Journal of Computational Physics 335 (2017) 327–351

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Geometric MCMC for infinite-dimensional inverse problems

Alexandros Beskos [a], Mark Girolami [b,c], Shiwei Lan [d,*], Patrick E. Farrell [e,f], Andrew M. Stuart [d,*]

[a] Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK
[b] Department of Mathematics, Imperial College London, London, SW7 2AZ, UK
[c] The Alan Turing Institute for Data Science, British Library, 96 Euston Road, London, NW1 2DB, UK
[d] Department of Computing + Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA
[e] Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK
[f] Center for Biomedical Computing, Simula Research Laboratory, Martin Linges vei 17, 1364 Fornebu, Norway

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form 8 December 2016
Accepted 13 December 2016
Available online 28 December 2016

Keywords:
Markov chain Monte Carlo
Local preconditioning
Infinite dimensions
Bayesian inverse problems
Uncertainty quantification

ABSTRACT

Bayesian inverse problems often involve sampling posterior distributions on infinite-dimensional function spaces. Traditional Markov chain Monte Carlo (MCMC) algorithms are characterized by deteriorating mixing times upon mesh-refinement, when the finite-dimensional approximations become more accurate. Such methods are typically forced to reduce step-sizes as the discretization gets finer, and thus are expensive as a function of dimension. Recently, a new class of MCMC methods with mesh-independent convergence times has emerged. However, few of them take into account the geometry of the posterior informed by the data. At the same time, recently developed geometric MCMC algorithms have been found to be powerful in exploring complicated distributions that deviate significantly from elliptic Gaussian laws, but are in general computationally intractable for models defined in infinite dimensions. In this work, we combine geometric methods on a finite-dimensional subspace with mesh-independent infinite-dimensional approaches. Our objective is to speed up MCMC mixing times, without significantly increasing the computational cost per step (for instance, in comparison with the vanilla preconditioned Crank–Nicolson (pCN) method). This is achieved by using ideas from geometric MCMC to probe the complex structure of an intrinsic finite-dimensional subspace where most data information concentrates, while retaining robust mixing times as the dimension grows by using pCN-like methods in the complementary subspace. The resulting algorithms are demonstrated in the context of three challenging inverse problems arising in subsurface flow, heat conduction and incompressible flow control. The algorithms exhibit up to two orders of magnitude improvement in sampling efficiency when compared with the pCN method.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* Corresponding authors.
E-mail addresses: a.beskos@ucl.ac.uk (A. Beskos), m.girolami@imperial.ac.uk (M. Girolami), slan@caltech.edu (S. Lan), patrick.farrell@maths.ox.ac.uk (P.E. Farrell), astuart@caltech.edu (A.M. Stuart).

http://dx.doi.org/10.1016/j.jcp.2016.12.041
0021-9991/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In this work we consider Bayesian inverse problems where the objective is to identify an unknown function parameter u which is an element of a separable Hilbert space $(\mathbb{X}, \langle\cdot,\cdot\rangle, |\cdot|)$. All probability measures on $\mathbb{X}$ in the rest of the paper are assumed to be defined on the standard Borel σ-algebra $\mathcal{B}(\mathbb{X})$. We are given finite-dimensional observations $y \in \mathbb{Y} = \mathbb{R}^m$, for $m \geq 1$, with u and y being connected via the mapping:

$$y = \mathcal{G}(u) + \eta\,, \quad \eta \sim f\,, \tag{1}$$

for some noise distribution f, with u representing the unknown parameter of a (non-linear) PDE and $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ the related forward solution operator for the PDE mapping u onto the data space $\mathbb{Y}$. In a Bayesian setting, a prior measure μ_0 is assigned to u. With a small abuse of notation, we denote also by f the density (assumed to exist) of the noise distribution with respect to the Lebesgue measure, thus we define the negative log-likelihood $\Phi : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ as:

$$\Phi(u; y) = -\log f\big\{\big(y - \mathcal{G}(u)\big); u\big\}\,,$$

with $f\{\cdot\,; u\}$ indicating the density function for a given u. Denoting by μ^y the posterior of u, and using Bayes' theorem, we get:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\,\exp(-\Phi(u; y))$$

for a normalizing constant $Z = \int_{\mathbb{X}} \exp(-\Phi(u; y))\mu_0(du)$ assumed positive and finite.

In this work we consider a Gaussian prior $\mu_0 = \mathcal{N}(0, \mathcal{C})$ with the covariance $\mathcal{C}$ being a positive, self-adjoint and trace-class operator on $\mathbb{X}$. Notice that the posterior μ^y can exhibit strongly non-Gaussian behavior, with finite-dimensional projections having complex non-elliptic contours, although the existence of a density with respect to μ_0 does imply near-Gaussianity for appropriate tail components of the target law μ^y.

Sampling from μ^y in the context of PDE-constrained inverse problems is typically a very challenging undertaking due to the high-dimensionality of the target, the non-Gaussianity of the posterior and the computational burden of repeated PDE solutions for evaluating the likelihood function at different parameters. It is now well-understood that traditional Metropolis-Hastings algorithms have deteriorating mixing times upon refinement of the mesh-size used in practice in the finite-dimensional projection of parameter u. This has prompted the recent development of a class of 'advanced' MCMC methods that avoid this deficiency, see for instance the line of works in [1–7]. The main difference of the new methodology compared to standard Metropolis-Hastings is that the algorithms are well-defined on the infinite-dimensional Hilbert space. This yields the important computational benefit of mesh-independent mixing times for the practical finite-dimensional algorithms ran on the computer.

This work makes a number of contributions. First, we generalize geometric MCMC methods – the simplified Riemannian manifold Metropolis-adjusted Langevin algorithm (MALA) of [8] and a Hamiltonian Monte-Carlo (HMC) extension of it – from finite to infinite dimensions. Unlike recent development of geometric methods including Stochastic Newton (SN) MCMC [9] and Riemannian manifold Hamiltonian Monte Carlo for large-scale PDE-constrained inverse problems [10], these proposed advanced MCMC algorithms are well-defined on the Hilbert space. They have the capacity to both explore complex probability structures and have robust mixing times in high dimensions. Our methodology can also be thought of as a generalization of the operator-weighted proposal of [4] or the dimension-independent likelihood informed (DILI) MCMC method of [7] which exploit the posterior curvature at a fixed point obtained via an optimizer or through adaptive averaging over samples; our methodology invokes position dependent curvatures to allow for more flexible geometric adaptation. We provide high-level conditions and rigorous proofs for the well-posedness of the new methods on infinite-dimensional Hilbert spaces. Second, we establish connections between MALA- and HMC-type algorithms in the infinite dimensional setting. HMC algorithms, viewed as multi-step generalizations of their MALA analogues, make big jumps that suppress random-walk behavior and can provide numerical advantages over MALA by substantially reducing mixing times. Third, we develop a straightforward dimension reduction methodology which renders the methods highly effective from a practical viewpoint. Our methods aim to adapt to the local curvature of the target and provide proposals which are appropriate for non-linear likelihood-informed subspaces. A simpler step is then developed for a complementary subspace obtained by truncating the Karhunen–Loève expansion of the Gaussian prior. Other such separation methods used in the non-geometric context (likelihood informed subspace [11, LIS] or the active subspace [12, AS]) could potentially be brought into our setting, though this requires further research. Lastly, we apply the geometric methods together with other main MCMC algorithms on three challenging inverse problems and contrast their efficiency. Two elliptic inverse problems, involving a groundwater flow and a thermal fin, aim to infer the coefficients of the elliptic PDEs (representing the permeability of a porous medium and the heat conductivity of a material respectively) from data taken at given locations of the forward solver. The third inverse problem involves an incompressible Navier–Stokes equation, with the objective to infer the inflow velocity given sparse observations from the downstream outlet boundary. To the best of our knowledge, it is the first successful application of geometric MCMC methods to non-linear infinite dimensional inverse problems and demonstration of their effectiveness in this field. We should mention here that an important paper in this context is [9] which introduced the Stochastic Newton (SN) method. Although the derivation of the algorithm was not infinite-dimensional, the authors do show that on linear

Gaussian problems the acceptance probability is one, an essential ingredient in the definition of an infinite-dimensional sampler. We also mention that the paper [13] generalizes the SN method by considering variants in which the Hessian is frozen at the maximum a posteriori (MAP) estimator, and low-rank approximations are employed; the methodology is applied to a non-linear ice sheet inverse problem with considerable success. The SN algorithm of [9] can be identified as a special case of our scheme and further details are given in Subsection 3.2).

The paper is organized as follows. Section 2 reviews the recently introduced MCMC methods on infinite-dimensional Hilbert spaces. Section 3 develops the new geometric MCMC methods and establishes their well-posedness under certain conditions. Section 4 applies the new methodology to a number of complex inverse problems and shows that use of information about the underlying geometry can provide significant computational improvements in the cost per unit sample. Section 3.1 concludes with a summary and a suggested path for several future investigations.

2. (Non-geometric) MCMC on Hilbert spaces

We review some of the advanced MCMC methods published in the literature, see e.g. [1–3] or [7] for recent contributions. For simplicity we drop y from the various terms involved, so we denote the posterior as $\mu(du)$ and the potential function as $\Phi(u)$. For target $\mu(du)$ and the various proposal kernels $Q(u, du')$ in the sequel, we define the bivariate law:

$$\nu(du, du') = \mu(du)\, Q(u, du') \,. \tag{2}$$

Following the theory of Metropolis-Hastings on general spaces [14], the acceptance probability $a(u, u')$ is non-trivial when $\nu(du, du') \simeq \nu^{\top}(du, du')$ with $\nu^{\top}$ denoting the symmetrization of ν, that is

$$\nu^{\top}(du, du') := \nu(du', du) \,. \tag{3}$$

The symbol ($\simeq$) denotes absolute continuity between probability measures. The acceptance probability is then:

$$a(u, u') = 1 \wedge \frac{d\nu^{\top}}{d\nu}(u, u') \,, \tag{4}$$

where $\alpha \wedge \beta$ denotes the minimum of $\alpha, \beta \in \mathbb{R}$.

The preconditioned Crank–Nicolson (pCN) method [15,1,3] is a modification of the standard random-walk Metropolis (RWM). The method is described in Algorithm 2.1 and involves a free parameter $\rho \in [0, 1)$ controlling the size of move from the current position. PCN is well-defined on the Hilbert space $\mathbb{X}$ with the proposal being prior-preserving, whereas standard

Algorithm 2.1 A single Markov step for pCN.

1. *Given current u, sample independently $\xi \sim \mathcal{N}(0, \mathcal{C})$ and propose:*

$$u' = \rho\, u + \sqrt{1 - \rho^2}\, \xi \,.$$

2. *Accept u' with probability $1 \wedge \exp\big\{ -\Phi(u') + \Phi(u) \big\}$, otherwise stay at u.*

RWM can only be defined on finite-dimensional discretization and has diminishing acceptance probability for fixed step-size and increasing resolution [16]. Thus, pCN mixes faster than RWM in high-enough dimensions and the disparity in mixing rates becomes greater upon mesh-refinement [3]. However, pCN in general does not use the data in the proposal and can exhibit strong diffusive behavior when exploring complex posteriors. We note here that some recent contributions [4–6] aim to adapt the pCN proposal to the covariance structure of the target.

One approach for developing data-informed methods is to take advantage of gradient information in a steepest-descent setting. Consider the Langevin SDE on the Hilbert space, preconditioned by some operator K:

$$\frac{du}{dt} = -\frac{1}{2} K \left\{ \mathcal{C}^{-1} u + D\Phi(u) \right\} + \sqrt{K}\, \frac{dW}{dt} \tag{5}$$

with $D\Phi(u)$ denoting the Fréchet derivative of Φ (or the corresponding element of the relevant dual space; we will be more precise when defining our new methods in the section 3) and W being the cylindrical Wiener process. We consider these dynamics under the setting $K = \mathcal{C}$, when scales are tuned to the prior. Formally, SDE (5) preserves the posterior μ and can be used as the basis for developing effective MCMC proposals [1,3]. [1] use the following semi-implicit Euler scheme to discretize the above SDE:

$$\frac{u' - u}{h} = -\frac{1}{2} \left\{ \frac{u + u'}{2} + \alpha\, \mathcal{C} D\Phi(u) \right\} + \sqrt{\frac{1}{h}}\, \xi \,, \quad \xi \sim \mathcal{N}(0, \mathcal{C}) \,, \tag{6}$$

for an algorithmic parameter $\alpha \equiv 1$ and some small step-size $h > 0$. This can be rewritten as:

$$u' = \rho\, u + \sqrt{1 - \rho^2}\, v \,, \quad v = \xi - \tfrac{\alpha\sqrt{h}}{2}\, \mathcal{C} D\Phi(u) \,, \quad \rho = (1 - \tfrac{h}{4})/(1 + \tfrac{h}{4}) \,. \tag{7}$$

Note that the image space $\mathrm{Im}(\mathcal{C}^{\frac{1}{2}})$ is comprised of all $u \in \mathbb{X}$ such that $\mathcal{N}(u, \mathcal{C}) \simeq \mathcal{N}(0, \mathcal{C})$, see e.g. [17]. Thus, following [1], under the assumption that $\mathcal{C}D\Phi(u) \in \mathrm{Im}(\mathcal{C}^{1/2})$, μ_0-a.s. in u, one can use Theorem 2.21 of [17] on translations of Gaussian measures on separable Hilbert spaces, to obtain the following Radon–Nikodym derivative (we denote by $Q(u, du')$ and $Q_0(u, du')$ the proposal kernels determined by (7) for $\alpha = 1$ and $\alpha = 0$, respectively):

$$\frac{dQ(u,\cdot)}{dQ_0(u,\cdot)}(u') = \exp\Big\{ -\tfrac{h}{8}\,|\mathcal{C}^{1/2}D\Phi(u)|^2 - \tfrac{\sqrt{h}}{2}\langle D\Phi(u), \tfrac{u'-\rho u}{\sqrt{1-\rho^2}}\rangle \Big\}\,. \tag{8}$$

The bivariate Gaussian law $\nu_0(du, du') := \mu_0(du)Q_0(u, du')$ is symmetric ($\nu_0 = \nu_0^\top$), thus one can obtain the Metropolis-Hastings ratio in the accept/reject (4) as $d\nu^\top/d\nu = (d\nu^\top/d\nu_0^\top)/(d\nu/d\nu_0)$. The complete method, labeled ∞-MALA (infinite-dimensional MALA), is defined in Algorithm 2.2.

Algorithm 2.2 A single Markov step for ∞-MALA.

1. *Given current u, sample independently $\xi \sim \mathcal{N}(0, \mathcal{C})$ and propose:*

$$u' = \rho\, u + \sqrt{1-\rho^2}\,\big\{\xi - \tfrac{\sqrt{h}}{2}\,\mathcal{C}D\Phi(u)\big\}$$

2. *Accept u' with probability $a(u, u') = 1 \wedge \frac{\kappa(u',u)}{\kappa(u,u')}$, where we have set:*

$$\kappa(u, u') = \frac{1}{Z}\,\exp\{-\Phi(u)\} \times \exp\Big\{ -\tfrac{h}{8}\,|\mathcal{C}^{1/2}D\Phi(u)|^2 - \tfrac{\sqrt{h}}{2}\langle D\Phi(u), \tfrac{u'-\rho u}{\sqrt{1-\rho^2}}\rangle \Big\}$$

otherwise stay at u.

Another likelihood-informed Metropolis-Hastings method involves exploiting Hamiltonian dynamics. Consider the Hamiltonian differential equation with mass matrix [1] equal to K^{-1}, that is:

$$\frac{d^2u}{dt^2} + K\big\{\mathcal{C}^{-1}u + D\Phi(u)\big\} = 0\,. \tag{9}$$

These dynamics, considered on the phase-space of (u, v), for the velocity $v = du/dt$, preserve the total energy:

$$H(u, v) = \Phi(u) + \tfrac{1}{2}\,\langle v, K^{-1}v\rangle\,.$$

From a probabilistic point of view, when initialized with $v \sim \mathcal{N}(0, K)$, the Hamiltonian dynamics (formally) preserve the target measure μ for any integration time, and thus they can form the basis for an MCMC method, termed Hybrid (or Hamiltonian) Monte-Carlo (HMC) [18,15]. [2] modify the standard HMC algorithm to develop an advanced method that is well-defined on the Hilbert space $\mathbb{X}$. We label this algorithm ∞-HMC (infinite-dimensional HMC). In more detail, setting again $K = \mathcal{C}$ the dynamics in (9) can be written in the standard form:

$$\frac{du}{dt} = v\,, \quad \frac{dv}{dt} = -u - \mathcal{C}D\Phi(u)\,. \tag{10}$$

Equation (10) gives rise to a semigroup that maps $(u(0), v(0)) \mapsto (u(t), v(t))$ and preserves the product measure $\mu \otimes \mu_0$ under regularity conditions on $\mathcal{C}$ and $D\Phi(u)$ [2]. Standard HMC synthesizes Euler steps on the two differential equations in (10) to produce an approximate symplectic integrator. In contrast, ∞-HMC makes use of the Strang splitting scheme:

$$du/dt = v\,, \quad dv/dt = -u\,; \tag{11}$$

$$du/dt = 0\,, \quad dv/dt = -\mathcal{C}D\Phi(u)\,, \tag{12}$$

and develops a Störmer–Verlet-type integrator [19,15] by synthesizing solvers of (11), (12) as follows, for some small $\varepsilon > 0$ and initial values (u_0, v_0):

$$\begin{aligned} v^- &= v_0 - \tfrac{\varepsilon}{2}\,\mathcal{C}D\Phi(u_0)\,; \\ \begin{bmatrix} u_\varepsilon \\ v^+ \end{bmatrix} &= \begin{bmatrix} \cos\varepsilon & \sin\varepsilon \\ -\sin\varepsilon & \cos\varepsilon \end{bmatrix} \begin{bmatrix} u_0 \\ v^- \end{bmatrix}\,; \\ v_\varepsilon &= v^+ - \tfrac{\varepsilon}{2}\,\mathcal{C}D\Phi(u_\varepsilon)\,. \end{aligned} \tag{13}$$

This scheme, referred to as a leapfrog step, gives rise to a map $\Psi_\varepsilon : (u_0, v_0) \mapsto (u_\varepsilon, v_\varepsilon)$. The algorithm proposes big jumps in the state space by synthesizing $I = \lfloor \tau/\varepsilon \rfloor$ leapfrog maps, for some time horizon $\tau > 0$. It can be shown that if $I = 1$ then

[1] The terminology 'mass matrix' used in Hamiltonian dynamical systems should not be confused with the same term used in finite element methods for PDEs.

∞-HMC coincides with ∞-MALA for particular choice of step-sizes (see more details in Subsection 3.3). ∞-HMC will many times manifest numerical advantages over ∞-MALA due to the longer, designated moves suppressing random walk behavior. ∞-HMC develops as shown in Algorithm 2.3, where for starting position and velocity (u, v) we have set $(u_i, v_i) = \Psi_\varepsilon^i(u, v)$, with Ψ_ε^i denoting the synthesis of i maps Ψ_ε, $0 \le i \le I$. Also, we denote by $\mathcal{P}_u$ the projection onto the u-argument. The derivation of the accept/reject rule is more involved than ∞-MALA, and requires again that $\mathcal{C}D\Phi(u) \in \mathrm{Im}(\mathcal{C}^{1/2})$, μ_0-a.s. in u; we refer the reader to [2]. We will provide full details on the accept/reject when developing the more general geometric version of ∞-HMC in Subsection 3.3.

Algorithm 2.3 A single Markov step for ∞-HMC.

1. *Given current u, sample independently $v \sim \mathcal{N}(0, \mathcal{C})$ and propose $u' = \mathcal{P}_u\{\Psi_\varepsilon^I(u, v)\}$.*
2. *Accept u' with probability $1 \wedge \exp\{-\Delta H(u, v)\}$ where we have set:*

$$\begin{aligned}\Delta H(u, v) &= H(\Psi_\varepsilon^I(u, v)) - H(u, v)\\ &\equiv \Phi(u_I) - \Phi(u_0) - \tfrac{\varepsilon^2}{8}\{|\mathcal{C}^{\frac12} D\Phi(u_I)|^2 - |\mathcal{C}^{\frac12} D\Phi(u_0)|^2\}\\ &\quad - \tfrac{\varepsilon}{2}\sum_{i=0}^{I-1}\big(\langle v_i, D\Phi(u_i)\rangle + \langle v_{i+1}, D\Phi(u_{i+1})\rangle\big)\end{aligned}$$

otherwise stay at u.

3. Geometric Metropolis-Hastings algorithms

Recall the assumed distribution of the data in (1). We will be more explicit here and for expository convenience assume Gaussian noise $\eta \sim \mathcal{N}_m(0, \Sigma)$, for some symmetric, positive-definite $\Sigma \in \mathbb{R}^{m\times m}$. Thus the target distribution is:

$$\frac{d\mu}{d\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u)) = \frac{1}{Z}\exp\{-\tfrac12 |y - \mathcal{G}(u)|_\Sigma^2\}$$

for some constant $Z > 0$, where we have considered the scaled inner product $\langle\cdot,\cdot\rangle_\Sigma = \langle\cdot, \Sigma^{-1}\cdot\rangle$. Below, we will define MCMC algorithms on the Hilbert space $\mathbb{X}$, and express conditions for their well-posedness in terms of the properties of the forward map $\mathcal{G} = (\mathcal{G}_k)_{k=1}^m : \mathbb{X} \mapsto \mathbb{R}^m$ which involves regularity properties of the underlying PDE in the given inverse problem.

We work with the eigenvectors and eigenvalues of the prior covariance operator $\mathcal{C}$, so that $\{\phi_j\}_{j\ge1}$ is an orthonormal basis of $\mathbb{X}$ and $\{\lambda_j^2\}_{j\ge1}$ a sequence of positive reals with $\sum\lambda_j^2 < \infty$ (this enforces the trace-class condition for $\mathcal{C}$), such that $\mathcal{C}\phi_j = \lambda_j^2\phi_j$, $j \ge 1$. We make the usual correspondence between an element u and its coordinates w.r.t. the basis $\{\phi_j\}_{j\ge1}$, that is $u = \sum_j u_j\phi_j \leftrightarrow \{u_j\}_{j\ge1}$. Using the standard Karhunen–Loève expansion of a Gaussian measure [20–22] we have the representation:

$$u \sim \mathcal{N}(0, \mathcal{C}) \iff u = \sum_{j=1}^\infty u_j\phi_j\,,\ u_j \sim \mathcal{N}(0, \lambda_j^2)\,,\ \text{ind. over } j \ge 1\,. \tag{14}$$

We define the Sobolev spaces corresponding to the basis $\{\phi_j\}$:

$$\mathbb{X}^s = \Big\{\{u_j\}_{j\ge1} : \sum j^{2s}|u_j|^2 < \infty\Big\}\,,\quad s \in \mathbb{R}\,,$$

so that $\mathbb{X}^0 \equiv \mathbb{X}$ and $\mathbb{X}^s \subset \mathbb{X}^{s'}$ if $s' < s$. Typically, we will have $\lambda_j = \Theta(j^{-\kappa})$ for some $\kappa > 1/2$ in the sense that $C_1\cdot j^{-\kappa} \le \lambda_j \le C_2\cdot j^{-\kappa}$ for all $j \ge 1$, for constants $C_1, C_2 > 0$. Thus, the prior (so also the posterior) concentrate on $\mathbb{X}^s$ for any $s < \kappa - 1/2$. Notice also that:

$$\mathrm{Im}(\mathcal{C}^{1/2}) = \mathbb{X}^\kappa\,.$$

Assumption 3.1 imposes some conditions on the gradient $D\Phi(u)$.

Assumption 3.1. *(i) $\lambda_j = \Theta(j^{-\kappa})$, for $\kappa > 1/2$.*
(ii) For some $\ell \in [0, \kappa - 1/2)$, the maps $\mathcal{G}_k : \mathbb{X}^\ell \mapsto \mathbb{R}$, $1 \le k \le m$, are Fréchet differentiable on $\mathbb{X}^\ell$ with derivatives $D\mathcal{G}_k \in \mathbb{X}^{-\ell}$.

We can assume that ℓ is arbitrarily close to $\kappa - 1/2$. We make the standard correspondence between the bounded linear operator $D\mathcal{G}_k(u)$ on $\mathbb{X}^\ell$ and an element of its dual space $D\mathcal{G}_k(u) \in \mathbb{X}^{-\ell}$ so that $D\mathcal{G}_k(u)(v) = \langle D\mathcal{G}_k(u), v\rangle$ for all $u, v \in \mathbb{X}^\ell$. We consider the derivative $D\mathcal{G}(u) = (D\mathcal{G}_1(u), \dots D\mathcal{G}_m(u)) \in \{\mathbb{X}^{-\ell}\}^m$, $u \in \mathbb{X}^\ell$. Under Assumption 3.1, mapping Φ is Fréchet differentiable on $\mathbb{X}^\ell$ with:

$$D\Phi(u) = D\mathcal{G}(u)\Sigma^{-1}(\mathcal{G}(u) - y) \in \mathbb{X}^{-\ell}\,.$$

3.1. Local Gaussian approximation of posterior

All three MCMC algorithms shown in Section 2 adjust scales in the proposal according to the prior covariance $\mathcal{C}$. Indeed, if the target distribution was simply μ_0, the proposal dynamics would equalize all scales and would also have acceptance probability equal to 1. However, one can get more effective algorithms if the geometry of the posterior itself is taken into consideration in the selection of step-sizes. We explore in this paper the idea of using a preconditioner $K = K(u)$ which will be location-specific in order to construct algorithms that are tuned to the local curvature of the posterior as pioneered in [8], and developed subsequently in other works, see e.g. [23].

Reviewing ∞-MALA and ∞-HMC methods presented in Section 2, the effect of the implicit method (6) and the splitting (11) used for ∞-MALA and ∞-HMC respectively is that the resulting scheme provides an 'ideal' proposal of acceptance probability 1 (respectively of the step-sizes h or ε) for the reference Gaussian measure $\mu_0 = \mathcal{N}(0, \mathcal{C})$. Thinking about the local-move ∞-MALA algorithm, if the negative log-density w.r.t. μ_0, $u \mapsto \Phi(u)$, is relatively flat locally around the current position u, then one can expect relatively high acceptance probability when proposing a move from u for the target μ itself, for a small step-size h. In general, it makes sense to attempt to obtain alternative (to the prior μ_0) Gaussian reference measures that deliver 'flattened' log-densities for the target μ. This leads naturally to the choice of *local* reference measures, as differently oriented elliptic contours can provide better proxies to the target contours at different parts of the state space.

We turn at this point to a finite-dimensional context (so $\mathbb{X} \equiv \mathbb{R}^n$ for some $n \geq 1$) and adopt an informal approach to avoid distracting technicalities. Assume that we are interested in the target posterior in the vicinity of $u_0 \in \mathbb{X}$. A second-order Taylor expansion of the log-target (up to an additive constant):

$$l(u) := -\Phi(u) - \tfrac{1}{2}\langle u, \mathcal{C}^{-1}u\rangle$$

around u_0 will give that:

$$\exp\{l(u)\} = \\ = c(u_0)\exp\big\{-\tfrac{1}{2}\big\langle u - m(u_0), [-D^2 l(u_0)](u - m(u_0))\big\rangle + \mathcal{O}(|u-u_0|^3)\big\}$$

for some easy-to-identify $m(u_0) \in \mathbb{X}$, $c(u_0) \in \mathbb{R}$ that depend on u_0. Thus, with the Gaussian law $\mathcal{N}(m(u_0), [-D^2 l(u_0)]^{-1})$ as new reference measure, the negative log-density (w.r.t. this Gaussian law) of the target μ will be equal to $c'(u_0) + \mathcal{O}(|u-u_0|^3)$ for some constant $c'(u_0) \in \mathbb{R}$, i.e., relatively flat in the vicinity of u_0. Following the discussion in the previous paragraph, we will aim to develop algorithms driven by these local reference measures. (Note that this local Gaussian reference measure coincides with the local Gaussian approximation used in the development of the Stochastic Newton method in [9].)

To be more specific, we will achieve the required effect by allowing for general location-specific preconditioner $K = K(u_0)$ with the choice of $K(u_0)^{-1}$ motivated by the structure of the negative Hessian $-D^2 l(u_0)$ at current position u_0. Thus, we will work with the local reference measure (in the vicinity of u_0):

$$\tilde{\mu}_0 = \mathcal{N}(m(u_0), K(u_0))$$

($m(u_0)$ cancels out in the subsequent developments and will not affect the algorithms) and the target distribution μ expressed as:

$$\frac{d\mu}{d\tilde{\mu}_0}(u) = c''(u_0)\exp\{-\widetilde{\Phi}(u; u_0)\}\,, \tag{15}$$

for some $c''(u_0) \in \mathbb{R}$, where we have defined the negative log-density:

$$\widetilde{\Phi}(u; u_0) := \Phi(u) + \tfrac{1}{2}\langle u, \mathcal{C}^{-1}u\rangle - \tfrac{1}{2}\big\langle u - m(u_0), K(u_0)^{-1}(u - m(u_0))\big\rangle\,,$$

indicating the discrepancy between the target and the local reference measure. We also write its derivative:

$$D\widetilde{\Phi}(u; u_0) = D\Phi(u) + \mathcal{C}^{-1}u - K(u_0)^{-1}(u - m(u_0))\,. \tag{16}$$

We will use the reference measures $\tilde{\mu}_0$ as drivers for the implicit scheme when deriving a local-move MALA algorithm. Similarly to Section 2, we will also define an HMC-type algorithm as an extension of the MALA version when we allow the synthesis of a number of local steps before applying the accept/reject.

3.2. ∞-mMALA

Recall the Langevin dynamics in (5) that gave rise (for $K = \mathcal{C}$) to ∞-MALA in Section 2. The above discussion, and re-expression of the target as in (15), suggest invoking dynamics of the type:

$$\frac{du}{dt} = -\frac{1}{2}K(u)\big\{\mathcal{C}^{-1}u + D\Phi(u)\big\} + \sqrt{K(u)}\,\frac{dW}{dt} \tag{17}$$

for a location-specific preconditioner $K(u)$ (its choice motivated in practice by the form of the inverse negative Hessian at the current position). Notice that these dynamics do not, in general, preserve the target μ as they omit the higher order (and computationally expensive) Christofell symbol terms, see e.g. [8] and the discussion in [24]. As noted with the study of 'Simplified MALA' in [8], the dynamics in (17) can still capture an important part of the local curvature structure of the target and can provide an effective balance between mixing and computational cost.

The time-discretization scheme develops as in the case of ∞-MALA, with the important difference that it will now be driven by the local reference measure $\tilde{\mu}_0$ rather than the prior. That is, we re-write (17) as follows:

$$\frac{du}{dt} = -\frac{1}{2}K(u)\left\{K(u)^{-1}(u-m(u)) + D\widetilde{\Phi}(u;u)\right\} + \sqrt{K(u)}\,\frac{dW}{dt} \tag{18}$$

and develop the semi-implicit scheme as follows:

$$\frac{u'-u}{h} = -\frac{1}{2}\left\{\frac{u+u'}{2} - m(u) + K(u)D\widetilde{\Phi}(u;u)\right\} + \sqrt{\frac{1}{h}}\,\xi\ ; \tag{19}$$
$$\xi \sim \mathcal{N}(0, K(u))\ .$$

Notice that $m(u)$ cancels out (simply apply operator $K(u)$ on both sides of (16), replace $u_0 \leftrightarrow u$ and use the obtained expression for $K(u)D\widetilde{\Phi}(u;u)$ here) and we can rewrite (19) in the general form:

$$\frac{u'-u}{h} = -\frac{1}{2}\left\{\frac{u'+u}{2} - g(u)\right\} + \sqrt{\frac{1}{h}}\,\xi\ , \quad \xi \sim \mathcal{N}(0, K(u))\ , \tag{20}$$

where we have defined:

$$g(u) = -K(u)\left\{(\mathcal{C}^{-1} - K^{-1}(u))u + D\Phi(u)\right\}\ . \tag{21}$$

Re-arranging terms, we can equivalently write:

$$u' = \rho\, u + \sqrt{1-\rho^2}\, v\ , \quad v = \xi + \tfrac{\sqrt{h}}{2}g(u)\ , \quad \xi \sim \mathcal{N}(0, K(u))\ , \tag{22}$$

for ρ defined as in (7).

Recall the steps for identifying the Metropolis-Hastings acceptance probability in (2)–(4) and the related notation for the involved bivariate measures. The following assumptions are sufficient for the well-posedness of the proposal (22) and for providing a non-trivial Radon–Nikodym derivative $(d\nu^{\mathsf{T}}/d\nu)(u,u')$ on the Hilbert space $\mathbb{X}$.

Assumption 3.2. *We have, μ_0-a.s. in $u \in \mathbb{X}$, that $K(u)$ is a self-adjoint, positive-definite and trace-class operator on Hilbert space $\mathbb{X}$, and it is such that:*

i) $\mathrm{Im}(K(u)^{1/2}) = \mathrm{Im}(\mathcal{C}^{\frac{1}{2}})(=\mathbb{X}^{\kappa})$;
ii) $\{\mathcal{C}^{-1/2}K(u)^{1/2}\}\{\mathcal{C}^{-1/2}K(u)^{1/2}\}^{\top} - I$ *is a Hilbert–Schmidt operator on* $\mathbb{X}$.

A linear, bounded operator $A : \mathbb{X} \mapsto \mathbb{X}$ is Hilbert–Schmidt if $\sum_j |A\phi_j|^2 < \infty$.

Assumption 3.3. $(K(u)\mathcal{C}^{-1} - I)u \in \mathrm{Im}(\mathcal{C}^{\frac{1}{2}})(=\mathbb{X}^{\kappa})$, μ_0-*a.s. in* u.

Corollary 3.4. *Under Assumptions 3.1–3.3, we have that* $g(u) \in \mathbb{X}^{\kappa}$.

Proof. Due to Assumption 3.3, it remains to show $K(u)D\Phi(u) \in \mathbb{X}^{\kappa}$. Note that $K(u)D\Phi(u) = \mathcal{C}^{1/2}R\,\mathcal{C}^{1/2}D\Phi(u)$ where $R = \{\mathcal{C}^{-1/2}K(u)^{1/2}\}\{\mathcal{C}^{-1/2}K(u)^{1/2}\}^{\top}$. Also, from Assumption 3.1, $\mathcal{C}^{1/2}D\Phi(u) \in \mathbb{X}^{\kappa-\ell} \subseteq \mathbb{X}$. So, $\mathcal{C}^{1/2}R\,\mathcal{C}^{1/2}D\Phi(u) \in \mathrm{Im}(\mathcal{C}^{1/2}) = \mathbb{X}^{\kappa}$. □

From the Feldman–Hajek theorem (see e.g. Theorem 2.23 in [17]), Assumption 3.2 and Corollary 3.4 are necessary and sufficient so that $\mathcal{N}(g(u), K(u)) \simeq \mathcal{N}(0, \mathcal{C})$, μ_0-a.s. in u. The following result gives the corresponding Radon–Nikodym derivative, which will then be used to illustrate the well-posedness of the MCMC algorithm and provide the Metropolis-Hastings acceptance probability.

Theorem 3.5. *Assumptions 3.1–3.3 imply that* $\mathcal{N}((\sqrt{h}/2)\,g(u), K(u)) \simeq \mathcal{N}(0, \mathcal{C})$, μ_0-*a.s. in* u, *with Radon–Nikodym derivative:*

$$\begin{aligned}\lambda(w;u) :&= \frac{d\mathcal{N}(\frac{\sqrt{h}}{2}\,g(u), K(u))}{d\mathcal{N}(0,\mathcal{C})}(w) = \frac{d\mathcal{N}(\frac{\sqrt{h}}{2}\,g(u), K(u))}{d\mathcal{N}(0,K(u))}(w) \times \frac{d\mathcal{N}(0,K(u))}{d\mathcal{N}(0,\mathcal{C})}(w)\\ &= \exp\left\{-\tfrac{h}{8}|K^{-\frac{1}{2}}(u)g(u)|^2 + \tfrac{\sqrt{h}}{2}\langle K^{-\frac{1}{2}}(u)g(u), K^{-\frac{1}{2}}(u)w\rangle\right\}\\ &\qquad \times \exp\left\{-\tfrac{1}{2}\langle w, (K^{-1}(u) - \mathcal{C}^{-1})w\rangle\right\} \cdot |\,\mathcal{C}^{1/2}K(u)^{-1/2}\,|\ .\end{aligned}$$

Proof. The first Radon–Nikodym derivative in the expression for $\lambda(w,u)$ is an application of Theorem 2.21 of [17] on translations of Gaussian measures. The second density is a formal expression of the ratio of two Gaussian measures. □

Remark 3.6. Note that due to the Hilbert–Schmidt property in Assumption 3.2, the term

$$\langle w, (K^{-1}(u) - \mathcal{C}^{-1})w\rangle - \log|\mathcal{C}\,K(u)^{-1}| \tag{23}$$

appearing in the expression for $\lambda(w,u)$ in Theorem 3.5 is a.s. finite under $w\sim\mu_0$ (μ_0-a.s. in $u\sim\mu_0$) as expected (since we assume existence of a density). For instance, the second moment of (23) is equal to (we use the standard representation on $\mathbb{R}^n$ by projecting onto the first n basis functions in $\{\phi_i\}$; we also denote by $\{\nu_{j,n}\}_{j=1}^n$ the eigenvalues of the projection $\{\mathcal{C}^{-1/2}K(u)^{1/2}\}\{\mathcal{C}^{-1/2}K(u)^{1/2}\}^*$ on $\mathbb{R}^{n\times n}$):

$$a_n := \Big\{\sum_{j=1}^n \big(\log\nu_{j,n} + \nu_{n,j}^{-1} - 1\big)\Big\}^2 + 2\sum_{j=1}^n(\nu_{n,j}^{-1}-1)^2$$

From the Hilbert–Schmidt assumption we have that $\sup_n\sum_{j=1}^n(1-\nu_{j,n})^2<\infty$, thus also $C_1\le\sum_{j,n}\nu_{j,n}\le C_2$, for constants $C_1,C_2>0$. Since $0\le(\log\nu_{j,n}+\nu_{n,j}^{-1}-1)\le C\,(1-\nu_{j,n})^2$ for some constant $C>0$, we have that $\sup_n a_n<\infty$.

Let $Q(u,du')$ being the proposal kernel derived from (22); we also consider the bivariate measure $\nu(du,du') = \mu(du)Q(u,du')$. Recall from (2)–(4) that obtaining the Metropolis-Hastings accept/reject rule requires finding the Radon–Nikodym derivative $d\nu^\top/d\nu$. Similarly to the derivation of ∞-MALA in Section 2 we consider now the bivariate Gaussian law $\nu_0(du,du') = \mu_0(du)Q_0(u,du')$ with $Q_0(u,du')$ as in (8). Recall we have the symmetry property $\nu_0\equiv\nu_0^\top$. Applying Theorem 3.5 we have:

$$\frac{d\nu}{d\nu_0}(u,u') = \frac{d\mu}{d\mu_0}(u)\cdot\frac{dQ(u,\cdot)}{dQ_0(u,\cdot)}(u') = \frac{1}{Z}\exp\{-\Phi(u)\}\cdot\lambda(\tfrac{u'-\rho u}{\sqrt{1-\rho^2}};u)\,. \tag{24}$$

We obtain the required density as $(d\nu^\top/d\nu) = [\,d\nu^\top/d\nu_0^\top\,]/[\,d\nu/d\nu_0\,]$. We can now define the complete method, labeled ∞-mMALA in Algorithm 3.7, (the small 'm' in the name stands for 'manifold').

Algorithm 3.7 A single Markov step for ∞-mMALA.

1. *Given current u, sample independently $\xi\sim\mathcal{N}(0,K(u))$ and propose:*

$$u' = \rho u + \sqrt{1-\rho^2}\big\{\xi + \tfrac{\sqrt{h}}{2}g(u)\big\}\,.$$

2. *Accept u' with probability $a(u,u') = 1\wedge\frac{\kappa(u',u)}{\kappa(u,u')}$, where we have set:*

$$\kappa(u,u') = \frac{1}{Z}\exp\{-\Phi(u)\}\times\lambda(\tfrac{u'-\rho u}{\sqrt{1-\rho^2}};u)$$

otherwise stay at u.

Remark 3.8. When $K(u)\equiv\mathcal{C}$, algorithms ∞-MALA and ∞-mMALA coincide.

In the following we let $\mathsf{H}(u)$ denote the posterior Hessian, computed from the negative log posterior:

$$\mathsf{H}(u) := \mathcal{C}^{-1} + D^2\Phi(u)\,;$$

since this is not necessarily positive-definite it is also of interest to consider a modification in which the non-positive and small eigenvalues are all shifted above a threshold, as in [9], and we use the same notation $\mathsf{H}(u)$ for this modification in order not to clutter notation. The following corollary connects our methodology with the Stochastic Newton (SN) MCMC method from [9]. We also recall that the paper [13] considered variants on this method where $\mathsf{H}(\cdot)$ is evaluated at the MAP point, and low rank approximations are employed.

Corollary 3.9. *When $\rho=0$ ($h=4$), ∞-mMALA coincides with the SN MCMC method.*

Proof. When $\rho=0$, we have $h=4$ from (7). The proposal (22) of ∞-mMALA becomes:

$$u'\sim\mathcal{N}(g(u),K(u))\,,\quad g(u) = u - K(u)(\mathcal{C}^{-1}u + D\Phi(u))\,,$$
$$K(u) = \mathsf{H}(u)^{-1} \tag{25}$$

which is exactly the proposal for the SN MCMC method defined in Section 2.3 of [9]. □

3.3. ∞-mHMC

Following the same direction as with ∞-mMALA, we now begin from the continuous-time Hamiltonian dynamics in (9), with a location-specific mass matrix:

$$\frac{d^2u}{dt^2} + K(u)\left\{\mathcal{C}^{-1}u + D\Phi(u)\right\} = 0\,. \tag{26}$$

For a splitting scheme driven by the local Gaussian reference measure $\tilde{\mu}_0$, we re-write the above dynamics as:

$$\frac{d^2u}{dt^2} + K(u)\left\{K(u)^{-1}(u - m(u)) + D\widetilde{\Phi}(u;u)\right\} = 0\,. \tag{27}$$

As with ∞-mMALA, $m(u)$ cancels out. Setting $du/dt = v$, we make use of the following splitting scheme:

$$du/dt = v\,, \quad dv/dt = -u\,; \tag{28}$$

$$du/dt = 0\,, \quad dv/dt = -K(u)\left\{(\mathcal{C}^{-1} - K^{-1}(u))\,u + D\Phi(u)\right\}. \tag{29}$$

Both (28), (29) can be solved analytically, the first by applying a rotation. Thus, we obtain the following approximate symplectic integrator of (26), for g as defined in (21):

$$\begin{aligned} v^- &= v_0 + \tfrac{\varepsilon}{2}\,g(u_0)\,; \\ \begin{bmatrix} u_\varepsilon \\ v^+ \end{bmatrix} &= \begin{bmatrix} \cos\varepsilon & \sin\varepsilon \\ -\sin\varepsilon & \cos\varepsilon \end{bmatrix} \begin{bmatrix} u_0 \\ v^- \end{bmatrix}; \\ v_\varepsilon &= v^+ + \tfrac{\varepsilon}{2}\,g(u_\varepsilon)\,. \end{aligned} \tag{30}$$

Equation (30) gives rise to the leapfrog map $\Psi_\varepsilon : (u_0, v_0) \mapsto (u_\varepsilon, v_\varepsilon)$. Given a time horizon τ and current position u, the MCMC mechanism proceeds by proposing:

$$u' = \mathcal{P}_u\left\{\Psi_\varepsilon^I(u,v)\right\}, \quad v \sim \mathcal{N}(0, K(u))\,,$$

for $I = \lfloor \tau/\varepsilon \rfloor$. Note that the dynamics in (26) do not preserve, in general, the target distribution μ (when initialized with $v \sim \mathcal{N}(0, K(u))$). Thus, there is no theoretical guarantee that the algorithm will give good acceptance probabilities for arbitrary time lengths τ with diminishing ε – an important property that characterizes non-local HMC algorithms. However, with properly chosen τ, ∞-mHMC, as a multi-step generalization of ∞-mMALA (see the similar discussion in Section 2 and the formal statement in Remark 3.13 below), is a valuable algorithm to be tested in applications, and in the numerical examples that follow it is indeed found in many cases to be superior than ∞-mMALA.

The following theorem is required for establishing the well-posedness of the developed algorithm. We define the probability measures on the phase-space:

$$S_0(du,dv) := \mathcal{N}(0,\mathcal{C})(du) \otimes \mathcal{N}(0,\mathcal{C})(dv)\,;$$

$$\tilde{S}_0(du,dv) := \mathcal{N}(0,\mathcal{C})(du) \otimes \mathcal{N}(\tfrac{\varepsilon}{2}g(u),\mathcal{C})(dv)\,;$$

$$S(du,dv) := \mu(du) \otimes \mathcal{N}(0,K(u))(dv)\,.$$

We also define the push-forward probability measures:

$$S^{(i)} := S \circ \Psi_\varepsilon^{-i}\,, \quad 1 \le i \le I\,.$$

For starting positions u_0, v_0, we set $(u_i, v_i) := \Psi_\varepsilon^i(u_0, v_0)$, $0 \le i \le I$.

Theorem 3.10.

(i) Under Assumptions 3.1–3.3, Theorem 3.5 implies the absolute continuity $S^{(i)} \simeq S_0$, for all indices $1 \le i \le I$, with Radon–Nikodym derivatives satisfying the recursion:

$$\frac{dS^{(i)}}{dS_0}(u_i,v_i) = \frac{dS^{(i-1)}}{dS_0}(u_{i-1},v_{i-1}) \cdot G(u_{i-1}, v_{i-1} + \tfrac{\varepsilon}{2}g(u_{i-1})) \cdot G(x_i,v_i)$$

where we have defined:

$$G(u,v) = \frac{d\tilde{S}_0}{dS_0}(u,v) = \exp\left\{-\tfrac{\varepsilon^2}{8}|\mathcal{C}^{-\frac{1}{2}}g(u)|^2 + \tfrac{\varepsilon}{2}\langle \mathcal{C}^{-\frac{1}{2}}g(u), \mathcal{C}^{-\frac{1}{2}}v\rangle\right\}.$$

(ii) From (i) we obtain that:

$$\frac{dS^{(I)}}{dS}(u_I, v_I) = \frac{(dS/dS_0)(u_0, v_0)}{(dS/dS_0)(u_I, v_I)} \times \prod_{i=1}^{I} G(u_{i-1}, v_{i-1} + \tfrac{\varepsilon}{2} g(u_{i-1})) \cdot G(u_i, v_i) \,.$$

We can re-write:

$$\log\left\{(dS^{(I)}/dS)(u_I, v_I)\right\} = \Delta H(u_0, v_0)$$

for the following quantity:

$$\begin{aligned}\Delta H(u_0, v_0) &= \Phi(u_I) - \Phi(u_0) + \tfrac{1}{2}\langle v_I, (K^{-1}(u_I) - \mathcal{C}^{-1}) v_I\rangle - \tfrac{1}{2}\langle v_0, (K^{-1}(u_0) - \mathcal{C}^{-1}) v_0\rangle \\ &\quad - \log|\mathcal{C}^{1/2} K^{-1/2}(u_I)| + \log|\mathcal{C}^{1/2} K^{-1/2}(u_0)| - \tfrac{\varepsilon^2}{8}\big(|\mathcal{C}^{-\frac{1}{2}} g(u_I)|^2 - |\mathcal{C}^{-\frac{1}{2}} g(u_0)|^2\big) \\ &\quad + \tfrac{\varepsilon}{2}\sum_{i=0}^{I-1}\big(\langle \mathcal{C}^{-1/2} g(u_i), \mathcal{C}^{-1/2} v_i\rangle + \langle \mathcal{C}^{-1/2} g(u_{i+1}), \mathcal{C}^{-1/2} v_{i+1}\rangle\big) \,.\end{aligned}$$

(iii) We have the identity:

$$\Delta H(u_0, v_0) \equiv H(u_I, v_I) - H(u_0, v_0)$$

for the energy function:

$$H(u, v) := \Phi(u) + \tfrac{1}{2}\langle u, \mathcal{C}^{-1} u\rangle + \tfrac{1}{2}\langle v, K(u)^{-1} v\rangle - \log|\mathcal{C}^{1/2} K(u)^{-1/2}| \,.$$

(iv) Given current position $u \in \mathbb{X}$, the Markov chain with proposed move:

$$u' = \mathcal{P}_u\left\{\Psi_\varepsilon^I(u, v)\right\} \,, \quad v \sim \mathcal{N}(0, K(u)) \,,$$

and acceptance probability:

$$a = 1 \wedge \exp\{-\Delta H(u, v)\}$$

preserves the target probability measure μ.

Proof. See Appendix A. □

We can now define the complete method, labeled ∞-mHMC, in Algorithm 3.11 below.

Algorithm 3.11 A single Markov step for ∞-mHMC.

1. *Given current u, sample independently $v \sim \mathcal{N}(0, K(u))$ and propose $u' = \mathcal{P}_u\left\{\Psi_\varepsilon^I(u, v)\right\}$.*
2. *Accept u' with probability $1 \wedge \exp\left\{-\Delta H(u, v)\right\}$ for the change of energy $\Delta H(u, v)$ defined in Theorem 3.10 (ii)–(iii), otherwise stay at u.*

Remark 3.12. When $K(u) \equiv \mathcal{C}$, algorithms ∞-HMC and ∞-mHMC coincide.

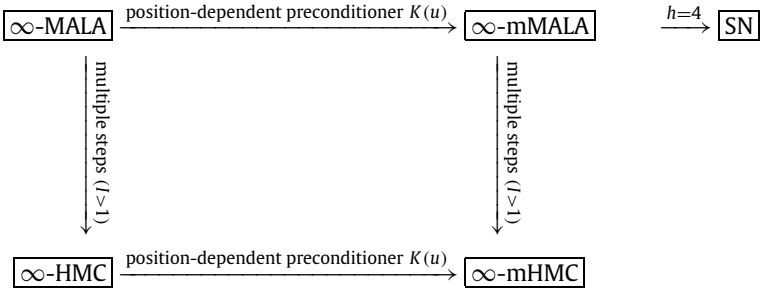
Corollary 3.13. *Assume that we allow for different step-sizes in the leapfrog scheme in (30): ε_1 in the first and third equation, and ε_2 in the second (the rotation). Recall the step-size h in the definition of ∞-mMALA. Then, if $I = 1$, and ε_1 and ε_2 are such that:*

$$\varepsilon_1^2 = h \,, \quad \cos\varepsilon_2 = \frac{1 - \varepsilon_1^2/4}{1 + \varepsilon_1^2/4} \,, \quad \sin\varepsilon_2 = \frac{\varepsilon_1}{1 + \varepsilon_1^2/4} \,, \tag{31}$$

algorithms ∞-mMALA and ∞-mHMC coincide.

Proof. See Appendix B. □

Remark 3.14. Following Corollary 3.9 and Corollary 3.13, the following plot illustrates graphically the connections between the various algorithms.



3.4. Split ∞-mMALA and ∞-mHMC

Following the discussion on optimal local Gaussian approximation in Subsection 3.2 or the metric tensor interpretation in [8], a typical choice of $K(u)^{-1}$ is the expectation over the data y given u of the negative Hessian of the log-target (this choice also guarantees positive-definiteness of $K(u)$), that is:

$$K(u)^{-1} = F(u) + \mathcal{C}^{-1}, \quad F(u) := \mathrm{E}_{y|u}[\, D\Phi(u;y) \otimes D\Phi(u;y)\,]\,. \tag{32}$$

Assuming a projection onto finite dimension $n \geq 1$, the operations of obtaining the operator $K(u)^{-1}$, applying it on a vector, inverting it or sampling from $\mathcal{N}(0, K(u))$ will typically have computational costs of order $\mathcal{O}(n^3)$ for each given current $u \in \mathbb{X}$. This can be prohibitively expensive when n is large, and could cause algorithms to be less efficient than simpler ones that use a constant mass matrix, when compared according to cost per independent sample. However, in a large class of inverse problem applications, the typical wave-length of the eigenfunctions of the covariance $\mathcal{C}$ decays as the eigenvalues decay (consider for example the periodic setting where $\mathcal{C}$ is an inverse fractional power of the Laplacian operator Δ). As a consequence, for typical observations which inform low frequencies, the information from the data spreads non-uniformly with respect to the coordinates $\{u_i\}$ of the unknown function parameter u, with most of it concentrating on the low-frequency coordinates. We will take advantage of this setting to recommend an effective choice of preconditioner $K(u)$.

Recall the orthonormal basis $\{\phi_j\}$ of $\mathbb{X}$ consisting of eigenfunctions of $\mathcal{C}$, and the isomorphism $\mathbb{X} \leftrightarrow \ell^2$ mapping $u \leftrightarrow \{u_j\}$ with $u = \sum_{j\geq 1} u_j\phi_j = \sum_{j\geq 1}\langle u, \phi_j\rangle\phi_j$. For a cut-off point $D_0 \geq 1$, we write $u = (u^t, u^r)$ with $u^t := u_{1:D_0}$ and residual part $u^r := u_{(D_0+1):\infty}$. We define the truncation operator T mapping

$$u \mapsto (u^t, 0, 0, \ldots) \tag{33}$$

with domain $\mathbb{X}^{-\ell}$. Balancing computational considerations with mixing effectiveness of the proposal move within the Metropolis-Hastings framework, we recommend using the following truncated Fisher information operator:

$$\tilde{F}(u) = \mathbb{E}_{y|u}[\, T\{D\Phi(u;y)\} \otimes T\{D\Phi(u;y)\}\,]\,. \tag{34}$$

Thus, we recommend the following choice:

$$K^{-1}(u) := \tilde{F}(u) + \mathcal{C}^{-1}\,. \tag{35}$$

Given that $\{\phi_j\}$ corresponds to the eigenfunctions of $\mathcal{C}$, operator $K(u)$ in (35) trivially satisfies Assumptions 3.2–3.3, as $\tilde{F}(u)$ only has a finite-size upper diagonal block of non-zero entries. Indeed, we can write:

$$K(u)^{-1} = \begin{pmatrix} \{K(u)^t\}^{-1} & 0 \\ 0 & \{K(u)^r\}^{-1} \end{pmatrix} = \begin{pmatrix} \tilde{F}(u)^t + \mathcal{C}^t & 0 \\ 0 & \mathcal{C}^r \end{pmatrix}\,,$$

with the truncations on the operators defined in the obvious way.

We label as *Split ∞-mMALA* and *Split ∞-mHMC* the corresponding MCMC methods resulting from the above choice of location specific preconditioner. The calculation of all required algorithmic quantities is now simplified, due to $K(u)$ being diagonal except for a finite-range of values. Indeed, in the case for instance of Split ∞-mMALA, the proposal may be written as:

$$(u^t, u^r)' = \rho\,(u^t, u^r) + \sqrt{1-\rho^2}\,\big\{(\xi^t, \xi^r) + \tfrac{\sqrt{h}}{2}(g(u)^t, g(u)^r)\big\}$$

where we have:

$$\xi^t \sim \mathcal{N}\big(0, K(u)^t\big)\,, \quad \xi^r \sim \mathcal{N}(0, \mathcal{C}^r)\,,$$
$$g(u)^t = -K(u)^t\,\big\{-\tilde{F}(u)^t\,u^t + D\Phi(u)^t\big\}\,, \quad g(u)^r = -\mathcal{C}^r D\Phi(u)^r\,.$$

Remark 3.15. Splitting the proposal into a likelihood-informed and a simpler step bears similarities with the 'intrinsic subspace' method in [7]. We stress however that our methodology develops geometric algorithms, in the sense that it employs location-specific curvature information. The development of the geometric methods in a general setting in the earlier sections (beyond the truncation we recommend here) is still necessary for mathematical rigorousness, and more importantly, for the numerical robustness to possibly high-dimensional 'intrinsic subspaces'. As previously discussed, the straightforward splitting implemented here works fairly well on a class of inverse problems we consider in Section 4. We should mention here that in a context where the data in the inverse problem possess such strong information that a faithful representation of u would require a large set of high-frequency coordinates, then more sophisticated likelihood-informed splitting methods, e.g. [11,12], could potentially be considered to help derive low-dimensional 'intrinsic subspaces'.

4. Numerical experiments

Our experiments involve simulation studies based on three physical inverse problems. The prior is in each case Gaussian on a Hilbert space $\mathbb{X}$. In this section we consider three inverse problems – the groundwater flow, the thermal fin heat conductivity and the laminar jet. The first two examples are implemented in MATLAB (r2015b) and the last one is implemented in FEniCS [25,26]. All computer codes are available at https://bitbucket.org/lanzithinking/geom-infmcmc. The necessary adjoint and tangent linearized versions of this solver are derived with the dolfin-adjoint package [27].

4.1. Prior specification

We will consider Hilbert spaces $\mathbb{X} \subseteq L^2(\mathcal{D};\mathbb{R})$, the latter denoting the space of real-valued squared-integrable functions on bounded open domains $\mathcal{D} \subset \mathbb{R}^d$, $d \geq 1$. We denote by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ the inner product and norm, respectively, of $L^2(\mathcal{D};\mathbb{R})$. In all of our examples we will construct the Karhunen–Loève (K-L) expansion (14) through eigenfunctions of the Laplacian. Specifically, we choose covariance operators on $\mathbb{X}$ of the form:

$$\sigma^2(\alpha\,\mathrm{I} - \Delta)^{-s} \tag{36}$$

for scale parameters $\alpha, \sigma^2 > 0$, 'smoothness' parameter $s \in \mathbb{R}$ and the Laplacian $\Delta = \sum_{j=1}^d \partial_j^2$.

In the first two numerical examples we have $d = 2$, and rectangular domain $\mathcal{D}$ of the form $[k_1, k_2] \times [l_1, l_2]$ for integers k_1, k_2, l_1, l_2. In this case, we will work with the orthonormal basis:

$$\phi_i(\mathbf{x}) = 2|\mathcal{D}|^{-1/2} \cos\left\{\pi\left(i_1 + \tfrac{1}{2}\right)x_1\right\} \cos\left\{\pi\left(i_2 + \tfrac{1}{2}\right)x_2\right\}, \quad i_1 \geq 0\,,\ i_2 \geq 0\,. \tag{37}$$

Thus, the Hilbert space will be (we set $I = \{i = (i_1, i_2) : i_1 \geq 0, i_2 \geq 0\}$):

$$\mathbb{X} = \mathrm{span}\{\phi_i;\, i \in I\} \equiv \left\{u \in L^2(\mathcal{D};\mathbb{R}) : u = \sum_{i\in I} u_i\phi_i\,,\ \sum_{i\in I} u_i^2 < \infty\right\}.$$

Guided by (36), we set the covariance operator $\mathcal{C}$ as:

$$\mathcal{C} = \sum_{i\in I} \lambda_i^2 \{\phi_i \otimes \phi_i\}\,; \quad \lambda_i^2 = \sigma^2\left\{\alpha + \pi^2\left(\left(i_1 + \tfrac{1}{2}\right)^2 + \left(i_2 + \tfrac{1}{2}\right)^2\right)\right\}^{-s}. \tag{38}$$

For $\mathcal{C}$ to be trace-class we require that $\sum_{i\in I} \lambda_i^2 < \infty$, that is $s > 1$.

In the third example we will have $d = 1$, $\mathcal{D} = [-1, 1]$ and use a prior covariance with the following orthonormal eigenfunctions and eigenvalues:

$$\phi_i(x) = \left(\tfrac{1}{\sqrt{2}}\right)^{\delta[\,i=0\,]} \cos(\pi i x)\,, \quad \lambda_i^2 = 2^{\,\delta[\,i=0\,]}\,\sigma^2\{\alpha + (\pi i)^2\}^{-s}\,, \quad i \geq 0\,, \tag{39}$$

where $\delta[\cdot]$ is the indicator of whether condition(s) in the square bracket being satisfied (1), or otherwise (0), and the trace-class property requires that $s > 1/2$.

For given orthonormal basis $\{\phi_i\,;\, i \in I\}$, we run MCMC algorithms to sample K-L coordinates $\{u_i := \langle u, \phi_i\rangle\,;\, i \in I_0 \subset I\}$ in the following experiments, for some chosen non-negative integer $|I_0|$. These coordinates can be viewed as projections of parameter function u onto K-L modes up to $|I_0|$. The splitting methods are implemented with Fisher operator truncated on the first D_0 of $|I_0|$ coordinates. The gradient $D\Phi(u)$ is obtained by one adjoint solver in addition to the forward solution to the relevant PDE; the metric action $\tilde{F}(u)\,v$ is obtained by another two extra adjoint (incremental) solvers for each $v \in \mathbb{X}$ [28].

4.2. Groundwater flow

We consider a canonical inverse problem involving the following elliptic PDE [29,30] defined on the unit square $\mathcal{D} = [0, 1]^2$:

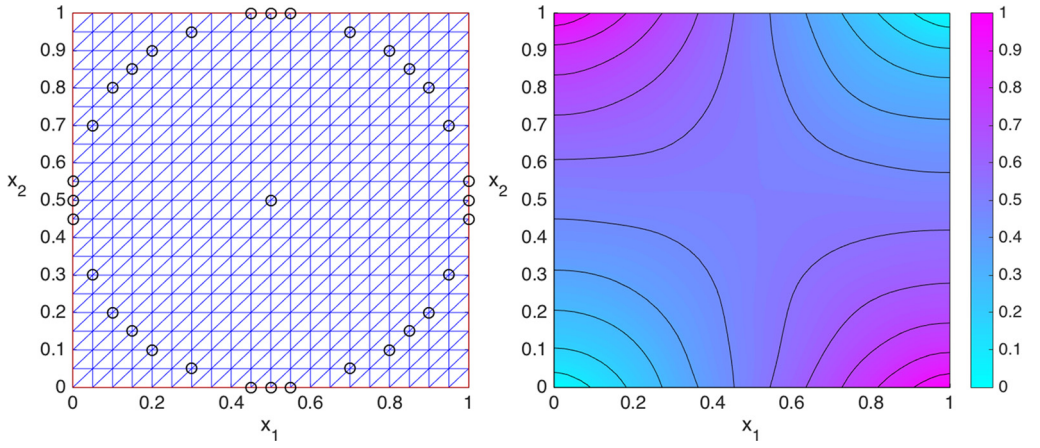


Fig. 1. Groundwater flow problem: the location of noisy observations (circles) on $[0,1]\times[0.1]$ (left) and the forward PDE solution under the true permeability $u^\dagger$ (right).

$$
\begin{aligned}
-\nabla\cdot(e^{u(\mathbf{x})}\nabla p(\mathbf{x})) &= 0\,;\\
p(\mathbf{x})|_{x_2=0} &= x_1\,;\\
p(\mathbf{x})|_{x_2=1} &= 1-x_1\,;\\
\left.\frac{\partial p(\mathbf{x})}{\partial x_1}\right|_{x_1=0} &= \left.\frac{\partial p(\mathbf{x})}{\partial x_1}\right|_{x_1=1} = 0\,.
\end{aligned}
\tag{40}
$$

This PDE serves as a simple model of steady-state flow in aquifers and other subsurface systems. The unknown parameter u represents the logarithm of permeability of the porous medium and p represents the hydraulic head function. The inverse problem involves inferring the log-permeability field $u=u(\mathbf{x})$ based on noisy observations, y, of $p=p(\mathbf{x})$.

We consider a Gaussian prior on $\mathbb{X}\subset L^2(\mathcal{D};\mathbb{R})$ with covariance $\mathcal{C}$ of eigen-structure $(\lambda_j^2,\phi_j)_{j\in I}$, as explained in Subsection 4.1. We pick hyper-parameter values $\alpha=0$, $s=1.1$, $\sigma^2=1$. To generate the data, we choose the true log-permeability field $u^\dagger$ via its coordinates $u_i^\dagger=\lambda_i^{1/2}\sin\left((i_1-1/2)^2+(i_2-1/2)^2\right)\cdot\delta\,[\,1\le i_1,i_2\le 10\,]$. In this setting, we solve the forward equation (40) on a 40×40 mesh and add Gaussian noise to 33 positions, $\mathbf{x}_n$, $1\le n\le 33$, of the true hydraulic head function $p^\dagger$ located on a circle and shown on the left panel of Fig. 1. In particular, we simulate data as:

$$y_n=p^\dagger(\mathbf{x}_n)+\varepsilon_n\,,\quad \varepsilon_n\sim\mathcal{N}(0,\sigma_y^2)\,,$$

with $\sigma_y^2=10^{-4}$. When running the MCMC algorithms, the posterior is approximated by projecting the coordinates on $I_0=\{i\in I: i_1\le 10,\ i_2\le 10\}$ and applying the PDE solver on a 20×20 mesh.

We run the MCMC algorithms: pCN, ∞-MALA, ∞-HMC, ∞-mMALA, ∞-mHMC, Split ∞-mMALA and Split ∞-mHMC. For the split methods we truncate at $i_1,i_2\le 5$ based on threshing the eigenvalues $\{\lambda_i^2\}$ of $\mathcal{C}$. Therefore we have $|I_0|=100$ and $D_0=25$ for this example. Each algorithm is run for 1.1×10^4 iterations, with the first 10^3-samples used for burn-in. HMC algorithms use a number of leapfrog steps chosen at random between 1 and 4. All steps-sizes were tuned to obtain acceptance probabilities of about 60%–70%.

Fig. 2 illustrates the posterior mean estimates of the permeability of the porous medium provided by the various algorithms. The estimates by pCN and ∞-MALA differ from the rest due to the bad convergence properties of these algorithms. Fig. 3 shows the traceplots and corresponding autocorrelation functions for the negative log-likelihood $\Phi(u)$ (or 'data-misfit') evaluated at the sample values; the various traces are vertically offset to allow for comparisons.

Table 1 compares the sampling efficiency of the various algorithms. Once more information is introduced (gradient or/and location-specific scales in the geometric methods) the mixing of the algorithms improves drastically. Even when the increased computational cost is taken under consideration, the overall effectiveness of Split ∞-mHMC, as measured by the minimal effective sample size (ESS) per CPU time (in secs), points to close to 4-fold improvement compared to pCN. In this example, the non-geometric methods ∞-MALA, ∞-HMC perform worse than pCN due to insufficient mixing improvement when weighted against the extra computations. The same holds for ∞-mHMC, clearly motivating in this case the significance of the truncation technique for reducing computational costs within Split ∞-mHMC.

Fig. 4 shows the first few data-misfit evaluations at the beginning of the algorithms. PCN exhibits strong diffusive behavior. The lower panel, where the horizontal axis corresponds to execution time, seems to indicate that maybe the various methods are not dramatically better than pCN in this case. Still, as mentioned above, the optimal speed-up against pCN is by a factor of 4. In the two subsequent, more complex, examples the speed-up factor will be much larger. Splitting methods with truncation number different from $D_0=25$ are also implemented: smaller D_0 causes the truncated Fisher operator

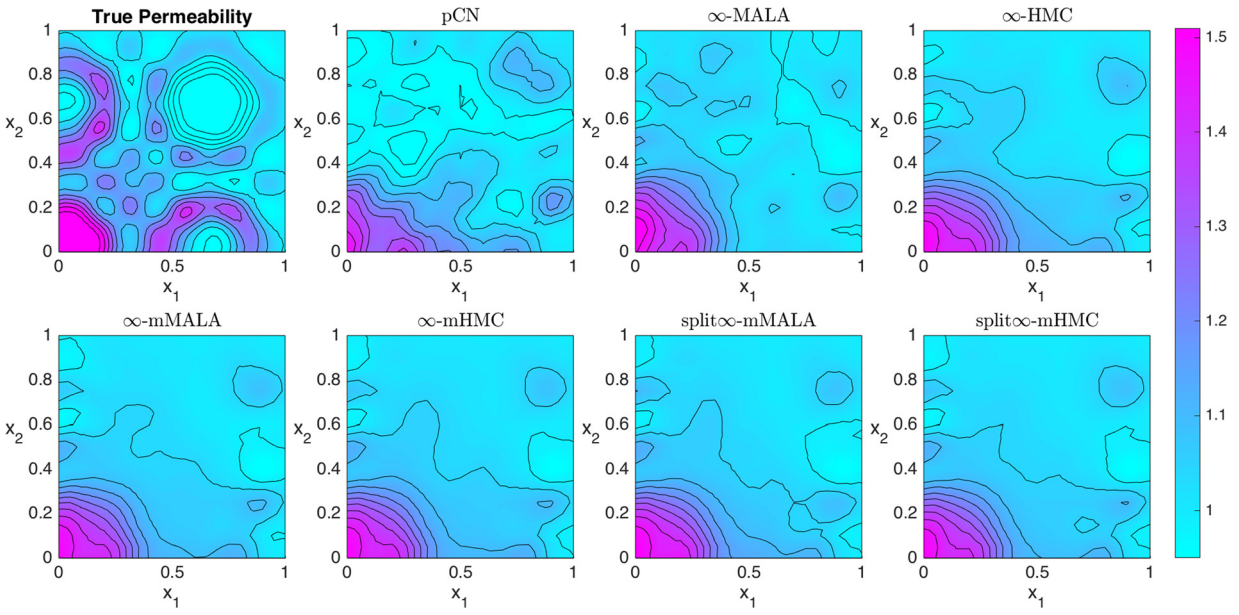


Fig. 2. Groundwater flow problem: the true permeability field $e^{u^\dagger}$ (upper left-most) and the posterior mean estimates provided by the various MCMC algorithms.

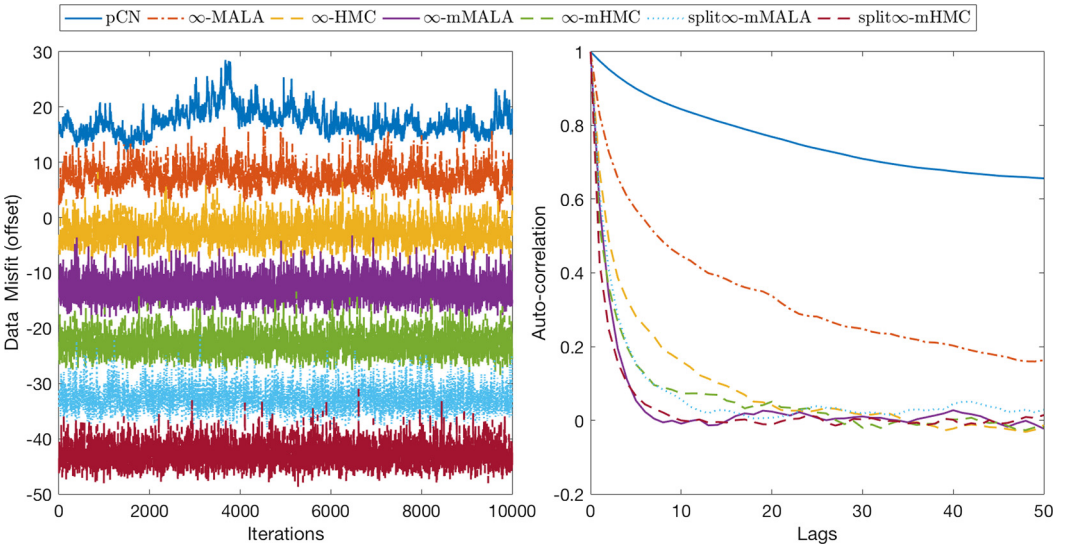


Fig. 3. Groundwater flow problem: the trace plots of data-misfit function (left panel, values have been offset for better comparison) and the corresponding acf plots (right panel).

Table 1
Sampling efficiency in the groundwater flow problem. Column labels are as follows. AP: average acceptance probability; s/iter: average seconds per iteration; ESS(min,med,max): minimum, median, maximum of Effective Sample Size across all posterior coordinates; min(ESS)/s: minimum ESS per second; spdup: speed-up relative to base pCN algorithm; PDEsolns: number of PDE solutions during execution.

Method	AP	s/iter	ESS(min,med,max)	minESS/s	spdup	PDEsolns
pCN	0.69	4.86E−03	(5.72, 17.23, 52.6)	0.118	1.00	11001
∞-MALA	0.71	2.23E−01	(27.15, 58.44, 138.93)	0.012	0.10	22002
∞-HMC	0.77	5.62E−01	(302.37, 461.03, 590.36)	0.054	0.46	54822
∞-mMALA	0.75	8.09E−01	(1422.11, 1747.68, 2051.5)	0.176	1.49	2222202
∞-mHMC	0.62	1.99E+00	(2514.45, 3667.88, 4438.35)	0.126	1.07	5562070
Split ∞-mMALA	0.67	3.20E−01	(654.22, 1078.15, 1283.37)	0.205	1.74	572052
Split ∞-mHMC	0.67	8.02E−01	(3641.2, 5230.48, 5746.96)	0.454	3.85	1434940

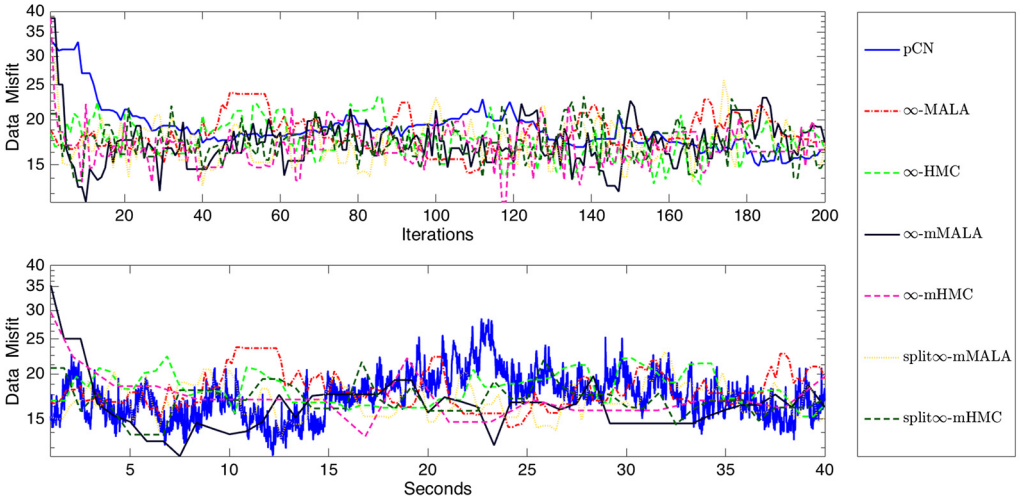


Fig. 4. Groundwater flow problem: trace plots of data-misfits before burn-in for the first 200 iterations (upper panel) and first 40 seconds (lower panel) respectively.

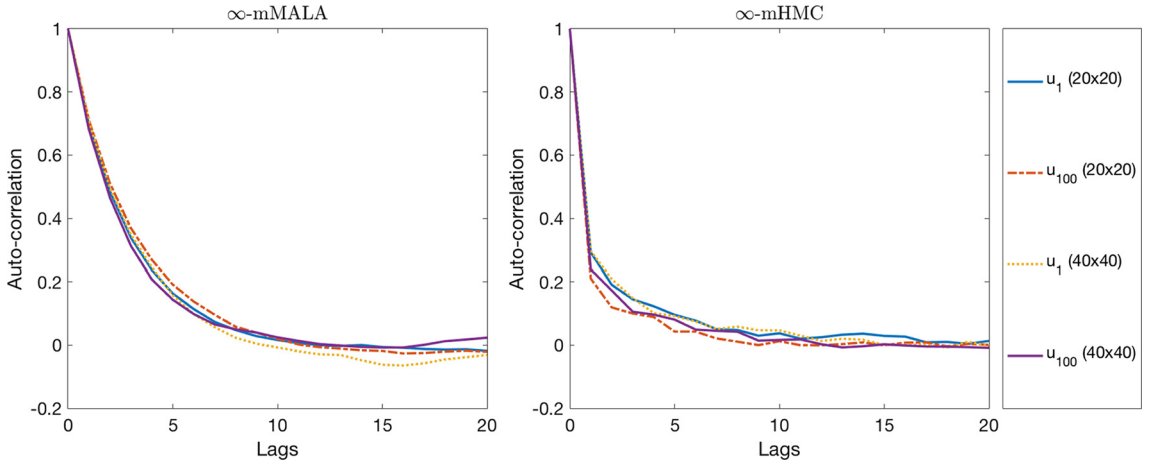


Fig. 5. Groundwater flow problem: auto-correlation function of selected samples (u_1, u_{25}, u_{100}) generated by ∞-mMALA (left) and ∞-mHMC (right) with forward PDE solving carried on 20×20 mesh and 40×40 mesh.

to lose useful information while larger D_0 negatively impacts the computational advantage. One can refer to Fig. 6 for illustration. Other results are omitted for brevity of exposition.

To verify mesh-independence of the proposed methods, we re-do the above inference with forward PDE solved on a refined, 40×40 mesh. Since the mesh-independence of non-geometric methods has been established in the literature [1, 3,2], and split algorithms are special cases of their full versions, we only compare the performance of ∞-mMALA (and ∞-mHMC) with PDE solved on 20×20 mesh and 40×40 mesh. For ∞-mMALA, the two implementations share the same acceptance rate 75% and their effective sample sizes (minimum, median, maximum) are $(1422.11, 1747.68, 2051.5)$ and $(1263.78, 1757.22, 2056.68)$ respectively. For ∞-mHMC, the two implementations have similar acceptance rates (62% and 61% respectively), and effective sample sizes $(2514.45, 3667.88, 4438.35)$ and $(2311.26, 3469.34, 4469.44)$ respectively. Fig. 5 illustrates that for both ∞-mMALA and ∞-mHMC, the auto-correlation functions of selected samples decay with lag but do not deteriorate under mesh refinement. This fact means that the number of MCMC steps to reach equilibrium is independent of the mesh [31]. Fig. 6 shows the close posterior mean estimates of the permeability field by ∞-mHMC with PDE solved on those two meshes (Similar result exists for ∞-mMALA but is omitted), which also qualitatively confirms the mesh-independence of ∞-mMALA and ∞-mHMC. The column wise comparison of estimates using different number of modes indicates that most posterior information is concentrated in the subspace formed by the first 25 eigen-directions.

4.3. Thermal fin

We now consider the following thermal fin model:

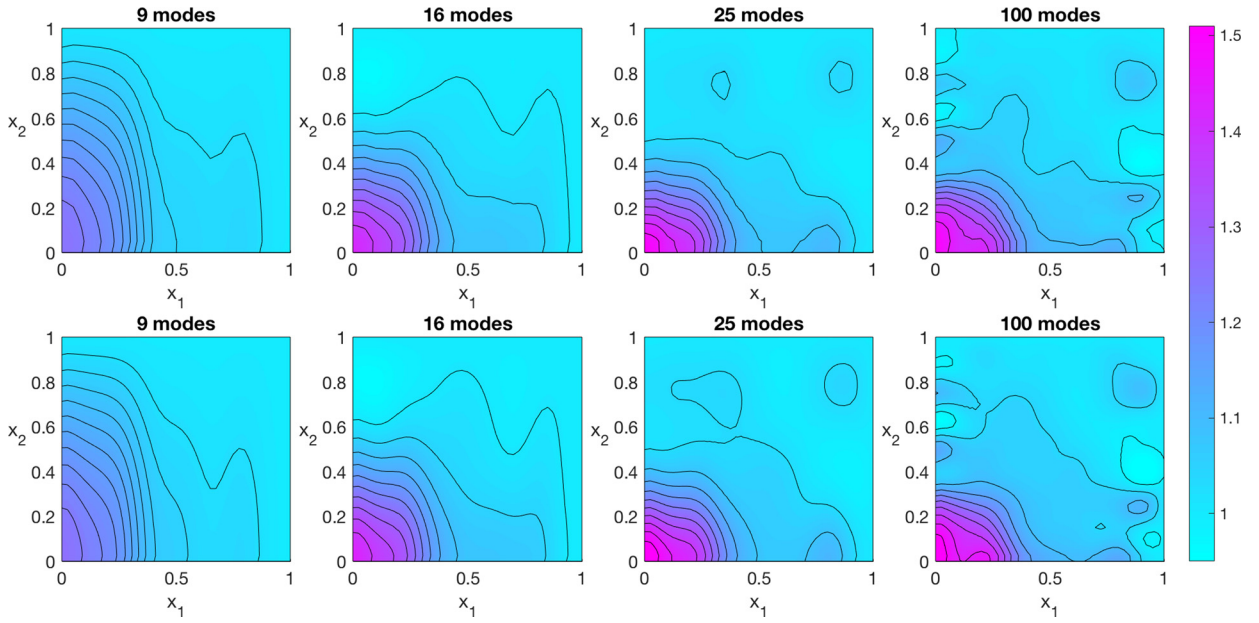


Fig. 6. Groundwater flow problem: estimated permeability field e^u using samples by ∞-mHMC with forward PDE solving carried on 20×20 mesh (upper row) and on 40×40 mesh (lower row). Each column corresponds to estimates with different number of modes (components of $\{u_i\}$).

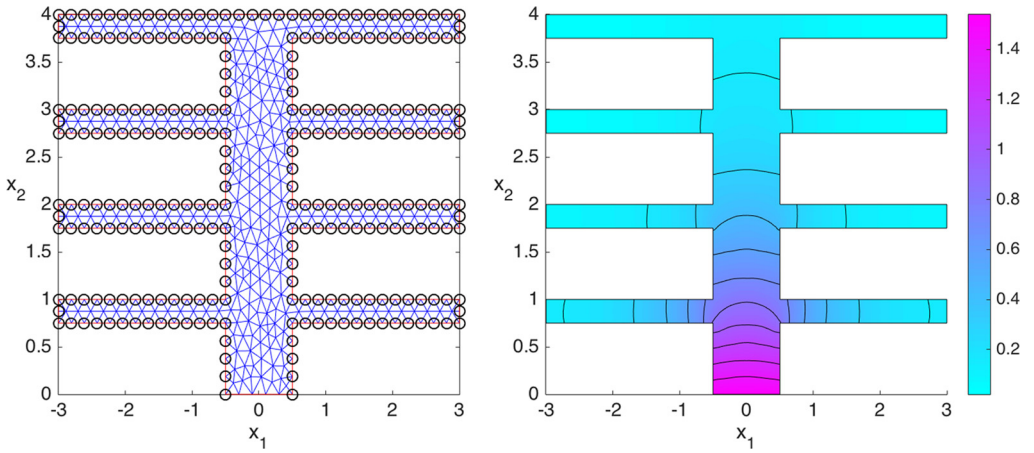


Fig. 7. Thermal fin problem: the location of observations (circles) (left panel) and the forward PDE solution $p^\dagger$ under the true parameter $u^\dagger$ (right panel).

$$
\begin{aligned}
-\nabla \cdot (e^{u(\mathbf{x})} \nabla p(\mathbf{x})) &= 0\,, && \mathbf{x} \in \mathcal{E}^0 = \text{Interior}(\mathcal{E})\,,\\
(e^{u(\mathbf{x})} \nabla p(\mathbf{x})) \cdot \mathbf{n} &= -Bi \cdot p(\mathbf{x})\,, && \mathbf{x} \in \partial\mathcal{E} \backslash \Gamma\,,\\
(e^{u(\mathbf{x})} \nabla p(\mathbf{x})) \cdot \mathbf{n} &= 1\,, && \mathbf{x} \in \Gamma = [-0.5, 0.5] \times \{0\}\,.
\end{aligned} \tag{41}
$$

These equations model the heat conduction over the non-convex domain $\mathcal{E}$ depicted in Fig. 7, where $\Gamma = [-0.5, 0.5] \times \{0\}$ is a part of the boundary $\partial\mathcal{E}$ on which the inflow heat flux is 1. For the rest of the boundary we assume Robin boundary conditions. Following [32], we set the Biot number to $Bi = 0.1$. The forward problem (41) provides the temperature p given the heat conductivity function e^u and the inverse problem involves reconstructing u from noisy observations of p. The complexity of the model domain makes this inverse problem more challenging than the previous groundwater flow problem.

The prior for u is obtained as explained at Subsection 4.1, for domain $\mathcal{D} = [-3, 3] \times [0, 4]$. We have chosen a rectangular domain $\mathcal{D}$ for u which contains the domain $\mathcal{E}$ of the PDE as a convenient way to construct the prior. However it should be mentioned that such a construction may introduce non-physical correlations between the fins; priors which are geometry-adapted could be used but would be more complicated to implement and maybe go beyond the scope of this paper. In this example, we set $\alpha = 0$, $s = 1.2$, $\sigma^2 = 1$ in the specification of $\mathcal{C}$. The true log-conductivity field $u^\dagger$ has coordinates $u_i^\dagger = \lambda_i^{1/2} \sin((i_1 - \frac{1}{2})^2 + (i_2 - \frac{1}{2})^2) \cdot \delta\,[\,i_1 \leq 10,\, i_2 \leq 10\,]$ and the simulated data are obtained by solving (41) on a triangular

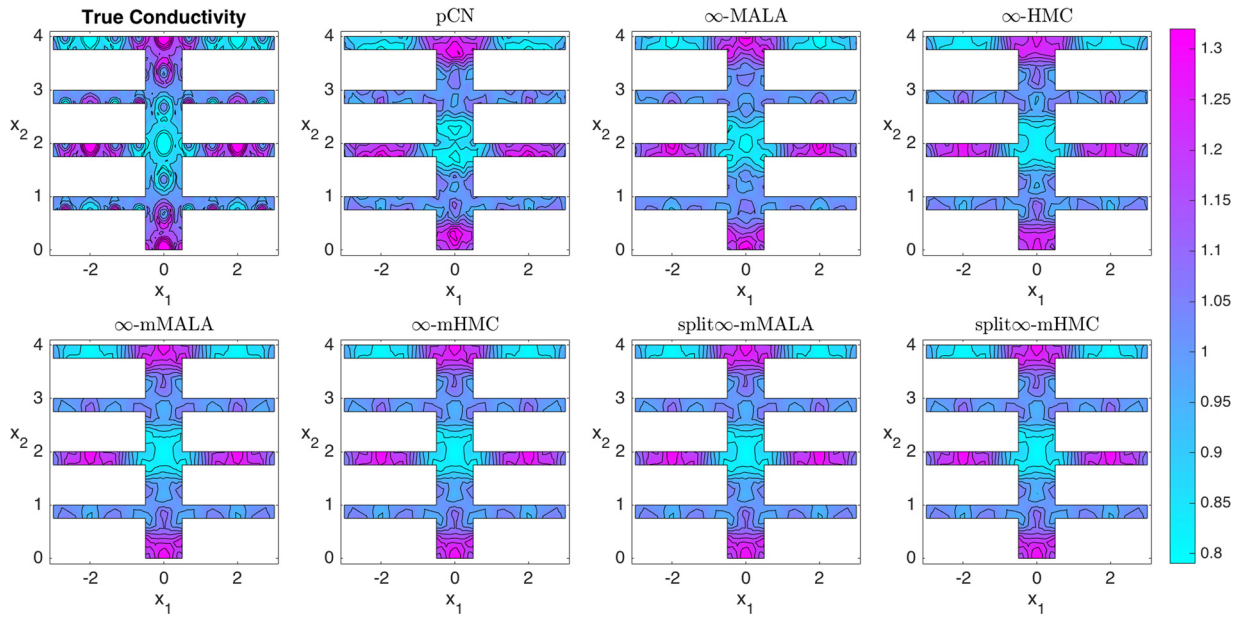


Fig. 8. Thermal fin problem: the true heat conductivity field $e^{u^\dagger}$ (upper left-most) and the posterior mean estimates obtained by the various MCMC algorithms.

Table 2
Sampling efficiency in the thermal fin problem. Column labels are as in Table 1.

Method	AP	s/iter	ESS(min, med, max)	minESS/s	spdup	PDEsolns
pCN	0.67	6.97E−03	(3.61, 8.67, 29.93)	0.052	1.00	11001
∞-MALA	0.70	9.60E−02	(5.52, 15.07, 33.91)	0.006	0.11	22002
∞-HMC	0.75	2.34E−01	(24.78, 81.13, 156.41)	0.011	0.20	55264
∞-mMALA	0.79	5.12E−01	(1729.28, 2224.8, 2474.28)	0.338	6.51	2222202
∞-mHMC	0.69	1.31E+00	(4018.07, 5679.26, 6956.14)	0.306	5.90	5582270
Split ∞-mMALA	0.77	1.53E−01	(1180.78, 1792.34, 2026.81)	0.770	14.87	572052
Split ∞-mHMC	0.72	3.85E−01	(5327.64, 7107.08, 8335.14)	1.384	26.70	1432704

mesh (left panel of Fig. 7) with discretization step-size $h_{max} = 0.1$. Then, $N = 262$ observations are taken along the Robin boundary $\partial\mathcal{E}\backslash\Gamma$ (we denote the positions of the observations $\{\mathbf{x}_n\}$, $1 \le n \le N$) and contaminated with Gaussian noise with mean zero and standard deviation $\sigma_y = 0.01 \cdot \max_{1\le n\le N}\{p^\dagger(\mathbf{x}_n)\}$, as in [32]. When running the MCMC algorithms, we project on the coordinates on $\{i_1, i_2 \le 10\}$, and use the same finite element construction as above. HMC algorithms use a number of leapfrog steps randomly chosen between 1 and 4. The split methods apply the geometric principle on $\{i_1, i_2 \le 5\}$. Thus similarly as the previous example, $|I_0| = 100$ and $D_0 = 25$.

In this example, there are ample data points (262) to provide enough information in inferring (100) unknown parameters, which is different from the previous example as an underdetermined elliptic inverse problem (inferring 100 unknown parameters from 33 data points) [22]. As shown in Fig. 8, the posterior mean estimates of heat conductivity are consistent across different algorithms and close to the truth. Due to having more informative data in this example, the posterior mean is closer to the truth than in the previous example (see Fig. 2). Table 2 and Fig. 9 compare the sampling efficiency of different algorithms. Notice that more than an order of magnitude of improvement is observed for Split ∞-mMALA and Split ∞-mHMC compared to pCN. In Fig. 10, pCN needs several iterations to reach the stationary stage. Notice that in this case also ∞-mHMC requires some time before reaching the stationary regime.

4.4. Laminar jet

We consider the 2D incompressible Navier–Stokes equation:

$$
\begin{aligned}
\text{Momentum}: \quad & -\operatorname{div}\left(\nu\left(\nabla\mathbf{u} + \nabla\mathbf{u}^\top\right)\right) + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p = 0\,, \\
\text{Continuity}: \quad & \operatorname{div}\mathbf{u} = 0\,, \\
& \mathbf{u}\cdot\mathbf{n} = -\theta(y)\,, \quad \boldsymbol{\sigma}_n \times \mathbf{n} = 0\,, \quad \text{on } \mathcal{I}\,, \\
& \boldsymbol{\sigma}_n + \beta\,(\mathbf{u}\cdot\mathbf{n})_-\,\mathbf{u} = \mathbf{0}\,, \quad \text{on } \mathcal{O}\,, \\
& \mathbf{u}\cdot\mathbf{n} = 0\,,\ \boldsymbol{\sigma}_n \times \mathbf{n} = 0\,, \quad \text{on } \mathcal{B}\,,
\end{aligned}
\tag{42}
$$

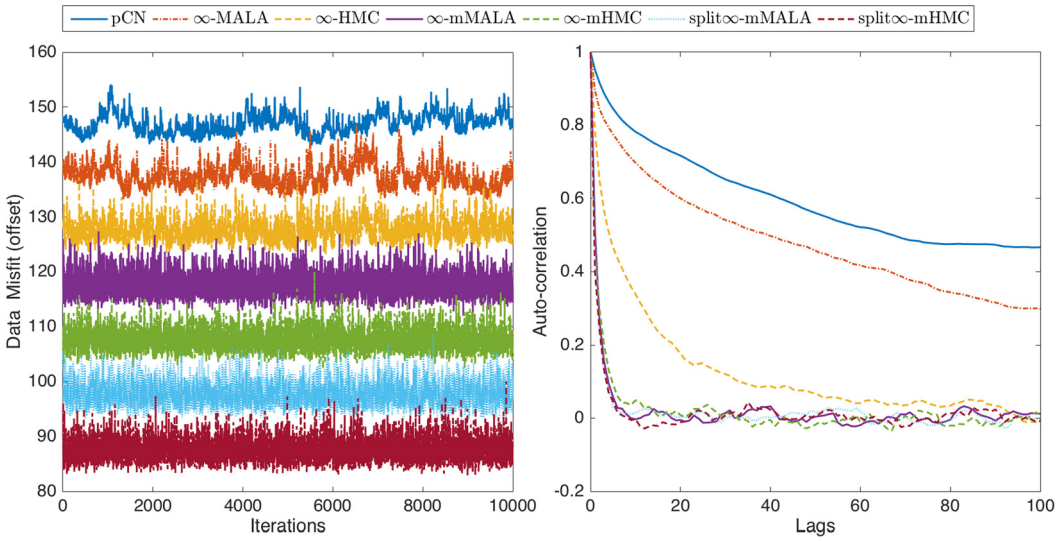


Fig. 9. Thermal fin problem: trace plots of data-misfit function (left panel, values have been offset for better comparison) and the corresponding acf functions (right panel).

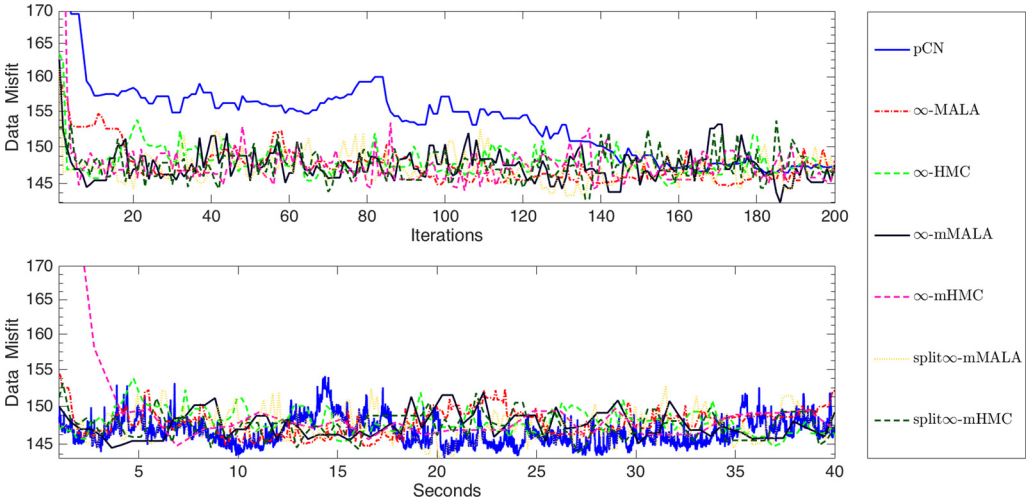


Fig. 10. Thermal fin problem: trace plots of data-misfits before burn-in for the first 200 iterations (upper) and the first 40 seconds (lower) respectively.

where $\mathbf{u} = (u_1, u_2)$ is the velocity, p is the pressure and $\nu > 0$ is the viscosity. Vector $\mathbf{n}$ denotes the unit normal to the mesh boundary and

$$\boldsymbol{\sigma}_n = -p\,\mathbf{n} + \nu\,(\nabla \mathbf{u} + \nabla \mathbf{u}^\top) \cdot \mathbf{n}$$

represents the boundary traction. Also,

$$(\mathbf{u} \cdot \mathbf{n})_- = (\mathbf{u} \cdot \mathbf{n} - |\mathbf{u} \cdot \mathbf{n}|)/2$$

and $\beta \in (0, 1]$ is the backflow stabilization parameter in [33]. This PDE models non-reacting turbulent jet dynamics. $\mathcal{I}, \mathcal{O}, \mathcal{B}$ denote the inlet, outlet and bounding sides respectively, to be described below.

We will describe a concrete simplified problem setting following [34]. The relevant domain $\mathcal{E}$ for the PDE is a rectangle with length $L_x = 10L$ and width $L_y = 8L$, with parameter L being a typical lengthscale of the (unknown) inlet velocity field; it is set to $L = 0.1$ in this experiment. The induced domain $\mathcal{E} = [0, 1] \times [-0.4, 0.4]$ is shown on the left panel of Fig. 11. We consider the following boundary conditions. At the inlet boundary $\mathcal{I} = \{x = 0,\ y \in (-L_y/2, L_y/2)\}$ we prescribe a normal velocity profile $\theta(y)$ and vanishing tangential stress. At the outflow boundary $\mathcal{O} = \{x = L_x\ , y \in (-L_y/2\ , L_y/2)\}$ we prescribe a traction-free condition plus an additional convective traction term to stabilize regions of possible backflow [33]. Finally, on the bounding sides $\mathcal{B} = \{x \in (0, L_x),\ y = \pm L_y/2\}$ we prescribe free-slip conditions. A typical solution is shown in the right panel of Fig. 11, where the heat map shows the pressure p and the arrows represent the velocity field $\mathbf{u}$. Note that

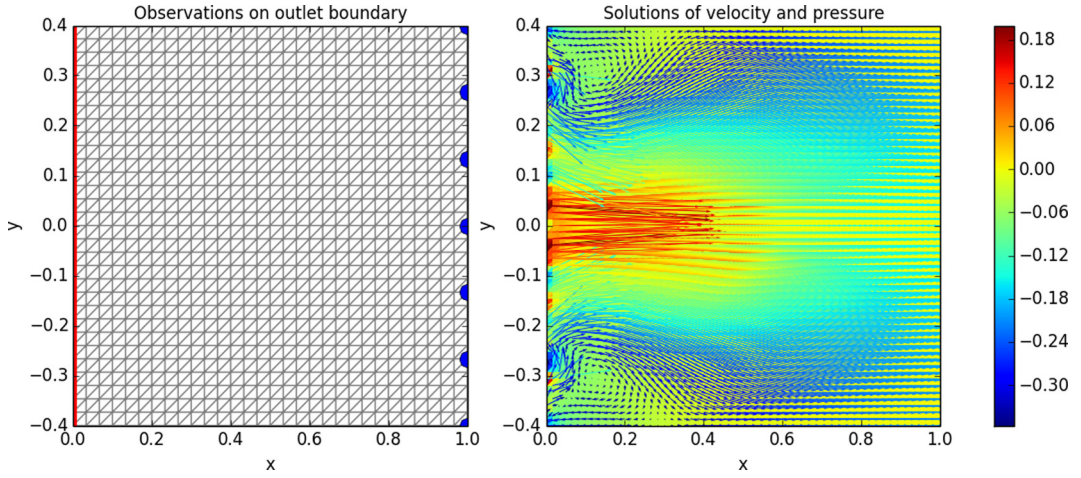


Fig. 11. Laminar jet problem: (left panel) the location of inlet velocity to be inferred (red line) and the measurement locations (blue dots); (right panel) the forward PDE solution with true unknown $\theta^\dagger$, with the heat map showing the pressure p and the arrows representing the velocity field $\mathbf{u}$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

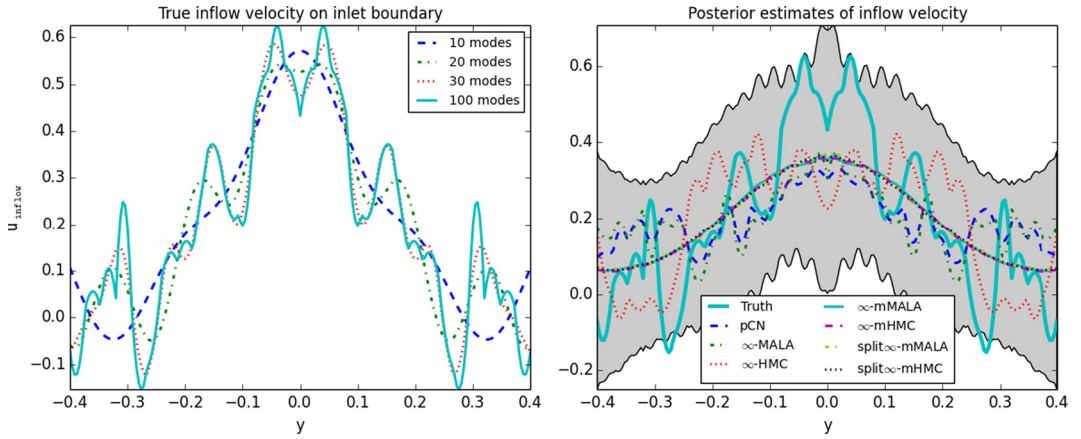


Fig. 12. Laminar jet problem: true inflow velocity for increasing number of frequencies (left panel); the true $\theta^\dagger$ used corresponded to the highest number shown, 100. Also, posterior mean estimates provided by the MCMC algorithms (right panel). Results of all geometric algorithms (with small 'm') agree with each other and the others (non-geometric methods) do not because they have not converged. The shaded region shows the 95% credible band constructed with samples from ∞-mHMC.

the color change along the inlet boundary reflects the persistence of high frequencies in the true inflow velocity profile (see also the left panel of Fig. 12).

Given an inflow velocity profile $\theta = \theta(y)$ on $\mathcal{I}$, the forward problem computes $\mathbf{u}(x, y)$, and $\varphi(y) = \mathbf{u}(L_x, y)$. The inverse problem aims to infer $\theta = \theta(y)$ given noisy observations of $\varphi(y)$ on the right boundary $\mathcal{O}$. We assume an 1D Gaussian prior on the super-domain $[-1, 1]$ as explained in Subsection 4.1. We choose hyper-parameters $\sigma = 0.5$, $\alpha = 1$ and $s = 0.8$. We obtain the true path $\theta^\dagger$ by sampling the coefficients $\theta_i^\dagger$, $1 \le i \le 100$, from the prior with $\theta_i^\dagger = 0$, $i > 100$. The true inflow velocity $\theta^\dagger$ on $[-0.4, 0.4]$ is shown at the left panel of Fig. 12. Note here negative values of $\theta(y)$ (around $y = \pm 0.4, \pm 0.3$) indicate backward flow, which also can be seen in the right panel of Fig. 11. We solve the Laminar equation for $\nu = 3 \times 10^{-2}$, $\beta = 0.3$ on a 60×60 mesh and obtain 7 observations from the velocity field at the locations indicated by blue dots on the left panel of Fig. 11, contaminated with Gaussian noise of variance $\sigma_{obs}^2 = 10^{-4}$. We stress here that this is a complex inverse problem due to the non-linearity of the forward PDE and the sparsity of observations. Each forward solution relies on an expensive Newton iteration with no clear theory about convergence of solutions when using different initializations. In this experiment, we choose the viscosity $\nu = 3 \times 10^{-2}$ as a compromise between reasonable convergence rate in the nonlinear solver, which favors larger ν, and obtaining interesting flow structure, which favors smaller ν. We also adopt the perspective of using a fixed initial position ($\theta_i = 0$ for all i here) for the Newton iteration every time the PDE dynamics are invoked, so that there is a well-defined map (on a given grid) from $\theta(y)$ to the likelihood of the observations. The required adjoints for gradient and metric-action (metric-vector product) are linear, and hence not too expensive to compute. The backflow stabilization term (in the 4th equation of (42)) involves taking the minimum of $\mathbf{u} \cdot \mathbf{n}$ with 0. This term is

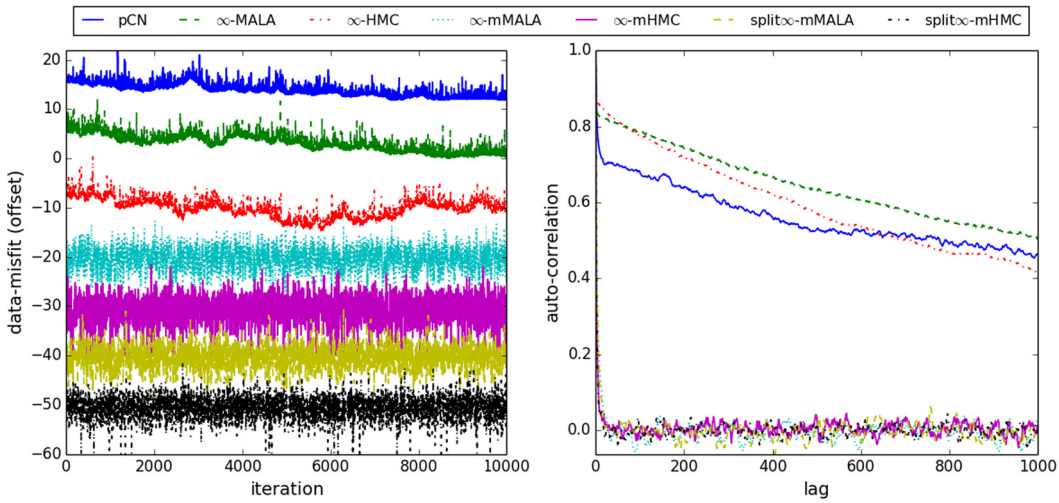


Fig. 13. Laminar jet problem: trace plots of the data-misfit function (left panel, values have been offset for better comparison) and the corresponding acf functions (right panel).

Table 3
Sampling Efficiency in the laminar jet problem. Column labels are as in Table 1.

Method	AP	s/iter	ESS(min, med, max)	minESS/s	spdup	PDEsolns
pCN	0.61	1.29	(5.24, 6.66, 13.33)	4.05E−04	1.00	22004
∞-MALA	0.66	1.68	(5.38, 6.62, 19.53)	3.21E−04	0.79	33005
∞-HMC	0.72	3.81	(5.41, 7.43, 16.44)	1.42E−04	0.35	82466
∞-mMALA	0.68	5.97	(1075.24, 2851.22, 3867.08)	1.80E−02	44.47	2233205
∞-mHMC	0.58	13.33	(2058.42, 3394.17, 4560.03)	1.54E−02	38.13	5575696
Split ∞-MMALA	0.57	3.66	(1079.55, 1805.89, 2395.13)	2.95E−02	72.82	693065
Split ∞-mHMC	0.60	6.88	(2749.63, 3974.36, 5498.03)	4.00E−02	98.67	1721694

non-differentiable wherever $\mathbf{u} \cdot \mathbf{n} = 0$, and thus the unknown-to-likelihood map is formally non-differentiable on the set $\{\mathbf{x} \in \mathcal{O}\,;\, \mathbf{u}(\mathbf{x}) \cdot \mathbf{n} = u_1(\mathbf{x}) = 0\}$. In future work we hope to extend geometric methods to such semi-smooth maps. However, we believe that this non-smoothness occurs on sets of measure zero in parameter space for the chosen PDE configuration, and hence poses no difficulties in practice when computing derivatives in geometric MCMC.

We run the various MCMC algorithms (all initialized at zero) for 1.1×10^4 iterations, treating the first 10^3 samples as burn-in. The posterior is obtained by stopping the K-L expansion for the prior at $|I_0| = 100$ and solving the PDE on a 30×30 mesh. The split-methods used location-specific scales up to $D_0 = 30$. HMC algorithms use a number of leapfrog steps randomly chosen between 1 and 4. We mention here an important practical consideration that arises when solving this problem. For almost all proposed states within MCMC the Newton solver converged. However with very low probability, and in almost all the experiments we ran, situations arise in which the proposed MCMC states led to divergence of the Newton solver. Whilst this might be ameliorated to some extent by different initializations of the Newton method, for reasons described above we have fixed the initialization. We deal with the divergence of Newton method in these situations by rejecting such proposals with probability 1, i.e. we remove these low probability states from the domain of the posterior.

Unlike the previous two PDE examples, none of the non-geometric methods converged to equilibrium due to requiring very small step-sizes ($\mathcal{O}(10^{-4})$) to provide non-negligible acceptance rates. The right panel of Fig. 12, shows the posterior means as estimated by the various MCMC algorithms. As expected, the estimate does not match the true inflow velocity $\theta^\dagger$ in the high frequencies due to limited amount of data. Note that the 95% credible band calculated with samples from ∞-mHMC is wide and covers most of the true inflow velocity (solid cyan line). Fig. 13 illustrates the extremely high auto-correlation of samples in the case of the non-geometric methods due to ineffective small step-sizes. The left panel indicates that non-geometric methods have not converged and the right panel shows high auto-correlation even at a lag of 1000. Table 3 shows that the proposed geometric methods yield almost 2 orders of magnitude improvement in sampling efficiency compared with pCN. Fig. 14 illustrates the first few data-misfit values according to different sampling methods. The upper plot shows pCN and ∞-MALA have not reached the center of the posterior, while ∞-HMC starts to approach it after 400 iterations. The lower plot verifies that this happens after 2500 seconds. It is also interesting to note that unlike other geometric methods, split ∞-mHMC takes about 450 iterations and 3000 seconds to enter the convergent region. All the above summaries confirm that geometric methods are advantageous in sampling efficiency.

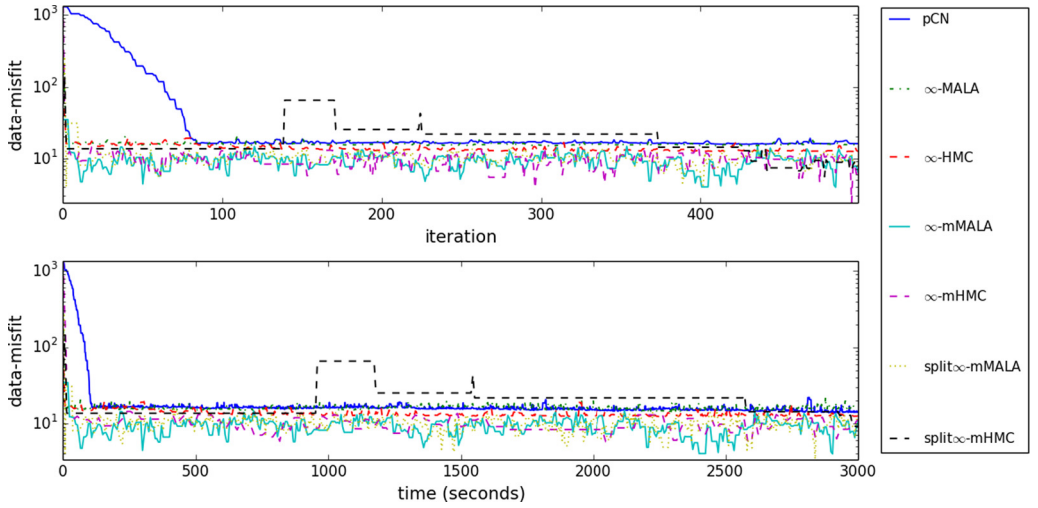


Fig. 14. Laminar jet problem: trace plots of data-misfits before burning-in for the first 500 iterations (upper panel) and the first 3000 seconds (lower panel) respectively.

5. Conclusion and discussion

This paper makes a number of contributions in the development of MCMC methods appropriate for the solution of inverse problems involving complex forward models with unknown parameters defined on infinite-dimensional Hilbert spaces. In particular: we generalize the simplified Riemannian manifold MALA of [8] from finite to infinite dimensions, and develop an HMC-version of the new method; we establish a connection between these infinite-dimensional geometric MALA and HMC algorithms; we develop a straightforward dimension reduction methodology which renders the methods highly effective in practice; we demonstrate the advantages of using HMC methods, built around ballistic motion, i.e. move with large step-size, that suppresses random walk behavior. All the algorithms are shown to be well-defined in the infinite dimensional setting, and three numerical studies demonstrate the effectiveness of the new methodology.

Some recent works have investigated incorporating information about the posterior within MCMC algorithms of mesh-independent mixing times, see e.g. [4] and the Dimension-Independent Likelihood-Informed MCMC in [7, DILI]. However, these approaches aim to make use of the curvature of the posterior at a fixed position (typically, the MAP, i.e. the maximizer of the posterior). The geometric methods defined here can be more appropriate for distributions with more complex non-Gaussian structures. In our laminar jet example for instance, Fig. 15 illustrates the non-Gaussianity of the posterior, thus incorporation of information about the local geometry can be beneficial in this context. Our methodology does not require pre-processing steps (e.g. finding the MAP and the Hessian at the MAP).

As mentioned in the main text, simplified manifold Langevin dynamics do not preserve the target distribution as they omit third order tensor terms, and can provide ineffective proposals for highly irregular targets (e.g. the banana-shaped distribution in [23] or the banana-biscuit-doughnut in [35]). In such cases, the multi-step HMC generalization will also be ineffective as the dynamics will soon drift away from the current energy contour, and have small acceptance probabilities. This consideration motivates a potential future development of infinite-dimensional MCMC methods that will incorporate full geometric information (including the third order tensor). The resulting method will be based on the full Riemannian manifold Langevin dynamics (say, on $\mathbb{R}^n$) [8]:

$$\frac{du}{dt} = -\frac{u}{2} + \frac{g(u)}{2} + \frac{dW^*}{dt} \tag{43}$$

where the Brownian motion W^* on the Riemannian manifold with metric tensor $G : \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$ has the form [8,36]:

$$dW^*(t)_i = |G(u)|^{-\frac{1}{2}} \sum_j \partial_j [G(u)^{-1}_{ij} |G(u)|^{\frac{1}{2}}] dt + [\sqrt{G(u)^{-1}} dW]_i \tag{44}$$

with $1 \le i \le n$, or the corresponding Lagrangian dynamics [23]:

$$\frac{du}{dt} = v\,, \quad \frac{dv}{dt} = -u + g^*(u)\,, \tag{45}$$

where $g^*(u)_i = g(u)_i - \mathrm{tr}[G(u)^{-1}\partial_i G(u)] - \Gamma^i_{k,l}(u) v^k v^l$. We have made use of the Christoffel symbols $\Gamma^i_{k,l}(u) = \frac{1}{2} g^{ij}[\partial_k g_{jl} + \partial_l g_{kj} - \partial_j g_{kl}]$, where g_{kl} denotes the (k,l)-th element of $G(u)$. Combining these dynamics with infinite-dimensional MCMC methodology will require some further research and is left for future work. Critically, one will need to carefully investigate the balance between improved mixing and the extra computational overheads.

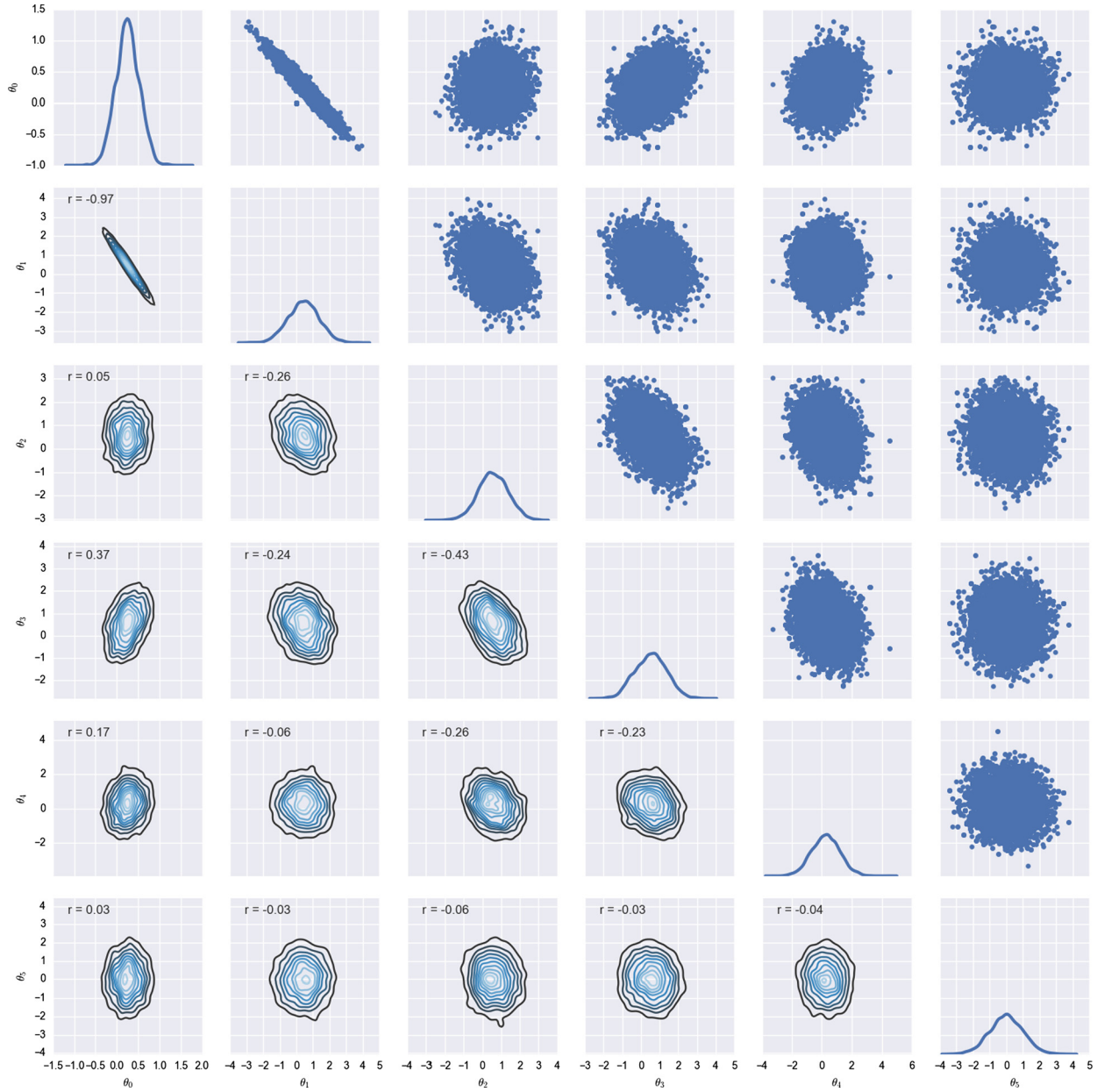


Fig. 15. Pair-wise marginal posterior distributions of the first 6 unknown frequencies of $\theta_1, \cdots, \theta_6$ in the laminar jet problem.

Future work will aim to incorporate alternative dimension reduction techniques such as Likelihood Informed Subspaces [LIS, 11, 7] or Active Subspaces [AS, 12, 37]. Fully geometric MCMC can then be employed in the finite dimensional 'intrinsic' subspace while its complement can be efficiently explored with relative simple methods like pCN or ∞-MALA. This merging of ideas will maybe enable us to make even better use of the geometric structure of the target within the MCMC algorithms.

Acknowledgements

We thank Claudia Schillings for her assistance in the development of adjoint codes for the groundwater flow problem and Umberto Villa for his assistance in the development of adjoint codes for the laminar jet problem. AB is supported by the Leverhulme Trust Prize. MG, SL and AMS are supported by the EPSRC program grant, Enabling Quantification of Uncertainty in Inverse Problems (EQUIP), EP/K034154/1 and the DARPA funded program Enabling Quantification of Uncertainty in Physical Systems (EQUiPS), contract W911NF-15-2-0121. MG is also supported by an EPSRC Established Career Research Fellowship, EP/J016934/2. PEF is supported by EPSRC grants EP/K030930/1 and EP/M019721/1, and a Center of Excellence

grant 179578 from the Research Council of Norway to the Center for Biomedical Computing at Simula Research Laboratory. AMS is also supported by an ONR grant N00014-17-1-2079.

Appendix A. Proof of Theorem 3.10

Proof. (i) Note that $S^{(i)} = S^{(i-1)} \circ \Psi_\varepsilon^{-1}$, and that $\Psi_\varepsilon = \Xi \circ R \circ \Xi$, where Ξ denotes the first or third map in (30) and R the second map (rotation). Thus, we have the equality $S^{(i)} = ((S^{(i-1)} \circ \Xi^{-1}) \circ R^{-1}) \circ \Xi^{-1}$. Notice that with this notation $\tilde{S}_0 \equiv S_0 \circ \Xi^{-1}$, so we have $G(u,v) = (d\tilde{S}_0/dS_0)(u,v) \equiv (d(S_0 \circ \Xi^{-1})/dS_0)(u,v)$. We proceed as follows:

$$
\begin{aligned}
&\frac{d((S^{(i-1)} \circ \Xi^{-1}) \circ R^{-1}) \circ \Xi^{-1}}{dS_0}(u_i, v_i) = \\
&= \frac{d(((S^{(i-1)} \circ \Xi^{-1}) \circ R^{-1}) \circ \Xi^{-1})}{d(S_0 \circ \Xi^{-1})}(u_i, v_i) \cdot \frac{d(S_0 \circ \Xi^{-1})}{dS_0}(u_i, v_i) \\
&= \frac{d((S^{(i-1)} \circ \Xi^{-1}) \circ R^{-1})}{dS_0}(\Xi^{-1}(u_i, v_i)) \cdot G(u_i, v_i) \\
&= \frac{d(S^{(i-1)} \circ \Xi^{-1})}{dS_0}(R^{-1}(\Xi^{-1}(u_i, v_i))) \cdot G(u_i, v_i) \\
&= \frac{d(S^{(i-1)} \circ \Xi^{-1})}{d(S_0 \circ \Xi^{-1})}(R^{-1}(\Xi^{-1}(u_i, v_i))) \\
&\qquad \times \frac{d(S_0 \circ \Xi^{-1})}{dS_0}(R^{-1}(\Xi^{-1}(u_i, v_i))) \cdot G(u_i, v_i) \\
&= \frac{dS^{(i-1)}}{dS_0}(u_{i-1}, v_{i-1}) \cdot G(u_{i-1}, v_{i-1} + \tfrac{\varepsilon}{2} g(u_{i-1})) \cdot G(u_i, v_i) \,.
\end{aligned}
$$

(ii) It is obtained from somewhat cumbersome, but straightforward algebraic calculations.

(iii) Same as (ii).

(iv) The proof is similar to the one of Theorem 3.1 in [38], but we include it here for completeness. The next position, u^n, of the Markov chain is (for a uniform $U \sim U[0,1]$):

$$u^n = \delta[\, U \le a(u_0, v_0)\,]\, u_I + \delta[\, U > a(u_0, v_0)\,]\, u_0 \,.$$

For continuous, bounded $f : \mathbb{X} \mapsto \mathbb{R}$, we take expectations on both sides to obtain:

$$\mathbb{E}[\, f(u^n)\,] = \mathbb{E}[\, a(\Psi_\varepsilon^{-I}(u_I, v_I))\, f(u_I)\,] - \mathbb{E}[\, a(u_0, v_0)\, f(u_0)\,] + \mathbb{E}[\, f(u_0)\,] \,.$$

Thus, it suffices to prove $\mathbb{E}[\, a(\Psi_\varepsilon^{-I}(u_I, v_I))\, f(u_I)\,] = \mathbb{E}[\, a(u_0, v_0)\, f(u_0)\,]$. Note now that (we sometimes stress the particular integrators in expectations/integrals by showing them explicitly as a subscript of $\mathbb{E}$ when needed):

$$
\begin{aligned}
\mathbb{E}[\, f(u_I)\, a(\Psi_\varepsilon^{-I}(u_I, v_I))\,] &\equiv \mathbb{E}_{S^{(I)}}[\, f(u_I)\, a(\Psi_\varepsilon^{-I}(u_I, v_I))\,] \\
&= \mathbb{E}_S[\, f(u_I)\, a(\Psi_\varepsilon^{-I}(u_I, v_I))\, e^{\Delta H(\Psi_\varepsilon^{-I}(u_I, v_I))}\,] \\
&= \mathbb{E}_S[\, f(u_I)\, (1 \wedge e^{\Delta H(\Psi_\varepsilon^{-I}(u_I, v_I))})\,] \\
&= \mathbb{E}_S[\, f(u_I) (1 \wedge e^{\Delta H(\Psi_\varepsilon^{-I}(u_I, -v_I))})\,] \,. \qquad \text{(A.1)}
\end{aligned}
$$

(For the 2nd equation we used the density $dS^{(I)}/dS$ we found in (ii) together with the identity in (iii); for the last equation, notice that (u_I, v_I) and $(u_I, -v_I)$ have the same law S.) Now, due to the symmetry property $\Psi_\varepsilon^I \circ M \circ \Psi_\varepsilon^I = M$ of the leapfrog operator (we have denoted by M the operator that 'flips' the sign of the velocity), we have that $\Psi_\varepsilon^{-I} \circ M = M \circ \Psi_\varepsilon^I$. Thus, we have:

$$
\begin{aligned}
&\Delta H(\Psi_\varepsilon^{-I}(u_I, -v_I))) = \Delta H(M \circ \Psi_\varepsilon^I(u_I, v_I))) \\
&\quad = H(M(u_I, v_I)) - H(M \circ \Psi_\varepsilon^I(u_I, v_I)) \equiv -\Delta H(u_I, v_I) \,,
\end{aligned}
$$

where in the last equation we used the fact that $H \circ M = H$ due to the energy H being quadratic in the velocity v. Using this in (A.1), we have obtained indeed that $\mathbb{E}[\, a(\Psi_\varepsilon^{-I}(u_I, v_I))\, f(u_I)\,] = \mathbb{E}[\, a(u_0, v_0)\, f(u_0)\,]$ as required. □

Appendix B. Proof of Corollary 3.13

Proof. For the given setting of the step-sizes (31), we first prove the coincidence of the proposals by ∞-mHMC and ∞-mMALA, that is, (30) reduces to (22). Noting that $u_0 = u$ and $v_0 = \xi \sim \mathcal{N}(0, K(u))$, with the first equation of (30) we have:

$$v^- = v_0 + \tfrac{\varepsilon_1}{2}\, g(u_0) \equiv v \ ,$$

with v as defined in the ∞-mMALA proposal in (22). Then, the definition of ρ in (7) and the setting (31) imply:

$$\rho = \frac{1 - h/4}{1 + h/4} = \cos\varepsilon_2 \, ; \quad \sqrt{1-\rho^2} = \frac{\sqrt{h}}{1+h/4} = \sin\varepsilon_2 \ .$$

Therefore, it follows from the second equation of (30), that the proposal, say u', of ∞-mHMC for one leapfrog step is equal to:

$$u' = u_{\varepsilon_2} = u_0 \cos\varepsilon_2 + v^- \sin\varepsilon_2 \equiv \rho\, u + \sqrt{1-\rho^2}\, v \ ,$$

with the term on the right hand side being the proposal from ∞-mMALA. Since the proposals coincide, the acceptance probabilities will also be the same, as they both apply the Metropolis-Hastings ratio. □

References

[1] A. Beskos, G. Roberts, A. Stuart, J. Voss, MCMC methods for diffusion bridges, Stoch. Dyn. 8 (03) (2008) 319–350.
[2] A. Beskos, F.J. Pinski, J.M. Sanz-Serna, A.M. Stuart, Hybrid Monte-Carlo on Hilbert spaces, Stoch. Process. Appl. 121 (2011) 2201–2230.
[3] S.L. Cotter, G.O. Roberts, A. Stuart, D. White, MCMC methods for functions: modifying old algorithms to make them faster, Stat. Sci. 28 (3) (2013) 424–446.
[4] K. Law, Proposals which speed up function-space MCMC, J. Comput. Appl. Math. 262 (2014) 127–138.
[5] F.J. Pinski, G. Simpson, A.M. Stuart, H. Weber, Algorithms for Kullback–Leibler approximation of probability measures in infinite dimensions, SIAM J. Sci. Comput. 37 (6) (2015) A2733–A2757.
[6] D. Rudolf, B. Sprungk, On a generalization of the preconditioned Crank–Nicolson Metropolis algorithm, arXiv preprint arXiv:1504.03461.
[7] T. Cui, K.J. Law, Y.M. Marzouk, Dimension-independent likelihood-informed MCMC, J. Comput. Phys. 304 (2016) 109–137.
[8] M. Girolami, B. Calderhead, Riemann manifold Langevin and Hamiltonian Monte Carlo methods, J. R. Stat. Soc. B 73 (2) (2011) 123–214 (with discussion).
[9] J. Martin, L.C. Wilcox, C. Burstedde, O. Ghattas, A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, SIAM J. Sci. Comput. 34 (3) (2012) A1460–A1487.
[10] T. Bui-Thanh, O. Ghattas, D. Higdon, Adaptive Hessian-based nonstationary Gaussian process response surface method for probability density approximation with application to Bayesian solution of large-scale inverse problems, SIAM J. Sci. Comput. 34 (6) (2012) A2837–A2871.
[11] T. Cui, J. Martin, Y.M. Marzouk, A. Solonen, A. Spantini, Likelihood-informed dimension reduction for nonlinear inverse problems, Inverse Probl. 30 (11) (2014) 114015.
[12] P.G. Constantine, Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies, SIAM, 2015.
[13] N. Petra, J. Martin, G. Stadler, O. Ghattas, A computational framework for infinite-dimensional Bayesian inverse problems. Part II: stochastic Newton MCMC with application to ice sheet flow inverse problems, SIAM J. Sci. Comput. 36 (4) (2014) A1525–A1555.
[14] L. Tierney, A note on Metropolis-Hastings kernels for general state spaces, Ann. Appl. Probab. 8 (1) (1998) 1–9.
[15] R.M. Neal, MCMC using Hamiltonian dynamics, in: S. Brooks, A. Gelman, G. Jones, X.L. Meng (Eds.), Handbook of Markov Chain Monte Carlo, Chapman and Hall/CRC, 2010.
[16] G.O. Roberts, A. Gelman, W.R. Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab. 7 (1) (1997) 110–120.
[17] G. Da Prato, J. Zabczyk, Stochastic Equations in Infinite Dimensions, vol. 152, Cambridge University Press, 2014.
[18] S. Duane, A.D. Kennedy, B.J. Pendleton, D. Roweth, Hybrid Monte Carlo, Phys. Lett. B 195 (2) (1987) 216–222.
[19] L. Verlet, Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard–Jones molecules, Phys. Rev. 159 (1) (1967) 98–103.
[20] R.J. Adler, The Geometry of Random Fields, Class. Appl. Math., vol. 62, Siam, 2010.
[21] V.I. Bogachev, Gaussian Measures, Math. Surv. Monogr., vol. 62, American Mathematical Soc., 1998.
[22] M. Dashti, A.M. Stuart, The Bayesian approach to inverse problems, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Handbook of Uncertainty Quantification, ISBN 978-3-319-12384-4, Springer, 2016, arXiv:1302.6989.
[23] S. Lan, V. Stathopoulos, B. Shahbaba, M. Girolami, Markov chain Monte Carlo from Lagrangian dynamics, J. Comput. Graph. Stat. 24 (2) (2015) 357–378.
[24] T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, M. Girolami, Langevin diffusions and the Metropolis-adjusted Langevin algorithm, Stat. Probab. Lett. 91 (2014) 14–19.
[25] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, G.N. Wells, The FEniCS project version 1.5, Archive of Numerical Software 3 (100) (2015) 9–23.
[26] A. Logg, K.-A. Mardal, G. Wells, Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book, vol. 84, Springer Science & Business Media, 2012.
[27] P.E. Farrell, D.A. Ham, S.W. Funke, M.E. Rognes, Automated derivation of the adjoint of high-level transient finite element programs, SIAM J. Sci. Comput. 35 (4) (2013) C369–C393.
[28] T. Bui-Thanh, M. Girolami, Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo, Inverse Probl. 30 (11) (2014) 114014.
[29] M. Dashti, A.M. Stuart, Uncertainty quantification and weak approximation of an elliptic inverse problem, SIAM J. Numer. Anal. 49 (6) (2011) 2524–2542.
[30] P.R. Conrad, Y.M. Marzouk, N.S. Pillai, A. Smith, Asymptotically exact MCMC algorithms via local approximations of computationally intensive models, J. Am. Stat. Assoc. 111 (516) (2016) 1591–1607.

[31] M. Hairer, A.M. Stuart, S.J. Vollmer, Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions, Ann. Appl. Probab. 24 (6) (2014) 2455–2490.
[32] T. Bui-Thanh, FEM-Based Discretization-Invariant MCMC Methods for PDE-constrained Bayesian Inverse Problems, Department of Aerospace Engineering and Engineering Mechanics, Institute for Computational Engineering and Sciences, The University of Texas at Austin, online manual Edition, July 2015.
[33] M. Esmaily Moghadam, Y. Bazilevs, T.-Y. Hsia, I. Vignon-Clementel, A. Marsden, A comparison of outlet boundary treatments for prevention of backflow divergence with relevance to blood flow simulations, Comput. Mech. 48 (3) (2011) 277–291.
[34] M. Klein, A. Sadiki, J. Janicka, Investigation of the influence of the Reynolds number on a plane jet using direct numerical simulation, Int. J. Heat Fluid Flow 24 (6) (2003) 785–794.
[35] S. Lan, T. Bui-Thanh, M. Christie, M. Girolami, Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems, J. Comput. Phys. 308 (2016) 81–101.
[36] K.L. Chung, Lectures from Markov Processes to Brownian Motion, A Series of Comprehensive Studies in Mathematics, vol. 249, Springer Science & Business Media, 2013.
[37] P.G. Constantine, C. Kent, T. Bui-Thanh, Accelerating MCMC with active subspaces, SIAM J. Sci. Comput. 38 (5) (2016) A2779–A2805.
[38] A. Beskos, K. Kalogeropoulos, E. Pazos, Advanced MCMC methods for sampling on diffusion pathspace, Stoch. Process. Appl. 123 (4) (2013) 1415–1453.