iteration at each step would generate unbiased estimators with astronomically large variances.

Whereas Calderhead (2014); Frenkel (2004); Schwedes & Calderhead (2021); Yang *et al.* (2018) generate multiple independent and identically distributed proposals, one may generate multiple proposals that share a more complex structure. Earlier in this section, we refer to the iterative strategy of Neal (2003) in the context of non-linear state space models, and Tjelmeland (2004) advances two additional proposal strategies designed to maintain exchangeability between proposals. The first of these strategies, ‘Proposal alternative 1’, (1) generates a random center for the proposal distribution that itself follows a, say, Gaussian distribution centered at the current state and (2) generates  $P$  proposals from the same distribution with updated center. The second strategy, ‘Proposal alternative 2’, cleverly enforces that the current state and all  $P$  proposals be equidistant from each other by iteratively generating and carefully manipulating all proposals. Two recent works advance the two proposal alternatives of Tjelmeland (2004). On the one hand, Luo & Tjelmeland (2019) extend the first proposal alternative to allow proposals to share a general acyclic graphical structure. On the other hand, Holbrook (2023a) develops a proposal mechanism that maintains equal distances between proposals; this *simplicial sampler* initializes all proposals as the vertices of a high-dimensional regular simplex and rotates these vertices according to the Haar distribution on the orthogonal group. The formal developments of Section 2 include general, structured proposal mechanisms, and Section 4 applies this theory to the structured proposals of Tjelmeland (2004) and Holbrook (2023a).

## 2. An abstract framework for pMCMC algorithms

In this section we introduce our abstract formulation of pMCMC algorithms along with rigorous conditions that guarantee that the generated Markov chains are ergodic or, even more, reversible. The section is organized as follows. First, Section 2.1 presents a generic definition of Metropolis-type Markov transition kernels in this multiproposal setting. Section 2.2 then establishes general conditions leading to the invariance or reversibility of such algorithms. Finally, Section 2.3 considers a special but still broad case of proposal mechanisms and identifies conditions under which corresponding acceptance probabilities assume a simplified proposal-independent expression. We refer to Appendix A for various measure-theoretic elements used herein and postpone all mathematical proofs to Appendix B.

## 2.1 Multiproposal MH Markov kernels

We formulate an abstract setting for multiproposal extended phase space MH kernels on general state spaces as follows. Throughout the manuscript, we denote by  $\text{Pr}(\mathcal{X})$  the set of all probability measures on a measurable space  $\mathcal{X}$ .

**DEFINITION 2.1** Let  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$  be measurable spaces and take  $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$  to be a Markov kernel. Namely, we suppose that  $\mathcal{V}(\mathbf{q}, d\mathbf{v}) \in \text{Pr}(Y)$  for each  $\mathbf{q} \in X$  and  $\mathbf{q} \rightarrow \mathcal{V}(\mathbf{q}, E)$  is a measurable map for each  $E \in \Sigma_Y$ . For each  $j = 0, 1, \dots, p$  consider measurable mappings  $S_j : X \times Y \rightarrow X \times Y$  and  $\alpha_j : X \times Y \rightarrow [0, 1]$  such that

$$\sum_{j=0}^p \int_Y \alpha_j(\mathbf{q}, \mathbf{v}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) = 1 \quad \text{for every } \mathbf{q} \in X. \quad (2.1)$$

Then, denoting  $\alpha = (\alpha_0, \dots, \alpha_p)$  and  $S = (S_0, \dots, S_p)$ , we define the *multiproposal MH Markov kernel given by  $(\alpha, S, \mathcal{V})$* ,  $P^{\alpha, S, \mathcal{V}} : X \times \Sigma_X \rightarrow [0, 1]$  as

$$P^{\alpha, S, \mathcal{V}}(\mathbf{q}, d\tilde{\mathbf{q}}) = \sum_{j=0}^p \int_Y \alpha_j(\mathbf{q}, \mathbf{v}) \delta_{\Pi_1 S_j(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}), \quad (2.2)$$

where  $\delta_{\Pi_1 S_j(\mathbf{q}, \mathbf{v})}$  is the Dirac measure concentrated at  $\Pi_1 S_j(\mathbf{q}, \mathbf{v})$  and  $\Pi_1$  denotes the projection operator onto the  $X$  component, i.e.,  $\Pi_1(\mathbf{q}, \mathbf{v}) = \mathbf{q}$  for all  $\mathbf{q} \in X, \mathbf{v} \in Y$ .

As described in Section 1, this setup may be seen as a multiproposal extension to the recent work of Glatt–Holtz *et al.* (2023). Algorithm 1 makes the underlying algorithmic procedure explicit. See also Definition 3.1 in Section 3 below for an extended abstract phase space pMCMC formalism that admits multiple jumps within a proposal set and directly generalizes the frameworks of Calderhead (2014); Tjelmeland (2004). Note that the class of kernels identified in Definition 2.1 encompasses a wide variety of algorithmic structures. This includes the multiproposal generalizations of random walk MH and pCN we describe in Section 4.1 and Section 4.2. In these settings, the  $S_j$ s represent coordinate exchanges as in (2.13), and the proposal kernels  $\mathcal{V}$  have the conditionally independent structure (2.12). On the other hand, Definition 2.1 also accommodates a variety of HMC-type algorithms where the  $S_j$ s are related to the numerical integration of a Hamiltonian system associated to the target measure (see Section 4.3).

We also note that the general multiproposal algorithm specified by the transition kernel (2.2) clearly differs from so-called hybrid MCMC strategies given by mixtures of kernels; see e.g. Robert & Casella (1999); Tierney (1994). In the latter, one specifies a set of probabilities  $a_1, \dots, a_m$  and Markov kernels  $P_1, \dots, P_m$ . At each iteration, one selects one of such kernels according to the given probabilities (which may only depend on the current state) and then draws a proposal sample from the chosen kernel. In contrast, under (2.2) with  $Y = X^p$ , one first samples a cloud of proposals  $\mathbf{v} = (\mathbf{q}_1, \dots, \mathbf{q}_p) \in X^p$  from the given proposal kernel  $\mathcal{V}(\mathbf{q}, \cdot)$ , and then selects one of  $\mathbf{q}, \mathbf{q}_1, \dots, \mathbf{q}_p$  according to probabilities  $\alpha_0, \dots, \alpha_p$  that depend on *all* of  $\mathbf{q}, \mathbf{q}_1, \dots, \mathbf{q}_p$ . See e.g. (4.22) and (4.35) below for concrete examples.

## 2.2 General criteria for invariance or reversibility

Our first result, Theorem 2.2, establishes a general set of conditions on the kernel  $\mathcal{V}$  and the mappings  $S_j, \alpha_j, j = 0, 1, \dots, p$ , under which the corresponding Markov kernel  $P^{\alpha, S, \mathcal{V}}$  maintains a given target

probability measure  $\mu$  on  $X$  invariant, or additionally that it satisfies detailed balance with respect to  $\mu$ . See Appendix B.1 for the proof.

**THEOREM 2.2** Let  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$  be measurable spaces. Fix a probability measure  $\mu$  on  $X$  and a Markov kernel  $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$ . Let  $\mathcal{M}$  be the probability measure on the product space  $X \times Y$  defined by

$$\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q}). \quad (2.3)$$

Fix any  $p \geq 1$ , and for each  $j = 0, 1, \dots, p$  consider the following statements for the given measurable mappings  $S_j : X \times Y \rightarrow X \times Y$  and  $\alpha_j : X \times Y \rightarrow [0, 1]$ :

(H1)  $S_j$  is an involution, i.e.  $S_j \circ S_j = I$ , for  $j = 0, 1, \dots, p$ ;

(H2) (2.1) holds and

$$\sum_{j=0}^p \int_Y \alpha_j(S_j(\mathbf{q}, \mathbf{v})) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mu(d\mathbf{q}); \quad (2.4)$$

(H3) (2.1) holds and, for every  $j = 0, 1, \dots, p$ ,<sup>1</sup>

$$\alpha_j(S_j(\mathbf{q}, \mathbf{v})) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \alpha_j(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}); \quad (2.5)$$

where in (2.4) and (2.5)  $S_j^* \mathcal{M}$  denotes the pushforward of  $\mathcal{M}$  under  $S_j$  (see (A1) below in Appendix A).

Then, under (H1) and (H2), it follows that the corresponding Markov kernel  $P^{\alpha, S, \mathcal{V}} : X \times \Sigma_X \rightarrow [0, 1]$  defined in (2.2) maintains  $\mu$  as an invariant measure, i.e.  $\mu P^{\alpha, S, \mathcal{V}} = \mu$ . Moreover, (H3) implies (H2), and under (H1) and (H3) the Markov kernel  $P^{\alpha, S, \mathcal{V}}$  additionally satisfies detailed balance with respect to  $\mu$ , i.e.

$$P^{\alpha, S, \mathcal{V}}(\mathbf{q}, d\tilde{\mathbf{q}})\mu(d\mathbf{q}) = P^{\alpha, S, \mathcal{V}}(\tilde{\mathbf{q}}, d\mathbf{q})\mu(d\tilde{\mathbf{q}}). \quad (2.6)$$

Algorithm 1 describes the general sampling procedure justified by Theorem 2.2.

**REMARK 2.3** To allow the algorithm the possibility of staying at the current state at any given step, one simply has to choose one of the involution mappings  $S_j, j = 0, 1, \dots, p$ , to be the identity.

In the next two corollaries, we present two possible definitions of sets of acceptance probabilities  $(\alpha_0, \dots, \alpha_p)$  satisfying property (H3), and hence (H2), in Theorem 2.2. Specifically, the  $\alpha_j$ s defined in Corollary 2.4 are of Barker-type (Barker, 1965), whereas the ones in Corollary 2.6 correspond to the MH type (Hastings, 1970; Metropolis *et al.*, 1953), cf. 2.8 and (2.10)–(2.11), respectively. See Appendix B.2 and Appendix B.3 for the proofs.

**COROLLARY 2.4** Let  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$  be measurable spaces. Fix any  $\mu \in \text{Pr}(X)$ , a Markov kernel  $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$ , and let  $\mathcal{M} \in \text{Pr}(X \times Y)$  be given by  $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q})$ . Further, let

---

<sup>1</sup> Equivalently,  $\int_{X \times Y} \varphi(\mathbf{q}, \mathbf{v}) \alpha_j(S_j(\mathbf{q}, \mathbf{v})) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \int_{X \times Y} \varphi(\mathbf{q}, \mathbf{v}) \alpha_j(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v})$  for every bounded and measurable function  $\varphi : X \times Y \rightarrow \mathbb{R}$ .

**Algorithm 1**

- 
- 1: Select the algorithm parameters:
- (i) The proposal kernel  $\mathcal{V}(\mathbf{q}, d\mathbf{v})$ .
  - (ii) The mappings  $S_j, \alpha_j, j = 0, 1, \dots, p$ , satisfying (H1) and (H2), or (H1) and (H3).
- 2: Choose  $\mathbf{q}^{(0)} \in X$ .
- 3: **for**  $k \geq 0$  **do**
- 4:     Sample  $\mathbf{v}^{(k+1)} \sim \mathcal{V}(\mathbf{q}^{(k)}, \cdot)$ .
- 5:     Compute  $S_j(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)})$ , for  $j = 0, 1, \dots, p$ .
- 6:     Set  $\mathbf{q}^{(k+1)}$  by drawing from  $(\Pi_1 S_0(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}), \dots, \Pi_1 S_p(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}))$  according to the probabilities  

$$(\alpha_0(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}), \dots, \alpha_p(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)})).$$
- 7:      $k \rightarrow k + 1$ .
- 8: **end for**
- 

$S_j : X \times Y \rightarrow X \times Y, j = 0, 1, \dots, p$ , be measurable mappings satisfying the involution assumption (H1) of Theorem 2.2. Assume additionally that, for every  $j = 0, 1, \dots, p$ ,

$$\sum_{k=0}^p (S_j \circ S_k)^* \mathcal{M}(E) = \sum_{k=0}^p S_k^* \mathcal{M}(E) \quad \text{for all } E \in \Sigma_{X \times Y}. \quad (2.7)$$

Also, let  $\alpha_j : X \times Y \rightarrow [0, 1], j = 0, \dots, p$ , be any measurable functions such that (2.1) holds and

$$\alpha_j(\mathbf{q}, \mathbf{v}) = \frac{dS_j^* \mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) \quad (2.8)$$

for  $(\sum_{j=0}^p S_j^* \mathcal{M})$ -a.e.  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ . Then, under this setting, it follows that condition (H3) of Theorem 2.2 is satisfied (and consequently also (H2)). Therefore, the associated Markov kernel  $P^{\alpha, S, \mathcal{V}} : X \times \Sigma_X \rightarrow [0, 1]$  given in (2.2), with  $\alpha_j$  as defined in (2.8), satisfies detailed balance with respect to  $\mu$  and thus maintains  $\mu$  as an invariant measure.

**REMARK 2.5** Under the additional assumption that  $S_j^* \mathcal{M} \ll \mathcal{M}$  for  $j = 0, \dots, p$ , (2.8) may be written as

$$\alpha_j(\mathbf{q}, \mathbf{v}) = \frac{\frac{dS_j^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v})}{\frac{dS_0^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) + \dots + \frac{dS_p^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v})}, \quad \text{for } j = 0, \dots, p. \quad (2.9)$$

On the other hand, the condition  $S_j^* \mathcal{M} \ll \mathcal{M}$  is not required for Corollary 2.4.

**COROLLARY 2.6** Let  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$  be measurable spaces. Fix any  $\mu \in \text{Pr}(X)$ , a Markov kernel  $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$ , and let  $\mathcal{M} \in \text{Pr}(X \times Y)$  be given by  $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q})$ . Further, set

$S_0 := I$ , and let  $S_j : X \times Y \rightarrow X \times Y$ ,  $j = 1, \dots, p$ , be measurable mappings satisfying assumption (H1) of Theorem 2.2 and such that  $S_j^* \mathcal{M} \ll \mathcal{M}$ . Consider a collection of weights  $\bar{\alpha}_j \in \mathbb{R}^+$ ,  $j = 1, \dots, p$ , satisfying  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ . Moreover, for each  $j = 0, 1, \dots, p$  define mappings  $\alpha_j : X \times Y \rightarrow [0, 1]$  given by

$$\alpha_j(\mathbf{q}, \mathbf{v}) := \bar{\alpha}_j \left[ 1 \wedge \frac{dS_j^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \right], \quad j = 1, \dots, p, \quad (2.10)$$

$$\alpha_0(\mathbf{q}, \mathbf{v}) := 1 - \sum_{j=1}^p \alpha_j(\mathbf{q}, \mathbf{v}), \quad (2.11)$$

for all  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ . Then, conditions (H2) and (H3) of Theorem 2.2 hold. Consequently, the associated Markov kernel  $P^{\alpha, S, \mathcal{V}} : X \times \Sigma_X \rightarrow [0, 1]$  given in (2.2), with  $\alpha_j$  as defined in (2.10)–(2.11), satisfies detailed balance with respect to  $\mu$  and thus maintains  $\mu$  as an invariant measure.

**REMARK 2.7** In the setting of Corollary 2.6, a natural choice of weights would be  $\bar{\alpha}_j = 1/p$ , for all  $j = 1, \dots, p$ . Under this specification, equations reminiscent of (2.10) and (2.11) appear in Calderhead (2014), albeit with no formal justification.

**REMARK 2.8** We notice that Corollary 2.6 encompasses Theorem 2.1 of Glatt–Holtz *et al.* (2023) for the case  $p = 1$ .

### 2.3 The case of conditionally independent proposals

Here, we specialize to a particular subclass of kernels falling under the wider umbrella of Definition 2.1. This setting encompasses and generalizes algorithms found in Tjelmeland (2004). Meanwhile, it permits the derivation of new Hilbert space-type algorithms. See Section 4.1 and Section 4.2 respectively below.

We specify the elements  $(\alpha, S, \mathcal{V})$  composing the kernel  $P^{\alpha, S, \mathcal{V}}$  in (2.2) of Definition 2.1 as follows. Let  $(X, \Sigma_X)$  be a measurable space, and fixing  $p \geq 1$ , we take  $Y = X^p$  to be the  $p$ -fold product of  $X$  that we endow with the standard product  $\sigma$ -algebra. Fixing any two Markov kernels  $Q, \bar{Q} : X \times \Sigma_X \rightarrow [0, 1]$ , we take

$$\mathcal{V}(\mathbf{q}_0, d\mathbf{v}) = \mathcal{V}(\mathbf{q}_0, d\mathbf{q}_1, \dots, d\mathbf{q}_p) = \int_X \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) \bar{Q}(\mathbf{q}_0, d\mathbf{q}), \quad \text{for } \mathbf{v} = (\mathbf{q}_1, \dots, \mathbf{q}_p) \in X^p, \quad (2.12)$$

and then consider the flip involutions  $S_j : X^{p+1} \rightarrow X^{p+1}$ ,  $j = 0, \dots, p$ , given by

$$S_0 := I, \quad \text{and} \quad S_j(\mathbf{q}_0, \mathbf{v}) = S_j(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) := (\mathbf{q}_j, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{q}_0, \mathbf{q}_{j+1}, \dots, \mathbf{q}_p), \quad (2.13)$$

for all  $(\mathbf{q}_0, \mathbf{v}) \in X \times X^p$  and  $j = 1, \dots, p$ . We refer to this class of proposal kernels defined by (2.12), (2.13) as having a *conditionally independent structure*. Viewed algorithmically, one makes such a proposal by drawing from  $\bar{Q}$  from around the current state and then using this new point we generate a cloud of  $p$  proposal points according to  $Q$  in a conditionally independent fashion. The main advantage of this proposal structure is that under suitable assumptions on the kernels  $Q, \bar{Q}$  it can lead to simplified and more

**Algorithm 2**

- 
- 1: Select the algorithm parameters:
- (i) The proposal kernels  $Q, \bar{Q} : X \times \Sigma_X \rightarrow [0, 1]$ .
  - (ii) The mappings  $\alpha_j, j = 0, 1, \dots, p$ , satisfying (H2) (invariance only) or (H3) (reversibility). Here,  $\mathcal{M}$ ,  $\mathcal{V}$ , and  $S_j, j = 0, 1, \dots, p$ , are as in (2.3), (2.12), and (2.13), respectively. Note that  $\alpha_j$  may always be specified as in (2.8) (Theorem 2.10) and see also (2.17), (2.18).
- 2: Choose  $\mathbf{q}_0^{(0)} \in X$ .
- 3: **for**  $k \geq 0$  **do**
- 4:     Sample  $\bar{\mathbf{q}} \sim \bar{Q}(\mathbf{q}_0^{(k)}, \cdot)$ .
- 5:     Sample  $\mathbf{q}_j^{(k+1)} \sim Q(\bar{\mathbf{q}}, \cdot)$  independently, for  $j = 1, \dots, p$ . Set  $\mathbf{v}^{(k+1)} := (\mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$ .
- 6:     Set  $\mathbf{q}_0^{(k+1)}$  by drawing from  $(\mathbf{q}_0^{(k)}, \mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$  according to the probabilities  $(\alpha_0(\mathbf{q}_0^{(k)}, \mathbf{v}^{(k+1)}), \dots, \alpha_p(\mathbf{q}_0^{(k)}, \mathbf{v}^{(k+1)}))$ .
- 7:      $k \rightarrow k + 1$ .
- 8: **end for**
- 

computationally efficient expressions for the acceptance probabilities  $\alpha_j, j = 0, 1, \dots, p$  as we illustrate in Theorem 2.11 and Corollary 2.12 below (see also Section 4.1). Algorithm 2 makes the associated sampling procedure precise.

**REMARK 2.9** The proposal formulation in (2.12) allows for the particular case where  $\mathbf{q}_1, \dots, \mathbf{q}_p$  are directly and independently drawn from a probability distribution  $Q(\mathbf{q}_0, \cdot)$ , by simply choosing  $\bar{Q}(\mathbf{q}_0, d\mathbf{q}) := \delta_{\mathbf{q}_0}(d\mathbf{q})$ . Note, however, that the condition (2.15) below typically leads to more computationally tractable acceptance probabilities. Compare for example (4.8) and (4.11) below in Section 4.1. In any case if we take  $Q(\mathbf{q}_0, \cdot)$  to be independent of  $\mathbf{q}_0$ , Algorithm 2 yields a multiproposal version of the standard independence sampler algorithm.

The following result states that under the choices of  $\mathcal{V}$  and  $S$  given in (2.12) and (2.13), and for a given target measure  $\mu$ , condition (2.7) of Corollary 2.4 is guaranteed to hold. Thus, by supplementing such  $\mathcal{V}$  and  $S$  with Barker-like acceptance probabilities  $\alpha$  as (2.8) one obtains a Markov transition kernel  $P^{\alpha, S, \mathcal{V}}$  as in (2.2) resulting from Algorithm 2 that satisfies detailed balance with respect to  $\mu$ .

**THEOREM 2.10** Fix any  $p > 0$ ,  $\mu \in \text{Pr}(X)$  and Markov kernels  $Q, \bar{Q} : X \times \Sigma_X \rightarrow [0, 1]$ . Define another Markov kernel  $\mathcal{V} : X \times \Sigma_{X^p} \rightarrow [0, 1]$  as in (2.12) and let  $\mathcal{M} \in \text{Pr}(X^{p+1})$  be given as  $\mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) = \mathcal{V}(\mathbf{q}_0, d\mathbf{v})\mu(d\mathbf{q}_0)$ . Further, let  $S_j : X^{p+1} \rightarrow X^{p+1}$  be the involution mappings defined in (2.13). Then, for every  $j = 0, 1, \dots, p$ , it holds that

$$\sum_{k=0}^p (S_j \circ S_k)^* \mathcal{M}(E) = \sum_{k=0}^p S_k^* \mathcal{M}(E) \quad \text{for all } E \in \Sigma_{X^{p+1}}. \quad (2.14)$$

Therefore, the associated Markov kernel  $P^{\alpha, S, \mathcal{V}} : X \times \Sigma_X \rightarrow [0, 1]$  given in (2.2), with  $\alpha_j$  as defined in (2.8), satisfies detailed balance with respect to  $\mu$ .

The proof of Theorem 2.10 is provided below in Section B.4.

The next theorem considers a particular setting in Theorem 2.10. Specifically, it identifies suitable conditions on the target measure  $\mu$  and the Markov kernels  $Q, \bar{Q}$  that imply that  $S_j^* \mathcal{M}$  is absolutely continuous with respect to  $\mathcal{M}$  and for which the Radon–Nikodym derivative  $dS_j^* \mathcal{M} / d\mathcal{M}$  assumes an explicit expression depending on  $\mu$  but which is independent of the kernels  $Q, \bar{Q}$ . This in turn yields in Corollary 2.12 simplified expressions for the acceptance probabilities  $\alpha_j, j = 0, 1, \dots, p$  from (2.8) and (2.10)–(2.11) that are independent of  $Q, \bar{Q}$ , a crucial property for effective applications.

**THEOREM 2.11** Fix any  $p > 0$ ,  $\mu \in \text{Pr}(X)$ , and Markov kernels  $Q, \bar{Q} : X \times \Sigma_X \rightarrow [0, 1]$ . Define another Markov kernel  $\mathcal{V} : X \times \Sigma_{X^p} \rightarrow [0, 1]$  as in (2.12). Take  $\mathcal{M} \in \text{Pr}(X^{p+1})$  given by  $\mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) = \mathcal{V}(\mathbf{q}_0, d\mathbf{v})\mu(d\mathbf{q}_0)$ , and let  $S_j : X^{p+1} \rightarrow X^{p+1}, j = 0, 1, \dots, p$ , be the involution mappings defined in (2.13).

Suppose  $\mu \ll \mu_0$  for some  $\sigma$ -finite measure  $\mu_0$  on  $X$ , with  $\frac{d\mu}{d\mu_0}(\mathbf{q}) > 0$  for a.e.  $\mathbf{q} \in X$ . Additionally, assume that  $Q$  and  $\bar{Q}$  satisfy the following balance-type condition

$$Q(\mathbf{q}, d\tilde{\mathbf{q}})\mu_0(d\mathbf{q}) = \bar{Q}(\tilde{\mathbf{q}}, d\mathbf{q})\mu_0(d\tilde{\mathbf{q}}). \quad (2.15)$$

Then, for every  $j = 0, 1, \dots, p$ ,  $S_j^* \mathcal{M} \ll \mathcal{M}$  and

$$\frac{dS_j^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}_0, \mathbf{v}) = \frac{d\mu}{d\mu_0}(\mathbf{q}_j) \left( \frac{d\mu}{d\mu_0}(\mathbf{q}_0) \right)^{-1} \quad \text{for } \mathcal{M} \text{-a.e. } (\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in X \times X^p. \quad (2.16)$$

See Appendix B.5 for the proof of Theorem 2.11. The next corollary follows immediately by plugging (2.16) into the expressions (2.9) and (2.10).

**COROLLARY 2.12** Under the assumptions of Theorem 2.11, it follows that by defining for any  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in X \times X^p$

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \frac{\frac{d\mu}{d\mu_0}(\mathbf{q}_j)}{\sum_{k=0}^p \frac{d\mu}{d\mu_0}(\mathbf{q}_k)}, \quad j = 0, 1, \dots, p, \quad (2.17)$$

or

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \bar{\alpha}_j \left[ 1 \wedge \frac{\frac{d\mu}{d\mu_0}(\mathbf{q}_j)}{\frac{d\mu}{d\mu_0}(\mathbf{q}_0)} \right], \quad j = 1, \dots, p; \quad \alpha_0(\mathbf{q}_0, \mathbf{v}) = 1 - \sum_{j=1}^p \alpha_j(\mathbf{q}_0, \mathbf{v}), \quad (2.18)$$

with  $\bar{\alpha}_j \in \mathbb{R}^+$  such that  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ , then the associated kernel  $P^{\alpha, \mathcal{V}} : X \times \Sigma_X \rightarrow [0, 1]$ , given as in (2.2), satisfies detailed balance with respect to  $\mu$ .

**REMARK 2.13** One may obtain kernels  $Q, \bar{Q}$  satisfying (2.15) by selecting any measurable  $f : X \times X \rightarrow \mathbb{R}^+$  such that

$$\int_X f(\mathbf{q}, \tilde{\mathbf{q}})\mu_0(d\tilde{\mathbf{q}}) = 1 = \int_X f(\tilde{\mathbf{q}}, \mathbf{q})\mu_0(d\tilde{\mathbf{q}}),$$

for any  $\mathbf{q} \in X$  and then defining  $Q, \bar{Q} : X \times \Sigma_X \rightarrow [0, 1]$  according to

$$Q(\mathbf{q}, d\tilde{\mathbf{q}}) = f(\mathbf{q}, \tilde{\mathbf{q}})\mu_0(d\tilde{\mathbf{q}}), \quad \bar{Q}(\mathbf{q}, d\tilde{\mathbf{q}}) = f(\tilde{\mathbf{q}}, \mathbf{q})\mu_0(d\tilde{\mathbf{q}}). \quad (2.19)$$

Of course, in the case that  $f$  is symmetric, namely that  $f(\mathbf{q}, \tilde{\mathbf{q}}) = f(\tilde{\mathbf{q}}, \mathbf{q})$  for every  $\mathbf{q}, \tilde{\mathbf{q}} \in X$ , then  $Q = \bar{Q}$ .

### 3. Incorporating multiple jumps between proposals

Whereas the previous section develops a rigorous framework for pMCMC methods that make a single jump from the current state to one of multiple proposals, Calderhead (2014) presents an algorithm that allows for multiple resamples from a generated proposal set. This section develops a rigorous ‘extended phase space’ multiple proposal, multiple jump formalism. In particular, we show that the setup considered in Calderhead (2014); Tjelmeland (2004) permits an involutive, abstract state space generalization inclusive of HMC and Hilbert space settings previously unaddressed. Indeed our formulation, culminating in Theorem 3.3 and Algorithm 3, justifies drawing multiple samples from a given proposal cloud in a wide variety of contexts as we sketch below in Section 4. Note that, from a practical standpoint, making multiple jumps within a proposal set may be beneficial, insofar as the computational burden is not much greater than the cost imposed for a single jump: one must always evaluate the target density at all proposal points regardless of the number of jumps one intends to make.

#### 3.1 An augmented extended phase space formulation

Just as for Definition 2.1, we let  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$  be measurable spaces and fix  $p \geq 1$ , the number of proposals at each step. We take

$$\mathbf{z} = (\mathbf{w}, k) = (\mathbf{q}, \mathbf{v}, k) \in \mathcal{Z} := X \times Y \times \{0, \dots, p\}, \quad (3.1)$$

and, as above, we define  $\Pi_1 : X \times Y \rightarrow X$  as  $\Pi_1(\mathbf{q}, \mathbf{v}) = \mathbf{q}$ , the projection onto the  $X$  coordinate.

We define transition kernels on  $\mathcal{Z}$  as follows:

**DEFINITION 3.1** For  $j = 0, 1, \dots, p$ , we consider

$$\gamma_j : X \times \Sigma_Y \rightarrow [0, 1] \quad \text{to be Markov kernels} \quad (3.2)$$

and take

$$S_j : X \times Y \rightarrow X \times Y \quad \text{to be measurable mappings.} \quad (3.3)$$

(i) Let  $\mathcal{R} : \mathcal{Z} \times \Sigma_{\mathcal{Z}} \rightarrow [0, 1]$  be the Markov kernel defined by

$$\mathcal{R}(\mathbf{q}, \mathbf{v}, k, d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}, d\tilde{k}) = S_k^*(\gamma_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) \delta_{\Pi_1 S_k(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}})) \delta_k(d\tilde{k}). \quad (3.4)$$

(ii) Let us furthermore suppose that for  $k, j = 0, \dots, p$  we have  $\alpha_{k,j} : X \times Y \rightarrow [0, 1]$  such that

$$\sum_{j=0}^p \alpha_{k,j}(\mathbf{q}, \mathbf{v}) = 1, \quad \text{for all } k = 0, \dots, p \text{ and } (\mathbf{q}, \mathbf{v}) \in X \times Y. \quad (3.5)$$

We then define  $\mathcal{A} : \mathcal{Z} \times \Sigma_{\mathcal{Z}} \rightarrow [0, 1]$  as the Markov kernel given by

$$\mathcal{A}(\mathbf{q}, \mathbf{v}, k, d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}, d\tilde{k}) = \delta_{(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) \sum_{j=0}^p \alpha_{k,j}(\mathbf{q}, \mathbf{v}) \delta_j(d\tilde{k}). \quad (3.6)$$

(iii) We denote the composition kernel

$$\mathcal{P}_1(\mathbf{z}, d\tilde{\mathbf{z}}) = \mathcal{P}(\mathbf{z}, d\tilde{\mathbf{z}}) := \mathcal{R}\mathcal{A}(\mathbf{z}, d\tilde{\mathbf{z}}) = \int_{\mathcal{Z}} \mathcal{A}(\hat{\mathbf{z}}, d\tilde{\mathbf{z}}) \mathcal{R}(\mathbf{z}, d\hat{\mathbf{z}}) \quad (3.7)$$

and furthermore iteratively define, for any  $n \geq 1$ ,

$$\mathcal{P}_n(\mathbf{z}, d\tilde{\mathbf{z}}) := \mathcal{P}_{n-1}\mathcal{A}(\mathbf{z}, d\tilde{\mathbf{z}}) = \int_{\mathcal{Z}} \mathcal{A}(\hat{\mathbf{z}}, d\tilde{\mathbf{z}}) \mathcal{P}_{n-1}(\mathbf{z}, d\hat{\mathbf{z}}) \quad (3.8)$$

(iv) Finally, we consider the projection operator  $\mathcal{E} : \mathcal{Z} \rightarrow X$  as

$$\mathcal{E}(\mathbf{q}, \mathbf{v}, k) = \Pi_1 S_k(\mathbf{q}, \mathbf{v}). \quad (3.9)$$

We thus specify the kernels  $\mathcal{R}$ ,  $\mathcal{A}$ ,  $\mathcal{P}_n$  and the projection  $\mathcal{E}$  by the triple  $(\mathcal{V}^e, S, \alpha)$  where  $\mathcal{V}^e = (\mathcal{V}_0, \dots, \mathcal{V}_p)$ ,  $S = (S_0, \dots, S_p)$  and  $\alpha = (\alpha_{j,k})_{j,k=0,\dots,p}$ . For a given  $(\mathcal{V}^e, S, \alpha)$  and any  $n \geq 1$ , (3.8) and (3.9) yields a sampling procedure by taking, for any  $m \geq 1$ ,  $\mathbf{r}^{(m)} \sim \mathcal{E}^*[(\mathcal{P}_n)^j \mathcal{P}_l](\mathbf{z}, d\mathbf{q})$  where  $m = nj + l$  for the appropriate  $j, l \geq 0$ , and  $(\mathcal{P}_n)^j$  denotes the  $j$ -fold composition of  $\mathcal{P}_n$ . Algorithm 3 describes this procedure.

**REMARK 3.2** For each  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ , let  $A(\mathbf{q}, \mathbf{v})$  be the  $(p+1) \times (p+1)$  real matrix with entries  $A_{k,j}(\mathbf{q}, \mathbf{v}) := \alpha_{k,j}(\mathbf{q}, \mathbf{v})$ ,  $k, j = 0, \dots, p$ . It is not difficult to show the following alternative expression for the kernel  $\mathcal{P}_n$ ,  $n \geq 1$ :

$$\begin{aligned} \mathcal{P}_n(\mathbf{z}, d\tilde{\mathbf{z}}) &= \mathcal{P}_n(\mathbf{q}, \mathbf{v}, k, d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}, d\tilde{k}) \\ &= \int_Y \delta_{S_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), \hat{\mathbf{v}})}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) \sum_{j=0}^p A_{k,j}^n(S_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), \hat{\mathbf{v}})) \delta_j(d\tilde{k}) \mathcal{V}_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\hat{\mathbf{v}}), \end{aligned} \quad (3.10)$$

where  $A_{k,j}^n(\mathbf{q}, \mathbf{v})$ ,  $k, j = 0, \dots, p$ , denote the entries of the matrix  $A^n(\mathbf{q}, \mathbf{v})$ , i.e., the  $n$ -fold composition of  $A(\mathbf{q}, \mathbf{v})$ . Now denote by  $A^\infty(\mathbf{q}, \mathbf{v})$  the matrix with entries  $A_{k,j}^\infty(\mathbf{q}, \mathbf{v}) = \alpha_j^\infty(\mathbf{q}, \mathbf{v})$  for all  $k, j = 0, \dots, p$ , where  $(\alpha_j^\infty(\mathbf{q}, \mathbf{v}))_{j=0,\dots,p}$  is the stationary vector of the finite state Markov chain with transition matrix  $A(\mathbf{q}, \mathbf{v})$ . Notice that one obtains by formally taking the limit  $n \rightarrow \infty$  in (3.10) the following Markov kernel

$$\mathcal{P}_\infty(\mathbf{z}, d\tilde{\mathbf{z}}) := \int_Y \delta_{S_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), \hat{\mathbf{v}})}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) \sum_{j=0}^p \alpha_j^\infty(S_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), \hat{\mathbf{v}})) \delta_j(d\tilde{k}) \mathcal{V}_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\hat{\mathbf{v}}),$$

which corresponds to the one-step transition kernel of the algorithm introduced in Calderhead (2014). Moreover, observe that when  $\alpha_{k,j}$  is  $k$ -independent then it follows from condition (3.5) that  $A^n = A$  for all  $n$ , so that  $A = A^\infty$ .

**Algorithm 3**


---

```

1: Select the algorithmic parameters
   (i)  $p \geq 1$  the number of elements generated in each proposal cloud.
   (ii) The proposal kernels  $\mathcal{V}^e = (\mathcal{V}_0, \dots, \mathcal{V}_p)$  as in (3.2).
   (iii) The mappings  $S = (S_0, \dots, S_p)$  as in (3.3)
   (iv) The transition probabilities  $\alpha = (\alpha_{k,j})_{j,k=0,\dots,p}$  as in (3.5).
   (v) The  $n \geq 1$  number of samples drawn per generated proposal cloud.

2: Choose an initial  $(\mathbf{q}^{(0)}, \mathbf{v}^{(0)}) \in X \times Y$  and  $k^{(0)} \in \{0, \dots, p\}$ .
3: Set  $\mathbf{r}^{(0)} := \Pi_1 S_{k^{(0)}}(\mathbf{q}^{(0)}, \mathbf{v}^{(0)})$ .
4: for  $j \geq 0$  do
5:   Set  $\bar{\mathbf{q}}^{(j+1)} := \mathbf{r}^{(nj)}$ .
6:   Sample  $\tilde{\mathbf{v}}^{(j+1)} \sim \mathcal{V}_{k^{(j)}}(\bar{\mathbf{q}}^{(j+1)}, d\mathbf{v})$ .
7:   Set  $(\mathbf{q}^{(j+1)}, \mathbf{v}^{(j+1)}) := S_{k^{(j)}}(\bar{\mathbf{q}}^{(j+1)}, \tilde{\mathbf{v}}^{(j+1)})$ .
8:   Set  $k_{cur} := k^{(j)}$ .
9:   for  $l = 1, \dots, n$  do
10:    Draw  $k_{nxt} \in \{0, \dots, p\}$  with the probabilities  $(\alpha_{k_{cur},0}(\mathbf{q}^{(j+1)}, \mathbf{v}^{(j+1)}), \dots, \alpha_{k_{cur},p}(\mathbf{q}^{(j+1)}, \mathbf{v}^{(j+1)}))$ .
11:    Set  $k_{cur} := k_{nxt}$ .
12:    Set  $\mathbf{r}^{(nj+l)} := \Pi_1 S_{k_{cur}}(\mathbf{q}^{(j+1)}, \mathbf{v}^{(j+1)})$ .
13:     $l \rightarrow l + 1$ .
14:  end for
15:  Set  $k^{(j+1)} = k_{cur}$ .
16:   $j \rightarrow j + 1$ .
17: end for

```

---

### 3.2 Main result

We turn to the main result of this section, Theorem 3.3. Consider a given target measure  $\mu \in \text{Pr}(X)$ . Having fixed  $\mathcal{V}^e = (\mathcal{V}_0, \dots, \mathcal{V}_p)$  and  $S = (S_0, \dots, S_p)$  as in (3.2) and (3.3), we denote

$$\mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}_j(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q}), \quad j = 0, \dots, p. \quad (3.11)$$

We then consider the following extended phase space measure

$$\mathcal{N}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}, d\tilde{k}) = \sum_{j=0}^p \frac{1}{p+1} S_j^* \mathcal{M}_j(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) \delta_j(d\tilde{k}). \quad (3.12)$$

Theorem 3.3 establishes conditions on the elements  $(\mathcal{V}^e, S, \alpha)$  such that  $\mathcal{N}$  is invariant under  $\mathcal{A}, \mathcal{R}$  and hence under  $\mathcal{P}_n$  for any  $n \geq 0$ . This result furthermore asserts that  $\mathcal{E}\mathcal{N} = \mu$  so that we indeed have conditions justifying Algorithm 3 as an unbiased sampling procedure.

**THEOREM 3.3** Let  $\mathcal{V}^e = (\mathcal{V}_0, \dots, \mathcal{V}_p)$  be a collection of Markov kernels as in (3.2),  $S = (S_0, \dots, S_p)$  measurable mappings as in (3.3) and  $\alpha = (\alpha_{k,j})_{k,j=0,\dots,p}$  acceptance probabilities as in (3.5). Let also  $\mu \in \text{Pr}(X)$  and define probability measures  $\mathcal{M}_j, j = 0, \dots, p$ , on  $X \times Y$  as in (3.11). We assume that

- (H1)  $S_j$  is an involution, namely we suppose that  $S_j \circ S_j = I$ , for  $j = 0, 1, \dots, p$ ; and  
(H2) (3.5) holds and

$$S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) = \sum_{k=0}^p \alpha_{k,j}(\mathbf{q}, \mathbf{v}) S_k^* \mathcal{M}_k(d\mathbf{q}, d\mathbf{v}), \quad \text{for each } j = 0, \dots, p. \quad (3.13)$$

Under these conditions, define the Markov kernels  $\mathcal{R}$ ,  $\mathcal{A}$  on  $\mathcal{X}$  as in (3.4) and (3.6), respectively. Take  $\mathcal{P}_n$  as in (3.7), (3.8) for any  $n \geq 1$ . Then:

- (i)  $\mathcal{N}$  is invariant under  $\mathcal{R}$ ;
- (ii)  $\mathcal{N}$  is invariant under  $\mathcal{A}$ ;
- (iii)  $\mathcal{N}$  is invariant under  $\mathcal{P}_n$  for any  $n \geq 1$ ;
- (iv)  $\mathcal{E}^* \mathcal{N} = \mu$ , and therefore  $\mathcal{E}^*(\mathcal{N} \mathcal{P}_n) = \mu$  for any  $n \geq 1$ .

The proof of Theorem 3.3 is found below in Appendix B.6.

**REMARK 3.4** Under (3.5), the condition on the acceptance probabilities (3.13) is implied by the following slightly stronger condition

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) S_k^* \mathcal{M}_k(d\mathbf{q}, d\mathbf{v}) = \alpha_{j,k}(\mathbf{q}, \mathbf{v}) S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}), \quad \text{for every } k, j = 0, \dots, p. \quad (3.14)$$

Next, we identify two examples of sets of acceptance probabilities  $\alpha_{k,j}$ ,  $k, j = 0, \dots, p$ , for which condition (H2) of Theorem 3.3 holds. In fact, both cases satisfy the stronger condition (3.14). The proof follows similarly as in Corollary 2.4 and Corollary 2.6, so we omit the details.

**COROLLARY 3.5** Take  $\mu, \mathcal{V}^e = (\mathcal{V}_0, \dots, \mathcal{V}_p), S = (S_0, \dots, S_p)$ , and  $\mathcal{M}_j, j = 0, \dots, p$ , as in Theorem 3.3. Consider the following definitions:

- (i) Let  $\alpha_{k,j} : X \times Y \rightarrow [0, 1]$ ,  $k, j = 0, \dots, p$ , be any measurable mappings such that (3.5) holds and

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) = \frac{dS_j^* \mathcal{M}_j}{d(S_0^* \mathcal{M}_0 + \dots + S_p^* \mathcal{M}_p)}(\mathbf{q}, \mathbf{v}), \quad (3.15)$$

for  $(S_0^* \mathcal{M}_0 + \dots + S_p^* \mathcal{M}_p)$ -a.e.  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ , and for every  $k = 0, \dots, p$ .

- (ii) Assume  $S_j^* \mathcal{M}_j \ll S_k^* \mathcal{M}_k$  for all  $k \neq j$ . Take  $\bar{\alpha}_{k,j} \in [0, 1]$ ,  $k, j = 0, \dots, p$ , such that  $\sum_{j=0}^p \bar{\alpha}_{k,j} \leq 1$  for all  $k = 0, \dots, p$ . Then, for each  $k, j = 0, \dots, p$  and  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ , define

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) = \begin{cases} \bar{\alpha}_{k,j} \left[ 1 \wedge \frac{dS_j^* \mathcal{M}_j}{dS_k^* \mathcal{M}_k}(\mathbf{q}, \mathbf{v}) \right] & \text{if } j \neq k, \\ 1 - \sum_{\substack{l=0 \\ l \neq k}}^p \alpha_{k,l}(\mathbf{q}, \mathbf{v}) & \text{if } j = k. \end{cases} \quad (3.16)$$

Then, under the conditions in (i) and (ii), it follows that both definitions (3.15) and (3.16) satisfy properties (3.5) and (3.14), and consequently also assumption (H2) of Theorem 3.3.

**REMARK 3.6** Note that any collection of  $k$ -independent  $\alpha_{k,j}$ 's that satisfy condition (3.13) must coincide  $(S_0^*\mathcal{M}_0 + \dots + S_p^*\mathcal{M}_p)$ -a.e. with the Barker-like expression in (3.15).

**REMARK 3.7** Similarly as in Corollary 2.12, we obtain that under an analogous conditionally independent framework from Theorem 2.11 the definitions in (3.15) and (3.16) assume a simplified expression. Indeed, take  $Y = X^p$  and suppose  $\mu \ll \mu_0$  for some  $\sigma$ -finite measure  $\mu_0$  on  $X$ , with  $\frac{d\mu}{d\mu_0}(\mathbf{q}) > 0$  for a.e.  $\mathbf{q} \in X$ . Assume that  $\mathcal{V}_0 = \dots = \mathcal{V}_p =: \mathcal{V}$ , with  $\mathcal{V}$  as in (2.12) for fixed Markov kernels  $Q, \bar{Q}$  satisfying (2.15), and let  $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q})$ . Moreover, take  $S_j, j = 0, \dots, p$  to be the flip involutions defined in (2.13). It thus follows from Theorem 2.11 that  $S_j^*\mathcal{M} \ll \mathcal{M}$  for  $j = 0, \dots, p$  and (2.16) holds, so that (3.15) and (3.16) reduce for any  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in X \times X^p$  and  $k, j = 0, \dots, p$  to

$$\alpha_{k,j}(\mathbf{q}_0, \mathbf{v}) = \frac{\frac{d\mu}{d\mu_0}(\mathbf{q}_j)}{\sum_{l=0}^p \frac{d\mu}{d\mu_0}(\mathbf{q}_l)}, \quad (3.17)$$

and

$$\alpha_{k,j}(\mathbf{q}_0, \mathbf{v}) = \begin{cases} \bar{\alpha}_{k,j} \left[ 1 \wedge \frac{\frac{d\mu}{d\mu_0}(\mathbf{q}_j)}{\frac{d\mu}{d\mu_0}(\mathbf{q}_k)} \right] & \text{if } j \neq k, \\ 1 - \sum_{\substack{l=0 \\ l \neq k}}^p \alpha_{k,l}(\mathbf{q}, \mathbf{v}) & \text{if } j = k, \end{cases} \quad (3.18)$$

respectively.

**REMARK 3.8** One might expect that, by setting the number of jumps  $n = 1$ , Algorithm 3 would essentially reduce to Algorithm 1. In fact, this seems to not be the case in general, owing crucially to step 7 in Algorithm 3. See Section 4.3, Remark 4.3, below for a specific example of this non-equivalence.

Nevertheless, there is a particular case where a suitable relationship between these two algorithms can be established. Indeed, assume the same setting from Remark 3.7, and let also  $\alpha_j = \alpha_{k,j}, k, j = 0, \dots, p$ , be the acceptance probabilities given as in (3.17). Then, denoting  $P = P^{\alpha, \mathcal{S}, \mathcal{V}}$  the corresponding kernel as defined in (2.2), it is not difficult to show that

$$P^m(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{q}}) = \mathcal{E}^*(\mathcal{P}_n)^m(\mathbf{q}, \mathbf{v}, k, \cdot)(d\tilde{\mathbf{q}}) \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in X \times X^p, \text{ and } m, n \in \mathbb{N}, \quad (3.19)$$

so that, setting in particular  $n = 1$  and  $k = 0$ , we have  $P^m(\mathbf{q}, d\tilde{\mathbf{q}}) = \mathcal{E}^*\mathcal{P}^m(\mathbf{q}, \mathbf{v}, 0, \cdot)(d\tilde{\mathbf{q}})$  for all  $(\mathbf{q}, \mathbf{v}) \in X \times X^p$  and  $m \in \mathbb{N}$ . To show (3.19), one first crucially notices that due to  $\alpha_{k,j}$  given in (3.17) being  $k$ -independent and also property (3.5) then it follows that the matrix  $A(\mathbf{q}, \mathbf{v})$  with entries  $\alpha_{k,j}(\mathbf{q}, \mathbf{v}), k, j = 0, \dots, p$ , satisfies  $A^n(\mathbf{q}, \mathbf{v}) = A(\mathbf{q}, \mathbf{v})$  for all  $n \in \mathbb{N}$ . Secondly, from the definitions of the flip involutions  $S_j, j = 0, \dots, p$ , in (2.13), and  $\alpha_j = \alpha_{k,j}$  given in (3.17), it follows immediately that, for any  $k, j = 0, \dots, p$  and  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in X \times X^p$ ,  $\alpha_j(S_k(\mathbf{q}_0, \mathbf{v}))$  is equal to:  $\alpha_k(\mathbf{q}_0, \mathbf{v})$  if  $j = 0$ ;  $\alpha_0(\mathbf{q}_0, \mathbf{v})$  if  $j = k$ ; and  $\alpha_j(\mathbf{q}_0, \mathbf{v})$  for  $j \in \{1, \dots, p\}$  with  $j \neq k$ . In essence, this nullifies the effect of step 7 in Algorithm 3, and ultimately implies that the samples  $\{\mathbf{q}^{(i)}\}_{i \in \mathbb{N}}$  generated by Algorithm 1 are equivalent to the chain  $\{\Pi_1 S_{k^{(i)}}(\mathbf{q}_0^{(i)}, \mathbf{v}^{(i)})\}_{i \in \mathbb{N}}$  derived from the samples  $\{(\mathbf{q}_0^{(i)}, \mathbf{v}^{(i)}, k^{(i)})\}_{i \in \mathbb{N}}$  generated by Algorithm 3.

#### 4. Applications for algorithm design

This section leverages our abstract formalisms developed in Section 2, Section 3 in service of the design and rigorous analysis of some concrete sampling algorithms. In Section 4.1 we provide a systematic treatment of multiproposal methods for continuous probability distributions on  $\mathbb{R}^N$ . This treatment generalizes previously observed finite dimensional methods from e.g. Tjelmeland (2004) to include non-symmetric proposal kernels. Next in Section 4.2 we derive a novel ‘Hilbert-space’ method, which we christen the multiproposal pCN (mpCN) sampler in Algorithms 5 and 6. As already previewed above in the introduction, mpCN extends the preconditioned Crank-Nicolson algorithm designed for infinite-dimensional target measures that are absolutely continuous with respect to a Gaussian base measure (Beskos *et al.*, 2008; Cotter *et al.*, 2013) to a multiproposal setting. Elsewhere in Section 4.3 we address applications for Hamiltonian-type sampling methods while in Section 4.4 we consider simplicial methods developed recently in Holbrook (2023a).

##### 4.1 Finite-dimensional multiproposal algorithms

In this subsection, we consider the particular case of finite-dimensional, continuously distributed measures in the algorithms presented in Section 2 and Section 3 above. We then rewrite the formulas for the acceptance probabilities previously introduced in terms of the associated probability densities, thus providing more directly applicable expressions. In all such formulas presented below, namely (4.2), (4.3), (4.7), (4.11), (4.12), (4.13) and (4.14), the acceptance probability is defined by the given expression at every point where the denominator is strictly positive, and otherwise the probability is assumed to be zero.

Take  $X = \mathbb{R}^N$  and  $Y = \mathbb{R}^D$  for some  $N, D > 0$ , endowed with their corresponding Borel  $\sigma$ -algebras. Consider a target distribution

$$\mu(d\mathbf{q}) = \pi(\mathbf{q})d\mathbf{q} \quad \text{for some density function } \pi : \mathbb{R}^N \rightarrow \mathbb{R}^+. \quad (4.1)$$

Following the framework from Section 2.2, take a Markov kernel  $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = g(\mathbf{q}, \mathbf{v})d\mathbf{v}$  for some measurable function  $g : \mathbb{R}^{N+D} \rightarrow \mathbb{R}^+$  with  $\int_{\mathbb{R}^D} g(\mathbf{q}, \mathbf{v})d\mathbf{v} = 1$  for all  $\mathbf{q} \in \mathbb{R}^N$ . It follows that

$$\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q}) = h(\mathbf{q}, \mathbf{v})d\mathbf{q}d\mathbf{v}, \text{ with } h(\mathbf{q}, \mathbf{v}) = g(\mathbf{q}, \mathbf{v})\pi(\mathbf{q}).$$

Then, given any  $C^1$  involution mappings  $S_j : \mathbb{R}^{N+D} \rightarrow \mathbb{R}^{N+D}$ ,  $j = 0, 1, \dots, p$ , it follows from (A3) that

$$S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = h(S_j(\mathbf{q}, \mathbf{v})) |\det \nabla S_j(\mathbf{q}, \mathbf{v})| d\mathbf{q}d\mathbf{v}, \quad j = 0, 1, \dots, p.$$

In this situation Corollary 2.4 and Corollary 2.6 yield the following general formulations. In the case of Corollary 2.4 the condition (2.7) translates to

$$\sum_{k=0}^p h(S_k \circ S_j(\mathbf{q}, \mathbf{v})) |\det \nabla(S_k \circ S_j)(\mathbf{q}, \mathbf{v})| d\mathbf{q}d\mathbf{v} = \sum_{k=0}^p h(S_k(\mathbf{q}, \mathbf{v})) |\det \nabla S_k(\mathbf{q}, \mathbf{v})| d\mathbf{q}d\mathbf{v},$$

for all  $j = 0, \dots, p$ . In this circumstance we obtain a reversible sampling scheme from Algorithm 1 by supplementing the input parameters  $\mathcal{V}$ ,  $(S_0, \dots, S_p)$  with acceptance probabilities  $\alpha_j$ ,  $j = 0, 1, \dots, p$ ,

given as in (2.8), which according to (A4) can be written here as

$$\alpha_j(\mathbf{q}, \mathbf{v}) = \frac{h(S_j(\mathbf{q}, \mathbf{v})) |\det \nabla S_j(\mathbf{q}, \mathbf{v})|}{\sum_{l=0}^p h(S_l(\mathbf{q}, \mathbf{v})) |\det \nabla S_l(\mathbf{q}, \mathbf{v})|}, \quad (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^N \times \mathbb{R}^D. \quad (4.2)$$

Alternatively, in the setting of Corollary 2.6 we assume that  $S_0 = I$  and consider the  $\alpha_j$ s in (2.10)–(2.11), given here by

$$\alpha_j(\mathbf{q}, \mathbf{v}) = \bar{\alpha}_j \left[ 1 \wedge \frac{h(S_j(\mathbf{q}, \mathbf{v})) |\det \nabla S_j(\mathbf{q}, \mathbf{v})|}{h(\mathbf{q}, \mathbf{v})} \right], \quad j = 1, \dots, p; \quad \alpha_0(\mathbf{q}, \mathbf{v}) := 1 - \sum_{j=1}^p \alpha_j(\mathbf{q}, \mathbf{v}) \quad (4.3)$$

for  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^N \times \mathbb{R}^D$ , where we recall that  $\bar{\alpha}_j \in [0, 1]$ ,  $j = 1, \dots, p$ , are any user defined weights satisfying  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ .

Let us now turn to the ‘conditionally independent’ setting of Section 2.3. We will provide the details for the case of Barker-type acceptance probabilities, à la (4.2), which we anticipate as being the most relevant in this particular setting. The reader will find the analogous reduction from (4.3) for the MH case to be direct, following from the same considerations.

In this conditionally independent situation we take  $Y = \mathbb{R}^{pN}$  and consider Markov kernels of the form

$$Q(\mathbf{q}, d\tilde{\mathbf{q}}) = f(\mathbf{q}, \tilde{\mathbf{q}}) d\mathbf{q} d\tilde{\mathbf{q}} \quad \text{and} \quad \bar{Q}(\mathbf{q}, d\tilde{\mathbf{q}}) = \bar{f}(\mathbf{q}, \tilde{\mathbf{q}}) d\mathbf{q} d\tilde{\mathbf{q}}. \quad (4.4)$$

Here,  $f : \mathbb{R}^{2N} \rightarrow \mathbb{R}^+$  and  $\bar{f} : \mathbb{R}^{2N} \rightarrow \mathbb{R}^+$  are measurable functions such that

$$\int_{\mathbb{R}^N} f(\mathbf{q}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}} = 1 = \int_{\mathbb{R}^N} \bar{f}(\mathbf{q}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}} \quad \text{for every } \mathbf{q} \in \mathbb{R}^N. \quad (4.5)$$

In this setting, the Markov kernel  $\mathcal{V}$  from (2.12) can be written as

$$\mathcal{V}(\mathbf{q}_0, d\mathbf{v}) = g(\mathbf{q}_0, \mathbf{v}) d\mathbf{v}, \quad \text{with } g(\mathbf{q}_0, \mathbf{v}) = \int_{\mathbb{R}^N} \prod_{i=1}^p f(\mathbf{q}, \mathbf{q}_i) \bar{f}(\mathbf{q}_0, \mathbf{q}) d\mathbf{q} \quad (4.6)$$

for all  $\mathbf{q}_0 \in \mathbb{R}^N$  and  $\mathbf{v} = (\mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^{pN}$ . We notice that the assumed conditions on  $f, \bar{f}$  imply that  $\int_{\mathbb{R}^{Np}} g(\mathbf{q}_0, \mathbf{v}) d\mathbf{v} = 1$  for all  $\mathbf{q}_0 \in \mathbb{R}^N$ , so that  $\mathcal{V}$  indeed defines a Markov kernel. We also fix  $S_j : \mathbb{R}^{(p+1)N} \rightarrow \mathbb{R}^{(p+1)N}$ ,  $j = 0, 1, \dots, p$ , to be the flip involutions defined in (2.13). Since these involutions are also linear, it is not difficult to see that each  $S_j$  is volume-preserving, i.e.  $|\det \nabla S_j(\mathbf{q}, \mathbf{v})| = 1$  for all  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{(p+1)N}$ .

We thus obtain the following particular expressions for the acceptance probabilities  $\alpha_j$ ,  $j = 0, 1, \dots, p$ , from (4.2) in this setting, which when input into Algorithm 2 yield a reversible sampling

procedure:

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \frac{\pi(\mathbf{q}_j) \int_{\mathbb{R}^N} \prod_{\substack{i=0 \\ i \neq j}}^p f(\mathbf{q}, \mathbf{q}_i) \bar{f}(\mathbf{q}_j, \mathbf{q}) d\mathbf{q}}{\sum_{l=0}^p \pi(\mathbf{q}_l) \int_{\mathbb{R}^N} \prod_{\substack{i=0 \\ i \neq l}}^p f(\mathbf{q}, \mathbf{q}_i) \bar{f}(\mathbf{q}_l, \mathbf{q}) d\mathbf{q}}, \quad (4.7)$$

for  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^N \times \mathbb{R}^{pN}$ . Note that we can also consider the particular case where  $\bar{Q}(\mathbf{q}, \cdot) = \delta_{\mathbf{q}}(\cdot)$ . Of course, in this case there is no density  $\bar{f}$  with respect to Lebesgue measure as  $\delta_{\mathbf{q}}$  is not continuously distributed. Nevertheless, it follows immediately from (2.12) that  $\mathcal{V}(\mathbf{q}_0, d\mathbf{v}) = \prod_{i=1}^p f(\mathbf{q}_0, \mathbf{q}_i) d\mathbf{v}$  for  $\mathbf{v} = (\mathbf{q}_1, \dots, \mathbf{q}_p)$ , so that (4.2) now reduces to

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \frac{\pi(\mathbf{q}_j) \prod_{\substack{i=0 \\ i \neq j}}^p f(\mathbf{q}_j, \mathbf{q}_i)}{\sum_{k=0}^p \pi(\mathbf{q}_k) \prod_{\substack{i=0 \\ i \neq k}}^p f(\mathbf{q}_k, \mathbf{q}_i)}, \quad (4.8)$$

again for any  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^N \times \mathbb{R}^{pN}$ .

The acceptance probabilities (4.7) and (4.8) are expensive or intractable to compute in many situations that highlights the salience of Theorem 2.11 here. Within this setting of Theorem 2.11, let us assume for simplicity that  $\mu_0(d\mathbf{q})$  is the Lebesgue measure  $d\mathbf{q}$ , so that  $d\mu/d\mu_0(\mathbf{q}) = \pi(\mathbf{q})$  for a.e.  $\mathbf{q} \in \mathbb{R}^N$ . Let us also assume that  $\pi(\mathbf{q}) > 0$  for a.e.  $\mathbf{q} \in \mathbb{R}^N$ . Moreover, we now suppose that the densities  $f, \bar{f}$  associated to  $Q, \bar{Q}$ , respectively, satisfy (2.15), or equivalently in this setting

$$\int_A \int_B f(\mathbf{q}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}} d\mathbf{q} = \int_B \int_A \bar{f}(\tilde{\mathbf{q}}, \mathbf{q}) d\mathbf{q} d\tilde{\mathbf{q}} \quad \text{for every Borel sets } A, B \subset \mathbb{R}^N. \quad (4.9)$$

Note that this condition, (4.9), is satisfied if we find any  $f$  such that

$$\int_{\mathbb{R}^N} f(\mathbf{q}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}} = 1 = \int_{\mathbb{R}^N} f(\tilde{\mathbf{q}}, \mathbf{q}) d\tilde{\mathbf{q}} \quad \text{for any } \mathbf{q} \in \mathbb{R}^N, \quad (4.10)$$

and then we set  $\bar{f}(\tilde{\mathbf{q}}, \mathbf{q}) := f(\mathbf{q}, \tilde{\mathbf{q}})$ . In particular, one may select  $f(\mathbf{q}, \tilde{\mathbf{q}}) = r(\mathbf{q} - \tilde{\mathbf{q}})$  for any probability density  $r : \mathbb{R}^N \rightarrow \mathbb{R}^+$ . In this case, the simplified acceptance probabilities  $\alpha_j, j = 0, 1, \dots, p$ , from (2.17) can be written as

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \frac{\pi(\mathbf{q}_j)}{\sum_{l=0}^p \pi(\mathbf{q}_l)}, \quad (4.11)$$

for  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^{(p+1)N}$ . Notice that in the special case  $\bar{f}(\tilde{\mathbf{q}}, \mathbf{q}) := f(\mathbf{q}, \tilde{\mathbf{q}})$ , and under (4.10), the formula (4.11) is immediately obtained from (4.7).

We highlight this particular case of interest arising out of Theorem 2.11 from (4.9) and (4.11) as Algorithm 4. This sampler is the basis for the numerical experiments carried out below in Section 5.1.

**REMARK 4.1** Algorithm 4 includes as a special case the proposal structure considered in (Tjelmeland, 2004, Section 3). Specifically the method in Tjelmeland (2004) corresponds to the special case  $Q = \bar{Q}$ ,

**Algorithm 4**

- 
- 1: Select the algorithmic parameters:
- (i) Any measurable function  $\bar{f}, f : \mathbb{R}^{2N} \rightarrow \mathbb{R}^+$  maintaining (4.9), which we use to define the proposal mechanism. In particular, one may consider any probability density  $r : \mathbb{R}^N \rightarrow \mathbb{R}^+$  and set  $\bar{f}(\mathbf{q}, \tilde{\mathbf{q}}) = r(\mathbf{q} - \tilde{\mathbf{q}})$ ,  $f(\mathbf{q}, \tilde{\mathbf{q}}) = r(\tilde{\mathbf{q}} - \mathbf{q})$ .
  - (ii) the number of proposals  $p \geq 1$ .
- 2: Choose  $\mathbf{q}_0^{(0)} \in \mathbb{R}^N$ .
- 3: **for**  $k \geq 0$  **do**
- 4:     Sample  $\tilde{\mathbf{q}}^{(k)} \sim \bar{f}(\mathbf{q}_0^{(k)}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}}$
- 5:     Sample  $\mathbf{q}_j^{(k+1)} \sim f(\tilde{\mathbf{q}}^{(k)}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}}$ , for  $j = 1, \dots, p$ . Set  $\mathbf{v}_0^{(k+1)} := (\mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$ .
- 6:     Set  $\mathbf{q}_0^{(k+1)}$  by drawing from  $(\mathbf{q}_0^{(k)}, \mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$  according to the probabilities  $(\alpha_0(\mathbf{q}_0^{(k)}, \mathbf{v}^{(k+1)}), \dots, \alpha_p(\mathbf{q}_0^{(k)}, \mathbf{v}^{(k+1)}))$  as defined in (4.11).
- 7:      $k \rightarrow k + 1$ .
- 8: **end for**
- 

so that  $f = \bar{f}$ , and where  $f(\mathbf{q}, \cdot)$  is taken as the density of a multivariate Gaussian centered at  $\mathbf{q}$ . Notice that since such an  $f = f(\mathbf{q}, \tilde{\mathbf{q}})$  is symmetric in the variables  $\mathbf{q}, \tilde{\mathbf{q}}$  it is clear that (4.10) holds trivially in this case. Aside from Theorem 2.11 being given in the broader context of general state spaces, yielding for example mpCN in Algorithm 5 below, it also shows that the simplified expression of acceptance probabilities in (4.11) can be obtained from arbitrary, even *non-symmetric* proposal densities  $f(\mathbf{q}, \tilde{\mathbf{q}}) := r(\mathbf{q} - \tilde{\mathbf{q}})$ , by choosing  $\bar{f}(\mathbf{q}, \tilde{\mathbf{q}}) := f(\tilde{\mathbf{q}}, \mathbf{q})$  as in Algorithm 4.

Finally, let us specialize the setting of Section 3 to the case of finite-dimensional continuously distributed measures. We start by considering again the spaces  $X = \mathbb{R}^N$ ,  $Y = \mathbb{R}^D$ , for  $N, D > 0$ , a target distribution  $\mu(d\mathbf{q}) = \pi(\mathbf{q})d\mathbf{q}$ , and generic involution mappings  $S_j : \mathbb{R}^{N+D} \rightarrow \mathbb{R}^{N+D}$ ,  $j = 0, \dots, p$ . Then, consider Markov proposal kernels  $\mathcal{V}_j(\mathbf{q}, d\mathbf{v}) = g_j(\mathbf{q}, \mathbf{v}) d\mathbf{v}$ ,  $j = 0, \dots, p$ . Here, each  $g_j : \mathbb{R}^{N+D} \rightarrow \mathbb{R}^+$  is a measurable function satisfying  $\int_{\mathbb{R}^D} g_j(\mathbf{q}, \mathbf{v}) d\mathbf{v} = 1$  for all  $\mathbf{q} \in \mathbb{R}^N$ . We define

$$\mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}_j(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q}) = h_j(\mathbf{q}, \mathbf{v}) d\mathbf{q} d\mathbf{v}, \quad \text{with } h_j(\mathbf{q}, \mathbf{v}) := g_j(\mathbf{q}, \mathbf{v})\pi(\mathbf{q}), \quad j = 0, \dots, p.$$

We thus obtain from (3.15) and (3.16) the following examples of acceptance probabilities that yield an unbiased sampling scheme according to Algorithm 3. Namely, for all  $k, j = 0, \dots, p$ :

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) = \frac{h_j(S_j(\mathbf{q}, \mathbf{v})) |\det \nabla S_j(\mathbf{q}, \mathbf{v})|}{\sum_{l=0}^p h_l(S_l(\mathbf{q}, \mathbf{v})) |\det \nabla S_l(\mathbf{q}, \mathbf{v})|}, \quad (4.12)$$

and

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) = \begin{cases} \bar{\alpha}_{k,j} \left[ 1 \wedge \frac{h_j(S_j(\mathbf{q}, \mathbf{v})) |\det \nabla S_j(\mathbf{q}, \mathbf{v})|}{h_k(S_k(\mathbf{q}, \mathbf{v})) |\det \nabla S_k(\mathbf{q}, \mathbf{v})|} \right] & \text{if } j \neq k, \\ 1 - \sum_{\substack{l=0 \\ l \neq k}}^p \alpha_{k,l}(\mathbf{q}, \mathbf{v}) & \text{if } j = k \end{cases} \quad (4.13)$$

for  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^N \times \mathbb{R}^D$ , and where we recall that  $\bar{\alpha}_{k,j} \in [0, 1]$ ,  $k, j = 0, \dots, p$ , are specified weights such that  $\sum_{j=0}^p \bar{\alpha}_{k,j} \leq 1$  for all  $k$ .

Let us now assume the particular setting described in Remark 3.7. Namely, take  $Y = \mathbb{R}^{pN}$ , assume for simplicity  $\mu_0$  is the Lebesgue measure, so that  $d\mu/d\mu_0(\mathbf{q}) = \pi(\mathbf{q})$  for a.e.  $\mathbf{q} \in \mathbb{R}^N$ , and suppose  $\pi(\mathbf{q}) > 0$  for a.e.  $\mathbf{q} \in \mathbb{R}^N$ . Also, assume  $\mathcal{V}_0 = \dots = \mathcal{V}_p =: \mathcal{V}$ , with  $\mathcal{V}$  as given in (4.6), and let  $S_j : \mathbb{R}^{(p+1)N} \rightarrow \mathbb{R}^{(p+1)N}$ ,  $j = 0, \dots, p$ , be the flip involutions defined in (2.13). Then, under this setting, (4.12) reduces to the same expression as in (4.11), cf. (3.17), whereas (4.13) reduces to

$$\alpha_{k,j}(\mathbf{q}_0, \mathbf{v}) = \begin{cases} \bar{\alpha}_{k,j} \left[ 1 \wedge \frac{\pi(\mathbf{q}_j)}{\pi(\mathbf{q}_k)} \right] & \text{if } j \neq k, \\ 1 - \sum_{\substack{l=0 \\ l \neq k}}^p \alpha_{k,l}(\mathbf{q}_0, \mathbf{v}) & \text{if } j = k \end{cases} \quad (4.14)$$

for  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^{(p+1)N}$ , cf. (3.18).

#### 4.2 Multiproposal pCN algorithms

We next develop a multiproposal version of the pCN (Beskos *et al.*, 2008; Cotter *et al.*, 2013). As in the previous subsection, we proceed by drawing on the conditionally independent formalism we developed above in Section 2.3 to extend the standard pCN proposal mechanism.

Recall that the pCN algorithm is a methodology built to resolve infinite-dimensional measures which are absolutely continuous with respect to a Gaussian base measure  $\mu_0$ . We therefore begin by briefly reviewing this Gaussian formalism as suits our purposes here; see, e.g., (Bogachev, 1998; Da Prato & Zabczyk, 2014) for a systematic treatment. Take  $X$  to be a real separable Hilbert space, with inner product and norm denoted by  $\langle \cdot, \cdot \rangle$  and  $|\cdot|$ , respectively. Consider any  $\tilde{\mathcal{C}} : X \rightarrow X$ , which is a trace-class, symmetric and strictly positive definite linear operator.<sup>2</sup> Let us recall that  $\nu_0$  is a Gaussian measure on  $X$  with mean  $\mathbf{m} \in X$  and covariance operator  $\tilde{\mathcal{C}}$ , denoted  $\nu_0 = N(\mathbf{m}, \tilde{\mathcal{C}})$ , if  $\varphi^* \mu_0$  is normally distributed for any bounded linear functional  $\varphi : X \rightarrow \mathbb{R}$  and

$$\int_X \langle \hat{\mathbf{q}}, \mathbf{q} \rangle \mu_0(d\hat{\mathbf{q}}) = \langle \mathbf{m}, \mathbf{q} \rangle, \quad \int_X \langle \hat{\mathbf{q}} - \mathbf{m}, \mathbf{q} \rangle \langle \hat{\mathbf{q}} - \mathbf{m}, \tilde{\mathbf{q}} \rangle \mu_0(d\hat{\mathbf{q}}) = \langle \tilde{\mathcal{C}} \mathbf{q}, \tilde{\mathbf{q}} \rangle, \quad \text{for any } \mathbf{q}, \tilde{\mathbf{q}} \in X.$$

Note that, as one would expect extending the finite-dimensional case, the associated characteristic function of  $\nu_0$  is

$$\xi \in X \mapsto \int_X \exp(i\langle \xi, \mathbf{q} \rangle) \nu_0(d\mathbf{q}) = \exp\left(i\langle \xi, \mathbf{m} \rangle - \frac{1}{2}\langle \tilde{\mathcal{C}}\xi, \xi \rangle\right). \quad (4.15)$$

A standard way to draw samples from such a  $\nu_0$  is to use a *Karhunen–Loëve expansion* as follows. According to the Hilbert–Schmidt theorem, we can find a complete orthonormal system  $\{\mathbf{e}_k\}_{k \geq 0}$  of eigenfunction of  $\tilde{\mathcal{C}}$  so that  $\tilde{\mathcal{C}}\mathbf{e}_k = \mu_k \mathbf{e}_k$  for any  $k \geq 0$ . Drawing an i.i.d sequence  $\{\xi_k\}_{k \geq 1}$  of normal

---

<sup>2</sup> In other words  $\tilde{\mathcal{C}}$  is bounded and linear such that  $\langle \tilde{\mathcal{C}}\mathbf{q}, \tilde{\mathbf{q}} \rangle = \langle \tilde{\mathcal{C}}\tilde{\mathbf{q}}, \mathbf{q} \rangle$ , for any  $\tilde{\mathbf{q}}, \mathbf{q} \in X$ ,  $\langle \tilde{\mathcal{C}}\mathbf{q}, \mathbf{q} \rangle > 0$  whenever  $\mathbf{q} \in X \setminus \{0\}$  and  $\sum_{k=1}^{\infty} \langle \tilde{\mathcal{C}}\mathbf{e}_k, \mathbf{e}_k \rangle < \infty$  for any complete orthonormal system  $\{\mathbf{e}_k\}_{k \geq 0}$ .

random variables in  $\mathbb{R}$  with mean zero and variance one, we find that

$$\mathbf{w} := \mathbf{m} + \sum_{k=1}^{\infty} \sqrt{\mu_k} \mathbf{e}_k \xi_k \sim v_0. \quad (4.16)$$

The pCN algorithm is used to sample from a target probability measure on  $X$  of the form

$$\mu(d\mathbf{q}) = \frac{1}{Z} e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q}), \quad Z = \int_X e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q}). \quad (4.17)$$

Here  $\mu_0 = N(0, \mathcal{C})$  with  $\mathcal{C}$  symmetric, positive and trace-class, and we suppose that  $\Phi : X \rightarrow \mathbb{R}$  is a potential function such that  $e^{-\Phi(\mathbf{q})}$  is  $\mu_0$ -integrable. The idea in Beskos *et al.* (2008); Cotter *et al.* (2013) is to develop a proposal kernel to sample from  $\mu$  by taking a Crank-Nicolson discretization of the following Ornstein–Uhlenbeck dynamics

$$d\mathbf{q} = -\frac{1}{2}\mathbf{q}dt + \sqrt{\mathcal{C}}dW, \quad (4.18)$$

where  $W$  is a cylindrical Brownian motion on  $X$  so that, for any  $t > s \geq 0$ ,  $\sqrt{\mathcal{C}}(W(t) - W(s)) = N(0, (t-s)\mathcal{C})$  (see e.g. Da Prato & Zabczyk, 2014). Here note that the preconditioned dynamics (4.18) maintains  $\mu_0$  as an invariant and indeed the choice of a Crank-Nicolson scheme for (4.18) is selected precisely to preserve this invariance under numerical discretization. Concretely, this yields a proposal kernel  $Q : X \times \mathcal{B}(X) \rightarrow [0, 1]$  given by

$$Q(\mathbf{q}_0, d\tilde{\mathbf{q}}_0) = F(\mathbf{q}_0, \cdot)^* \mu_0(d\tilde{\mathbf{q}}_0) \sim N(\rho\mathbf{q}_0, (1-\rho^2)\mathcal{C}) \quad (4.19)$$

where  $F : X \times X \rightarrow X$  is defined as

$$F(\mathbf{q}, \mathbf{w}) = \rho\mathbf{q} + \sqrt{1-\rho^2}\mathbf{w}, \quad (4.20)$$

for some tuning parameter  $\rho \in [0, 1]$ . Note that  $\rho$  is given in terms of the time step  $\delta$  from the Crank-Nicolson discretization of (4.18) as  $\rho = (4-\delta)/(4+\delta)$ . Note moreover that proposals can be generated from (4.19), (4.20) by making use of an expansion of the form (4.16).

We derive our multiproposal version of the pCN algorithm based on  $Q$  in (4.19) to sample from  $\mu$  as in (4.17) as a direct corollary of Theorem 2.11. Fix the number of samples per step as  $p > 0$  and take  $Y = X^p$ , i.e. the  $p$ -fold product of  $X$ . After (2.12), we take  $\mathcal{V} : X \times \mathcal{B}(X^p) \rightarrow [0, 1]$  as

$$\mathcal{V}(\mathbf{q}_0, d\mathbf{v}) = \mathcal{V}(\mathbf{q}_0, d\mathbf{q}_1, \dots, d\mathbf{q}_p) := \int_X \prod_{k=1}^p Q(\mathbf{q}, d\mathbf{q}_k) Q(\mathbf{q}_0, d\mathbf{q}), \quad (4.21)$$

where  $\mathbf{v} = (\mathbf{q}_1, \dots, \mathbf{q}_p)$  and with  $Q$  given by (4.19). We also consider involution mappings  $S_j : X \times X^p \rightarrow X \times X^p$ ,  $j = 0, \dots, p$ , of the form (2.13), namely the  $S_j$ s are the coordinate flip operators:  $S_0 := I$ ;  $S_j(\mathbf{q}_0, \mathbf{v}) := S_j(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) := (\mathbf{q}_j, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{q}_0, \mathbf{q}_{j+1}, \dots, \mathbf{q}_p)$ ,  $j = 1, \dots, p$ .

**Algorithm 5** (Multiproposal pCN (mpCN))

- 
- 1: Select the algorithm parameters:
    - (i)  $\rho \in [0, 1]$ .
    - (ii) the number of proposals  $p \geq 1$ .
  - 2: Choose  $\mathbf{q}^{(0)} \in X$ .
  - 3: **for**  $k \geq 0$  **do**
  - 4:   Sample  $\mathbf{w}^{(k+1)} \sim \mu_0$  (cf. (4.16)).
  - 5:   Compute  $\mathbf{q} := \rho\mathbf{q}^{(k)} + \sqrt{1 - \rho^2}\mathbf{w}^{(k+1)}$ .
  - 6:   Sample  $\mathbf{w}_j^{(k+1)} \sim \mu_0$  independently for  $j = 1, \dots, p$ .
  - 7:   Compute  $\mathbf{q}_j^{(k+1)} = \rho\mathbf{q} + \sqrt{1 - \rho^2}\mathbf{w}_j^{(k+1)}, j = 1, \dots, p$ . Set  $\mathbf{v}^{(k+1)} := (\mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$ .
  - 8:   Draw  $\mathbf{q}^{(k+1)}$  from the set  $(\mathbf{q}^{(k)}, \mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$  according to the probabilities  $(\alpha_0(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}), \dots, \alpha_p(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}))$  as defined in (4.22).
  - 9:    $k \rightarrow k + 1$ .
  - 10: **end for**
- 

Regarding the assumptions in Theorem 2.11, notice from (4.17) that  $\mu \ll \mu_0$  and  $d\mu/d\mu_0(\mathbf{q}) = e^{-\Phi(\mathbf{q})} > 0$  for a.e.  $\mathbf{q} \in X$ . Moreover, (2.15) reduces in our situation to showing that  $\mu_0$  is in detailed balance with respect to  $Q$ , which can easily be verified via e.g. equivalence of characteristic functionals. Indeed, it follows from (4.15) that for any  $\xi, \tilde{\xi} \in X$

$$\begin{aligned} \int_{X \times X} \exp(i(\langle \xi, \mathbf{q} \rangle + \langle \tilde{\xi}, \tilde{\mathbf{q}} \rangle)) Q(\mathbf{q}, d\tilde{\mathbf{q}}) \mu_0(d\mathbf{q}) &= \int_{X \times X} \exp(i(\langle \xi, \mathbf{q} \rangle + \langle \tilde{\xi}, \rho\mathbf{q} + \sqrt{1 - \rho^2}\tilde{\mathbf{q}} \rangle)) \mu_0(d\tilde{\mathbf{q}}) \mu_0(d\mathbf{q}) \\ &= \exp\left(-\frac{1}{2}(\langle \mathcal{C}\xi, \xi \rangle + 2\rho\langle \mathcal{C}\xi, \tilde{\xi} \rangle + \langle \mathcal{C}\tilde{\xi}, \tilde{\xi} \rangle)\right). \end{aligned}$$

An analogous calculation produces the same result for  $Q(\tilde{\mathbf{q}}, d\mathbf{q})\mu_0(d\tilde{\mathbf{q}})$ , allowing us to conclude that indeed  $Q(\tilde{\mathbf{q}}, d\mathbf{q})\mu_0(d\tilde{\mathbf{q}}) = Q(\mathbf{q}, d\tilde{\mathbf{q}})\mu_0(d\mathbf{q})$ . According to Theorem 2.10, Theorem 2.11 and Corollary 2.12, we may thus obtain a reversible sampling scheme by supplementing the above choices of  $\mathcal{V}$  and  $S = (S_0, \dots, S_p)$  with acceptance probabilities given e.g. as in (2.17), written here as

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \frac{\exp(-\Phi(\mathbf{q}_j))}{\sum_{k=0}^p \exp(-\Phi(\mathbf{q}_k))}, \quad (\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in X \times X^p, \quad j = 0, \dots, p, \quad (4.22)$$

with  $\Phi$  as appears in (4.17). Alternatively, one could also consider  $\alpha_j$  as defined in (2.18), but here we restrict our attention to the definition (4.22) in view of avoiding potentially poor behavior of the MH-type  $\alpha_j$  in (2.18) for a large number  $p$  of proposals. See, however, Section 6 below.

Algorithm 5 summarizes the reversible sampling procedure that follows from Algorithm 2 with these choices.

Under the framework of Section 3, we may also consider an extension of Algorithm 5 by allowing for resampling among the cloud of proposals at each iteration, as in Algorithm 3. Specifically, we take  $\alpha_{k,j} = \alpha_j$  as defined in (4.22) for all  $k, j = 0, \dots, p$ , together with Markov kernels  $\mathcal{V}_0 = \dots = \mathcal{V}_p = \mathcal{V}$ ,

**Algorithm 6** (Multiproposal pCN with Proposal Resampling)

---

```

1: Select the algorithm parameters:
   (i)  $\rho \in [0, 1]$ .
   (ii) the number of proposals  $p \geq 1$ .
   (iii) the number of resamples per proposal cloud  $n \geq 1$ .
2: Choose  $\mathbf{q}^{(0)} \in X$ .
3: for  $k \geq 0$  do
4:   Sample  $\mathbf{w}^{(k+1)} \sim \mu_0$  (cf. (4.16)).
5:   Compute  $\mathbf{q} := \rho\mathbf{q}^{(k)} + \sqrt{1 - \rho^2}\mathbf{w}^{(k+1)}$ .
6:   Sample  $\mathbf{w}_j^{(k+1)} \sim \mu_0$  independently for  $j = 1, \dots, p$ .
7:   Compute  $\mathbf{q}_j^{(k+1)} = \rho\mathbf{q} + \sqrt{1 - \rho^2}\mathbf{w}_j^{(k+1)}$ ,  $j = 1, \dots, p$ . Set  $\mathbf{v}^{(k+1)} := (\mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$ .
8:   for  $l = 1, \dots, n$  do
9:     Draw  $\mathbf{r}^{(nk+l)}$  from the set  $(\mathbf{q}^{(k)}, \mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$  with the probabilities
        $(\alpha_0(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}), \dots, \alpha_p(\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)}))$  defined in (4.22).
10:     $l \rightarrow l + 1$ .
11:   end for
12:   Set  $\mathbf{q}^{(k+1)} := \mathbf{r}^{(n(k+1))}$ .
13:    $k \rightarrow k + 1$ .
14: end for

```

---

with  $\mathcal{V}$  as in (4.21), and again the flip involutions  $S_j, j = 0, \dots, p$ , as in (2.13). In this case, it follows from Remark 3.8 that (3.19) holds, so that we may write this particular case of Algorithm 3 as in Algorithm 6. Notice that by setting the number of resamples as  $n = 1$  the procedure indeed coincides with Algorithm 5.

### 4.3 Multiproposal HMC algorithms

In this section, we present a few instances of Hamiltonian Monte Carlo (HMC)-like algorithms based on multiple proposals, restricting our attention to the finite-dimensional case for simplicity. We emphasize that the general scope of Algorithm 1 and Algorithm 3 allow for the possibility of a variety of other HMC-like sampling schemes under different choices of algorithmic parameters and acceptance probabilities, which we will explore in future work.

We proceed by briefly recalling some generalities regarding HMC algorithms, following a similar presentation as in our recent contribution (Glatt–Holtz *et al.*, 2023). For more complete details, we refer to e.g. the pioneering works (Duane *et al.*, 1987; Neal, 1993) and also to Bou-Rabee & Sanz-Serna (2018); Hairer *et al.* (2006); Leimkuhler & Reich (2004); Neal (2011).

Let us consider spaces  $X = Y = \mathbb{R}^N$  and fix a target measure of the form

$$\mu(d\mathbf{q}) = \frac{1}{Z} e^{-\Phi(\mathbf{q})} d\mathbf{q}, \quad Z = \int_{\mathbb{R}^N} e^{-\Phi(\mathbf{q})} d\mathbf{q}, \quad (4.23)$$

for some potential function  $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}$  that we assume to be in  $C^1(\mathbb{R}^N)$  and such that  $e^{-\Phi(\mathbf{q})} \in L^1(\mathbb{R}^N)$ . An HMC algorithm samples from this measure by first selecting a Hamiltonian function

$\mathcal{H} : \mathbb{R}^{2N} \rightarrow \mathbb{R}$  such that the marginal of the associated Gibbs measure

$$\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \frac{1}{Z_{\mathcal{H}}} e^{-\mathcal{H}(\mathbf{q}, \mathbf{v})} d\mathbf{q} d\mathbf{v}, \quad Z_{\mathcal{H}} = \int_{\mathbb{R}^{2N}} e^{-\mathcal{H}(\mathbf{q}, \mathbf{v})} d\mathbf{q} d\mathbf{v} \quad (4.24)$$

with respect to the ‘position’ variable  $\mathbf{q}$  coincides with the target measure  $\mu$  in (4.23). We may thus write such Hamiltonian function in a general form as

$$\mathcal{H}(\mathbf{q}, \mathbf{v}) = \Phi(\mathbf{q}) + \Psi(\mathbf{q}, \mathbf{v}), \quad (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N},$$

for some  $C^1$  function  $\Psi : \mathbb{R}^{2N} \rightarrow \mathbb{R}$  with  $\int_{\mathbb{R}^N} e^{-\Psi(\mathbf{q}, \mathbf{v})} d\mathbf{v} = 1$  for all  $\mathbf{q} \in \mathbb{R}^N$ . Under this definition,  $\mathcal{M}$  can be written as

$$\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q}), \quad \text{for } \mathcal{V}(\mathbf{q}, d\mathbf{v}) = e^{-\Psi(\mathbf{q}, \mathbf{v})} d\mathbf{v}, \quad (4.25)$$

where it follows by construction that  $\mathcal{V}$  is a Markov kernel.

In association with such  $\mathcal{H}$ , one considers the following Hamiltonian dynamic for the pair  $\mathbf{y} = (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$

$$\frac{d\mathbf{y}}{dt} = J^{-1} \nabla \mathcal{H}(\mathbf{y}), \quad \mathbf{y}(0) = (\mathbf{q}_0, \mathbf{v}_0), \quad (4.26)$$

for some  $2N \times 2N$  real matrix  $J$  that is invertible and antisymmetric.<sup>3</sup> This dynamic leaves  $\mathcal{H}$  invariant and preserves volume elements<sup>4</sup>. As such,  $\mathcal{M}$  is an invariant measure under the associated flow. This in turn implies that the projected dynamic of the  $\mathbf{q}$ -variable leaves invariant the  $\mathbf{q}$ -marginal of  $\mathcal{M}$ , namely  $\mu$ .

Of course, it is typically intractable to exactly solve (4.26). One must instead resort to a suitable numerical integrator maintaining certain indispensable geometric properties. For a chosen time step size  $\delta > 0$ , such a scheme  $\{\hat{S}_{j,\delta}(\mathbf{q}_0, \mathbf{v}_0)\}_{j \in \mathbb{N}}$  yields an approximation of the solution of (4.26) at times  $j\delta$ ,  $j \in \mathbb{N}$ .

For the given step size  $\delta > 0$ , one builds  $\{\hat{S}_{j,\delta}\}_{j \in \mathbb{N}}$  starting from the single integration step  $\Xi_\delta : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$ , which is in general supposed to be an invertible  $C^1$  mapping. We require that  $\Xi_\delta$  satisfies the following standard geometric properties (see, e.g., Hairer *et al.*, 2006; Leimkuhler & Reich, 2004):

(P1)  $\Xi_\delta$  is *volume-preserving*, i.e.

$$|\det \nabla \Xi_\delta(\mathbf{q}, \mathbf{v})| = 1 \quad \text{for every } (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}; \quad (4.27)$$

---

<sup>3</sup> Typical choices for  $J$  and  $\Psi$  are given by  $J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$  and  $\Psi(\mathbf{q}, \mathbf{v}) = \frac{1}{2} \langle M(\mathbf{q})^{-1} \mathbf{v}, \mathbf{v} \rangle + \frac{1}{2} \ln((2\pi)^N \det(M(\mathbf{q})))$ , where  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product, for some ‘mass’ matrix  $M$ . In this case,  $\Psi(\mathbf{q}, \mathbf{v})$  corresponds to the negative log-density of the Gaussian  $N(0, M(\mathbf{q}))$  in  $\mathbb{R}^N$ .

<sup>4</sup> The solution map is actually symplectic, a stronger property that implies the preservation of volumes. See (Glatt–Holtz *et al.*, 2023), Definition 4.1) in our context.

(P2)  $\Xi_\delta$  is *reversible* with respect to the ‘momentum’-flip involution

$$R(\mathbf{q}, \mathbf{v}) := (\mathbf{q}, -\mathbf{v}), \quad (4.28)$$

i.e.

$$R \circ \Xi_\delta(\mathbf{q}, \mathbf{v}) = \Xi_\delta^{-1} \circ R(\mathbf{q}, \mathbf{v}) \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}. \quad (4.29)$$

Or, equivalently,  $(R \circ \Xi_\delta)^2 = I$ , i.e.  $R \circ \hat{S}$  is an involution.

We now define

$$\hat{S}_{j,\delta} = \Xi_\delta^j \text{ for } j \in \mathbb{Z}, \quad (4.30)$$

where  $\Xi_\delta^j$  denotes the  $j$ -fold composition of  $\Xi_\delta$  for  $j \geq 1$ , the identity map for  $j = 0$  and the  $j$ -fold composition of  $\Xi_\delta^{-1}$  for  $j \leq -1$ . Note that it follows immediately that for  $j \in \mathbb{N}$ ,  $\hat{S}_{j,\delta}$  maintains the same volume preservation and reversibility properties à la (4.27), (4.29).

As a concrete example, we recall that when the Hamiltonian is written in the separable form  $\mathcal{H}(\mathbf{q}, \mathbf{v}) = \Phi(\mathbf{q}) + \Psi(\mathbf{v})$  and  $J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$ , the associated Hamiltonian dynamic is  $d\mathbf{q}/dt = \nabla \Psi(\mathbf{v})$ ,  $d\mathbf{v}/dt = -\nabla \Phi(\mathbf{q})$ . In this case, the classical leapfrog integrator is defined as

$$\Xi_\delta := \Xi_{\delta/2}^{(1)} \circ \Xi_\delta^{(2)} \circ \Xi_{\delta/2}^{(1)}, \quad (4.31)$$

where  $\Xi^{(1)}$  and  $\Xi^{(2)}$  are the exact solution operators of  $d\mathbf{q}/dt = 0$ ,  $d\mathbf{v}/dt = -\nabla \Phi(\mathbf{q})$ , and  $d\mathbf{q}/dt = \nabla \Psi(\mathbf{v})$ ,  $d\mathbf{v}/dt = 0$ , respectively. Namely,

$$\Xi_t^{(1)}(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, \mathbf{v} + t\nabla \Phi(\mathbf{q})) \quad \text{and} \quad \Xi_t^{(2)}(\mathbf{q}, \mathbf{v}) = (\mathbf{q} + t\nabla \Psi(\mathbf{v}), \mathbf{v}) \quad (4.32)$$

for all  $t \in \mathbb{R}$ . This classical leapfrog scheme is the basis for many of the common HMC implementations.

Fixing now  $\delta > 0$ ,  $p \geq 1$  and an integration scheme defined by  $\Xi_\delta$  maintaining (P1),(P2), we now construct a multiproposal sampling procedure starting from the setting of Section 2.1, Algorithm 1. Our approach makes use of a sequence of integration steps as the multiple proposals at each iteration. For this, we define the mappings, cf. (4.28), (4.30),

$$S_0 := I, \quad S_j := R \circ \hat{S}_{j,\delta}, \quad \text{for } j = 1, \dots, p. \quad (4.33)$$

Each of these maps are involutions according to property (P2) above. Property (P1) furthermore implies that each of these maps  $S_j$  are volume preserving so that, cf. (4.3),

$$\frac{dS_j^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \exp(\mathcal{H}(\mathbf{q}, \mathbf{v}) - \mathcal{H}(S_j(\mathbf{q}, \mathbf{v}))). \quad (4.34)$$

**Algorithm 7** (Multiproposal HMC (mHMC))

---

1: Select the algorithm parameters:

- (i) The ‘momentum’ kernel  $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = e^{-\Psi(\mathbf{q}, \mathbf{v})} d\mathbf{v}$ .
- (ii) The time step size  $\delta > 0$  and the corresponding integration step  $\Xi_\delta : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$  satisfying (P1), (P2) used for approximating the Hamiltonian dynamic (4.26) (see (4.31), (4.32)).
- (iii) The number of integration steps  $p \geq 1$ .
- (iv) The acceptance weights  $\bar{\alpha}_j \in [0, 1]$  such that  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ .

2: Choose  $\mathbf{q}^{(0)} \in \mathbb{R}^N$ .

3: **for**  $k \geq 0$  **do**

4:     Sample  $\mathbf{v}^{(k+1)} \sim \mathcal{V}(\mathbf{q}^{(k)}, \cdot)$ .

5:     Set  $(\bar{\mathbf{q}}^{(k,0)}, \bar{\mathbf{v}}^{(k,0)}) := (\mathbf{q}^{(k)}, \mathbf{v}^{(k+1)})$ .

6:     **for**  $j = 0, \dots, p - 1$  **do**

7:         Compute  $(\bar{\mathbf{q}}^{(k,j+1)}, \bar{\mathbf{v}}^{(k,j+1)}) := \Xi_\delta(\bar{\mathbf{q}}^{(k,j)}, \bar{\mathbf{v}}^{(k,j)})$ .

8:          $j \rightarrow j + 1$ .

9:     **end for**

10:    Set  $\mathbf{q}^{(k+1)}$  by drawing from  $(\mathbf{q}^{(k)}, \bar{\mathbf{q}}^{(k,1)}, \dots, \bar{\mathbf{q}}^{(k,p)})$  according to the probabilities  $(\alpha_0^{(k)}, \dots, \alpha_p^{(k)})$  where

$$\alpha_j^{(k)} := \bar{\alpha}_j \left[ 1 \wedge \exp(\mathcal{H}(\bar{\mathbf{q}}^{(k,0)}, \bar{\mathbf{v}}^{(k,0)}) - \mathcal{H}(\bar{\mathbf{q}}^{(k,j)}, -\bar{\mathbf{v}}^{(k,j)})) \right] \text{ for } j = 1, \dots, p, \text{ and } \alpha_0^{(k)} := 1 - \sum_{j=1}^p \alpha_j^{(k)}.$$

11:      $k \rightarrow k + 1$ .

12: **end for**

---

Therefore, it follows from Corollary 2.6 that we may obtain a reversible sampling algorithm by supplementing  $\mathcal{V}$  from (4.25) and these mappings  $S_j$  with the following acceptance probabilities

$$\alpha_j(\mathbf{q}, \mathbf{v}) = \bar{\alpha}_j \left[ 1 \wedge e^{\mathcal{H}(\mathbf{q}, \mathbf{v}) - \mathcal{H}(S_j(\mathbf{q}, \mathbf{v}))} \right], \quad j = 1, \dots, p; \quad \alpha_0(\mathbf{q}, \mathbf{v}) = 1 - \sum_{j=1}^p \alpha_j(\mathbf{q}, \mathbf{v}), \quad (4.35)$$

for all  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$ , where  $\bar{\alpha}_j$  are given weights such that  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ . Algorithm 7 summarizes the resulting reversible procedure that follows as a particular case of Algorithm 1 with these choices.

**REMARK 4.2** Note that one could also consider in Algorithm 7 acceptance probabilities  $\alpha_j, j = 0, \dots, p$ , in the corresponding Barker form (2.8). However, to guarantee reversibility or even just ergodicity in this case according to Corollary 2.4, the summation condition (2.7) would be imposed, thus introducing a seemingly restrictive assumption on the numerical integrator.

Furthermore, under the framework of Section 3, we can also construct a multiproposal HMC algorithm that allows for proposal resampling. In this case, we choose in Algorithm 3 Markov kernels  $\mathcal{V}_0 = \dots = \mathcal{V}_p =: \mathcal{V}$ , with  $\mathcal{V}$  as in (4.25), and again consider the involution mappings  $S_j$  from the integration of (4.26) given in (4.33). From Corollary 3.5, we may supplement these choices with e.g.

either of the following definitions of acceptance probabilities to yield an unbiased sampling algorithm, cf. (4.12)–(4.13):

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) = \frac{e^{-\mathcal{H}(S_j(\mathbf{q}, \mathbf{v}))}}{\sum_{l=0}^p e^{-\mathcal{H}(S_l(\mathbf{q}, \mathbf{v}))}}, \quad (4.36)$$

or

$$\alpha_{k,j}(\mathbf{q}, \mathbf{v}) = \begin{cases} \bar{\alpha}_{k,j} \left[ 1 \wedge e^{\mathcal{H}(S_k(\mathbf{q}, \mathbf{v})) - \mathcal{H}(S_j(\mathbf{q}, \mathbf{v}))} \right] & \text{if } j \neq k, \\ 1 - \sum_{\substack{l=0 \\ l \neq k}}^p \alpha_{k,l}(\mathbf{q}, \mathbf{v}) & \text{if } j = k \end{cases} \quad (4.37)$$

for all  $k, j = 0, \dots, p$  and  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$ , where in (4.37)  $\bar{\alpha}_{k,j} \in [0, 1]$ ,  $k, j = 0, \dots, p$ , are chosen weights satisfying  $\sum_{j=0}^p \bar{\alpha}_{k,j} \leq 1$  for all  $k$ .

In fact, due to step 7 in Algorithm 3, at each iteration the procedure requires the calculation of  $\alpha_{k,j}(S_m(\mathbf{q}, \mathbf{v})) = \alpha_{k,j}(R \circ \hat{S}_{m,\delta}(\mathbf{q}, \mathbf{v}))$  for some  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$  and some  $m \in \mathbb{N}$ . Note that, according to (P2) above, we have

$$S_j \circ S_m = R \circ \hat{S}_{j,\delta} \circ R \circ \hat{S}_{m,\delta} = \hat{S}_{j,\delta}^{-1} \circ \hat{S}_{m,\delta} = (\Xi_\delta^j)^{-1} \circ \Xi_\delta^m = (\Xi_\delta^{-1})^j \circ \Xi_\delta^m = \Xi_\delta^{m-j},$$

for all  $m, j = 0, \dots, p$ , where recall that we denote  $\Xi_\delta^{-l} := (\Xi_\delta^{-1})^l$  for all  $l \in \mathbb{N}$ . Therefore, it follows respectively under (4.36) and (4.37) that

$$\alpha_{k,j}(S_m(\mathbf{q}, \mathbf{v})) = \alpha_{k,j}(R \circ \hat{S}_{m,\delta}(\mathbf{q}, \mathbf{v})) = \frac{\exp(-\mathcal{H}(\Xi_\delta^{m-j}(\mathbf{q}, \mathbf{v})))}{\sum_{l=0}^p \exp(-\mathcal{H}(\Xi_\delta^{m-l}(\mathbf{q}, \mathbf{v})))}, \quad (4.38)$$

or

$$\alpha_{k,j}(S_m(\mathbf{q}, \mathbf{v})) = \alpha_{k,j}(R \circ \hat{S}_{m,\delta}(\mathbf{q}, \mathbf{v})) = \begin{cases} \bar{\alpha}_{k,j} \left[ 1 \wedge e^{\mathcal{H}(\Xi_\delta^{m-k}(\mathbf{q}, \mathbf{v})) - \mathcal{H}(\Xi_\delta^{m-j}(\mathbf{q}, \mathbf{v}))} \right] & \text{if } j \neq k, \\ 1 - \sum_{\substack{l=0 \\ l \neq k}}^p \alpha_{k,l}(\mathbf{q}, \mathbf{v}) & \text{if } j = k \end{cases} \quad (4.39)$$

for all  $k, j, m = 0, \dots, p$  and  $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$ . Algorithm 8 details the sampling procedure resulting from these observations.

**REMARK 4.3** In contrast to the situation for the multiproposal pCN algorithms in Section 4.2, cf. Remark 3.8, it is clear by comparing the previous two algorithms that Algorithm 8 with  $n = 1$  and under the choice of  $\alpha_{k,j}$  in (4.39) yields a different procedure than Algorithm 7.

#### 4.4 Simplicial samplers

In this section, we cast into the framework of Section 2 and Section 3 a class of algorithms recently introduced in Holbrook (2023a) (see also Tjelmeland (2004)), where proposals are generated as the vertices of a fixed regular simplex after a random rescaling of edge lengths, random rotation and translation by the current state. We make the algorithmic setting more precise as follows.

**Algorithm 8** (Multiproposal HMC with Proposal Resampling)

---

1: Select the algorithmic parameters

- (i) The “momentum” kernel  $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = e^{-\Psi(\mathbf{q}, \mathbf{v})} d\mathbf{v}$ .
- (ii) The time step size  $\delta > 0$  and the corresponding integration step  $\Xi_\delta : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$  satisfying (P1), (P2) used in the scheme  $\hat{S}_{k,\delta} = \Xi_\delta^k$  approximating the Hamiltonian dynamic (4.26) (see (4.31), (4.32)).
- (iii) The number of integration steps  $p \geq 1$ .
- (iv) If using the mechanism (4.39), specify the weights  $\bar{\alpha}_j \in [0, 1], j = 1, \dots, p$ , such that  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ .
- (v) The  $n \geq 1$  number of samples drawn per generated proposal cloud.

2: Choose an initial  $(\mathbf{q}^{(0)}, \mathbf{v}^{(0)}) \in \mathbb{R}^{2N}$  and  $k^{(0)} \in \{0, \dots, p\}$ .

3: Set  $\mathbf{r}^{(0)} := \Pi_1 \hat{S}_{k^{(0)}, \delta}(\mathbf{q}^{(0)}, \mathbf{v}^{(0)})$ .

4: **for**  $i \geq 0$  **do**

5:     Set  $\mathbf{q}^{(i+1)} := \mathbf{r}^{(ni)}$ .

6:     Sample  $\mathbf{v}^{(i+1)} \sim \mathcal{V}_{k^{(i)}}(\mathbf{q}^{(i+1)}, d\mathbf{v})$ .

7:     Set  $(\bar{\mathbf{q}}^{(k^{(i)}, k^{(i)})}, \bar{\mathbf{v}}^{(k^{(i)}, k^{(i)})}) := (\mathbf{q}^{(i+1)}, \mathbf{v}^{(i+1)})$ .

8:     **for**  $j = 0, \dots, k^{(i)} - 1$  **do**

9:         Compute  $(\bar{\mathbf{q}}^{(k^{(i)}, k^{(i)} - j - 1)}, \bar{\mathbf{v}}^{(k^{(i)}, k^{(i)} - j - 1)}) := \Xi_\delta(\bar{\mathbf{q}}^{(k^{(i)}, k^{(i)} - j)}, \bar{\mathbf{v}}^{(k^{(i)}, k^{(i)} - j)})$ .

10:          $j \rightarrow j + 1$ .

11:     **end for**

12:     **for**  $j = k^{(i)}, \dots, p - 1$  **do**

13:         Compute  $(\bar{\mathbf{q}}^{(k^{(i)}, j + 1)}, \bar{\mathbf{v}}^{(k^{(i)}, j + 1)}) := \Xi_\delta^{-1}(\bar{\mathbf{q}}^{(k^{(i)}, j)}, \bar{\mathbf{v}}^{(k^{(i)}, j)})$ .

14:          $j \rightarrow j + 1$ .

15:     **end for**

16:     Set  $k_{cur} := k^{(i)}$ .

17:     **for**  $m = 1, \dots, n$  **do**

18:         Draw  $k_{nxt} \in \{0, \dots, p\}$  with the probabilities  $(\alpha_{k_{cur}, 0}^i, \dots, \alpha_{k_{cur}, p}^i)$  where

$$\alpha_{k_{cur}, j}^i := \frac{\exp(-\mathcal{H}(\bar{\mathbf{q}}^{(k^{(i)}, j)}, \bar{\mathbf{v}}^{(k^{(i)}, j)}))}{\sum_{l=0}^p \exp(-\mathcal{H}(\bar{\mathbf{q}}^{(k^{(i)}, l)}, \bar{\mathbf{v}}^{(k^{(i)}, l)}))}, \quad j = 0, \dots, p,$$

or

$$\alpha_{k_{cur}, j}^i = \begin{cases} \bar{\alpha}_{k_{cur}, j} \left[ 1 \wedge \exp \left( \mathcal{H}(\bar{\mathbf{q}}^{(k^{(i)}, k_{cur})}, \bar{\mathbf{v}}^{(k^{(i)}, k_{cur})}) - \mathcal{H}(\bar{\mathbf{q}}^{(k^{(i)}, j)}, \bar{\mathbf{v}}^{(k^{(i)}, j)}) \right) \right] & \text{if } j \neq k_{cur}, \\ 1 - \sum_{\substack{l=0 \\ l \neq k_{cur}}}^p \alpha_{k_{cur}, l}^i(\mathbf{q}, \mathbf{v}) & \text{if } j = k_{cur}. \end{cases}$$

19:         Set  $k_{cur} := k_{nxt}$ .

20:         Set  $\mathbf{r}^{(ni+m)} := \bar{\mathbf{q}}^{(k^{(i)}, k_{cur})}$ .

21:          $m \rightarrow m + 1$ .

22:     **end for**

23:     Set  $k^{(i+1)} = k_{cur}$ .

24:      $i \rightarrow i + 1$ .

25: **end for**

---

Fix  $N \geq 1$  and a continuously distributed target measure  $\mu \in \text{Pr}(\mathbb{R}^N)$ , namely  $\mu(d\mathbf{q}) = \pi(\mathbf{q})d\mathbf{q}$  for some density function  $\pi : \mathbb{R}^N \rightarrow \mathbb{R}^+$ . Fix also  $p \leq N$  and a  $p$ -simplex in  $\mathbb{R}^N$  with equidistant vertices  $\mathbf{w}_1, \dots, \mathbf{w}_p, \mathbf{0} \in \mathbb{R}^N$ , where we set  $\mathbf{0}$  as one of the vertices for convenience. Specifically,  $\mathbf{w}_1, \dots, \mathbf{w}_p$  are linearly independent vectors that we assume for simplicity to have unit norm and pairwise distance, namely

$$\|\mathbf{w}_j\| = 1 \text{ for } j = 1, \dots, p; \text{ and } \|\mathbf{w}_j - \mathbf{w}_k\| = 1, \text{ for all } j, k = 1, \dots, p, j \neq k. \quad (4.40)$$

The corresponding  $p$ -simplex is defined as the convex hull of  $\mathbf{w}_1, \dots, \mathbf{w}_p, \mathbf{0}$ , so that for, e.g.,  $p = 1, 2, 3$  it is given respectively by a line, a triangle, and a tetrahedron.

To perform random rotations and rescalings of this fixed simplex, we set the following additional framework. Denote by  $\mathcal{O}_N$  the  $N$ -dimensional orthogonal group, i.e., the group formed by all real  $N \times N$  orthogonal matrices  $Q$ , namely  $QQ^T = I$ . We endow  $\mathcal{O}_N$  with the Borel  $\sigma$ -algebra of  $\mathbb{R}^{N \times N} \simeq \mathbb{R}^{N^2}$ . Since  $\mathcal{O}_N$  is a locally compact topological group, it admits a unique normalized left Haar measure  $\mathcal{H}_N$  on  $\mathcal{O}_N$  (Folland, 1999; Halmos, 2013; Stewart, 1980). Here by normalized we mean precisely  $\mathcal{H}_N(\mathcal{O}_N) = 1$ , and we recall that a left Haar measure is defined as a nonzero left-invariant Radon measure, where the left-invariance property means that

$$\mathcal{H}_N(Q(E)) = \mathcal{H}_N(E) \quad \text{for all } Q \in \mathcal{O}_N \text{ and Borel set } E \subset \mathcal{O}_N, \quad (4.41)$$

or, equivalently,  $(Q^T)^* \mathcal{H}_N = \mathcal{H}_N$  for every  $Q \in \mathcal{O}_N$ .

Let us additionally fix a probability distribution of edge lengths  $\lambda > 0$  given by a measure  $\nu(d\lambda)$  in  $\mathbb{R}^+$ . We define a mapping  $F : \mathbb{R}^N \times \mathbb{R}^+ \times \mathcal{O}_N \rightarrow \mathbb{R}^{Np}$  that for each  $(\mathbf{q}, \lambda, Q) \in \mathbb{R}^N \times \mathbb{R}^+ \times \mathcal{O}_N$  yields a rescaling of the vertices  $\mathbf{w}_1, \dots, \mathbf{w}_p$  by  $\lambda$ , applies the orthogonal transformation  $Q$ , and then shifts the output vertices  $\lambda Q \mathbf{w}_1, \dots, \lambda Q \mathbf{w}_p \in \mathbb{R}^N$  by  $\mathbf{q}$ . Namely,

$$F(\mathbf{q}, \lambda, Q) = \mathbf{v} := (\mathbf{q} + \lambda Q \mathbf{w}_1, \dots, \mathbf{q} + \lambda Q \mathbf{w}_p) \quad \text{for all } (\mathbf{q}, \lambda, Q) \in \mathbb{R}^N \times \mathbb{R}^+ \times \mathcal{O}_N.$$

From a current state  $\mathbf{q} \in \mathbb{R}^N$ , one then draws independently  $\lambda \sim \nu$ ,  $Q \sim \mathcal{H}_N$ , and proposes  $p$  new states given by  $F(\mathbf{q}, \lambda, Q)$ . This proposal procedure can be written explicitly in terms of the following Markov kernel

$$\mathcal{V}(\mathbf{q}, d\mathbf{v}) = F(\mathbf{q}, \cdot)^*(\nu \otimes \mathcal{H}_N)(d\mathbf{v}) \quad \text{for } \mathbf{q} \in \mathbb{R}^N, \mathbf{v} \in \mathbb{R}^{Np}. \quad (4.42)$$

Next, similarly as in (2.13), we define involution mappings  $S_j : \mathbb{R}^{N(p+1)} \rightarrow \mathbb{R}^{N(p+1)}$ ,  $j = 0, \dots, p$ , given as the coordinate flip operators

$$S_0 := I; \quad S_j(\mathbf{q}_0, \mathbf{v}) := S_j(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) := (\mathbf{q}_j, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{q}_0, \mathbf{q}_{j+1}, \dots, \mathbf{q}_p) \quad (4.43)$$

for all  $(\mathbf{q}_0, \mathbf{v}) \in \mathbb{R}^{N(p+1)}$  and  $j = 1, \dots, p$ .

To complete the parameter setup from Algorithm 1, we consider acceptance probabilities  $\alpha_j$ ,  $j = 0, \dots, p$ , given as either the Barker type (2.8) or the MH type (2.10)–(2.11). Before providing the explicit interpretation of these formulas within this special setting, we justify these choices of  $\alpha_j$  by noticing that condition (2.7) in Corollary 2.4 and also the absolute continuity condition  $S_j^* \mathcal{M} \ll \mathcal{M}$

from Corollary 2.6 are indeed satisfied. These statements follow from the next two propositions, whose proofs are presented in Appendix B.7 and Appendix B.8, respectively.

**PROPOSITION 4.4** Fix any  $\mu \in \text{Pr}(\mathbb{R}^N)$ . Let  $\mathcal{V}$  be the Markov kernel in (4.42), and define the measure  $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q})$  in  $\mathbb{R}^{N(p+1)}$ . Let also  $S_j : \mathbb{R}^{N(p+1)} \rightarrow \mathbb{R}^{N(p+1)}$ ,  $j = 0, \dots, p$ , be the flip involution mappings defined in (4.43). Then, we have the following equivalence of measures:

$$\sum_{k=0}^p (S_j \circ S_k)^* \mathcal{M} = \sum_{k=0}^p S_k^* \mathcal{M} \quad (4.44)$$

for all  $j = 0, \dots, p$ .

**PROPOSITION 4.5** Let  $\mu \in \text{Pr}(\mathbb{R}^N)$  be given as  $\mu(d\mathbf{q}) = \pi(\mathbf{q})d\mathbf{q}$  for some density function  $\pi : \mathbb{R}^N \rightarrow \mathbb{R}^+$ . Set  $\mathcal{V}$  to be the Markov kernel in (4.42), and define the measure  $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\pi(\mathbf{q})d\mathbf{q}$  in  $\mathbb{R}^{N(p+1)}$ . Let also  $S_j : \mathbb{R}^{N(p+1)} \rightarrow \mathbb{R}^{N(p+1)}$ ,  $j = 0, \dots, p$ , be the flip involution mappings defined in (4.43). Then,  $S_j^* \mathcal{M} \ll \mathcal{M}$  for all  $j = 0, \dots, p$ , and

$$\frac{dS_j^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}_0, \mathbf{v}) = \frac{\pi(\mathbf{q}_j)}{\pi(\mathbf{q}_0)} \quad \text{for } \mathcal{M}\text{-a.e. } (\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^N \times \mathbb{R}^{pN}. \quad (4.45)$$

It thus follows from Corollary 2.4 and Corollary 2.6 that we can obtain a reversible sampling algorithm by combining the above choices of  $\mathcal{V}$  and  $S_j$ ,  $j = 0, \dots, p$ , in (4.42)–(4.43) with acceptance probabilities  $\alpha_j$ ,  $j = 0, \dots, p$ , given either by (2.9) or (2.10)–(2.11), which according to (4.45) are written here respectively as

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \frac{\pi(\mathbf{q}_j)}{\sum_{l=0}^p \pi(\mathbf{q}_l)}, \quad j = 0, \dots, p, \quad (4.46)$$

and

$$\alpha_j(\mathbf{q}_0, \mathbf{v}) = \bar{\alpha}_j \left[ 1 \wedge \frac{\pi(\mathbf{q}_j)}{\pi(\mathbf{q}_0)} \right], \quad j = 1, \dots, p; \quad \alpha_0(\mathbf{q}, \mathbf{v}) = 1 - \sum_{j=1}^p \alpha_j(\mathbf{q}_0, \mathbf{v}), \quad (4.47)$$

for all  $(\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in \mathbb{R}^{N(p+1)}$  such that the expression in the denominator in each case is strictly positive, and zero otherwise. As before,  $\bar{\alpha}_j \in [0, 1]$ ,  $j = 1, \dots, p$ , in (4.47) are user specified weights such that  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ . Under this setting, Algorithm 9 describes the reversible procedure for sampling from  $\mu(d\mathbf{q}) = \pi(\mathbf{q})d\mathbf{q}$  in  $\mathbb{R}^N$  as a special case of Algorithm 1.

Furthermore, we can also obtain a simplicial sampler algorithm allowing for multiple selected states from the cloud of proposals at each iteration as in Algorithm 3. For this, we set  $\mathcal{V}_0 = \dots = \mathcal{V}_p := \mathcal{V}$ , with  $\mathcal{V}$  as in (4.42), and  $(S_0, \dots, S_p)$  as in (4.43), along with acceptance probabilities  $\alpha_{k,j}$ ,  $k, j = 0, \dots, p$ , that can be defined for example as in (3.15) or (3.16). Note that, similarly as in (4.46) and (4.47), these formulas can be written explicitly in terms of the density  $\pi$  from the target  $\mu$  by invoking (4.45). Moreover, it follows from the same arguments as in Remark 3.8 that by setting the number of jumps  $n = 1$  in the resulting algorithm would reduce it to Algorithm 9. We omit further details.

**Algorithm 9**


---

1: Select the algorithm parameters:

- (i) Linearly independent vectors  $\mathbf{w}_1, \dots, \mathbf{w}_p \in \mathbb{R}^N$ , for some  $p \leq N$ , satisfying (4.40) to form a  $p$ -simplex with vertices  $(\mathbf{w}_1, \dots, \mathbf{w}_p, \mathbf{0})$ .
- (ii) The edge length distribution  $v(d\lambda)$ .
- (iii) If using the mechanism (4.47), select the weights  $\bar{\alpha}_j \in [0, 1]$  for  $j = 1, \dots, p$  such that  $\sum_{j=1}^p \bar{\alpha}_j \leq 1$ .

2: Choose  $\mathbf{q}_0^{(0)} \in X$ .

3: **for**  $k \geq 0$  **do**

4:   Sample  $\lambda^{(k+1)} \sim v$ .

5:   Sample  $Q \sim \mathcal{H}_N$ .

6:   Compute  $\mathbf{q}_j^{(k+1)} = \mathbf{q}_0^{(k)} + \lambda Q \mathbf{w}_j$  for  $j = 1, \dots, p$ . Set  $\mathbf{v}^{(k+1)} := (\mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$ .

7:   Set  $\mathbf{q}_0^{(k+1)}$  by drawing from  $(\mathbf{q}_0^{(k)}, \mathbf{q}_1^{(k+1)}, \dots, \mathbf{q}_p^{(k+1)})$  according to the probabilities  $(\alpha_0(\mathbf{q}_0^{(k)}, \mathbf{v}^{(k+1)}), \dots, \alpha_p(\mathbf{q}_0^{(k)}, \mathbf{v}^{(k+1)}))$  as defined in (4.46) or (4.47).

8:    $k \rightarrow k + 1$ .

9: **end for**

---

## 5. Numerical case studies

This section collects various case studies that illustrate some of the potential scope of the pMCMC algorithms we derive above in Section 4. Section 5.1 considers two case studies for Algorithm 4 focusing on the case when  $p$ , the number of proposals per step, is taken to be very large. These examples illustrate the promise of pMCMC methods in the context of modern computational architectures while illuminating its efficacy in addressing highly multimodal problems. In Section 5.2 we present a series of examples in the framework of Bayesian statistical inverse problems. In order to provide some preliminary studies of Algorithm 5 we focus on the case of Gaussian priors for our unknown parameter so that we are resolving target measures of the form (4.17). Here two of our examples confront PDE constrained problems. These problems are naturally high (infinite) dimensional featuring complex correlation geometry and multimodal structure and where, particularly in our second example, gradient based methods may be impractical or costly to employ.

### 5.1 High-performance computing

The general structure of parallel MCMC algorithms leverages parallel computing resources that are becoming increasingly available and easy to use. Whereas Holbrook (2023b) considers implementation of parallel MCMC algorithms using quantum computers, here we consider parallelization using conventional parallel computing resources. We first compare the efficiency of sequential and massively parallel implementations of Algorithm 4 using an Intel Xeon CPU with 8 cores and 26GB RAM and an Nvidia Tesla P100 GPU with 3,584 cores, 16GB RAM and 730 GB/s memory bandwidth. Since Algorithm 4 enables simplified acceptance probabilities (Holbrook, 2023a), its two main computational bottlenecks are the generation of  $p$  proposals  $\mathbf{q}_j$  and the calculation of  $p$  target values  $\pi(\mathbf{q}_j)$ . Within the TENSORFLOW (Abadi *et al.*, 2016) framework, the RANDOM module accelerates the former task with

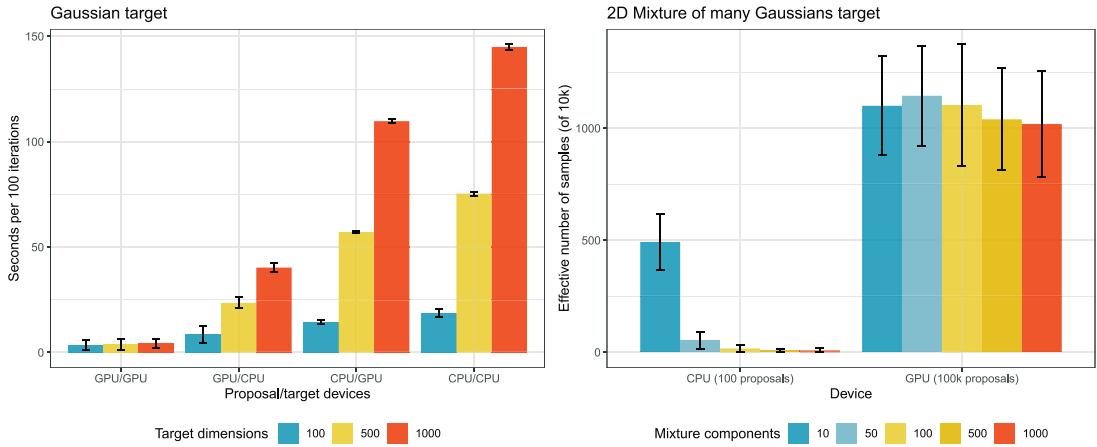


FIG. 1. Parallelization study displaying means and 95% confidence intervals. Left: we use a graphics processing unit (GPU) to accelerate embarrassingly parallel proposal and/or target evaluation steps for Algorithm 4 and compare to conventional central processing unit (CPU) implementation. Here, we fix proposal count to 100,000. Right: we demonstrate how parallelization across many target components can allow for more proposals when working with a fixed timing budget. Here, GPU uses 100,000 proposals while CPU uses 100. Crucially, both take roughly the same amount of time per iteration.

GPU powered implementations of the parallel PRNG algorithms of Salmon *et al.* (2011). For the latter task, the DISTRIBUTIONS library (Dillon *et al.*, 2017) enables GPU powered implementations of popular probability density functions that parallelize across evaluation arguments and density components when appropriate. All experiments use the isotropic Gaussian proposal kernel  $Q(\mathbf{q}, d\bar{\mathbf{q}})$  with variance tuned to obtain a 50% acceptance rate.

Figure 1 shows results from two parallelization experiments that demonstrate the speedups associated with such GPU implementations compared to sequential implementations on a CPU. In the first experiment, we record the number of seconds required for 100 iterations of Algorithm 4 using 100,000 proposals with parallelized proposal and/or acceptance steps for centered/isotropic multivariate Gaussian targets of 100, 500 and 1000 dimensions. In general, higher-dimensional targets require more computations, but the burden of dimensionality is significantly greater for CPU implementations. We find that both proposal and acceptance steps benefit from parallelization but that parallelizing proposals is more beneficial at these scales. In the second experiment, we maintain a fixed time budget for parallel MCMC implementations that use CPU (with 100 proposals) or GPU (with 100,000 proposals) for both proposal and acceptance steps. The target distributions are 2-dimensional mixtures of 10, 50, 100, 500 and 1000 standard Gaussians centered at **0**, **10**, **20**, etc. (Figure C13). We monitor effective sample sizes after 10,000 MCMC iterations. In this context, the target evaluations scale linearly in the number of mixture components, but the DISTRIBUTIONS library enables parallelization across proposals and mixture components. The upshot is a thousand-fold increase in the number of proposals when operating under a fixed time budget, and this increase in proposals leads to significantly better exploration of the heavily multimodal targets.

Next, we compare the same parallel MCMC algorithm to the massively parallel implementation of independent random walk Metropolis Markov chains to see if the latter, embarrassingly parallel approach is sufficient. Importantly, we implement both strategies within TENSORFLOW using the same GPU. An immediate difference between the two approaches is that the memory burden for independent chains

quickly accrues. For example, collecting  $S$  states each from  $C$  chains with  $D$  dimensional targets requires  $O(SCD)$  memory. In high-dimensional settings with thousands of chains, we therefore find it necessary to severely thin the chains, saving a relatively small number of states. Figure 2 shows results from just such a comparison, targeting centered, isotropic and high-dimensional Gaussians. We run 10,000 parallel chains but find it necessary to remove 99 states for every 100. An adverse dependence of Monte Carlo estimator accuracy on initial positions demonstrates that the act of combining such a large number of chains fails to make up for the weaknesses of individual chains despite their being tuned for optimal scaling (Gelman *et al.*, 1997). On the other hand, the parallel MCMC algorithm achieves accuracy when run for the same number of iterations. Figure 3 shows similar results using the 2-dimensional mixture of many Gaussians targets from the previous experiment (Figure 1). Thanks to the low-dimensional nature of the problem, we are able to save all 10,000 states of 10,000 parallel chains, but this approach fails nonetheless. A harbinger of this failure is that the parallel chains achieve a potential scale reduction factor of 2.30 (Gelman & Rubin, 1992). Again, the parallel MCMC algorithm performs well for the same number of MCMC iterations and, free from memory constraints, is able to use as many as 100,000 proposals within each iteration.

## 5.2 Statistical inversion with infinite-dimensional unknowns

We next present three case studies involving target measures arising from ill-posed inverse problems framed in the setting of Bayesian statistical inversion, Dashti & Stuart (2017); Kaipio & Somersalo (2006); Stuart (2010). Here we are particularly focused on PDE constrained problems with infinite-dimensional unknowns with a particular emphasis on fluid measurement problems in two different settings we advocated recently in Borggaard *et al.* (2020); Borggaard *et al.* As noted above, while such problems must be truncated to fit on a computer, study in the full infinite-dimensional setting has yielded MCMC algorithms with beneficial convergence results in the high-dimensional limit (Beskos *et al.*, 2008, 2011; Cotter *et al.*, 2013; Hairer *et al.*, 2014).

Before delving into the specific details of each of the three problems let us first recall some generalities of the statistical inversion framework. Suppose we are trying to recover an unknown parameter  $\mathbf{q}$  sitting in a parameter space  $X$ , typically a separable Hilbert space. Our parameter is observed through a forward model  $\mathcal{G} : X \rightarrow \mathbb{R}^k$  and is subjected to an additive observational noise  $\eta$ . Namely, we would like to estimate  $\mathbf{q} \in X$  from the observation model

$$\mathcal{Y} = \mathcal{G}(\mathbf{q}) + \eta. \quad (5.1)$$

By placing a prior probability  $\mu_0$  on  $X$  for our unknown  $\mathbf{q}$  and assuming that  $\eta$  is continuously distributed with density  $p_\eta$ , Bayes theorem (see (Dashti & Stuart, 2017, Borggaard *et al.*, 2020)) uniquely determines the posterior probability measure for  $\mathbf{q}$  given observed data  $\mathcal{Y}$  as

$$\mu(d\mathbf{q}) := \frac{1}{Z} p_\eta(\mathcal{Y} - \mathcal{G}(\mathbf{q})) \mu_0(d\mathbf{q}), \quad \text{where } Z = \int_X p_\eta(\mathcal{Y} - \mathcal{G}(\tilde{\mathbf{q}})) \mu_0(d\tilde{\mathbf{q}}). \quad (5.2)$$

We view this posterior  $\mu$  as an optimal solution of the ill-posed inverse problem (5.1).

As noted above, we are interested in estimating unknown parameters appearing in a partial differential equation. Typically we consider the situation where

$$\mathcal{G} = \mathcal{O} \circ \hat{S}, \text{ where } \hat{S} \text{ is a ‘PDE parameter to solution map’ and } \mathcal{O} \text{ is an ‘observation operator’}. \quad (5.3)$$

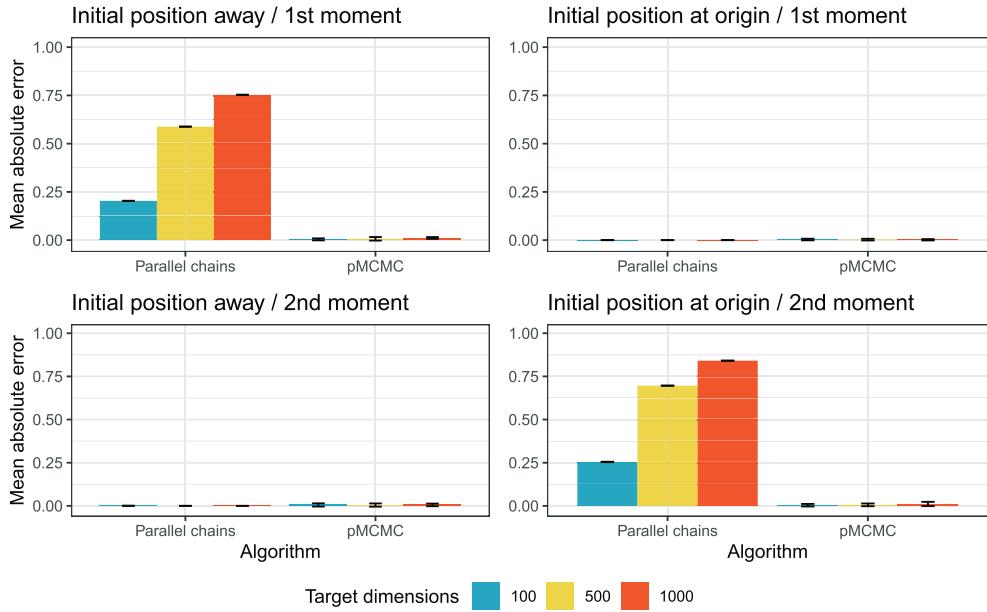


FIG. 2. Means and 95% confidence intervals for mean absolute estimation error from 10k parallel MH chains versus a single parallel MCMC chain produced by Algorithm 4 using 100k proposals. Both target isotropic Gaussians of 100, 500 and 1000 dimensions using 10k MCMC iterations. Estimator from 10k parallel chains demonstrates adverse dependence on starting point: starting at distribution mean  $\mathbf{0}$  harms second moment estimator, but starting at  $\mathbf{1}$  harms first moment estimator. The difference between chain and proposal counts arises from memory constraints on the multi-chain approach and does not appreciably change results.

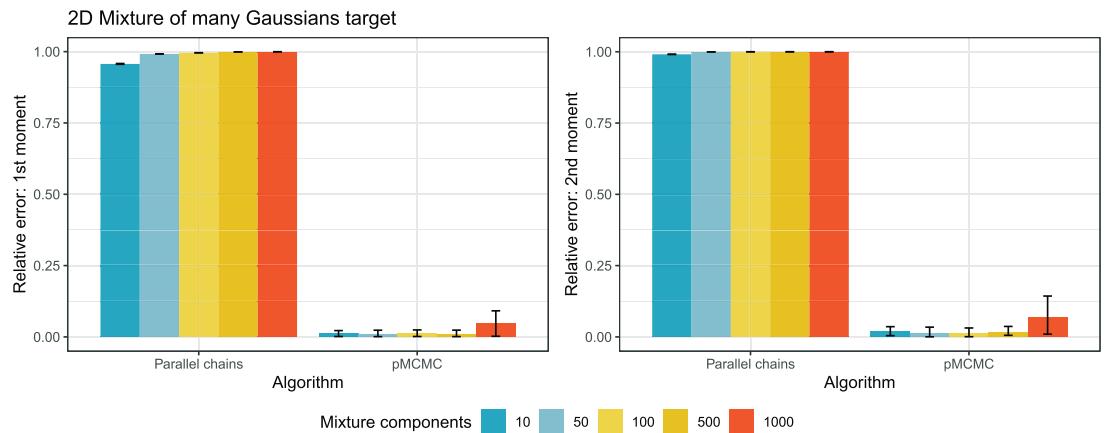


FIG. 3. Means and 95% empirical intervals for mean absolute estimation error from 10k parallel MH chains versus a single parallel MCMC chain produced by Algorithm 4 using 10k proposals. Both algorithms target mixtures of isotropic Gaussians using 10k MCMC iterations. To facilitate comparison between errors associated with the estimation of first and second moments, we divide mean absolute errors by true moment values.

This second operator  $\mathcal{O}$  projects the PDE solution down to a finite-dimensional space. Notice that in the natural case of a Gaussian prior  $\mu_0$  and e.g. Gaussian observational errors  $\eta_j$ , (5.2) takes the form (1.6), which falls in the scope of Algorithm 5.

**5.2.1 A finite-dimensional toy model.** Our first statistical inverse problem has a finite-dimensional unknown parameter space, but was designed as a toy model for much more complex and computationally involved PDE inverse problems we consider below in Section 5.2.2 and Section 5.2.3. Our goal here is to have a ‘table-top’ statistical experiment that mimics some of the structure (and indeed produces statistics reminiscent of) these more involved problems but which can be fully resolved on a laptop in a matter of hours.

### Problem specification.

We let  $X$ , our unknown parameter space, be the collection of skew symmetric matrices acting on  $\mathbb{R}^d$  so that  $m := \dim(X) = d(d - 1)/2$ . We then take  $\hat{S} = \hat{S}_{\mathbf{g}, \kappa} : \mathbb{R}^m \rightarrow \mathbb{R}^d$  to be the solution of

$$(A_{\mathbf{q}} + \kappa I)\mathbf{x} = \mathbf{g}, \quad (5.4)$$

namely  $\mathbf{x}(A_{\mathbf{q}}) = \hat{S}(\mathbf{q}) := (A_{\mathbf{q}} + \kappa I)^{-1}\mathbf{g}$ . Here  $\kappa > 0$  and  $\mathbf{g} = (g_1, \dots, g_d) \in \mathbb{R}^d$  are fixed and known parameters in our model. Regarding the observation procedure  $\mathcal{O} : \mathbb{R}^d \rightarrow \mathbb{R}^k$  we consider resolving some of the components of  $\hat{S}(\mathbf{q}) = (\hat{S}(\mathbf{q})_1, \dots, \hat{S}(\mathbf{q})_d)$ , namely

$$\mathcal{O}(\hat{S}(\mathbf{q})) = (\langle \hat{S}(\mathbf{q}), \mathbf{z}_1 \rangle, \dots, \langle \hat{S}(\mathbf{q}), \mathbf{z}_k \rangle)$$

for some directions  $\mathbf{z}_1, \dots, \mathbf{z}_k \in \mathbb{R}^d$ .  $\mathcal{O}$  could be taken to be nonlinear in general, for example in the case of the norm  $\mathcal{O}(\hat{S}(A_{\mathbf{q}})) = \|\hat{S}(A_{\mathbf{q}})\|$ . Note that (5.4) explicitly mimics some of the structure of a steady state version of the advection-diffusion problem, (5.8), we consider further on below.

Turning now to the specifics of the numerical study carried out here we let  $d = 4$  so that the dimension of the unknown parameter space is  $\dim(X) = 6$ . Regarding the observational noise we assume  $\eta \sim N(0, \sigma_\eta^2 I_k)$ . We observe the first two components of  $\hat{S}(A_{\mathbf{q}})$  (i.e  $k = 2$ ) so that our target measure from (5.2) takes the form

$$\mu(d\mathbf{q}) = \frac{e^{-\Phi(\mathbf{q})}}{Z} \mu_0(\mathbf{q}), \quad \text{where } \Phi(\mathbf{q}) := \frac{1}{2\sigma_\eta^2} \left( (y_1 - \hat{S}(\mathbf{q})_1)^2 + (y_2 - \hat{S}(\mathbf{q})_2)^2 \right) \quad (5.5)$$

where the prior  $\mu_0$  is a centered Gaussian with a diagonal covariance specified by two parameters  $\sigma \in \mathbb{R}$  and  $\gamma > 0$  as

$$\mu_0 = N(0, \text{diag}(\sigma^2, \sigma^2 2^{-\gamma}, \sigma^2 3^{-\gamma}, \sigma^2 4^{-\gamma}, \sigma^2 5^{-\gamma}, \sigma^2 6^{-\gamma})).$$

Regarding the choice of parameters defining (5.5) we selected

$$\sigma_\eta^2 = 2, \quad y_1 = 4.601, \quad y_2 = 18.021, \quad \sigma^2 = 5, \quad \gamma = 1.5, \quad (5.6)$$

and to specify  $\hat{S}$  from (5.4) we considered

$$\kappa = .1, \quad \mathbf{g} = (0, 0, 5, 2). \quad (5.7)$$

Note that, regarding the selection of  $y_1$  and  $y_2$ , we found a value of  $\bar{A}$  through a Monte Carlo search such that  $\mathbf{x}(-\bar{A})_1 \approx \mathbf{x}(\bar{A})_1 = y_1$  and  $\mathbf{x}(-\bar{A})_2 \approx \mathbf{x}(\bar{A})_2 = y_2$  up to an error of approximately .09.

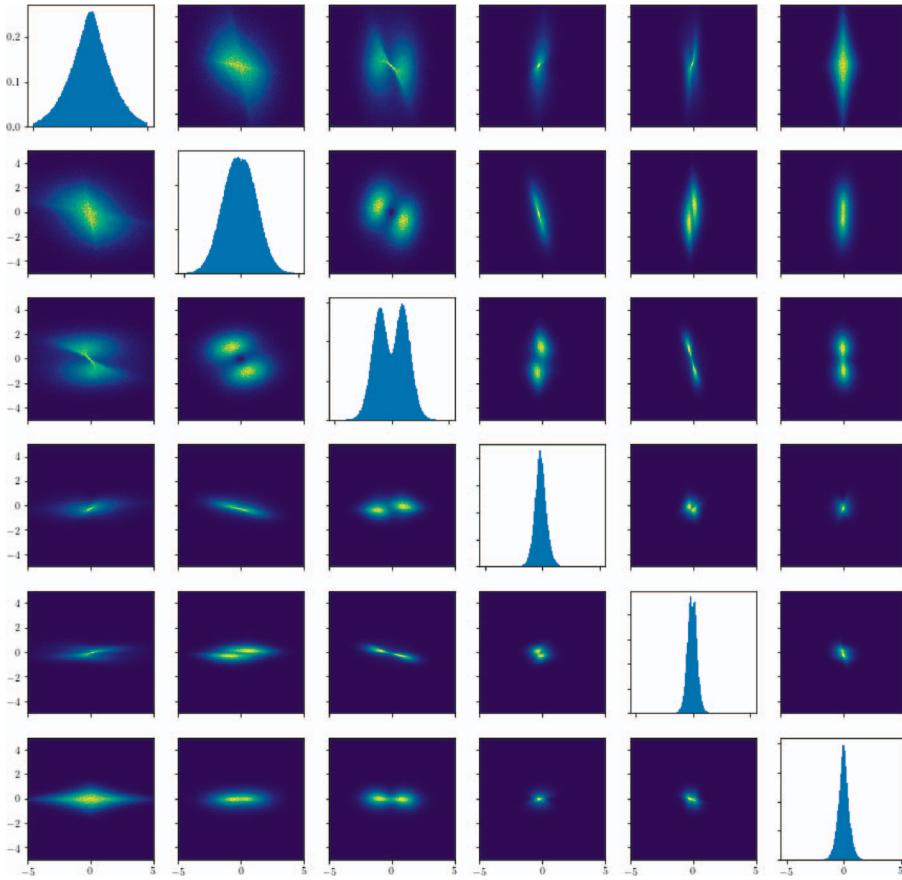


FIG. 4. Histogram of the one and two dimensional marginals of  $\mu$  defined by (5.5) relative to the parameters (5.6) and (5.7). While relatively low-dimensional, this target measure provides an interesting test case exhibiting both ill-conditioned geometry and multimodality. In fact, two different flavors of multimodality are exhibited: the dimensions 2-3 marginal exhibits energetic multimodality while the dimensions 1-3 marginal exhibits entropic multimodality.

However, from a preliminary, non-systematic, exploration of the parameter space, we experimentally found covariance structures that were broadly similar topologically with regard to the covariance structures appearing in the resulting target measures across a variety of different parameter values in comparison to those chosen above.

### Numerical results.

As a benchmark we fully resolved (5.5) using a traditional pCN sampler with  $10^8$  samples taking the algorithmic parameter  $\rho = .99$  that yielded an approximate acceptance ratio of .22. As a sanity check we also ran mpCN out to  $10^7$  samples under the algorithmic parameter  $\rho = .6, p = 100$ . These two benchmark runs produced very similar statistics. A histogram representing the one and two dimensional marginals of the fully resolved target are pictured in Figure 4.

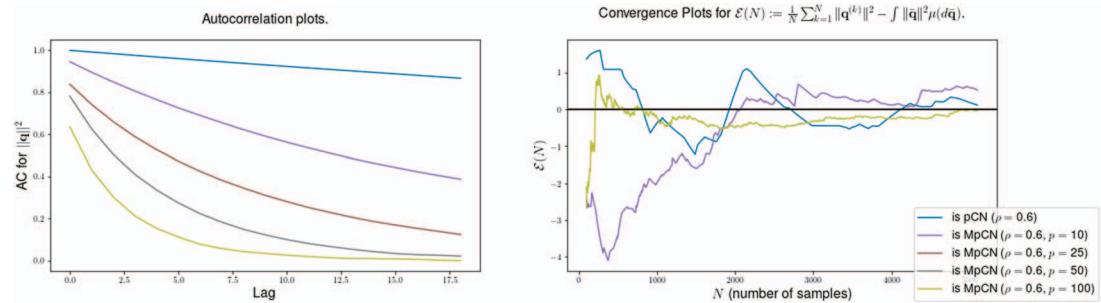


FIG. 5. Two different illustrations of improving per-sample performance for mpCN (Algorithm 5) for the observable  $\Psi(\mathbf{q}) = \|\mathbf{q}\|^2$ . The left panel provides autocorrelation plots while the right panel measures the convergence of ergodic averages.

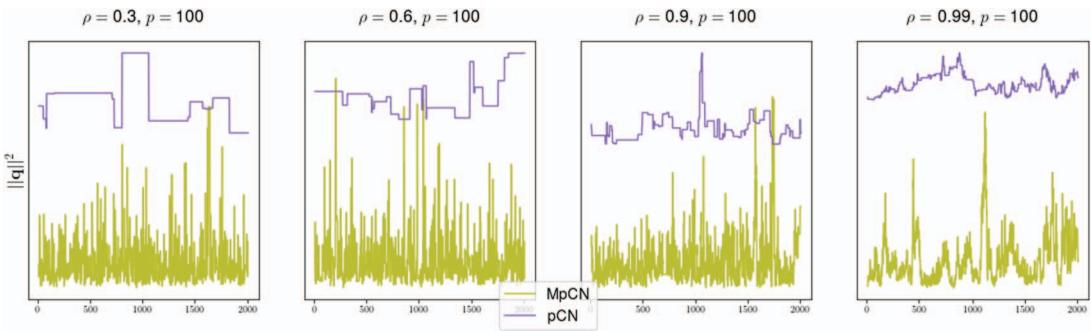


FIG. 6. Time series plots comparing mpCN (Algorithm 5) with  $p = 100$  and pCN for various values of  $\rho$  for the observable  $\Psi(\mathbf{q}) = \|\mathbf{q}\|^2$ .

To provide an assessment of the efficacy of mpCN (Algorithm 5) for this model (5.5) we ran mpCN out to  $10^5$  samples over a range of algorithmic parameter values  $\rho \in [0, 1]$  and  $p \geq 1$  namely:

$$\rho \in \{.3, .4, .5, .6, .7, .8, .9, .95, .99\} \quad p \in \{10, 25, 50, 100\}.$$

Unsurprisingly performance improved on a per sample basis for fixed  $\rho$  as  $p$  increases as illustrated in Figure 5.

An unambiguous heuristic for the selection of an optimal choice for  $\rho$  for a given value of  $p$  remains unclear. What is evident from our numerical experiments is that, firstly,  $\rho$  should be chosen to be rather more aggressive in comparison to its single proposal pCN counterpart. Secondly, it appears that effective chains are produced over a fairly robust range of values for  $\rho$ . These observations are illustrated in Figure 6 and Figure 7. For further illustrations of these points, see Figure C14, Figure C15 and Figure C16 in Appendix C below.

**5.2.2 Estimation of velocity fields from a passive solute.** Our second model problem, developed previously in Borggaard *et al.* (2020), involves the estimation of a time independent, divergence free vector field  $\mathbf{q}$  from the sparse measurement of a solute passively advected and diffusing in the fluid medium. By choosing observations that leverage symmetries in the problem, we generate a posterior measure with complex correlation and multimodal structures that make efficient sampling difficult.

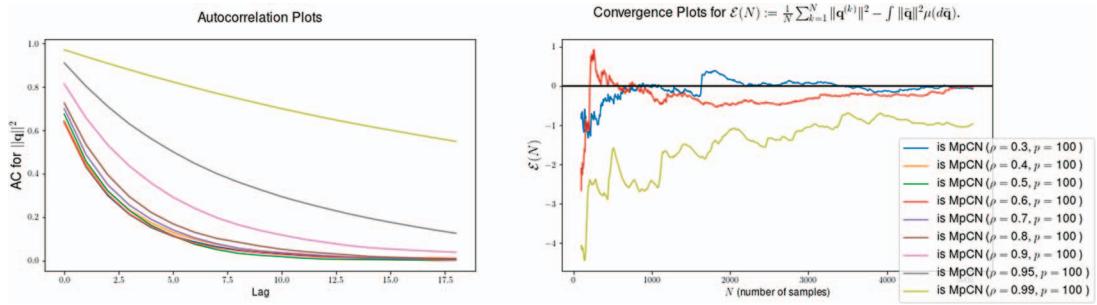


FIG. 7. Convergence plots comparing mpCN (Algorithm 5) with  $p = 100$  and pCN for various values of  $\rho$  over different observables.

### Problem specification.

The basic physical model for this situation is the advection-diffusion equation

$$\partial_t \theta + \mathbf{q} \cdot \nabla \theta = \kappa \Delta \theta, \quad (5.8)$$

where  $\theta$  represents the concentration of the solute. We consider (5.8) in the case of a periodic two dimensional domain  $\mathbb{T}^2 = [0, 1]^2$  with a known initial condition  $\theta(0, x) = \theta_0(x)$  (i.e. initial solute concentration) and diffusivity parameter  $\kappa > 0$ .

To develop concrete examples in this setting we need to specify the prior for the velocity field  $\mathbf{q}$ , the solute observation procedure  $\mathcal{O}$  in (5.3) and the observation noise structure  $\eta$  in (5.1). We consider the unknown parameter space as the Sobolev space of divergence free, mean free vector fields (see, e.g., Temam, 1995) defined as

$$H_{div}^s := \left\{ \mathbf{q} : \mathbb{T}^2 \rightarrow \mathbb{R}^2 : \mathbf{q}(x) := \sum_{\mathbf{k} \in \mathbb{Z}^2} q_{\mathbf{k}} \frac{\mathbf{k}^\perp}{\|\mathbf{k}\|_2} e^{2\pi i \mathbf{k} \cdot \mathbf{x}}, \bar{q}_{\mathbf{k}} = -q_{-\mathbf{k}}, \|\mathbf{q}\|_{H^s} < \infty \right\}, \quad (5.9)$$

where  $q_{\mathbf{k}}$  are complex numbers with  $\bar{q}_{\mathbf{k}}$  representing the complex conjugate and

$$\|\mathbf{q}\|_{H^s}^2 := \sum_{\mathbf{k} \in \mathbb{Z}^2} \|\mathbf{k}\|^{2s} |q_{\mathbf{k}}|^2. \quad (5.10)$$

The Sobolev spaces  $H^s$  of mean free scalar fields are defined analogously. When  $s > 1$  we have that  $\hat{S}(\mathbf{q}, \theta_0) = \theta(\mathbf{q}, \theta_0)$ , the solution of (5.8) corresponding to any  $\mathbf{q} \in H_{div}^s$  and  $\theta_0 \in H^s$ , is an element in  $C([0, T], H^s)$  i.e. the continuous functions from  $[0, T]$  taking values in  $H^s$ . Note that this solution map  $\hat{S}$  depends continuously on  $\mathbf{q}, \theta_0$  in the standard topologies; see Borggaard (2020) for a proof of these well-posedness claims.

Regarding the prior  $\mu_0$ , we consider a centered Gaussian with covariance operator defined as (see Section 4.2 above):

$$\mathcal{C}\mathbf{q} = \sum_{\mathbf{k} \in \mathbb{Z}^2} E_{\mathbf{k}} \left( \int_{\mathbb{T}^2} \mathbf{e}_{\mathbf{k}}(\mathbf{x}) \cdot \mathbf{q}(\mathbf{x}) d\mathbf{x} \right) \mathbf{e}_{\mathbf{k}} \quad \text{where } E_{\mathbf{k}} = \frac{E(\|\mathbf{k}\|)}{2\pi \|\mathbf{k}\|}. \quad (5.11)$$

Here  $\mathbf{e}_{\mathbf{k}}(\mathbf{x}) := \frac{\mathbf{k}^\perp}{\|\mathbf{k}\|} \exp(2\pi i \mathbf{k} \cdot \mathbf{x})$  and

$$E(k) = E_0 \sum_{j=0}^N \left( \frac{k}{2^{j/2}} \right)^4 \exp \left( -\frac{3k^2}{2^{j+1}} \right) 2^{-j\xi/2}. \quad (5.12)$$

The form of  $E$  in (5.12) determines the ‘energy spectrum’ we are placing on the prior for the unknown velocity field  $\mathbf{q}$  and was inspired by a turbulence model due to Kraichnan; see Kraichnan (1968, 1991). The model is specified by the three parameters,  $E_0$ ,  $\xi$ ,  $N$  where  $E_0$  is related to the overall energy of turbulent flow,  $N$  determines the range of scales  $k \in [0, 2^N]$  in an ‘inertial range’ where  $\xi$  is the exponent in a power law decay over this range. In our example we specify

$$E_0 = 1, \quad N = 20, \quad \xi = 3/2. \quad (5.13)$$

To make the sampling computationally tractable, we truncate the unknown (5.10) to  $0 < \|\mathbf{k}\| \leq 8$ , yielding a 196-dimensional sample space. See Figure C17 below in Appendix C for a visualization of a typical draw from this prior.

Regarding the observational data specifying our posterior measure we exploit symmetries in the parameter to solution maps specified by (5.8) to obtain a posterior with a complex, highly non-Gaussian structure; see Figure C18 below and Borggaard *et al.* (2020) for further details. We consider

$$\theta_0(x, y) = \frac{1}{2} - \frac{1}{4} \cos(2\pi x) - \frac{1}{4} \cos(2\pi y), \quad \mathbf{q}^*(x, y) = [8 \cos(2\pi y), 8 \cos(2\pi x)],$$

and notice that

$$\theta(t, \mathbf{x}_i, \mathbf{q}^*, \theta_0) = \theta(t, \mathbf{x}_i, -\mathbf{q}^*, \theta_0) \quad \text{for any } t > 0 \text{ and } i = 1, 2 \text{ where } \mathbf{x}_1 = [0, 0], \mathbf{x}_2 = [\frac{1}{2}, \frac{1}{2}].$$

With this in mind we take as our data

$$\mathcal{Y} := \{\theta(t_k, \mathbf{x}_i, \mathbf{q}^*, \theta_0)\}_{i=1,2, t_k=0.001k, k=1, \dots, 50}$$

Regarding the observation noise we consider the situation at which each data point is subject to an independent and identically distributed Gaussian error  $N(0, \sigma_\eta)$  with  $\sigma_\eta = 2^{-3}$ .

## Numerical results.

Figure 8 shows the results of a parameter study for mpCN, Algorithm 5, in  $\rho$  and in  $p$  the number of proposals. Two 250,000-sample chains were run for several  $\rho$  values ranging from very aggressive step sizes (low  $\rho$ ) to very conservative step sizes (high  $\rho$ ) for the original ‘vanilla’ pCN algorithm and for multiproposal pCN with 8, 16 and 64 proposals. The figure shows the average across these two chains for each case. The plot on the left shows the acceptance rate for the chains; we see that adding proposals of course increases the acceptance rate for most of the range of  $\rho$  values, although the effect diminishes as the number of proposals grows large.

It is perhaps worth noting that the acceptance rate for  $p = 8$  was lower than for vanilla pCN when  $\rho$  was set to be very low (overly-aggressive); likely this is because use of the Barkerized acceptance ratio, (4.22), rather than the traditional MH ratio increases the chance of rejection of a given proposal. The right-hand plot shows the number of samples required to achieve an effective sample – i.e., the number of samples divided by the effective sample size (ESS), where ESS is calculated according to (Gelman *et al.*, 2014, Section 11.5)—computed with the unnormalized posterior density (the log-prior plus the log-likelihood) as the estimand of interest and using the final 100,000 samples from each chain. The results show that using multiple proposals reduced the number of samples required to generate a statistically independent sample by nearly an order of magnitude, from roughly 200 samples in the best cases for pCN to approximately 50 samples for 16 proposals and 35 samples for 64 proposals, respectively. Also,

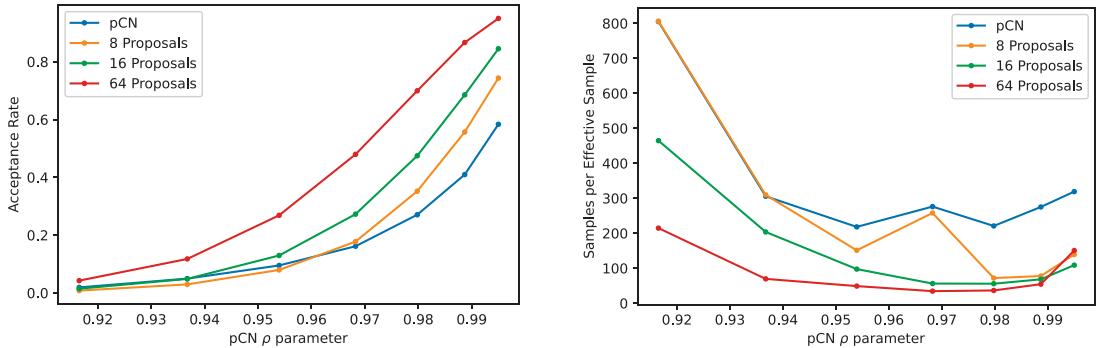


FIG. 8. Sweep of tuning parameters  $\rho$  and  $p$  for the advection-diffusion problem with 250,000 samples (average values across two chains) generated using mpCN (Algorithm 5) and pCN. Left: Acceptance rate. Right: Samples per effective sample of the (unnormalized) posterior density (lower is better).

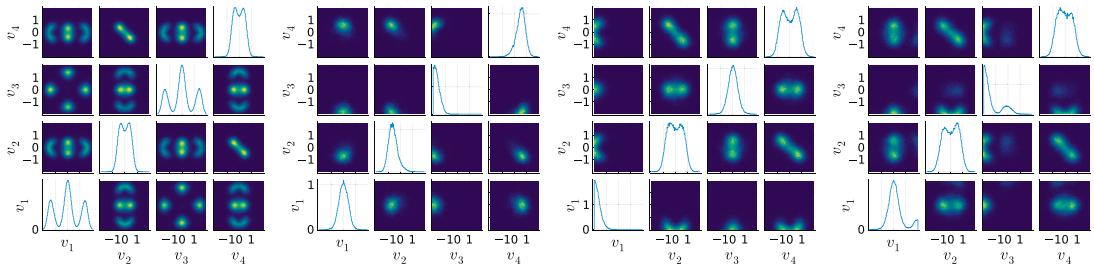


FIG. 9. Posterior two-dimensional histograms for the first few components of the advection-diffusion problem. From left: (1) ‘Truth’ from Borggaard *et al.* (2020); (2) pCN ( $\rho = 0.980$ ) after 250,000 samples; (3) and (4) two chains of multiproposal pCN (Algorithm 5) ( $\rho = 0.968$ , 64 proposals) after 250,000 samples.

when the step size was well-tuned to an acceptance rate of 30 – 50% even 8 or 16 proposals seemed to be enough to substantially improve performance according to this metric.

Based on the results shown in Figure 8, we selected ‘optimal’  $\rho$  values of 0.980 for vanilla pCN and 0.968 for multiproposal pCN with 64 proposals. Figure 9 compares the results of the two-dimensional histograms for the first few components of the posterior measure for these  $\rho$  values compared with the ‘true’ posterior (computed via many long chains) from Borggaard *et al.* (2020), which is shown in the plot on the left. The result a pCN chain is show in the second figure from the left and two 64-proposal multiproposal pCN chains are shown in the figures on the right. We see that through 250,000 samples the pCN chain has yet to explore the multimodal structure that characterizes the posterior; a second pCN chain was run for this  $\rho$  value and yielded the same result. The two multiproposal pCN chains, by contrast, appear to have fully explored the ‘gentle’ multimodality in the second and fourth components; one of the two chains has also made the larger jump between modes in the first and third components, although the balance between modes has not yet been achieved. These results show the promise of multiproposal methods for exploring complex posterior measures in high-dimensional spaces.

**5.2.3 Estimation of boundary shape characteristics in a rotating Stokes flow.** Our final statistical inversion problem is a shape estimation problem from Borggaard *et al.* (2023) wherein the unknown parameter  $\mathbf{q}$  specifies the shape of a domain upon which a system of PDEs is defined. The forward map

TABLE 1 *Parameter choices for the Stokes problem.*

| Parameter  | Value   | Parameter                               | Value |
|--|---|---|-------|
| Observations, $\mathcal{O}$ (5.3)                | Sectoral Scalar Variance (5.21)                                       | Sampling dimension, $K$                 | 320   |
| Data, $\mathcal{Y} = (y_1, y_2, y_3, y_4)$ (5.1) | $y_1 = y_3 = 0.4, y_2 = y_4 = 0$                                      | Angular Velocity, $\bar{\omega}$ (5.15) | 10    |
| Prior, $\mu_0$ (5.2)                             | $a_{2k-1}, a_{2k} \sim N(0, k^{-2s-1})$                               | Mean radius, $a_0$ (5.20)               | 1.0   |
| Noise, $\eta$ (5.1)                              | $N(0, \sigma_\eta^2 I), \sigma_\eta = 0.05$                           | Diffusion, $\kappa$ (5.14)              | 1.0   |
| Radius Constraints (5.18)                        | $r_{\min} = 0.5, r_{\max} = 1.5, R = 2$                               | Number of B-splines                     | 160   |
| Source, $f(\mathbf{x})$ (5.14)                   | $4 \exp[-(\mathbf{x} - \mathbf{x}_0)^2/100], \mathbf{x}_0 = (1.5, 1)$ | Range for $c, \epsilon$                 | 0.1   |

therefore requires solving the PDEs on an irregular domain, making the derivation of an appropriate adjoint method a difficult task. Furthermore, typical implementation will involve the use of a meshing algorithm, often third-party and black box to re-mesh the domain at each iteration step, ruling out use of automatic differentiation algorithms. This problem is therefore a natural fit for exploration of gradient-free methods such as Algorithm 5.

### Problem specification.

We specify our shape estimation problem in the statistical inversion framework, (5.1), as follows. Start with the forward map  $\mathcal{G}$ . Consider a steady (i.e. time-independent) Stokes flow coupled to an advection-diffusion equation describing the equilibrium concentration of solute sitting passively in this flow. The governing PDEs for this situation are

$$\nu \Delta \mathbf{u} = \nabla p, \quad \nabla \cdot \mathbf{u} = 0, \quad \mathbf{u} \cdot \nabla \theta = \kappa \Delta \theta + f \quad \text{in } \mathcal{D}_{\mathbf{q}}. \quad (5.14)$$

The unknown parameter  $\mathbf{q}$  determines the shape of an annular domain  $\mathcal{D}_{\mathbf{q}} \subset \mathbb{R}^2$  with a circular outer boundary on  $\Gamma_{\mathbf{q}}^o$  and unknown inner boundary  $\Gamma_{\mathbf{q}}^i$ . Here  $\mathbf{u} : \mathcal{D}_{\mathbf{q}} \rightarrow \mathbb{R}^2, p : \mathcal{D}_{\mathbf{q}} \rightarrow \mathbb{R}$  represent the fluid velocity and pressure and  $\theta : \mathcal{D}_{\mathbf{q}} \rightarrow \mathbb{R}$  is the concentration of the solute. The physical constants  $\nu, \kappa > 0$  are the viscosity and diffusivity while  $f$  is a source continuously injecting solute into the system. Regarding boundary conditions for our problem, we posit

$$\mathbf{u} = \bar{\omega} \mathbf{x}^\perp \text{ on } \Gamma_{\mathbf{q}}^o \quad \text{and} \quad \mathbf{u} = 0 \text{ on } \Gamma_{\mathbf{q}}^i, \quad \theta = 0 \text{ on } \Gamma_{\mathbf{q}}^o \quad \text{and} \quad \nabla \theta \cdot \hat{\mathbf{n}} = 0 \text{ on } \Gamma_{\mathbf{q}}^i, \quad (5.15)$$

that is the fluid is being rotated at its outer boundary with a rate  $\bar{\omega} \in \mathbb{R}$ . Here  $\hat{\mathbf{n}}$  is the outward normal to  $\Gamma_{\mathbf{q}}^i$  so that  $\nabla \theta \cdot \hat{\mathbf{n}} = 0$  represents an insulation condition if we interpret  $\theta$  as the temperature of the fluid. The precise values of the various physical parameters  $\nu, \kappa, \bar{\omega}$  as well as the form of  $f$ , which we used for our test example are given in Table 1 below.

We specify our fluid domain  $\mathcal{D}_{\mathbf{q}}$  as a function of  $\mathbf{q}$  as follows. We suppose that the parameter space  $X$  is given by the collection of  $2\pi$  periodic functions with Sobolev regularity  $s > 0$  namely

$$X = H^s := \left\{ \mathbf{q} : [0, 2\pi] \rightarrow \mathbb{R} : \mathbf{q}(x) := \sum_{k=1}^{\infty} (a_{2k-1} \cos(kx) + a_{2k} \sin(kx)), \|\mathbf{q}\|_{H^s} < \infty \right\} \quad (5.16)$$

where

$$\|\mathbf{q}\|_{H^s}^2 := \frac{1}{2} \sum_{k=1}^{\infty} k^{2s} (a_{2k-1}^2 + a_{2k}^2). \quad (5.17)$$

Here  $s$  will determine the ‘degree of smoothness’ of the inner boundary  $\Gamma_{\mathbf{q}}^i$ . To define the map associating each  $\mathbf{q} \in X$  its associated  $\mathcal{D}_{\mathbf{q}}$ , we let

$$\mathcal{D} := \{\mathbf{x} \in \mathbb{R}^2 : r_{\min} < |\mathbf{x}| \leq R\}. \quad (5.18)$$

Our domains  $\mathcal{D}_{\mathbf{q}}$  are defined so that they always contain the subdomain

$$\mathcal{D}^0 := \{\mathbf{x} \in \mathbb{R}^2 : r_{\max} \leq |\mathbf{x}| \leq R\}, \quad (5.19)$$

for some  $0 < r_{\min} < r_{\max} < R$ . We denote by  $(\phi, |\cdot|) : \mathcal{D} \rightarrow [0, 2\pi) \times [r_{\min}, R]$  the conversion to polar coordinates. To ensure that the inner radius does not decrease below  $r_{\min}$  or extend too far into the domain, we next fix a ‘clamping’ function  $c : \mathbb{R} \rightarrow (r_{\min}, r_{\max})$  to be a smooth strictly increasing function with

$$\lim_{t \rightarrow -\infty} c(t) = r_{\min}, \quad \lim_{t \rightarrow \infty} c(t) = r_{\max}.$$

To limit the effect of clamping on the shapes, we used quadratic interpolation to develop a smooth  $c$  that changes the radius only within a given range  $\epsilon$  of either boundary  $r_{\min}$  or  $r_{\max}$ ; the precise form of this interpolant is given in (Borggaard *et al.*, 2023, Section 3.2.2). With  $c$  in place, the radius of the inner boundary for a given angle  $\phi$  is then given by  $c(\mathbf{q}(\phi))$ . Then given any  $\mathbf{q} : \mathbb{R} \rightarrow \mathbb{R}$  that is sufficiently smooth and  $2\pi$  periodic, we consider the domain

$$\mathcal{D}_{\mathbf{q}} := \{\mathbf{x} \in \mathbb{R}^2 : c(a_0 + \mathbf{q}(\phi(\mathbf{x}))) \leq |\mathbf{x}| \leq R\}, \quad (5.20)$$

where  $a_0$  is a fixed mean radius and denote the boundary as

$$\Gamma_{\mathbf{q}}^i := \{\mathbf{x} \in \mathbb{R}^2 : c(a_0 + \mathbf{q}(\phi(\mathbf{x}))) = |\mathbf{x}|\}, \quad \Gamma^o := \{\mathbf{x} \in \mathbb{R}^2 : |\mathbf{x}| = R\}.$$

This procedure to construct annular domains  $\mathcal{D}_{\mathbf{q}}$  is visualized in Figure C19.

Having specified in (5.14), (5.15) and (5.20) the ‘parameter to solution map’  $\hat{S}$  portion of the forward map  $\mathcal{G}$  in (5.3), it remains to describe our observation procedure  $\mathcal{O}$ . In fact there are a number of different physically interesting possibilities  $\mathcal{O}$  that yield interesting statistics; see Borggaard *et al.* (2023) for further details. For our purposes here we restrict our attention to a procedure based on the observation of *scalar variance* by quadrant. Scalar variance is an important metric for mixing; low scalar variance indicates that  $\theta$  has been well mixed while high scalar variance indicates that  $\theta$  has been ‘trapped’ somewhere in the region.

To make this precise define  $\phi_j = j\pi/2$  for  $j = 0, \dots, 4$  and define the quadrants

$$\Omega_j(\mathcal{D}_{\mathbf{q}}) = \{\mathbf{x} \in \mathcal{D}_{\mathbf{q}} : \phi_{j-1} < \phi(\mathbf{x}) \leq \phi_j\}$$

for  $j = 1, \dots, 4$ . Then we define the four observations to be the average scalar variance for each of the four quadrants, i.e.,

$$\mathcal{O}_j(\mathbf{u}, \theta, \mathcal{D}_{\mathbf{q}}) := \frac{1}{|\mathcal{D}_{\mathbf{q}}|} \int_{\Omega_j(\mathcal{D}_{\mathbf{q}})} \left( \theta(\mathbf{x}; \mathbf{u}) - \frac{1}{|\mathcal{D}_{\mathbf{q}}|} \int_{\mathcal{D}_{\mathbf{q}}} \theta(\mathbf{y}; \mathbf{u}) d\mathbf{y} \right)^2 d\mathbf{x}, \quad (5.21)$$

where  $|\mathcal{D}_{\mathbf{q}}|$  is the area of the domain  $\mathcal{D}_{\mathbf{q}}$  so that the observations sum to the average scalar variance across the whole domain.

For the data,  $\mathcal{Y} = (y_1, y_2, y_3, y_4)$  in (5.1), we choose high scalar variance (indicating that the scalar should be trapped) in the first and third quadrants and low scalar variance (indicating that the scalar should be mixed) in the second and fourth quadrants. For the observational noise  $\eta$  in (5.1) we assume

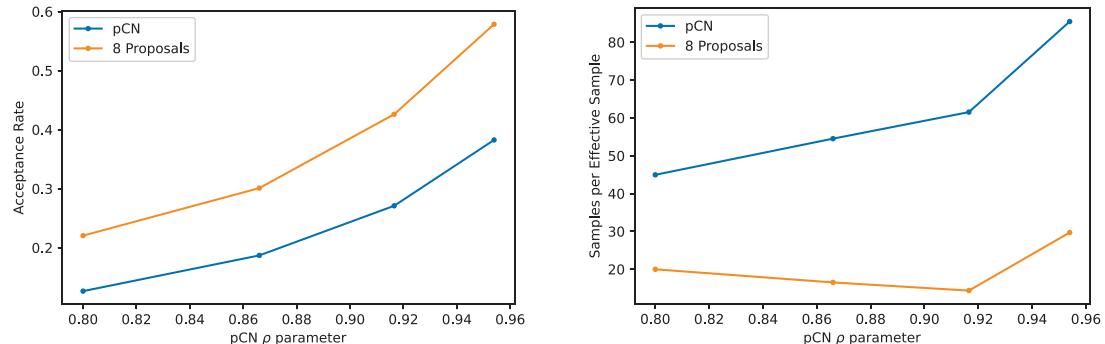


FIG. 10. Sweep of tuning parameters  $\rho$  and  $p$  for the Stokes problem with 10,000 samples generated using mpCN (Algorithm 5) and pCN. Left: Acceptance rate. Right: Samples per effective sample of the (unnormalized) posterior density (lower is better).

that each sectorial observation is perturbed by an independent mean zero Gaussian noise. See Table 1 for precise parameter values for  $\mathcal{Y}$  and the distribution of  $\eta$ .

To parameterize the inner boundary, we let the components  $\{a_k\}_{k=1}^\infty$  be our unknown parameters. For the prior measure on these parameters, we choose  $\mu_0 = N(0, C)$  where  $C$  is a diagonal operator such that components  $a_k$ ,  $k \geq 1$ , are mutually independent and distributed as  $a_{2k-1}, a_{2k} \sim N(0, k^{-2s-1})$ . Using (5.17), this ensures that draws from the prior are almost surely elements of  $H^{s'}$  for any  $s' < s$ :

$$\mathbb{E}\|\mathbf{q}\|_{H^s}^2 = \mathbb{E}\left[\frac{1}{2} \sum_{k=1}^{\infty} k^{2s}(a_{2k-1}^2 + a_{2k}^2)\right] = \frac{1}{2} \sum_{k=1}^{\infty} k^{2s} \mathbb{E}(a_{2k-1}^2 + a_{2k}^2) = \sum_{k=1}^{\infty} k^{-1-2(s-s')} < \infty. \quad (5.22)$$

To compute the forward map  $\mathcal{G}(\mathbf{q})$  numerically from a given (finite) array of components  $\{a_k\}_{k=1}^K$ , we (1) compute the Fourier expansion (5.16), (2) compute the optimal B-spline approximation to the Fourier expansion, (3) clamp via  $c$  to get a B-spline representation of the inner boundary  $\Gamma_q^i$ , (4) mesh the interior of the domain  $\mathcal{D}_q$  with Gmsh (Geuzaine & Remacle, 2009), a popular meshing software, (5) solve the Stokes and advection-diffusion PDEs (5.14), (5.15) via a finite element solver, and (6) compute the observations by computing the integrals in (5.21). Here step (2) is included because a B-spline representation is required by Gmsh. Additional details are available in Borggaard *et al.* (2023); code implementing the full solver and MCMC routines are publicly available at <https://github.com/jborggaard/BayesianShape>. Figure C20 shows plots of the solution to the coupled Stokes and advection-diffusion equations; the example was taken from a sample from one of the multiproposal pCN chains described below.

The remaining problem parameters are summarized in Table 1.

## Numerical results.

PDE solves for the Stokes problem (5.14)–(5.15) are, due to the irregular boundary conditions that change for each choice of parameter, quite a bit more expensive than for the advection-diffusion example, (5.8) covered in Section 5.2.2. For this reason, we limit the number of proposals to eight and the chains to 10,000 samples each. We begin by presenting the acceptance rate and samples per effective sample by MCMC method and pCN  $\rho$  parameter, which are shown in Figure 10. Each value shown is the average across two chains. To minimize the effects of burn in, the effective sample size in this case was computed on the final 5,000 samples of the 10,000-sample chains. As in the analogous figure for the advection-diffusion example (Figure 8), this shows that multiproposal pCN yields higher acceptance rates and requires fewer samples to achieve an effective sample for a given choice of  $\rho$ .

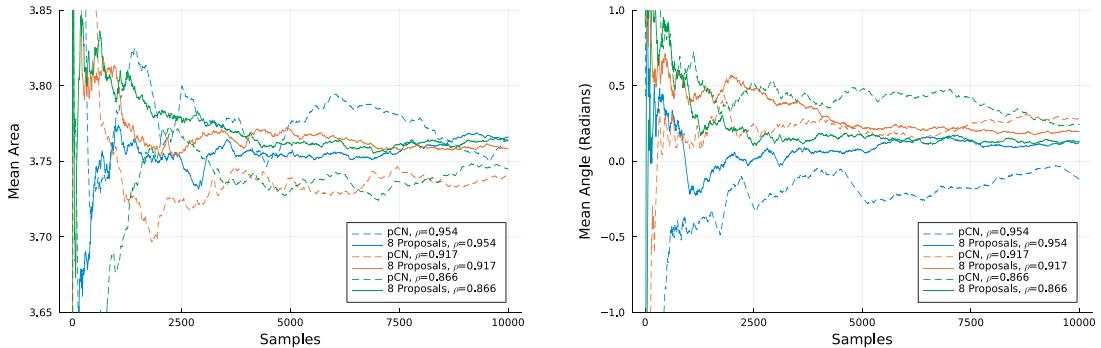


FIG. 11. Running averages of inner boundary enclosed area (left) and primary angle (right) by  $\rho$  (color) and MCMC method (pCN dashed and multiproposal pCN (Algorithm 5) solid).

We now consider convergence of two quantities of interest that provide key information about the boundaries that make up the posterior. The first is the area enclosed by the inner boundary, which describes the size of the domain  $\mathcal{D}_q$ . The second is the principal angle associated with the inner boundary, which is given by the orientation of its first Fourier component, i.e.,  $\arctan(Im(z)/Real(z))$  where  $z = \int_0^{2\pi} e^{i\phi} c(b(x)) d\phi$ . This scalar quantity describes the polar orientation where the radius of the inner boundary tends to be largest. Figure 11 shows running averages for these two scalar quantities; the results are given for both pCN and multiproposal pCN with eight proposals, with two chains run for each of three values of pCN's  $\rho$  parameter. The results in each case show that the multiproposal chains have, for the most part, converged to a steady state about halfway through the 10,000 sample chains; the pCN chains, meanwhile, still deviate quite a bit from each other and what appears to be the true value.

Finally, we can consider the computed distributions on the shapes for each MCMC method. Figure 12 shows radius quantiles by angle for pCN and multiproposal ( $p = 8$ ) pCN chains; to generate the figure, we compute the radius of the inner boundary at each angle for each sample and then tabulate the quantiles of the set of radii for that angle. The bowtie shape of the boundaries are a result of the desire to ‘trap’ the scalar  $\theta$  in the first and third quadrants as dictated by the data. The two quantile plots are largely similar, although the pCN (left) plot shows significant divergence in the tails (e.g., the 90th percentile in the third quadrant) as a result of the slower convergence.

## 6. Outlook

Moving forward, the present work suggests a number of avenues for future research. First, certain foundational questions remain to be addressed stemming from our analysis in Section 2 and Section 3. It is clear that these two formalisms, namely Algorithm 1 and Algorithm 3, do not entirely coincide while nevertheless demonstrate a significant overlapping scope. This overlapping scope includes conditionally independent proposal structures encompassing important examples, i.e., Algorithm 4 and Algorithm 5. See Remark 3.8. In any case, a systematic account of the relationship between the formulations leading to Algorithm 1, Algorithm 3 remain elusive at the time of writing. Note also that Theorem 3.3 only addresses invariance (i.e., lack of bias) with respect to the target whereas Theorem 2.2 and the other results in Section 2 provide comprehensive conditions for reversibility. Thus, this question of reversibility for the framework in Section 3 and its significance at the level of the extended (indexed) phase space remain to be addressed.

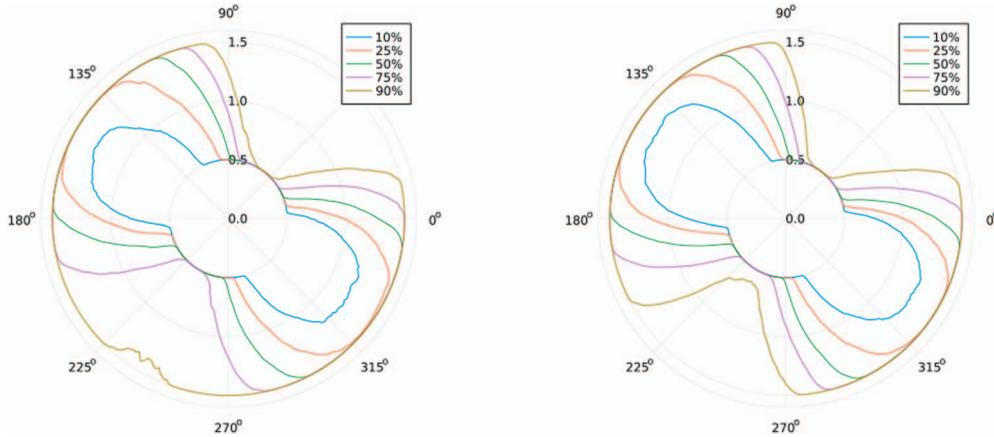


FIG. 12. Radius quantiles for 10,000 samples with  $\rho = 0.917$ . Left: pCN. Right: multiproposal pCN (Algorithm 5) with 8 proposals.

Leaving aside these general considerations, we would like to emphasize the scope for novel concrete pMCMC algorithm design and development from our formalisms (and other potential generalizations of these formalisms), which we initiated above in Section 4 is far from exhausted. For example, mpCN (Algorithm 5, Algorithm 6) suggests that a number of other such ‘infinite-dimensional’ sampling methods, namely  $\infty$ HMC and  $\infty$ MALA, (Beskos *et al.*, 2011; Cotter *et al.*, 2013), can be fruitfully extended to the multiproposal setting. We would also like to emphasize that Section 4.3 only scratches the surface as far as the range of possible formulations for multiproposal variants of the Hamiltonian (Hybrid) Monte Carlo paradigm.

Note that our high-performance computing studies suggest the potential of using GPUs to power pMCMC implementations with massive numbers of proposals, and Holbrook (2023b) shows that quantum computing also provides speedups for pMCMC. From an engineering perspective, our extended phase space formulation of pMCMC may motivate further high-performance computing methods that support gradient based pMCMC such as that of Section 4.3. From a theoretical perspective, the same extended phase space formulation may extend to other pMCMC algorithms motivated by novel computational paradigms. For example, it has been demonstrated that parallel, asynchronous updates of different parameters within Gibbs sampling is theoretically valid so long as one can bound latency between updates (Terenin *et al.*, 2020). If each multiproposal constitutes a (many-pronged) fork in the road, we are interested in pMCMC algorithms that employ a branching structure, accepting multiple proposals at once and conditionally assigning multiple individual iterations to new nodes. Such a viral program could prove useful on large computing resources with nodes constantly going on- and off-line.

Another outstanding question is that of optimal scaling. Here, one has additional degrees of freedom to consider beyond those of the traditional random walk Monte Carlo analysis (Gelman *et al.*, 1997). Optimal scaling should depend on the number of proposals, and the optimal number of proposals should depend on the specific form of the acceptance probabilities (e.g., (2.8) or (2.10)). A useful analysis should account for multimodal structure of the target distribution, i.e., the number of modes and their mutual distances relative to individual scales.

Related to this question of optimal scaling is the establishment of mixing rates and geometric ergodicity for various pMCMC algorithms applied to specific target distributions. Ultimately one would like to rigorously quantify the improvements in mixing rates on a per sample basis in comparison to traditional single proposal methods. Here, one might suspect that mixing improves monotonically with

the number of proposals but any rigorous justification of this conjecture appears to be far from reach at present. More tractably than simply establishing mixing rates for, e.g., Algorithm 5 one may appeal to the strategies from Glatt-Holtz & Mondaini (2021); Hairer *et al.* (2014) as a fruitful starting point. These works use a weak Harris theorem and a generalized coupling argument to establish geometric ergodicity for, respectively, pCN and HMC in infinite-dimensional Hilbert spaces.

Finally, rigorous investigation is outside the scope of this manuscript, but intuition suggests that the truncation of acceptance probabilities in MH-type mechanism (2.18) might result in poor performance in the large proposal limit. An in depth study of the relative benefits of the MH-type (2.18) and Barker-type (2.17) acceptance probabilities remains for future work.

## Acknowledgements

Our efforts are supported under the grants NSF-DMS-1816551 and NSF-DMS-2108790 (N.E.G.H.); NSF-DMS-2108791 (J.A.K.); NIH-K25-AI153816, NSF-DMS-2152774 and NSF-DMS-2236854 (A.J.H.); and NSF-DMS-2009859 and NSF-DMS-2239325 (CFM). The authors thank J. Borggaard for help in formulating and preparing the code for the Stokes example presented in Section 5.2.3. We would also like to express our appreciation to H. Riggs for clarifying some issues related to macrodata refinement. The authors acknowledge Advanced Research Computing at Virginia Tech (<https://arc.vt.edu>) for providing computational resources that have contributed to the results reported within this paper.

## REFERENCES

- ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G., ISARD, M., et al. (2016) {TensorFlow} A system for {large-scale} machine learning. *The 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283.
- ALIPRANTIS, C. D. & BORDER, K. C. (2013) *Infinite Dimensional Analysis: A Hitchhiker’s Guide*. Berlin and Heidelberg GmbH & Company KG: Springer.
- C. ANDRIEU, A. LEE, and S. LIVINGSTONE. A general perspective on the Metropolis-Hastings kerne I. *arXiv preprint, arXiv:2012.14881*, 2020.
- BARKER, A. A. (1965) Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Aust. J. Phys.*, **18**, 119–133.
- BESKOS, A., PINSKI, F. J., SANZ-SERNA, J. M. & STUART, A. M. (2011) Hybrid Monte Carlo on Hilbert spaces. *Stoch. Process. Appl.*, **121**, 2201–2230.
- BESKOS, A., ROBERTS, G., STUART, A. & VOSS, J. (2008) MCMC methods for diffusion bridges. *Stoch. Dynam.*, **08**, 319–350.
- BOGACHEV, V. I. (1998) *Gaussian Measures. Number 62*, American Mathematical Society.
- BOGACHEV, V. I. (2007) *Measure Theory, Volume 1*. Springer Science & Business Media.
- BORGGAARD, J., GLATT-HOLTZ, N. & KROMETIS, J. (2020) A Bayesian approach to estimating background flows from a passive scalar. *SIAM/ASA J. Uncertain. Quantification*, **8**, 1036–1060.
- BORGGAARD, J., GLATT-HOLTZ, N. E. & KROMETIS, J. (2023) A statistical framework for domain shape estimation in Stokes flows. *Inverse Prob.*, **39**, 085009 Publisher: IOP Publishing.
- BORGGAARD, N. & J. AND GLATT-HOLTZ AND J. KROMETIS. (2020) On Bayesian consistency for flows observed through a passive scalar. *Ann. Appl. Probab.*, **30**, 1762–1783.
- BOU-RABEE, N. & SANZ-SERNA, J. M. (2018) Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.*, **27**, 113–206.
- BROCKWELL, A. E. (2006) Parallel Markov chain Monte Carlo simulation by pre-fetching. *J. Comput. Graph. Stat.*, **15**, 246–261.
- CALDERHEAD, B. (2014) A general construction for parallelizing Metropolis-Hastings algorithms. *Proc. Natl. Acad. Sci.*, **111**, 17408–17413.

- COTTER, S. L., ROBERTS, G. O., STUART, A. M. & WHITE, D. (2013) MCMC methods for functions: modifying old algorithms to make them faster. *Stat. Sci.*, **28**, 424–446.
- DA PRATO, G. & ZABCZYK, J. (2014) *Stochastic Equations in Infinite Dimensions*. Cambridge University Press.
- DASHTI, M. & STUART, A. M. (2017) The Bayesian approach to inverse problems. *Handbook of Uncertainty Quantification*. Springer, pp. 311–428.
- DELMAS, J.-F. & JOURDAIN, B. (2009) Does waste recycling really improve the multi-proposal Metropolis–Hastings algorithm? An analysis based on control variates. *J. Appl. Probab.*, **46**, 938–959.
- J.V. DILLON, I. LANGMORE, D. TRAN, E. BREVDO, S. VASUDEVAN, D. MOORE, B. PATTON, A. ALEMI, M. HOFFMAN, and R.A. SAUROUS. Tensorflow distributions. *arXiv preprint, arXiv:1711.10604*, 2017.
- DUANE, S., KENNEDY, A. D., PENDLETON, B. J. & ROWETH, D. (1987) Hybrid monte carlo. *Phys. Lett. B*, **195**, 216–222.
- FOLLAND, G. B. (1999) *Real Analysis: Modern Techniques and Their Applications, Volume 40*. John Wiley & Sons.
- FRENKEL, D. (2004) Speed-up of Monte Carlo simulations by sampling of rejected states. *Proc. Natl. Acad. Sci.*, **101**, 17571–17575.
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. & RUBIN, D. B. (2014) *Bayesian Data Analysis*, third edn. CRC Press.
- GELMAN, A., GILKS, W. R. & ROBERTS, G. O. (1997) Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, **7**, 110–120.
- GELMAN, A. & RUBIN, D. B. (1992) Inference from iterative simulation using multiple sequences. *Stat. Sci.*, **7**, 457–472.
- GEUZAIN, C. & REMACLE, J.-F. (2009) GMSH: A three-dimensional finite element mesh generator with built-in pre-and post-processing facilities. *Int. J. Numer. Methods. Eng.*, **79**, 1309–1331.
- GIROLAMI, M. & CALDERHEAD, B. (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B*, **73**, 123–214.
- GLATT-HOLTZ, N., KROMETIS, J. & MONDAINI, C. (2023) On the accept–reject mechanism for Metropolis–Hastings algorithms. *Ann. Appl. Probab.*, **33**, 5279–5333.
- GLATT-HOLTZ, N. E. & MONDAINI, C. F. (2021) Mixing rates for Hamiltonian Monte Carlo algorithms in finite and infinite dimensions. *Stoch. Partial Differential Equations Anal. Comput.*, **10**, 1318–1391.
- HAARIO, H., SAKSMAN, E. & TAMMINEN, J. (2001) An adaptive Metropolis algorithm. *Bernoulli*, **7**, 223–242.
- HAIRER, E., LUBICH, C. & WANNER, G. (2006) *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations, Volume 31*. Springer Science & Business Media.
- HAIRER, M., STUART, A. M. & VOLLMER, S. J. (2014) Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, **24**, 2455–2490.
- HALMOS, P. R. (2013) *Measure Theory, Volume 18*. Springer.
- W.K. HASTINGS. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. 1970.
- HOLBROOK, A. J., JI, X. & SUCHARD, M. A. (2022a) Bayesian mitigation of spatial coarsening for a Hawkes model applied to gunfire, wildfire and viral contagion. *Ann. Appl. Stat.*, **16**, 573–595.
- HOLBROOK, A. J., JI, X. & SUCHARD, M. A. (2022b) From viral evolution to spatial contagion: a biologically modulated Hawkes model. *Bioinformatics*, **38**, 1846–1856.
- HOLBROOK, A. J., LEMEY, P., BAELE, G., DELLICOUR, S., BROCKMANN, D., RAMBAUT, A. & SUCHARD, M. A. (2021a) Massive parallelization boosts big Bayesian multidimensional scaling. *J. Comput. Graph. Stat.*, **30**, 11–24.
- HOLBROOK, A. J., LOEFFLER, C. E., FLAXMAN, S. R. & SUCHARD, M. A. (2021b) Scalable Bayesian inference for self-excitatory stochastic processes applied to big American gunfire data. *Stati. Comput.*, **31**, 1–15.
- HOLBROOK, A. J. (2023a) Generating MCMC proposals by randomly rotating the regular simplex. *J. Multivariate Anal.*, **194**, 105106.
- HOLBROOK, A. J. (2023b) A quantum parallel Markov chain Monte Carlo. *J. Comput. Graph. Stat.*, **32**, 1402–1415.
- KAPIO, J. & SOMERSALO, E. (2006) *Statistical and Computational Inverse Problems, Volume 160*. Springer Science & Business Media.
- KRAICHNAN, R. H. (1968) Small-scale structure of a scalar field convected by turbulence. *Phys. Fluids*, **11**, 945–953.
- KRAICHNAN, R. H. (1991) Stochastic modeling of isotropic turbulence. *New Perspect. Turbulence*, 1–54.

- LEIMKUHLER, B. & REICH, S. (2004) *Simulating Hamiltonian Dynamics. Number 14.* Cambridge University Press.
- LIU, J. S., LIANG, F. & WONG, W. H. (2000) The multiple-try method and local optimization in Metropolis sampling. *J. Amer. Stat. Assoc.*, **95**, 121–134.
- LUO, X. & TJELMELAND, H. (2019) A multiple-try Metropolis–Hastings algorithm with tailored proposals. *Comput. Stat.*, **34**, 1109–1133.
- O. MANGOUBI, N.S. PILLAI, and A. SMITH. Does Hamiltonian Monte Carlo mix faster than a random walk on multimodal densities? *arXiv preprint, arXiv:1808.03230*, 2018.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.
- NEAL, R. (2011) MCMC using Hamiltonian dynamics. *Handbook of Markov chain Monte Carlo*, **2**, 2.
- NEAL, R. M. (1993) Probabilistic inference using Markov chain Monte Carlo methods. *Department of Computer Science*. ON, Canada: University of Toronto Toronto.
- NEAL, R. M. (1996) Sampling from multimodal distributions using tempered transitions. *Stat. Comput.*, **6**, 353–366.
- R.M. NEAL. Markov chain sampling for non-linear state space models using embedded hidden Markov models. *arXiv preprint, math/0305039*, 2003.
- NEKLYUDOV, K., WELLING, M., EGOROV, E. & VETROV, D. (2020) Involutive MCMC: a unifying framework. *International Conference on Machine Learning*. PMLR, pp. 7273–7282.
- ROBERT, C. P. & CASELLA, G. (1999) *Monte Carlo Statistical Methods, Volume 2*. Springer.
- ROBERTS, G. O. & STRAMER, O. (2002) Langevin diffusions and Metropolis-Hastings algorithms. *M Comput. Appl. Probab.*, **4**, 337–357.
- RUDOLF, D. & SPRUNGK, B. (2020) On a metropolis–hastings importance sampling estimator. *Electr. J. Stat.*, **14**, 857–889.
- SALMON, J. K., MORAES, M. A., DROR, R. O. & SHAW, D. E. (2011) Parallel random numbers: as easy as 1, 2, 3. *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–12.
- SCHUSTER, I. & KLEBANOV, I. (2020) Markov chain importance sampling—a highly efficient estimator for MCMC. *J. Comput. Graph. Stat.*, **30**, 260–268.
- SCHWEDES, T. & CALDERHEAD, B. (2021) Rao-Blackwellised parallel MCMC. *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 3448–3456.
- SCOTT, S. L. (2002) Bayesian methods for Hidden Markov models: Recursive computing in the 21st century. *J. Amer. Stat. Assoc.*, **97**, 337–351.
- STEWART, G. W. (1980) The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM J. Numer. Anal.*, **17**, 403–409.
- STUART, A. M. (2010) Inverse problems: a Bayesian perspective. *Acta Numer.*, **19**, 451–559.
- TEMAM, R. (1995) *Navier–Stokes Equations and Nonlinear Functional Analysis*. SIAM.
- TERENIN, A., SIMPSON, D. & DRAPER, D. (2020) Asynchronous Gibbs sampling. *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 144–154.
- TIERNEY, L. (1994) Markov chains for exploring posterior distributions. *Ann. Stat.*, 1701–1728.
- TIERNEY, L. (1998) A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, **8**, 1–9.
- TJELMELAND, H. (2004) Using all Metropolis–Hastings proposals to estimate mean values. Technical report. Citeseer.
- YANG, S., CHEN, Y., BERNTON, E. & LIU, J. S. (2018) On parallelizable Markov Chain Monte Carlo algorithms with waste-recycling. *Stat. Comput.*, **28**, 1073–1081.

## A. Appendix: Some measure-theoretic tools

This first appendix gathers together for the convenience of the reader some measure theoretic elements used extensively throughout the manuscript. We refer to e.g. [Aliprantis & Border \(2013\)](#); [Bogachev \(2007\)](#); [Folland \(1999\)](#) for further general background.

Let  $(\mathcal{X}, \Sigma_{\mathcal{X}})$  and  $(\mathcal{Y}, \Sigma_{\mathcal{Y}})$  be measurable spaces. Given a measurable function  $\phi : \mathcal{X} \rightarrow \mathcal{Y}$  and a measure  $v$  on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$ , the *pushforward* of  $v$  by  $\phi$ , denoted  $\phi^*v$ , is the measure on  $\mathcal{Y}$  defined as

$$\phi^*v(E) := v(\phi^{-1}(E)) \quad \text{for all } E \in \Sigma_{\mathcal{Y}}. \quad (\text{A.1})$$

In a statistical context, we notice that if  $\mathbf{w}$  is a random variable with probability distribution given by a measure  $v$  then  $\phi(\mathbf{w})$  is another random variable that is distributed as  $\phi^*v$ .

For any measurable functions  $\phi_1, \phi_2$  defined on appropriate spaces so that the composition  $\phi_1 \circ \phi_2$  makes sense, it follows immediately from (A.1) that

$$(\phi_1 \circ \phi_2)^*v = \phi_1^*(\phi_2^*v).$$

We also recall the following change of variables formula regarding pushforward measures. Namely, given a  $(\phi^*v)$ -integrable function  $\psi : \mathcal{Y} \rightarrow \mathbb{R}$ , i.e.  $\psi \in L^1(\phi^*v)$ , it follows that the composition  $\psi \circ \phi : \mathcal{X} \rightarrow \mathbb{R}$  belongs to  $L^1(v)$  and

$$\int_{\mathcal{Y}} \psi(\mathbf{w}) \phi^*v(d\mathbf{w}) = \int_{\mathcal{X}} \psi(\phi(\mathbf{w})) v(d\mathbf{w}). \quad (\text{A.2})$$

In the particular case when  $\mathcal{X} = \mathbb{R}^N$  and  $v$  is any Borel measure on  $\mathbb{R}^N$  that is absolutely continuous with respect to Lebesgue measure, namely

$$v(d\mathbf{w}) = \pi(\mathbf{w}) d\mathbf{w}$$

for some density function  $\pi : \mathbb{R}^N \rightarrow \mathbb{R}$ , then for any diffeomorphism  $\phi : \mathbb{R}^N \rightarrow \mathbb{R}$ , we have

$$\phi^*v(d\mathbf{w}) = \pi(\phi^{-1}(\mathbf{w})) |\det \nabla \phi^{-1}(\mathbf{w})| d\mathbf{w}. \quad (\text{A.3})$$

Further, recall that a measure  $v$  on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$  is *absolutely continuous* with respect to another measure  $\rho$  on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$ , denoted  $v \ll \rho$ , if  $v(E) = 0$  whenever  $\rho(E) = 0$ , for  $E \in \Sigma_{\mathcal{X}}$ . If  $v$  and  $\rho$  are two sigma-finite measures on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$  such that  $v \ll \rho$  then there exists a  $\rho$ -almost unique function  $dv/d\rho \in L^1(\rho)$  such that

$$v(E) = \int_E \frac{dv}{d\rho}(\mathbf{w}) \rho(d\mathbf{w}), \quad E \in \Sigma_{\mathcal{X}},$$

called the *Radon-Nikodym derivative* of  $v$  with respect to  $\rho$ . Moreover, given sigma-finite measures  $v$ ,  $\rho$  and  $\gamma$  on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$  such that  $v \ll \rho$  and  $\rho \ll \gamma$ , it follows that  $v \ll \gamma$  and

$$\frac{dv}{d\gamma}(\mathbf{w}) = \frac{dv}{d\rho}(\mathbf{w}) \frac{d\rho}{d\gamma}(\mathbf{w}) \quad \text{for } \gamma \text{-a.e. } \mathbf{w} \in \mathcal{X}.$$

In particular, if  $v_1, v_2$  and  $\rho$  are sigma-finite measures on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$  with  $v_1 \ll \rho$  and  $v_2 \ll \rho$ , so that

$$v_1(d\mathbf{w}) = \phi_1(\mathbf{w}) \rho(d\mathbf{w}), \quad v_2(d\mathbf{w}) = \phi_2(\mathbf{w}) \rho(d\mathbf{w}),$$

where  $\phi_1 = dv_1/d\rho$  and  $\phi_2 = dv_2/d\rho$ , and if  $\phi_2 > 0$   $\rho$ -a.e., then  $v_1 \ll v_2$  and

$$\frac{dv_1}{dv_2}(\mathbf{w}) = \frac{\phi_1(\mathbf{w})}{\phi_2(\mathbf{w})} \quad \text{for } \rho \text{-a.e. } \mathbf{w} \in \mathcal{X}. \quad (\text{A.4})$$

Finally, under the above definitions the following identities can be easily verified, see (Glatt-Holtz *et al.*, 2023, Section 2.1). Firstly, given a measurable and invertible mapping  $\phi : \mathcal{X} \rightarrow \mathcal{X}$  with

measurable inverse  $\phi^{-1} : \mathcal{X} \rightarrow \mathcal{X}$ , and sigma-finite measures  $\nu$  and  $\rho$  on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$  with  $\nu \ll \rho$ , it follows that  $\phi^*\nu \ll \phi^*\rho$  and

$$\frac{d\phi^*\nu}{d\phi^*\rho}(\mathbf{w}) = \frac{d\nu}{d\rho}(\phi^{-1}(\mathbf{w})) \quad \text{for } (\phi^*\rho) \text{-a.e. } \mathbf{w} \in \mathcal{X}. \quad (\text{A.5})$$

Secondly, if  $\nu$  is a sigma-finite measure on  $(\mathcal{X}, \Sigma_{\mathcal{X}})$ , and  $\phi_i : \mathcal{X} \rightarrow \mathcal{X}$ ,  $i = 1, \dots, n$ , a sequence of measurable and invertible functions with measurable inverses  $\phi_i^{-1} : \mathcal{X} \rightarrow \mathcal{X}$  such that  $\phi_i^*\nu \ll \nu$  for all  $i = 1, \dots, n$ , then  $(\phi_n \circ \dots \circ \phi_1)^*\mu \ll \mu$  and

$$\frac{d(\phi_n \circ \dots \circ \phi_1)^*\mu}{d\mu}(\mathbf{w}) = \frac{d\phi_n^*\mu}{d\mu}(\mathbf{w}) \prod_{i=1}^{n-1} \frac{d\phi_i^*\mu}{d\mu}((\phi_n \circ \dots \circ \phi_{i+1})^{-1}(\mathbf{w})) \quad \text{for } \mu \text{-a.e. } \mathbf{w} \in \mathcal{X}. \quad (\text{A.6})$$

## B. Appendix: Rigorous Proofs

This appendix gathers together the rigorous proofs for all the claims made above in Section 2, Section 3 and Section 4.

### B.1 Proof of Theorem 2.2

To show that (H1) and (H2) imply  $\mu P^{\alpha, S, \mathcal{V}} = \mu$ , notice that for any bounded and measurable function  $\varphi : X \rightarrow \mathbb{R}$  we have

$$\begin{aligned} \int_X \varphi(\mathbf{q}) \mu P^{\alpha, S, \mathcal{V}}(d\mathbf{q}) &= \int_X \varphi(\mathbf{q}) \int_X P^{\alpha, S, \mathcal{V}}(\tilde{\mathbf{q}}, d\mathbf{q}) \mu(d\tilde{\mathbf{q}}) \\ &= \sum_{j=0}^p \int_X \int_X \int_Y \varphi(\mathbf{q}) \alpha_j(\tilde{\mathbf{q}}, \mathbf{v}) \delta_{\Pi_1 S_j(\tilde{\mathbf{q}}, \mathbf{v})}(d\mathbf{q}) \mathcal{V}(\tilde{\mathbf{q}}, d\mathbf{v}) \mu(d\tilde{\mathbf{q}}) \\ &= \sum_{j=0}^p \int_X \int_Y \varphi(\Pi_1 S_j(\tilde{\mathbf{q}}, \mathbf{v})) \alpha_j(\tilde{\mathbf{q}}, \mathbf{v}) \mathcal{M}(d\tilde{\mathbf{q}}, d\mathbf{v}) \\ &= \sum_{j=0}^p \int_X \int_Y \varphi(\tilde{\mathbf{q}}) \alpha_j(S_j(\tilde{\mathbf{q}}, \mathbf{v})) S_j^* \mathcal{M}(d\tilde{\mathbf{q}}, d\mathbf{v}) = \int_X \varphi(\tilde{\mathbf{q}}) \mu(d\tilde{\mathbf{q}}). \end{aligned}$$

Since  $\varphi$  is arbitrary, this implies  $\mu P^{\alpha, S, \mathcal{V}} = \mu$  as desired.

For the second part of the statement, the fact that (H3) implies (H2) follows immediately upon summing (2.5) over  $j = 0, \dots, p$ , taking the integral over  $Y$  and invoking (2.1). For (2.6), it suffices to show that for every bounded and measurable function  $\psi : X \times X \rightarrow \mathbb{R}$ , we have

$$\sum_{j=0}^p \int_X \int_Y \psi(\mathbf{q}, \Pi_1 \circ S_j(\mathbf{q}, \mathbf{v})) \alpha_j(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \sum_{j=0}^p \int_X \int_Y \psi(\Pi_1 \circ S_j(\tilde{\mathbf{q}}, \mathbf{v}), \tilde{\mathbf{q}}) \alpha_j(\tilde{\mathbf{q}}, \mathbf{v}) \mathcal{M}(d\tilde{\mathbf{q}}, d\mathbf{v}). \quad (\text{B.1})$$

Invoking (H1) and changing variables, cf. (A.2), we obtain that

$$\begin{aligned}
& \sum_{j=0}^p \int_X \int_Y \psi(\mathbf{q}, \Pi_1 \circ S_j(\mathbf{q}, \mathbf{v})) \alpha_j(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\
&= \sum_{j=0}^p \int_X \int_Y \psi(\Pi_1 \circ S_j \circ S_j(\mathbf{q}, \mathbf{v}), \Pi_1 \circ S_j(\mathbf{q}, \mathbf{v})) \alpha_j(S_j \circ S_j(\mathbf{q}, \mathbf{v})) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\
&= \sum_{j=0}^p \int_X \int_Y \psi(\Pi_1 \circ S_j(\mathbf{q}, \mathbf{v}), \mathbf{q}) \alpha_j(S_j(\mathbf{q}, \mathbf{v})) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}). \tag{B.2}
\end{aligned}$$

Now from (H3) we deduce (B.1), completing the proof.

## B.2 Proof of Corollary 2.4

We first notice that  $S_j^* \mathcal{M} \ll (S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})$ , for each  $j = 0, \dots, p$ , so that the Radon-Nikodym derivative in (2.8) is well-defined. Moreover, clearly

$$\sum_{j=0}^p \frac{dS_j^* \mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) = \frac{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) = 1 \tag{B.3}$$

for  $(\sum_{j=0}^p S_j^* \mathcal{M})$ -a.e.  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ , which implies that  $\alpha_j, j = 0, \dots, p$ , are well-defined, and condition (2.1) from (H3) in Theorem 2.2 holds.

It remains to verify (2.5). Since  $S_j \circ S_j = I$ , it follows from (A.5) that  $\mathcal{M} \ll S_j^*(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})$  and

$$\frac{d\mathcal{M}}{dS_j^*(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) = \frac{dS_j^* \mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(S_j(\mathbf{q}, \mathbf{v})) = \alpha_j(S_j(\mathbf{q}, \mathbf{v}))$$

for  $S_j^*(\sum_{k=0}^p S_k^* \mathcal{M})$ -a.e.  $(\mathbf{q}, \mathbf{v}) \in X \times Y$ . But from our assumption (2.7), we have

$$S_j^*(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M}) = (S_j \circ S_0)^* \mathcal{M} + \dots + (S_j \circ S_p)^* \mathcal{M} = S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M}.$$

Hence,  $\mathcal{M} \ll S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M}$  and

$$\frac{d\mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) = \alpha_j(S_j(\mathbf{q}, \mathbf{v})) \quad \text{for } \left( \sum_{j=0}^p S_j^* \mathcal{M} \right) \text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

We thus obtain

$$\begin{aligned}
\alpha_j(S_j(\mathbf{q}, \mathbf{v})) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \frac{d\mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\
&= \frac{d\mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) \frac{dS_j^* \mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) (S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})(d\mathbf{q}, d\mathbf{v}) \\
&= \frac{dS_j^* \mathcal{M}}{d(S_0^* \mathcal{M} + \dots + S_p^* \mathcal{M})}(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \alpha_j(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}),
\end{aligned}$$

so that (2.5) also holds. Therefore, condition (H3) of Theorem 2.2 is satisfied. This concludes the proof.

### B.3 Proof of Corollary 2.6

Clearly, condition (2.1) follows immediately from the definition of  $\alpha_0$ . Regarding (2.5), the case  $j = 0$  is readily verified since  $S_0 = I$ . For  $j = 1, \dots, p$ ,

$$\begin{aligned}\alpha_j(S_j(\mathbf{q}, \mathbf{v}))S_j^*\mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \bar{\alpha}_j \left( 1 \wedge \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(S_j(\mathbf{q}, \mathbf{v})) \right) \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v})\mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \bar{\alpha}_j \left[ \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \wedge \left( \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(S_j(\mathbf{q}, \mathbf{v})) \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \right) \right] \mathcal{M}(d\mathbf{q}, d\mathbf{v}).\end{aligned}$$

From (A.6) and since  $S_j \circ S_j = I$ , we deduce that

$$\begin{aligned}\alpha_j(S_j(\mathbf{q}, \mathbf{v}))S_j^*\mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \bar{\alpha}_j \left[ 1 \wedge \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(S_j(\mathbf{q}, \mathbf{v})) \right] \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v})\mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \bar{\alpha}_j \left[ \frac{dS_j^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \wedge 1 \right] \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \alpha_j(\mathbf{q}, \mathbf{v})\mathcal{M}(d\mathbf{q}, d\mathbf{v}),\end{aligned}$$

so that (2.5) also holds in this case. This concludes the proof.

### B.4 Proof of Theorem 2.10

Recall that  $\sigma$ -finite measures on  $(X^p, \Sigma_{X^p})$  are uniquely determined by their evaluation on cylinder sets of the form  $E = A_0 \times \dots \times A_p$  where  $A_j \in \Sigma_X$ . Thus, it suffices to verify that

$$\sum_{k=0}^p \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v})(S_j \circ S_k)^*\mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) = \sum_{k=0}^p \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v})S_k^*\mathcal{M}(d\mathbf{q}_0, d\mathbf{v}), \quad (\text{B.4})$$

for every  $j = 0, 1, \dots, p$ , and all  $\varphi : X^{p+1} \rightarrow \mathbb{R}$  of the form

$$\varphi(\mathbf{q}_0, \mathbf{v}) = \prod_{l=0}^p \varphi_l(\mathbf{q}_l), \quad (\mathbf{q}_0, \mathbf{v}) = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \in X^{p+1},$$

with bounded and measurable functions  $\varphi_l : X \rightarrow X$ ,  $l = 0, 1, \dots, p$ .

Clearly, (B.4) holds for  $j = 0$ , so we may assume from now on that  $j \in \{1, \dots, p\}$ . Now for any such  $j$ , noting that  $S_k$  is an involution for each  $k = 0, 1, \dots, p$  we infer that (B.4) reduces to showing that

$$\sum_{\substack{k=1 \\ k \neq j}}^p \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v})(S_j \circ S_k)^*\mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) = \sum_{\substack{k=1 \\ k \neq j}}^p \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v})S_k^*\mathcal{M}(d\mathbf{q}_0, d\mathbf{v}). \quad (\text{B.5})$$

Regarding the right-hand side of (B.5), we obtain by changing variables, (A.2), and recalling the definition of  $\mathcal{M}$  that for every  $k = 1, \dots, p$

$$\begin{aligned}
\int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) S_k^* \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) &= \int_{X^{p+1}} \varphi(S_k(\mathbf{q}_0, \mathbf{v})) \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) \\
&= \int_{X^{p+1}} \varphi(\mathbf{q}_k, \mathbf{q}_1, \dots, \mathbf{q}_{k-1}, \mathbf{q}_0, \mathbf{q}_{k+1}, \dots, \mathbf{q}_p) \int_X \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0) \\
&= \int_{X^{p+2}} \varphi_0(\mathbf{q}_k) \varphi_k(\mathbf{q}_0) \prod_{\substack{l=1 \\ l \neq k}}^p \varphi_l(\mathbf{q}_l) \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0) \tag{B.6} \\
&= \int_{X^2} \varphi_k(\mathbf{q}_0) \left( \int_X \varphi_0(\mathbf{q}_k) Q(\mathbf{q}, d\mathbf{q}_k) \right) \prod_{\substack{l=1 \\ l \neq k}}^p \left( \int_X \varphi_l(\mathbf{q}_l) Q(\mathbf{q}, d\mathbf{q}_l) \right) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0)
\end{aligned}$$

$$\begin{aligned}
&= \int_{X^2} \varphi_k(\mathbf{q}_0) Q\varphi_0(\mathbf{q}) \prod_{\substack{l=1 \\ l \neq k}}^p Q\varphi_l(\mathbf{q}) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0) \\
&= \int_X \varphi_k(\mathbf{q}_0) \int_X \prod_{\substack{l=0 \\ l \neq k}}^p Q\varphi_l(\mathbf{q}) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0), \tag{B.7}
\end{aligned}$$

where we recall the notation  $\tilde{Q}\phi(\mathbf{q}) = \int \phi(\mathbf{q}') \tilde{Q}(\mathbf{q}, d\mathbf{q}')$  for the action of a Markov kernel  $\tilde{Q}$  on a measurable function  $\phi$ .

Turning to the left-hand side of (B.5), it follows once again by change of variables and the definition of  $\mathcal{M}$  that for every  $k = 1, \dots, p$  with  $k \neq j$ ,

$$\begin{aligned}
\int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) (S_j \circ S_k)^* \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) &= \int_{X^{p+1}} \varphi(S_j \circ S_k(\mathbf{q}_0, \mathbf{v})) \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) \\
&= \int_{X^{p+2}} \varphi_0(\mathbf{q}_j) \varphi_j(\mathbf{q}_k) \varphi_k(\mathbf{q}_0) \prod_{\substack{l=1 \\ l \neq k, j}}^p \varphi_l(\mathbf{q}_l) \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0) \\
&= \int_{X^2} \varphi_k(\mathbf{q}_0) \left( \int_X \varphi_0(\mathbf{q}_j) Q(\mathbf{q}, d\mathbf{q}_j) \right) \left( \int_X \varphi_j(\mathbf{q}_k) Q(\mathbf{q}, d\mathbf{q}_k) \right) \prod_{\substack{l=1 \\ l \neq k, j}}^p \left( \int_X \varphi_l(\mathbf{q}_l) Q(\mathbf{q}, d\mathbf{q}_l) \right) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0) \\
&= \int_X \varphi_k(\mathbf{q}_0) \int_X \prod_{\substack{l=0 \\ l \neq k}}^p Q\varphi_l(\mathbf{q}) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu(d\mathbf{q}_0). \tag{B.8}
\end{aligned}$$

Comparing (B.8) with (B.7) and summing yields (B.5) and hence (B.4). This concludes the proof of (2.14). The second part of the statement follows immediately from Corollary 2.4.

### B.5 Proof of Theorem 2.11

We need to establish the absolute continuity  $S_j^* \mathcal{M} \ll \mathcal{M}$  and (2.16) for  $j = 0, \dots, p$ . Since  $S_0 = I$ , clearly this holds for  $j = 0$ . Now we consider  $j \in \{1, \dots, p\}$ . Let  $\mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}) := \mathcal{V}(\mathbf{q}_0, d\mathbf{v})\mu_0(d\mathbf{q}_0)$ . We first claim that assumption (2.15) implies that  $S_j^* \mathcal{M}_0 = \mathcal{M}_0$ . Indeed, take any measurable and bounded function  $\varphi : X^{p+1} \rightarrow \mathbb{R}$ . By change of variables, (A.2), and the definitions of  $\mathcal{V}$  and  $S_j$ ,  $j = 1, \dots, p$ , in (2.12)–(2.13) it follows that

$$\begin{aligned} \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) S_j^* \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}) &= \int_{X^{p+1}} \varphi(S_j(\mathbf{q}_0, \mathbf{v})) \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}) \\ &= \int_{X^{p+1}} \varphi(\mathbf{q}_j, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{q}_0, \mathbf{q}_{j+1}, \dots, \mathbf{q}_p) \int_X \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu_0(d\mathbf{q}_0). \end{aligned}$$

From (2.15) and Fubini's theorem, we thus obtain

$$\begin{aligned} \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) S_j^* \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}) &= \int_{X^{p+1}} \varphi(\mathbf{q}_j, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{q}_0, \mathbf{q}_{j+1}, \dots, \mathbf{q}_p) \int_X \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) Q(\mathbf{q}, d\mathbf{q}_0) \mu_0(d\mathbf{q}) \\ &= \int_{X^{p+2}} \varphi(\mathbf{q}_j, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{q}_0, \mathbf{q}_{j+1}, \dots, \mathbf{q}_p) Q(\mathbf{q}, \cdot)^{\otimes(p+1)}(d\mathbf{q}_j, d\mathbf{q}_1, \dots, d\mathbf{q}_{j-1}, d\mathbf{q}_0, d\mathbf{q}_{j+1}, \dots, d\mathbf{q}_p) \mu_0(d\mathbf{q}) \\ &= \int_{X^{p+2}} \varphi(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) Q(\mathbf{q}, \cdot)^{\otimes(p+1)}(d\mathbf{q}_0, d\mathbf{q}_1, \dots, d\mathbf{q}_p) \mu_0(d\mathbf{q}) \\ &= \int_{X^{p+2}} \varphi(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) Q(\mathbf{q}, d\mathbf{q}_0) \mu_0(d\mathbf{q}) \\ &= \int_{X^{p+2}} \varphi(\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_p) \prod_{i=1}^p Q(\mathbf{q}, d\mathbf{q}_i) \bar{Q}(\mathbf{q}_0, d\mathbf{q}) \mu_0(d\mathbf{q}_0) = \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}), \end{aligned}$$

where  $Q(\mathbf{q}, \cdot)^{\otimes(p+1)}$  denotes the  $(p+1)$ -fold product of the probability measure  $Q(\mathbf{q}, \cdot)$ . Since  $\varphi : X^{p+1} \rightarrow \mathbb{R}$  is an arbitrary measurable and bounded function, we deduce that  $S_j^* \mathcal{M}_0 = \mathcal{M}_0$ .

Next, again for any such function  $\varphi$ , we obtain by invoking the assumption  $\mu \ll \mu_0$  that

$$\begin{aligned} \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) S_j^* \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) &= \int_{X^{p+1}} \varphi(S_j(\mathbf{q}_0, \mathbf{v})) \mathcal{V}(\mathbf{q}_0, d\mathbf{v}) \mu(d\mathbf{q}_0) \\ &= \int_{X^{p+1}} \varphi(S_j(\mathbf{q}_0, \mathbf{v})) \frac{d\mu}{d\mu_0}(\mathbf{q}_0) \mathcal{V}(\mathbf{q}_0, d\mathbf{v}) \mu_0(d\mathbf{q}_0) \\ &= \int_{X^{p+1}} \varphi(S_j(\mathbf{q}_0, \mathbf{v})) \frac{d\mu}{d\mu_0}(\Pi_1(\mathbf{q}_0, \mathbf{v})) \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}) \\ &= \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) \frac{d\mu}{d\mu_0}(\Pi_1 \circ S_j(\mathbf{q}_0, \mathbf{v})) S_j^* \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}). \end{aligned}$$

Since  $S_j^* \mathcal{M}_0 = \mathcal{M}_0$ , then

$$\begin{aligned} \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) S_j^* \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}) &= \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) \frac{d\mu}{d\mu_0}(\Pi_1 \circ S_j(\mathbf{q}_0, \mathbf{v})) \mathcal{M}_0(d\mathbf{q}_0, d\mathbf{v}) \\ &= \int_{X^{p+1}} \varphi(\mathbf{q}_0, \mathbf{v}) \frac{d\mu}{d\mu_0}(\Pi_1 \circ S_j(\mathbf{q}_0, \mathbf{v})) \left( \frac{d\mu}{d\mu_0}(\mathbf{q}_0) \right)^{-1} \mathcal{M}(d\mathbf{q}_0, d\mathbf{v}). \end{aligned}$$

This implies that  $S_j^* \mathcal{M} \ll \mathcal{M}$  and (2.16) holds, which concludes the proof.

### B.6 Proof of Theorem 3.3

Let  $\Phi : \mathcal{Z} \rightarrow \mathbb{R}$  be any bounded measurable function. Regarding the first item, (i),

$$\begin{aligned} \int_{\mathcal{Z}} \Phi(\tilde{\mathbf{z}}) \mathcal{N} \mathcal{A}(d\tilde{\mathbf{z}}) &= \int_{\mathcal{Z}} \Phi(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}, \tilde{k}) \int_{\mathcal{Z}} \mathcal{A}(\mathbf{q}, \mathbf{v}, k, d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}, d\tilde{k}) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk) \\ &= \int_{\mathcal{Z}} \Phi(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}, \tilde{k}) \int_{\mathcal{Z}} S_k^*(\mathcal{V}_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) \delta_{\Pi_1 S_k(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}})) \delta_k(d\tilde{k}) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk) \\ &= \int_{\mathcal{Z}} \Phi(S_k(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}), \tilde{k}) \int_{\mathcal{Z}} \mathcal{V}_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) \delta_{\Pi_1 S_k(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \delta_k(d\tilde{k}) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk) \\ &= \int_{\mathcal{Z}} \int_Y \Phi(S_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), \tilde{\mathbf{v}}), k) \mathcal{V}_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk) \\ &= \sum_{j=0}^p \frac{1}{p+1} \int_{\mathcal{Z}} \int_Y \Phi(S_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), \tilde{\mathbf{v}}), k) \mathcal{V}_k(\Pi_1 S_k(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) \delta_j(dk) \\ &= \sum_{j=0}^p \frac{1}{p+1} \int_Y \int_{X \times Y} \Phi(S_j(\Pi_1 S_j(\mathbf{q}, \mathbf{v}), \tilde{\mathbf{v}}), j) \mathcal{V}_j(\Pi_1 S_j(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) \\ &= \sum_{j=0}^p \frac{1}{p+1} \int_Y \int_{X \times Y} \Phi(S_j(\Pi_1(\mathbf{q}, \mathbf{v}), \tilde{\mathbf{v}}), j) \mathcal{V}_j(\Pi_1(\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{v}}) \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) \\ &= \sum_{j=0}^p \frac{1}{p+1} \int_Y \int_{X \times Y} \Phi(S_j(\mathbf{q}, \tilde{\mathbf{v}}), j) \mathcal{V}_j(\mathbf{q}, d\tilde{\mathbf{v}}) \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q}) \\ &= \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(S_j(\mathbf{q}, \tilde{\mathbf{v}}), j) \mathcal{V}_j(\mathbf{q}, d\tilde{\mathbf{v}}) \mu(d\mathbf{q}) \\ &= \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\mathbf{q}, \tilde{\mathbf{v}}, j) S_j^* \mathcal{M}_j(d\mathbf{q}, d\tilde{\mathbf{v}}) = \int_{\mathcal{Z}} \Phi(\mathbf{q}, \tilde{\mathbf{v}}, k) \mathcal{N}(d\mathbf{q}, d\tilde{\mathbf{v}}, dk), \end{aligned}$$

as desired. Turning to the second item we have

$$\int_{\mathcal{Z}} \Phi(\tilde{\mathbf{z}}) \mathcal{N} \mathcal{A}(d\tilde{\mathbf{z}}) = \int_{\mathcal{Z}} \Phi(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}, \tilde{k}) \int_{\mathcal{Z}} \mathcal{A}(\mathbf{q}, \mathbf{v}, k, d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}, d\tilde{k}) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk)$$

$$\begin{aligned}
&= \sum_{j=0}^p \int_{\mathcal{Z}} \Phi(\mathbf{q}, \mathbf{v}, j) \alpha_{k,j}(\mathbf{q}, \mathbf{v}) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk) \\
&= \sum_{j=0}^p \sum_{l=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\mathbf{q}, \mathbf{v}, j) \alpha_{l,j}(\mathbf{q}, \mathbf{v}) S_l^* \mathcal{M}_l(d\mathbf{q}, d\mathbf{v}) \\
&= \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\mathbf{q}, \mathbf{v}, j) \sum_{l=0}^p \alpha_{l,j}(\mathbf{q}, \mathbf{v}) S_l^* \mathcal{M}_l(d\mathbf{q}, d\mathbf{v}) \\
&= \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\mathbf{q}, \mathbf{v}, j) S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}),
\end{aligned}$$

where we used (3.13) for the final equality. The third claim follows immediately from the first and second one.

Turning to the final claim, (iv), we have, for any  $\Phi : X \rightarrow \mathbb{R}$  bounded and measurable,

$$\begin{aligned}
\int_X \Phi(\mathbf{q}) \mathcal{E}^* \mathcal{N}(d\mathbf{q}) &= \int_{\mathcal{Z}} \Phi(\mathcal{E}(\mathbf{q}, \mathbf{v}, k)) \mathcal{N}(d\mathbf{q}, d\mathbf{v}, dk) = \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\mathcal{E}(\mathbf{q}, \mathbf{v}, j)) S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) \\
&= \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\Pi_1 S_j(\mathbf{q}, \mathbf{v})) S_j^* \mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) \\
&= \sum_{j=0}^p \frac{1}{p+1} \int_{X \times Y} \Phi(\mathbf{q}) \mathcal{V}_j(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q}) = \int_X \Phi(\mathbf{q}) \mu(d\mathbf{q}).
\end{aligned}$$

The proof is complete.

### B.7 Proof of Proposition 4.4

Since  $S_0 = I$  then (4.44) is clearly satisfied for  $j = 0$ . Let us thus assume from now on that  $j \in \{1, \dots, p\}$ . Since  $S_j$  is an involution, namely  $S_j^2 = I$ , then (4.44) reduces to showing that

$$\sum_{\substack{k=1 \\ k \neq j}}^p (S_j \circ S_k)^* \mathcal{M} = \sum_{\substack{k=1 \\ k \neq j}}^p S_k^* \mathcal{M}. \quad (\text{B.9})$$

Fix  $j \in \{1, \dots, p\}$  and  $k \in \{1, \dots, p\}$  with  $k \neq j$ . Let  $\varphi : \mathbb{R}^{N(p+1)} \rightarrow \mathbb{R}$  be any measurable and bounded function. It follows by change of variables, (A.2), that

$$\begin{aligned}
\int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) (S_j \circ S_k)^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_j \circ S_k(\mathbf{q}, \mathbf{v})) F(\mathbf{q}, \cdot)^*(v \otimes \mathcal{H}_N)(d\mathbf{v}) \mu(d\mathbf{q}) \\
&= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_j \circ S_k(\mathbf{q}, F(\mathbf{q}, \lambda, Q))) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}) \\
&= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_j \circ S_k(\mathbf{q}, \mathbf{q} + \lambda Q\mathbf{w}_1, \dots, \mathbf{q} + \lambda Q\mathbf{w}_p)) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}).
\end{aligned}$$

Assuming without loss of generality that  $j \geq k$ , it thus follows from the definition of  $(S_0, \dots, S_p)$  in (4.43) that

$$\begin{aligned} & \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) (S_j \circ S_k)^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q} + \lambda Q \mathbf{w}_j, \mathbf{q} + \lambda Q \mathbf{w}_1, \dots, \mathbf{q} + \lambda Q \mathbf{w}_{k-1}, \mathbf{q}, \mathbf{q} + \lambda Q \mathbf{w}_{k+1}, \dots, \\ & \quad \mathbf{q} + \lambda Q \mathbf{w}_{j-1}, \mathbf{q} + \lambda Q \mathbf{w}_k, \mathbf{q} + \lambda Q \mathbf{w}_{j+1}, \dots, \mathbf{q} + \lambda Q \mathbf{w}_p) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}). \end{aligned} \quad (\text{B.10})$$

Now let  $A \in \mathcal{O}_N$  be the reflection matrix between the vertices  $\mathbf{w}_j$  and  $\mathbf{w}_k$ , so that  $A\mathbf{w}_j = \mathbf{w}_k$ ,  $A\mathbf{w}_k = \mathbf{w}_j$ , and  $A = A^T$ . Concretely,  $A$  can be written as  $I - 2(\mathbf{w}_j - \mathbf{w}_k) \otimes (\mathbf{w}_j - \mathbf{w}_k)$ , namely  $A\mathbf{w} = \mathbf{w} - 2\langle \mathbf{w}, \mathbf{w}_j - \mathbf{w}_k \rangle (\mathbf{w}_j - \mathbf{w}_k)$  for all  $\mathbf{w} \in \mathbb{R}^N$ , where  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product. From the assumed simplex structure in (4.40), it follows immediately that  $A$  leaves all other vertices invariant, namely  $A\mathbf{w}_l = \mathbf{w}_l$  for all  $l \in \{1, \dots, p\}$  with  $l \neq j$  and  $l \neq k$ . For any fixed  $Q \in \mathcal{O}_N$ , we then define  $\tilde{Q} := QAQ^T$ . Notice that  $\tilde{Q} \in \mathcal{O}_N$ , and from the left-invariant property of the Haar measure  $\mathcal{H}_N$ , (4.41), we have  $(\tilde{Q}^T)^* \mathcal{H}_N = \mathcal{H}_N$ . Using this fact in (B10), we obtain by changing variables and invoking the properties of  $A$  that

$$\begin{aligned} & \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) (S_j \circ S_k)^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_j, \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_1, \dots, \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_{k-1}, \mathbf{q}, \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_{k+1}, \dots, \\ & \quad \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_{j-1}, \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_k, \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_{j+1}, \dots, \mathbf{q} + \lambda \tilde{Q} Q \mathbf{w}_p) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}) \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q} + \lambda Q A \mathbf{w}_j, \mathbf{q} + \lambda Q A \mathbf{w}_1, \dots, \mathbf{q} + \lambda Q A \mathbf{w}_{k-1}, \mathbf{q}, \mathbf{q} + \lambda Q A \mathbf{w}_{k+1}, \dots, \\ & \quad \mathbf{q} + \lambda Q A \mathbf{w}_{j-1}, \mathbf{q} + \lambda Q A \mathbf{w}_k, \mathbf{q} + \lambda Q A \mathbf{w}_{j+1}, \dots, \mathbf{q} + \lambda Q A \mathbf{w}_p) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}) \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q} + \lambda Q \mathbf{w}_k, \mathbf{q} + \lambda Q \mathbf{w}_1, \dots, \mathbf{q} + \lambda Q \mathbf{w}_{k-1}, \mathbf{q}, \mathbf{q} + \lambda Q \mathbf{w}_{k+1}, \dots, \\ & \quad \mathbf{q} + \lambda Q \mathbf{w}_{j-1}, \mathbf{q} + \lambda Q \mathbf{w}_j, \mathbf{q} + \lambda Q \mathbf{w}_{j+1}, \dots, \mathbf{q} + \lambda Q \mathbf{w}_p) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}) \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_k(\mathbf{q}, F(\mathbf{q}, \lambda, Q))) v(d\lambda) \mathcal{H}_N(dQ) \mu(d\mathbf{q}) \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_k(\mathbf{q}, \mathbf{v})) F(\mathbf{q}, \cdot)^*(v \otimes \mathcal{H}_N)(d\mathbf{v}) \mu(d\mathbf{q}) = \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) S_k^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}). \end{aligned}$$

Since  $\varphi$  is arbitrary, we deduce that  $(S_j \circ S_k)^* \mathcal{M} = S_k^* \mathcal{M}$ , which shows (B9) and concludes the proof.

### B.8 Proof of Proposition 4.5

We first notice that since  $S_0 = I$ , then (4.45) is readily satisfied when  $j = 0$ . Let us now fix  $j \in \{1, \dots, p\}$ . For any bounded and measurable function  $\varphi : \mathbb{R}^{N(p+1)} \rightarrow \mathbb{R}$ , we have

$$\begin{aligned} \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_j(\mathbf{q}, \mathbf{v})) \pi(\mathbf{q}) F(\mathbf{q}, \cdot)^*(\nu \otimes \mathcal{H}_N)(d\mathbf{v}) d\mathbf{q} \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(S_j(\mathbf{q}, F(\mathbf{q}, \lambda, Q))) \pi(\mathbf{q}) \nu(d\lambda) \mathcal{H}_N(dQ) d\mathbf{q} \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q} + \lambda Q \mathbf{w}_j, \mathbf{q} + \lambda Q \mathbf{w}_1, \dots, \mathbf{q} + \lambda Q \mathbf{w}_{j-1}, \mathbf{q}, \mathbf{q} + \lambda Q \mathbf{w}_{j+1}, \dots, \mathbf{q} + \lambda Q \mathbf{w}_p) \pi(\mathbf{q}) \nu(d\lambda) \mathcal{H}_N(dQ) d\mathbf{q}. \end{aligned}$$

With the change of variables  $\tilde{\mathbf{q}} = \mathbf{q} + \lambda Q \mathbf{w}_j$ , it follows that

$$\begin{aligned} \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, \tilde{\mathbf{q}} + \lambda Q(\mathbf{w}_1 - \mathbf{w}_j), \dots, \tilde{\mathbf{q}} + \lambda Q(\mathbf{w}_{j-1} - \mathbf{w}_j), \tilde{\mathbf{q}} - \lambda Q \mathbf{w}_j, \tilde{\mathbf{q}} + \lambda Q(\mathbf{w}_{j+1} - \mathbf{w}_j), \dots, \\ &\quad \tilde{\mathbf{q}} + \lambda Q(\mathbf{w}_p - \mathbf{w}_j)) \pi(\tilde{\mathbf{q}} - \lambda Q \mathbf{w}_j) \nu(d\lambda) \mathcal{H}_N(dQ) d\tilde{\mathbf{q}}. \end{aligned} \tag{B.11}$$

Let us now consider the reflection matrix  $A$  of  $\mathbf{w}_j$  about the origin, so that  $A\mathbf{w}_j = -\mathbf{w}_j$  and  $A = A^T$ . This can be written explicitly as  $A = I - 2\mathbf{w}_j \otimes \mathbf{w}_j$ , namely  $A\mathbf{w} = \mathbf{w} - 2\langle \mathbf{w}, \mathbf{w}_j \rangle \mathbf{w}_j$  for all  $\mathbf{w} \in \mathbb{R}^N$ , where  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product. Then, assumption (4.40) regarding the simplex  $(\mathbf{w}_1, \dots, \mathbf{w}_p, \mathbf{0})$  implies that  $A\mathbf{w}_l = \mathbf{w}_l - \mathbf{w}_j$  for all  $l \in \{1, \dots, p\}$  with  $l \neq j$ .

Next, similarly as in the proof of Proposition 4.4, we define  $\tilde{Q} = QAQ^T \in \mathcal{O}_N$  and notice that  $(\tilde{Q}^T)^* \mathcal{H}_N = \mathcal{H}_N$  by the left-invariance property of  $\mathcal{H}_N$ , (4.41). From (B.11), we thus obtain after change of variables and the properties of  $A$  that

$$\begin{aligned} \int_{\mathbb{R}^{N(p+1)}} \varphi(\mathbf{q}, \mathbf{v}) S_j^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, \tilde{\mathbf{q}} + \lambda \tilde{Q}Q(\mathbf{w}_1 - \mathbf{w}_j), \dots, \tilde{\mathbf{q}} + \lambda \tilde{Q}Q(\mathbf{w}_{j-1} - \mathbf{w}_j), \tilde{\mathbf{q}} - \lambda \tilde{Q}Q \mathbf{w}_j, \tilde{\mathbf{q}} + \lambda \tilde{Q}Q(\mathbf{w}_{j+1} - \mathbf{w}_j), \dots, \\ &\quad \tilde{\mathbf{q}} + \lambda \tilde{Q}Q(\mathbf{w}_p - \mathbf{w}_j)) \pi(\tilde{\mathbf{q}} - \lambda \tilde{Q}Q \mathbf{w}_j) \nu(d\lambda) \mathcal{H}_N(dQ) d\tilde{\mathbf{q}} \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, \tilde{\mathbf{q}} + \lambda QA(\mathbf{w}_1 - \mathbf{w}_j), \dots, \tilde{\mathbf{q}} + \lambda QA(\mathbf{w}_{j-1} - \mathbf{w}_j), \tilde{\mathbf{q}} - \lambda QA \mathbf{w}_j, \tilde{\mathbf{q}} + \lambda QA(\mathbf{w}_{j+1} - \mathbf{w}_j), \dots, \\ &\quad \tilde{\mathbf{q}} + \lambda QA(\mathbf{w}_p - \mathbf{w}_j)) \pi(\tilde{\mathbf{q}} - \lambda QA \mathbf{w}_j) \nu(d\lambda) \mathcal{H}_N(dQ) d\tilde{\mathbf{q}} \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, \tilde{\mathbf{q}} + \lambda Q \mathbf{w}_1, \dots, \tilde{\mathbf{q}} + \lambda Q \mathbf{w}_{j-1}, \tilde{\mathbf{q}} + \lambda Q \mathbf{w}_j, \tilde{\mathbf{q}} + \lambda Q \mathbf{w}_{j+1}, \dots, \tilde{\mathbf{q}} + \lambda Q \mathbf{w}_p) \pi(\tilde{\mathbf{q}} + \lambda Q \mathbf{w}_j) \\ &\quad \nu(d\lambda) \mathcal{H}_N(dQ) d\tilde{\mathbf{q}} \\ &= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, F(\tilde{\mathbf{q}}, \lambda, Q)) \pi(\Pi_1 S_j(\tilde{\mathbf{q}}, F(\tilde{\mathbf{q}}, \lambda, Q))) \nu(d\lambda) \mathcal{H}_N(dQ) d\tilde{\mathbf{q}} \end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, \mathbf{v}) \pi(\Pi_1 S_j(\tilde{\mathbf{q}}, \mathbf{v})) F(\tilde{\mathbf{q}}, \cdot)^*(v \otimes \mathcal{H}_N)(d\mathbf{v}) d\tilde{\mathbf{q}} \\
&= \int_{\mathbb{R}^{N(p+1)}} \varphi(\tilde{\mathbf{q}}, \mathbf{v}) \frac{\pi(\Pi_1 S_j(\tilde{\mathbf{q}}, \mathbf{v}))}{\pi(\tilde{\mathbf{q}})} \mathcal{M}(d\tilde{\mathbf{q}}, d\mathbf{v}).
\end{aligned}$$

Since  $\varphi$  is arbitrary, we conclude that

$$\frac{dS_j^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \frac{\pi(\Pi_1 S_j(\mathbf{q}, \mathbf{v}))}{\pi(\mathbf{q})} \quad \text{for } \mathcal{M} \text{-a.e. } (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{N(p+1)}.$$

This shows (4.45) and completes the proof.

## C. Appendix: Supplementary Figures

This appendix collects various additional materials from our case studies described in Section 5.

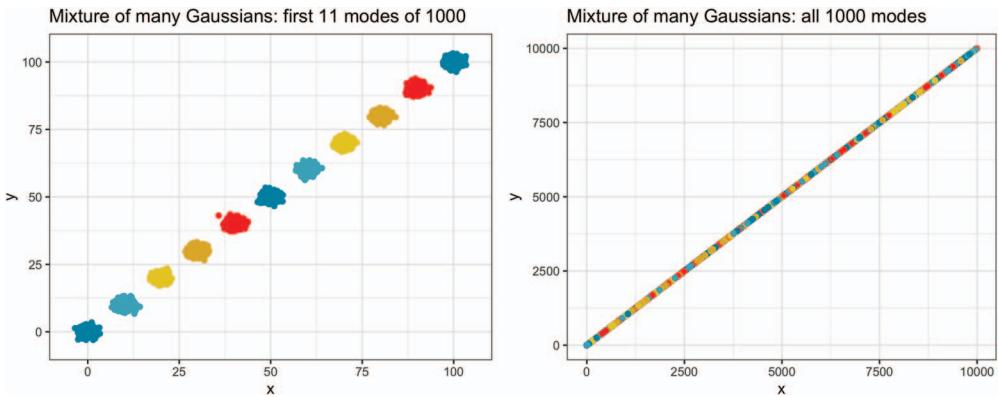


FIG. C13. Mixture of many Gaussians

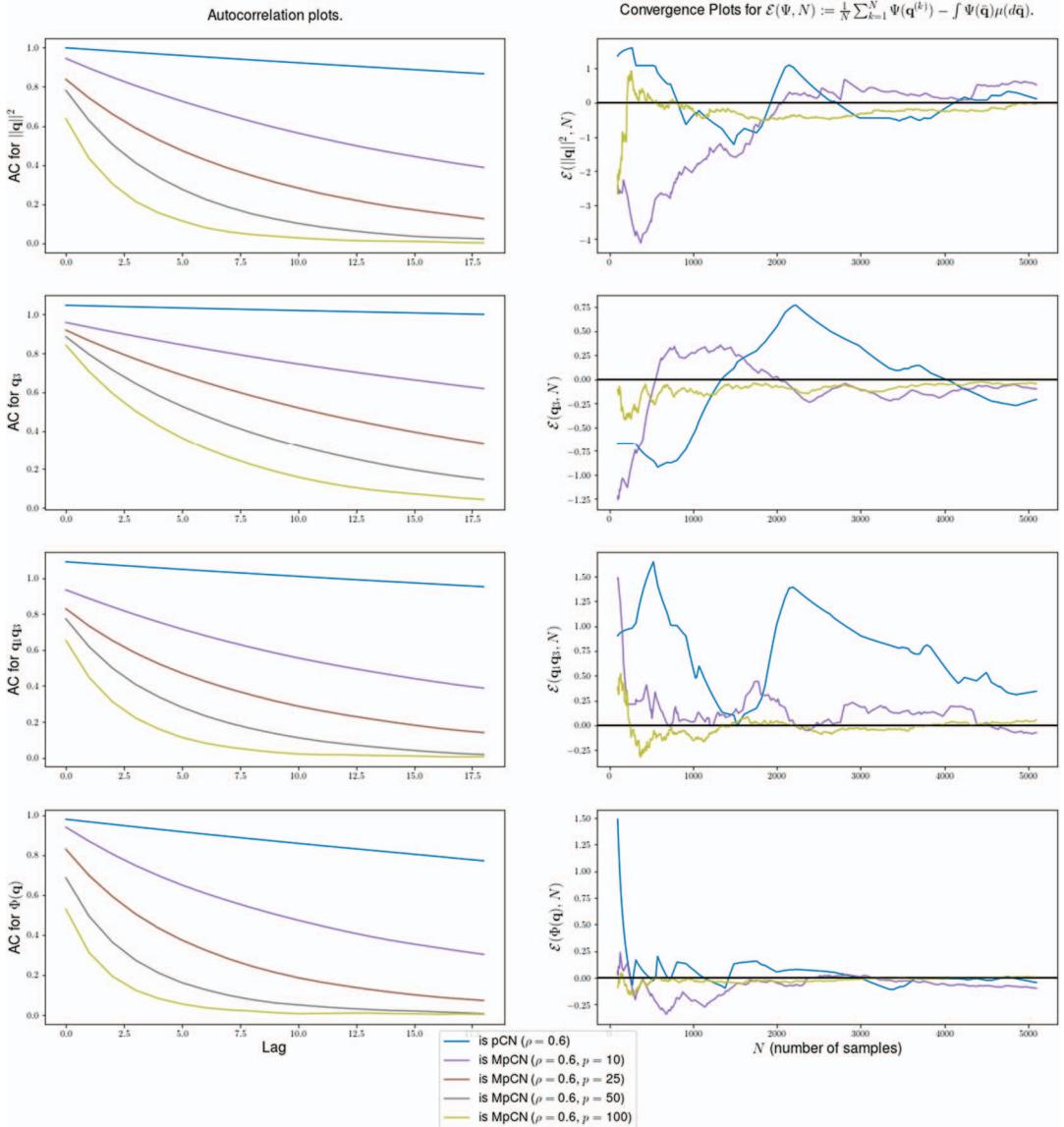


FIG. C14. Autocorrelation and convergence plots for pCN and mpCN for increasing values of  $p$  over various observables.

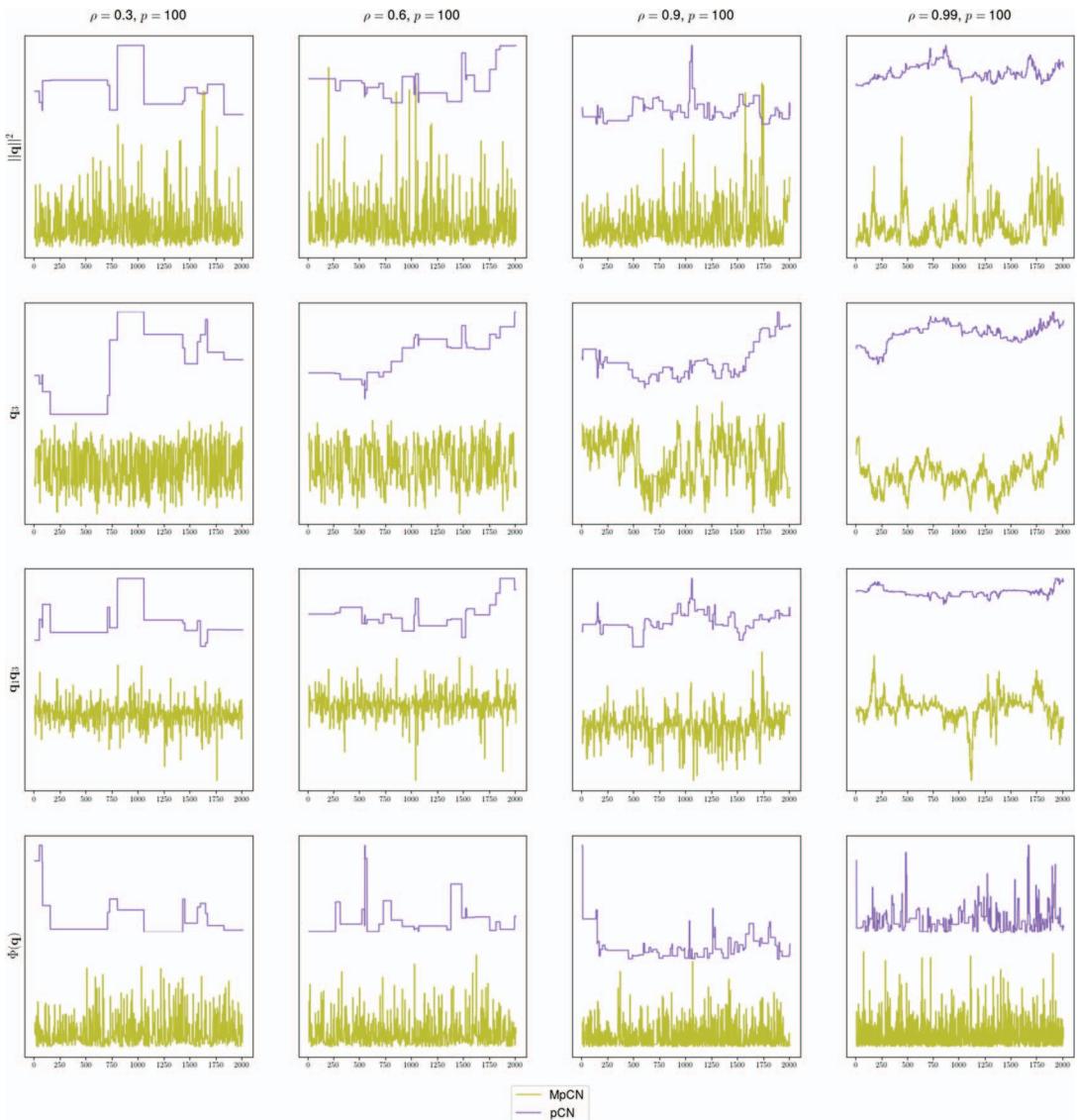


FIG. C15. Time series plots comparing mpCN with  $p = 100$  and pCN for various values of  $\rho$  for different observables.

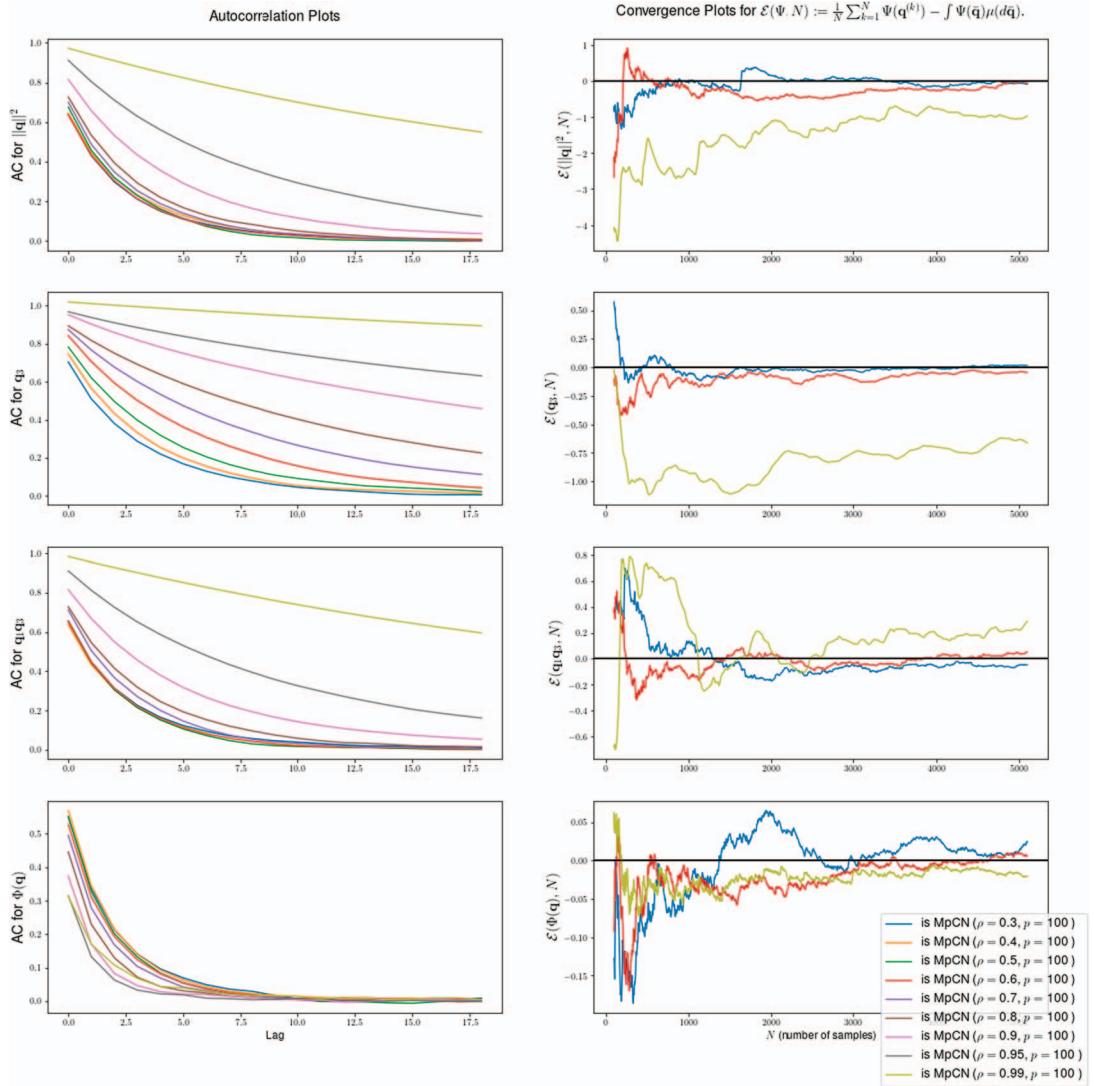


FIG. C16. Autocorrelation and convergence plots for ergodic averages plots comparing mpCN with  $p = 100$  over various values of  $\rho$  for different observables.

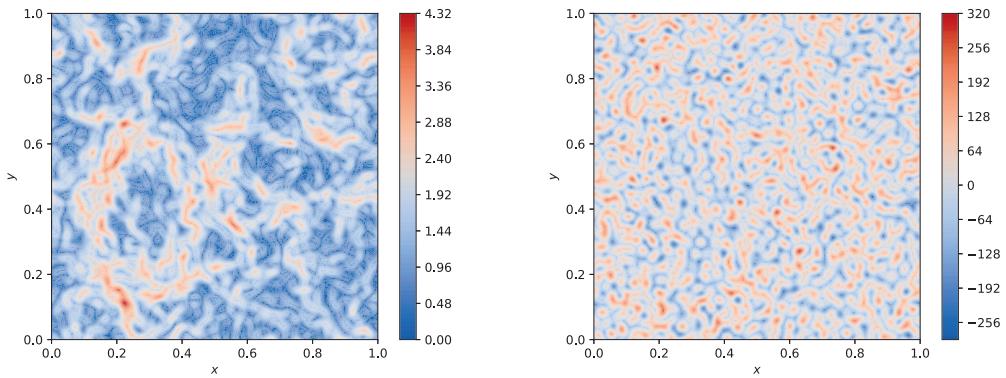


FIG. C17. Typical draw from the prior  $\mu_0 = N(0, \mathcal{C})$  with  $\mathcal{C}$  defined by (5.11). The right panel is  $\text{curl}(\mathbf{q}(\mathbf{x})) = (\partial_{x_2} q_1 - \partial_{x_1} q_2)(\mathbf{x})$  while the left panel is  $\|\mathbf{q}(\mathbf{x})\|$ .

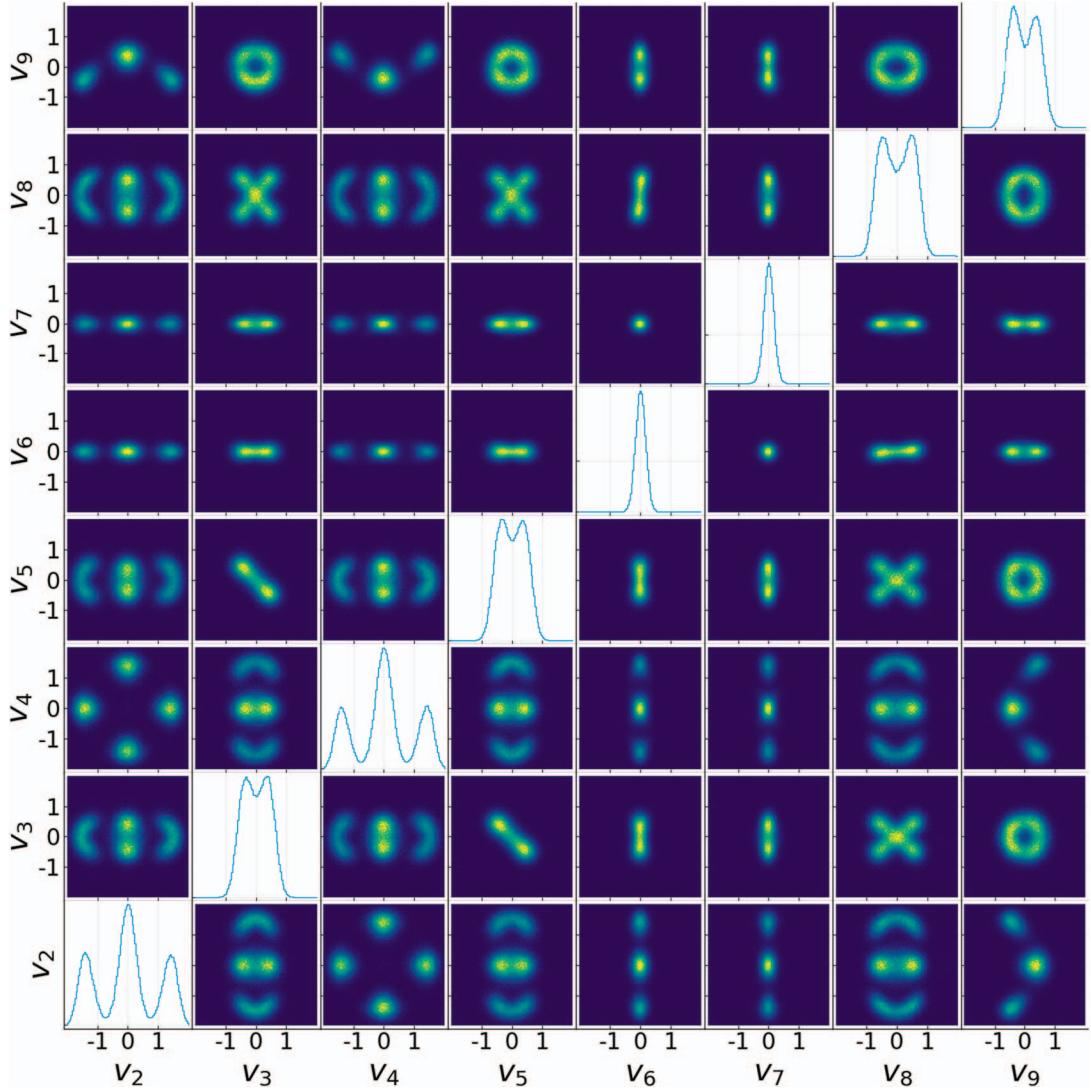


FIG. C18. A grid of histogram of the (fully resolved) two dimensional marginal onto various low frequencies of the posterior measure. More precisely the  $i - j$ th off diagonal is  $F_{i,j}^\# \mu$  where  $F_{i,j}(\mathbf{q}) = [\int_{\mathbb{T}^2} \mathbf{f}_i(\mathbf{x}) \cdot \mathbf{q}(\mathbf{x}) d\mathbf{x}, \int_{\mathbb{T}^2} \mathbf{f}_j(\mathbf{x}) \cdot \mathbf{q}(\mathbf{x}) d\mathbf{x}]$  with  $\mathbf{f}_2(x, y) = [0, \cos(2\pi y)], \mathbf{f}_3(x, y) = [0, -\sin(2\pi y)], \mathbf{f}_4(x, y) = [\cos(2\pi x), 0], \mathbf{f}_5(x, y) = [-\sin(2\pi x), 0], \mathbf{f}_6(x, y) = [0, \cos(4\pi y)], \mathbf{f}_7(x, y) = [0, -\sin(4\pi y)], \mathbf{f}_8(x, y) = [\cos(2\pi x), \cos(2\pi y)], \mathbf{f}_9(x, y) = [-\sin(2\pi x), \sin(2\pi y)]$ .

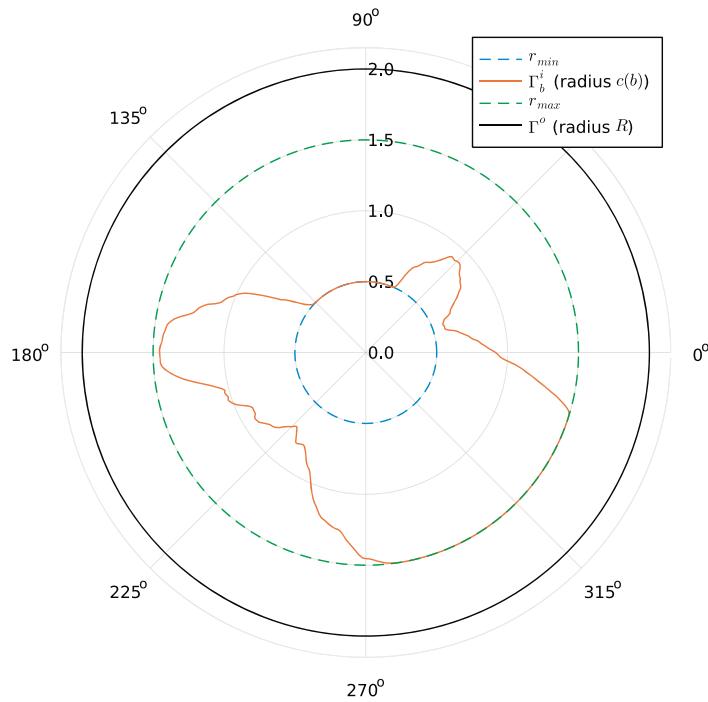


FIG. C19. Diagram of the radius constraints involved in constructing the domain  $\mathcal{D}_q$ .

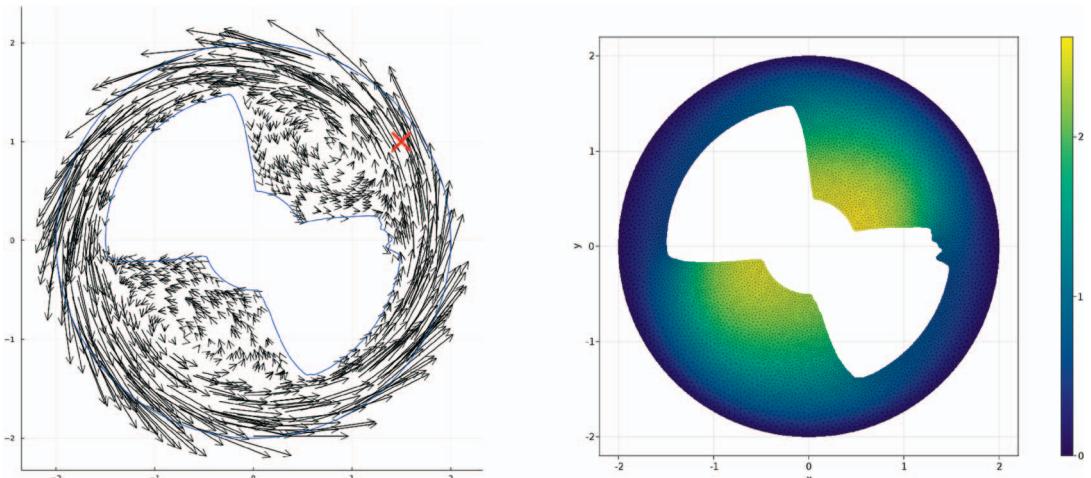


FIG. C20. Solutions of the coupled Stokes and advection-diffusion PDEs. Left: Quiver plot of the solution to the Stokes PDE. The red X marks the point at which the scalar is injected into the system. Right: Contour plot of the solution to the advection-diffusion equation associated with the Stokes flow in the left-hand plot.