FOUNDATIONS OF COMPUTATIONAL MATHEMATICS

The Journal of the Society for the Foundations of **Computational** Mathematics

CrossMark

On a Generalization of the Preconditioned Crank–Nicolson Metropolis Algorithm

Daniel Rudolf[1] · **Björn Sprungk**[2]

Abstract Metropolis algorithms for approximate sampling of probability measures on infinite dimensional Hilbert spaces are considered, and a generalization of the preconditioned Crank–Nicolson (pCN) proposal is introduced. The new proposal is able to incorporate information on the measure of interest. A numerical simulation of a Bayesian inverse problem indicates that a Metropolis algorithm with such a proposal performs independently of the state-space dimension and the variance of the observational noise. Moreover, a qualitative convergence result is provided by a comparison argument for spectral gaps. In particular, it is shown that the generalization inherits geometric convergence from the Metropolis algorithm with pCN proposal.

✉ Daniel Rudolf
daniel.rudolf@uni-goettingen.de

Björn Sprungk
bjoern.sprungk@mathematik.tu-chemnitz.de

1   Institut für Mathematische Stochastik, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany

2   Technische Universität Chemnitz, Reichenhainer Str. 41, 09107 Chemnitz, Germany

🖄 Springer

1 Introduction

Consider a target probability distribution μ defined on a possibly infinite dimensional separable Hilbert space $\mathcal{H}$. It is of interest to sample from this probability measure and assumed that there is a density of μ w.r.t. a Gaussian reference measure μ_0 on $\mathcal{H}$ given by

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u)), \qquad u \in \mathcal{H}. \tag{1}$$

Here $\Phi \colon \mathcal{H} \to \mathbb{R}_+$ is a measurable function and $Z = \int_{\mathcal{H}} \exp(-\Phi(u)) \, \mu_0(\mathrm{d}u)$ the normalizing constant. Such probability measures μ arise as posterior distributions in Bayesian inference with μ_0 as a Gaussian prior. Common examples in infinite dimensional spaces are inferring spatially distributed properties of porous media or stock prices.

Unfortunately, the fact that the normalizing constant Z is typically unknown and that Φ is only available in the form of function evaluations makes it difficult to sample μ directly. But Markov chains and in particular Metropolis–Hastings (MH) algorithms are applicable for approximate sampling. These algorithms consist of a proposal and an acceptance/rejection step. A state is proposed by a proposal kernel, but it is only accepted with a certain probability which depends on $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}$. The authors of [1] suggested a modification of a Gaussian random walk proposal which is μ_0-reversible. The latter property leads to a well-defined MH algorithm in infinite dimensional Hilbert spaces, see also [30]. This proposal was later [4] referred to as *preconditioned Crank–Nicolson (pCN) proposal*. Remarkably, the Markov chain of the resulting pCN Metropolis algorithm has *dimension-independent sampling efficiency*, see [4,13]. This is a significant advantage compared to earlier, popular MH algorithms whose performance usually deteriorates with increasing state-space dimension [4,13,25].

We extend the pCN proposal to incorporate information about the target measure μ. Such an adaption might account for the anisotropy of the covariance of μ or the local curvature of Φ. Intuitively, the resulting Markov chain has on average a larger step size and, thus, explores the state space faster. This idea is not entirely new. It is already mentioned in [29] where it is suggested to choose the covariance of the proposal adapted to the target measure. Later in [11] the authors explain how to propose new states using general local metric tensors. Moreover, in [22] the Hessian of the negative log density Φ of μ is employed as local curvature information to design a stochastic Newton MH method in finite dimensions and in [5,17] a Gauss–Newton variant for capturing global curvature in an infinite dimensional setting is outlined.

Our approach for adapting the proposal to the target measure μ has a similar motivation as the proposals considered in [5,17]. It comes from a local linearization of the unknown-to-observable map in Bayesian inverse problems. This suggests a particular form for approximating the covariance of the target measure, namely $(C + \Gamma)^{-1}$, where C denotes the covariance of the reference measure μ_0 and Γ is a suitable self-adjoint and positive operator. We then consider the class of Gaussian proposals with covariance $C_\Gamma = (C + \Gamma)^{-1}$. By enforcing μ_0-reversibility we derive our class of generalized pCN (gpCN) proposal kernels P_Γ.

In a numerical simulation the resulting Metropolis algorithm seems to perform independently of dimension and variance. Here variance independence refers to the variance of the observational noise, which affects the covariance of the target distribution μ. Particularly, if the variance of the noise decreases the measure μ becomes more concentrated. Our numerical experiments also indicate that other popular MH or random walk algorithms perform worse, i.e., variance dependent.

Moreover, we present a convergence result for the gpCN Metropolis based on spectral gaps. It is well known, see [24], that for Markov chains with reversible transition kernels K a strictly positive spectral gap, denoted $\mathrm{gap}(K) > 0$, is equivalent to a form of geometric ergodicity. The latter roughly means that, in an appropriate setting, the distribution of the nth step of a Markov chain converges exponentially fast to its stationary measure. We refer to Sect. 2.1 for precise definitions and further details.

Our main theoretic result, stated in Theorem 20, is as follows. Let us assume that the transition kernel M_0 of the pCN algorithm has a positive spectral gap, i.e., $\mathrm{gap}(M_0) > 0$. Then, for any $\varepsilon > 0$ there is an explicitly given probability measure μ_R such that

$$\|\mu - \mu_R\|_{\mathrm{tv}} \leq \varepsilon \quad \text{and} \quad \mathrm{gap}(M_{\Gamma,R}) > 0$$

where $M_{\Gamma,R}$ denotes the transition kernel of the gpCN Metropolis algorithm targeting the measure μ_R and $\|\cdot\|_{\mathrm{tv}}$ is the total variation distance, see (3).

The key for the proof is a new comparison theorem for spectral gaps of Markov chains generated by MH algorithms. In order to apply this comparison argument we show that the proposal kernels of the pCN and gpCN Metropolis are equivalent and that their Radon–Nikodym derivative belongs to an L_p-space for a $p > 1$. We note that in [13] under additional assumptions on the density function $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}$ it is proven that there exists a strictly positive spectral gap of the pCN Metropolis. Thus, in this setting the gpCN Metropolis algorithm targeting μ_R also converges exponentially.

The remainder of the paper is organized as follows. In Sect. 2 we state the precise framework, recall preliminary facts and give a brief introduction to Markov chain Monte Carlo and MH algorithms including the pCN Metropolis algorithm. The gpCN Metropolis algorithm is motivated and defined in Sect. 3. Particularly, in Sect. 3.3 we illustrate its superior performance compared to other popular MH algorithms. In Sect. 4 we state a general result for comparing spectral gaps of MH algorithms and then apply it to the gpCN and pCN Metropolis. Section 5 provides an outlook to gpCN algorithms in infinite dimensions which use Gaussian proposals with state-dependent covariance. For the convenience of the reader we recall some facts about Gaussian measures in "Gaussian Measures" appendix and relegate more technical proofs to "Proofs" appendix.

2 Preliminaries

Let $\mathcal{H}$ be a separable Hilbert space with inner-product and norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$. By $\mathcal{B}(\mathcal{H})$ we denote the corresponding Borel σ-algebra and by $\mathcal{L}(\mathcal{H})$ the set of all bounded, linear operators $A \colon \mathcal{H} \to \mathcal{H}$. Further, we have a Gaussian measure

$\mu_0 = N(0, C)$ on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. Here and in the remainder of the paper $C: \mathcal{H} \to \mathcal{H}$ denotes a nonsingular *covariance operator* on $\mathcal{H}$, i.e., a bounded, self-adjoint and positive trace class operator with $\ker C = \{0\}$. By μ we denote the probability measure of interest on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ given through the density defined in (1). Typically, the desired distribution is complicated and the density only known up to a constant, which makes direct sampling from μ difficult. This is the reason why Markov chains are used for approximate sampling according to μ.

2.1 Markov Chains and Spectral Gaps

We give a short introduction to Markov chains and Markov chain Monte Carlo (MCMC) methods on general state spaces. We call a mapping $K: \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0, 1]$ a *transition kernel* if $K(x, \cdot)$ is a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ for each $x \in \mathcal{H}$ and $K(\cdot, A)$ is a measurable function for each $A \in \mathcal{B}(\mathcal{H})$. Then, a *Markov chain with transition kernel* K is a sequence of random variables $(X_n)_{n \in \mathbb{N}}$, mapping from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, satisfying

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \mathbb{P}(X_{n+1} \in A \mid X_n) = K(X_n, A)$$

almost surely for all $A \in \mathcal{B}(\mathcal{H})$. Most properties of a Markov chain can be expressed as properties of its transition kernel. For example, we say the transition kernel K is *μ-reversible* if

$$K(u, \mathrm{d}v)\, \mu(\mathrm{d}u) = K(v, \mathrm{d}u)\, \mu(\mathrm{d}v) \tag{2}$$

in the sense of measures on $\mathcal{H} \times \mathcal{H}$. This property is also known as the *detailed balance condition*, and it implies that the distribution μ is a stationary or invariant probability measure of a Markov chain with transition kernel K, i.e., if $X_1 \sim \mu$ then also $X_2 \sim \mu$.

Each μ-reversible transition kernel K on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ induces a *Markov operator*, which we shall also denote by K, given by

$$Kf(u) = \int_{\mathcal{H}} f(v)\, K(u, \mathrm{d}v), \quad f \in L_2(\mu),$$

where

$$L_2(\mu) = \left\{ f: \mathcal{H} \to \mathbb{R} \mid \|f\|_{2,\mu} := \left(\int_{\mathcal{H}} |f(u)|^2\, \mu(\mathrm{d}u) \right)^{1/2} < \infty \right\},$$

is the Hilbert space of measurable, square integrable functions with respect to μ. By the μ-reversibility we have that $K: L_2(\mu) \to L_2(\mu)$ is a bounded and self-adjoint linear operator. We also introduce the closed subspace

$$L_2^0(\mu) = \left\{ f \in L_2(\mu) \mid \int_{\mathcal{H}} f(u)\, \mu(\mathrm{d}u) = 0 \right\}$$

of $L_2(\mu)$ and the operator norm

$$\|K\|_\mu = \sup_{f \in L_2^0(\mu),\, f \neq 0} \frac{\|Kf\|_{2,\mu}}{\|f\|_{2,\mu}}$$

for $K : L_2^0(\mu) \to L_2^0(\mu)$. Let $\mathrm{spec}(K \mid L_2^0(\mu))$ denote the spectrum of K on $L_2^0(\mu)$. Then, we also have

$$\|K\|_\mu = \sup\left\{ |\lambda| : \lambda \in \mathrm{spec}\left(K \mid L_2^0(\mu)\right) \right\}.$$

We define the *spectral gap of K* (w.r.t. μ) by $\mathrm{gap}(K) = 1 - \|K\|_\mu$. This is an important quantity which can be used to formulate conditions ensuring an exponentially fast convergence of the distribution of X_n to μ. To be more precise, we introduce the total variation distance of two probability measures ν_1, ν_2 on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ by

$$\|\nu_1 - \nu_2\|_{\mathrm{tv}} := \sup_{A \in \mathcal{B}(\mathcal{H})} |\nu_1(A) - \nu_2(A)|. \tag{3}$$

Let ν be the initial distribution of the Markov chain, i.e., $X_1 \sim \nu$. Then, with

$$K^n(u, A) = \int_{\mathcal{H}} K^{n-1}(v, A)\, K(u, \mathrm{d}v), \quad A \in \mathcal{B}(\mathcal{H}),$$

for $n \in \mathbb{N}$, the distribution of X_{n+1} is given by

$$\nu K^n(A) = \int_{\mathcal{H}} K^n(u, A)\, \nu(\mathrm{d}u).$$

In the setting above it is well known, see [24, Proposition 2.2], that $\|K\|_\mu < 1$, or equivalently $\mathrm{gap}(K) > 0$, holds, iff the transition kernel is $L_2(\mu)$-geometrically ergodic. Here by $L_2(\mu)$-geometric ergodicity we mean that there exists a number $r \in [0, 1)$ such that for any probability measure ν, which has a density $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_2(\mu)$ w.r.t μ, there is a constant $C_\nu < \infty$ such that

$$\|\nu K^n - \mu\|_{\mathrm{tv}} \leq C_\nu\, r^n, \quad n \in \mathbb{N}.$$

If the distribution of X_n converges to μ, then the Markov chain $(X_n)_{n \in \mathbb{N}}$ can be used for approximate sampling from μ. This leads to Markov chain Monte Carlo methods for the computation of expectations. The mean $\mathbb{E}_\mu(f)$ of a function $f : \mathcal{H} \to \mathbb{R}$ w.r.t μ can then be approximated by the time average

$$S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0})$$

where n is the sample size and n_0 a burn-in parameter to decrease the influence of the initial distribution. The spectral gap of K of the Markov chain $(X_n)_{n \in \mathbb{N}}$ can then be

applied to assess the error of the time average $S_{n,n_0}(f)$. We assume $\text{gap}(K) > 0$ and mention two results. The first is rather classical and due to Kipnis and Varadhan [16]. If the initial distribution is μ and $f \in L_2(\mu)$, then the error $\sqrt{n}(S_{n,n_0}(f) - \mathbb{E}_\mu(f))$ converges weakly to $N(0, \sigma_{f,K}^2)$ with

$$
\sigma_{f,K}^2 = \left\langle (I + K)(I - K)^{-1}(f - \mathbb{E}_\mu(f)), (f - \mathbb{E}_\mu(f)) \right\rangle_\mu \leq \frac{2\,\|f\|_{2,\mu}^2}{\text{gap}(K)}
$$

where $\langle \cdot, \cdot \rangle_\mu$ denotes the inner-product in $L_2(\mu)$. The second result is more recent and provides a nonasymptotic bound for the mean square error. We have

$$
\sup_{\|f\|_4 \leq 1} \mathbb{E}\left| S_{n,n_0}(f) - \mathbb{E}_\mu(f) \right|^2 \leq \frac{2}{n \cdot \text{gap}(K)} + \frac{C_\nu\,\|K\|_\mu^{n_0}}{n^2 \cdot \text{gap}(K)^2}
$$

with $\|f\|_4 = \left(\int_{\mathcal{H}} |f(u)|^4\,\mu(\mathrm{d}u)\right)^{1/4}$ and a number $C_\nu \geq 0$ depending on the initial distribution ν. We refer to [26] for details.

This shows that $\text{gap}(K)$ is a crucial quantity in the study of Markov chains and the numerical analysis of MCMC methods.

2.2 Metropolis Algorithm with pCN Proposal

In this work we focus on Markov chains generated by the Metropolis algorithm. This algorithm employs a transition kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ for proposing new states which we shall denote by P and call *proposal kernel*. Moreover, let $\alpha \colon \mathcal{H} \times \mathcal{H} \to [0, 1]$ be a measurable function denoting the *acceptance probability*. Then, a transition of a Markov chain $(X_n)_{n \in \mathbb{N}}$ generated by the Metropolis algorithm can be represented in algorithmic form:

1. Given the current state $X_n = u$, draw independently a sample v of a random variable $V \sim P(u, \cdot)$ and a sample a of a random variable $A \sim \text{Unif}[0, 1]$.
2. If $a < \alpha(u, v)$, then set $X_{n+1} = v$, otherwise set $X_{n+1} = u$.

The transition kernel of such a Markov chain is then

$$
M(u, \mathrm{d}v) = \alpha(u, v)P(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \int_{\mathcal{H}} (1 - \alpha(u, w))\,P(u, \mathrm{d}w) \tag{4}
$$

and we call it *Metropolis kernel*. It is well known, see [30], that M is reversible w.r.t. μ if $\alpha(\cdot, \cdot)$ is chosen as

$$
\alpha(u, v) = \min\left\{ 1, \frac{\mathrm{d}\eta^\perp}{\mathrm{d}\eta}(u, v) \right\}, \qquad u, v \in \mathcal{H}, \tag{5}
$$

where $\frac{\mathrm{d}\eta^\perp}{\mathrm{d}\eta}$ denotes the Radon–Nikodym derivative of the measures

$$
\eta(\mathrm{d}u, \mathrm{d}v) := P(u, \mathrm{d}v)\,\mu(\mathrm{d}u) \quad \text{and} \quad \eta^\perp(\mathrm{d}u, \mathrm{d}v) := P(v, \mathrm{d}u)\,\mu(\mathrm{d}v),
$$

which we assume to exist. For finite dimensional state spaces the condition of absolute continuity of $\eta^\perp$ w.r.t. η is often easily satisfied. However, for infinite dimensional state spaces this becomes a real issue, since there measures tend to be mutually singular. As pointed out in [1,4] a possible way to ensure the existence of $\frac{\mathrm{d}\eta^\perp}{\mathrm{d}\eta}$ is to choose a proposal kernel P which is μ_0-reversible, i.e.,

$$P(u, \mathrm{d}v)\, \mu_0(\mathrm{d}u) = P(v, \mathrm{d}u)\, \mu_0(\mathrm{d}v). \tag{6}$$

Then, due to the fact that $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}$ and $\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu}$ exist, see (1), it follows that

$$\frac{\mathrm{d}\eta^\perp}{\mathrm{d}\eta}(u, v) = \frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(v)\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu}(u) = \exp(\Phi(u) - \Phi(v)) \tag{7}$$

and, hence, $\alpha(u, v) = \min\{1, \exp(\Phi(u) - \Phi(v))\}$.

We next introduce the Metropolis algorithm with the *preconditioned Crank–Nicolson* (pCN) proposal, see also [4] for details. The pCN proposal kernel arises from a discretization of an Ornstein–Uhlenbeck process with invariant measure μ_0 and takes the form

$$P_0(u, \cdot) = N\left(\sqrt{1 - s^2}u, s^2 C\right) \tag{8}$$

where $s \in [0, 1]$ denotes a variance or step size parameter. It is straightforward to verify that P_0 is μ_0-reversible. Namely, by applying (34) from "Gaussian Measures" appendix we deduce

$$P_0(u, \mathrm{d}v)\, \mu_0(\mathrm{d}u) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} C & \sqrt{1 - s^2}C \\ \sqrt{1 - s^2}C & C \end{bmatrix}\right) = P_0(v, \mathrm{d}u)\, \mu_0(\mathrm{d}v).$$

In the following we call the resulting Metropolis algorithm with proposal P_0 simply pCN Metropolis algorithm or pCN Metropolis and denote its Metropolis kernel by M_0.

Next, we generalize the pCN Metropolis algorithm to allow for proposal kernels which employ a different covariance structure than the covariance of μ_0.

3 Metropolis with gpCN Proposals

In recent years many authors have proposed and pursued the idea to construct proposals which try to exploit certain geometric features of the target measure, see for example [5,11,17,22].

We consider generalized pCN (gpCN) proposals which aim to adapt to the covariance structure of the target measure μ. We motivate our gpCN proposal and show that it is well defined in function spaces and illustrate its superior performance in a simple but common setting.

3.1 Motivation from Bayesian Inference

We briefly recall the Bayesian framework for inverse problems and refer to [9] for an overview and to [28] for a comprehensive introduction to the topic.

Assume X is a random variable on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ with distribution $\mu_0 = N(0, C)$. Here μ_0 is called the *prior* distribution and describes our initial uncertainty about X. Let Y be a random variable on $\mathbb{R}^m$ given by

$$Y = G(X) + \varepsilon \tag{9}$$

with a continuous map $G \colon \mathcal{H} \to \mathbb{R}^m$ and $\varepsilon \sim N(0, \Sigma)$, independent of X, with $\Sigma \in \mathbb{R}^{m \times m}$. The variable Y models an observable quantity depending on X via the map G which is perturbed by additive noise ε. Then, given some observation $y \in \mathbb{R}^m$ of Y we want to infer X, i.e., we are interested in the conditional distribution of X given the event $Y = y$. We denote this conditional distribution by μ and call it *posterior* distribution. In particular, in this setting μ admits a representation of the form (1) with

$$\Phi(u) = \frac{1}{2} |y - G(u)|^2_{\Sigma^{-1}} \tag{10}$$

where $|x|^2_{\Sigma^{-1}} = x^{\mathrm{T}} \Sigma^{-1} x$ for $x \in \mathbb{R}^m$.

A special situation appears if $G(u) = Lu + b$ with a linear mapping $L \colon \mathcal{H} \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$. Then, it is known from [21] that $\mu = N(\hat{m}, \widehat{C})$ with

$$\hat{m} = CL^*(LCL^* + \Sigma)^{-1}(y - b), \qquad \widehat{C} = (C^{-1} + L^* \Sigma^{-1} L)^{-1}, \tag{11}$$

where L^* denotes the adjoint operator of L. If we want to sample approximately from a Gaussian target measure $\mu = N(\hat{m}, \widehat{C})$ by Metropolis algorithms with Gaussian proposals, it seems beneficial to employ $s^2 \widehat{C}$ as proposal covariance, see for example [17,25,29]. Intuitively, since then the Gaussian proposal possesses the same principal directions and the same ratio of variances as the Gaussian target measure, the proposed states should be accepted more often than for other proposals. See also Fig. 1 for an illustration. This leads to a higher average acceptance probability and, thus, a faster exploration of the state space.

The affine case indicates how we can construct good Gaussian proposal kernels if the map G is nonlinear but smooth. For a fixed $u_0 \in \mathcal{H}$ local linearization leads to

$$G(u) = G(u_0) + \nabla G(u_0)\,(u - u_0) + r(u)$$

with a remainder term $r(u) \in \mathbb{R}^m$. For a sufficiently smooth G the remainder r is small (in a neighborhood of u_0), so that

$$\widetilde{G}(u) = G(u_0) + \nabla G(u_0)\,(u - u_0)$$
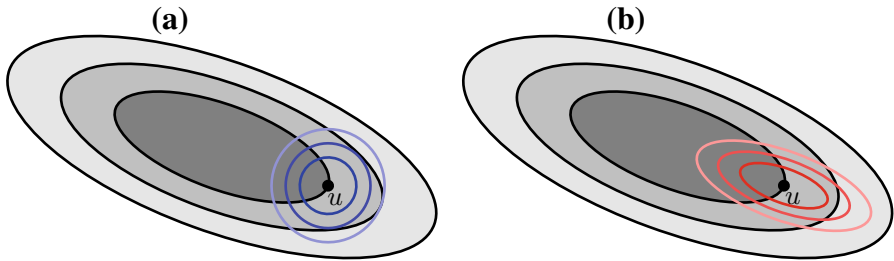
(a) **(b)**

Fig. 1 For a Gaussian target measure $\mu = N(m, \widehat{C})$ and current state u the region of acceptance $\{v : \alpha(u, v) = 1\}$ (*dark gray region*) and two regions of possible rejection $\{v : \underline{p} \leq \alpha(u, v) < \overline{p} \leq 1\}$ (*lighter gray regions*) are displayed. Moreover, we present the contour lines (*blue* and *red*, resp.) of Gaussian proposals $N(u, s^2 C)$ with covariance $C = I$ in part (**a**) and target covariance $C = \widehat{C}$ in part (**b**) (Color figure online)

is close to $G(u)$ (in a neighborhood of u_0). The substitution of G by $\widetilde{G}$ in the model (9) leads to a Gaussian target measure $\widetilde{\mu} = N(\widetilde{m}, \widetilde{C})$ with covariance

$$\widetilde{C} = \left(C^{-1} + L^* \Sigma^{-1} L\right)^{-1}, \qquad L = \nabla G(u_0).$$

By the fact that G and $\widetilde{G}$ are close, we also have that the measures μ and $\widetilde{\mu}$ are close as well. Then, it is reasonable to use $\widetilde{C}$ in the covariance operator of the proposal in a Metropolis algorithm. Of course, there might be other choices besides a simple linearization of G at one point. For example, averaging linearizations at several points $u_1, \ldots, u_n \in \mathcal{H}$ leads to

$$\widetilde{C} = \left(C^{-1} + \frac{1}{N} \sum_{n=1}^{N} L_n^* \Sigma^{-1} L_n\right)^{-1}, \qquad L_n = \nabla G(u_n).$$

Natural candidates for the points $u_1, \ldots, u_N$ are samples according to the prior or samples taken from a short run of a preliminary Markov chain with the posterior as stationary measure, cf. the adaptive method in [5, Section 3.4]. One could also think of a state-dependent covariance $C(u)$. This motivates the study of proposals which use covariances of the form $C_\Gamma = (C^{-1} + \Gamma)^{-1}$ for suitably chosen operators Γ.

3.2 Well-Defined gpCN Proposals

In this section we introduce the gpCN proposal kernel and prove that the Metropolis algorithm with this proposal is well defined in the sense that it leads to a μ-reversible transition kernel.

For this we introduce the set $\mathcal{L}_+(\mathcal{H})$ of all bounded, self-adjoint and positive linear operators $\Gamma : \mathcal{H} \to \mathcal{H}$. We define the operators

$$C_\Gamma := (C^{-1} + \Gamma)^{-1}, \qquad \Gamma \in \mathcal{L}_+(\mathcal{H}), \tag{12}$$

motivated in Sect. 3.1, where C denotes the covariance operator of the prior measure $\mu_0 = N(0, C)$, for which we also use the equivalent representation

$$C_\Gamma = C^{1/2} (I + H_\Gamma)^{-1} C^{1/2}, \qquad H_\Gamma := C^{1/2} \Gamma C^{1/2}. \tag{13}$$

In the following we prove that C_Γ can be considered as covariance operator.

Proposition 1 *Let C be a nonsingular covariance operator on $\mathcal{H}$, $\Gamma \in \mathcal{L}_+(\mathcal{H})$ and C_Γ with H_Γ given as in (13). Then $H_\Gamma \in \mathcal{L}_+(\mathcal{H})$ is trace class and C_Γ is also a nonsingular covariance operator on $\mathcal{H}$.*

Proof That $H_\Gamma \in \mathcal{L}_+(\mathcal{H})$ follows by construction. Furthermore, since H_Γ is a composition of two Hilbert–Schmidt and one bounded operator, $C^{1/2}$ and Γ, respectively, it is trace class [6, Proposition 1.1.2]. Since H_Γ is self-adjoint and compact, we have from Fredholm operator theory that the operator $I + H_\Gamma$ is invertible iff $\ker H_\Gamma = \{0\}$. The latter is the case since H_Γ is positive which implies $\langle (I + H_\Gamma)u, u \rangle \geq \langle u, u \rangle$. Hence, the inverse $(I + H_\Gamma)^{-1}$ exists and, moreover, $(I + H_\Gamma)^{-1} \in \mathcal{L}_+(\mathcal{H})$ with $\|(I + H_\Gamma)^{-1}\| \leq 1$. The self-adjointness and positivity of C_Γ follow immediately and since C_Γ is a composition of two nonsingular Hilbert–Schmidt operators and a nonsingular bounded operator, $C^{1/2}$ and $(I + H)^{-1}$, respectively, it is trace class and nonsingular as well. $\qquad \square$

By Proposition 1 we can use the covariance operator C_Γ for constructing proposal kernels. Specifically, we consider

$$P(u, \cdot) = N(Au, s^2 C_\Gamma), \qquad s \in [0, 1), \ \Gamma \in \mathcal{L}_+(\mathcal{H}), \tag{14}$$

where $A \colon \mathcal{H} \to \mathcal{H}$ denotes a suitably chosen bounded linear operator on $\mathcal{H}$. Here A should be chosen such that P is μ_0-reversible, which means that a Metropolis kernel with proposal P is μ-reversible, see Sect. 2.2. By applying (34) we obtain in this setting

$$P(u, \mathrm{d}v)\, \mu_0(\mathrm{d}u) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} C & CA^* \\ AC & ACA^* + s^2 C_\Gamma \end{bmatrix} \right)$$

and

$$P(v, \mathrm{d}u)\, \mu_0(\mathrm{d}v) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} ACA^* + s^2 C_\Gamma & AC \\ CA^* & C \end{bmatrix} \right).$$

Thus, for satisfying (6) we need to choose A so that

$$AC = CA^*, \qquad ACA^* + s^2 C_\Gamma = C. \tag{15}$$

By straightforward calculation we obtain as the formal solution to (15)

$$A = A_\Gamma = C^{1/2} \sqrt{I - s^2 (I + H_\Gamma)^{-1}} C^{-1/2}. \tag{16}$$

The following lemma ensures that this choice of A yields a well-defined bounded linear operator on $\mathcal{H}$.

Lemma 2 *Let the assumptions of Proposition 1 be satisfied and let $s \in [0, 1)$. Then* (16) *defines a bounded linear operator $A_\Gamma : \operatorname{Im} C^{1/2} \to \mathcal{H}$.*

The well definedness of $A_\Gamma : \operatorname{Im} C^{1/2} \to \mathcal{H}$ follows rather easily, whereas its boundedness is not trivial. Namely, one easily can construct a bounded $B \in \mathcal{L}(\mathcal{H})$ such that $C^{1/2} B C^{-1/2}$ is unbounded on $\operatorname{Im} C^{1/2}$. Since the proof of Lemma 2 is rather technical, it is postponed to "Proof of Lemma 2" appendix.

Lemma 2 allows us now to extend A_Γ to $\mathcal{H}$ by continuation, because the Cameron–Martin space $\operatorname{Im} C^{1/2}$ is a dense subspace of $\mathcal{H}$. For simplicity we denote this continuous extension again by $A_\Gamma : \mathcal{H} \to \mathcal{H}$.

Definition 3 (*gpCN proposal*) For $s \in [0, 1)$ and $\Gamma \in \mathcal{L}_+(\mathcal{H})$ the *generalized pCN proposal kernel* is given by

$$P_\Gamma(u, \cdot) := N(A_\Gamma u, s^2 C_\Gamma). \tag{17}$$

For the zero operator $\Gamma = 0$ we recover the pCN proposal. By Lemma 2 and the arguments given in Sect. 2.2 we obtain the following important result.

Corollary 4 *Let $\mu_0 = N(0, C)$ and μ be given by* (1). *Let the assumptions of Lemma 2 be satisfied. Then, a gpCN proposal kernel P_Γ given by* (17) *and an acceptance probability $\alpha(u, v) = \min\{1, \exp(\Phi(u) - \Phi(v))\}$ induce a μ-reversible Metropolis kernel denoted by M_Γ.*

For simplicity we also call the Metropolis algorithm with transition kernel M_Γ just *gpCN Metropolis*. There are connections of the gpCN Metropolis to other recently developed Metropolis algorithms for general Hilbert spaces which also use more sophisticated choices for the proposal than the pCN proposal. The following two remarks address these connections.

Remark 5 The gpCN proposals form a subclass of the *operator weighted proposals* introduced in [5,17]. The particular form of the gpCN proposal allows us to derive properties such as boundedness of the "proposal mean operator" A_Γ and the convergence of the resulting Markov chain, see Sect. 4. These issues were left open in [5,17].

Remark 6 In [23] the authors compute a Gaussian measure $\mu_* = N(m_*, C_*)$ which comes closest to μ w.r.t. the Kullback–Leibler distance. The admissible class of Gaussian measures considered there is closely related to our parameterized proposal covariances C_Γ, although their class of Gaussian measures is slightly larger. The measure μ_* is then used to construct a proposal kernel $P_*(u, \cdot) = N(m_* + \sqrt{1 - s^2}(u - m_*), s^2 C_*)$ for Metropolis algorithms. Note that P_* is not μ_0-reversible but μ_*-reversible, since it is a pCN proposal given the prior μ_*. In order to obtain a μ-reversible Metropolis kernel the authors need to adapt the acceptance probability by including terms of $\frac{d\mu_*}{d\mu_0}$, cf. Sect. 5. Thus, the authors of [23] also use a different

covariance operator than the prior covariance in a pCN proposal in order to increase the efficiency of the Metropolis algorithm. The difference to our approach is the way they ensure the μ-reversibility of the algorithm. They keep the mean of the original pCN proposal and modify the acceptance probability, whereas we modify also the mean of the proposal to maintain its μ_0-reversibility and, therefore, can leave the acceptance probability unchanged.

3.3 Numerical Illustrations

We illustrate the gpCN Metropolis algorithm for approximating samples of a posterior distribution in Bayesian inference. In particular, we compare different Metropolis algorithms and investigate which of these perform independently of the state-space dimension and of the variance of the involved noise.

We consider the same setting and inference problem as in [23, Section 6.1]. Assume noisy observations $y_j = p(0.2j) + \varepsilon_j$ with $j = 1, \ldots, 4$, of the solution p of

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{u(x)}\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = 0, \qquad p(0) = 0, \ p(1) = 2, \tag{18}$$

on $D = [0, 1]$ are given and we want to infer u. Here the ε_j are independent realizations of the normal distribution $N(0, \sigma_\varepsilon^2)$. We place a Gaussian prior $N(0, \Delta^{-1})$ with $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ on the completion $\mathcal{H}_c$ of $H_0^1(D) \cap H^2(D)$ in $L^2(D)$. Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and let $U : \Omega \to \mathcal{H}_c \subset L^2(D)$ be a random function with distribution $N(0, \Delta^{-1})$. This allows us to represent the random function U as

$$U(\omega)(x) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty}\xi_k(\omega)\sin(k\pi x), \qquad \xi_k \sim N(0, k^{-2}), \tag{19}$$

$\mathbb{P}$-a.s. where all random variables ξ_k are independent. Thus, inference on u is equivalent to inference on $\boldsymbol{\xi} = (\xi_k)_{k\in\mathbb{N}}$. This leads to the prior μ_0 for $\boldsymbol{\xi}$ on $\mathcal{H} := \ell^2(\mathbb{R})$ given by $\mu_0 = N(0, C)$ with $C = \mathrm{diag}\{k^{-2} : k \in \mathbb{N}\}$. Further, we denote by μ the resulting conditional distribution of $\boldsymbol{\xi}$ given the observed data $y_1, \ldots, y_4$. The measure μ is given by a density of the form (1) with Φ as in (10) where $\Sigma = \sigma_\varepsilon^2 I$ and $G(\boldsymbol{\xi})$ is the mapping

$$\boldsymbol{\xi} \mapsto u(\cdot, \boldsymbol{\xi}) \mapsto p(\cdot, \boldsymbol{\xi}) \mapsto (p(0.2j, \boldsymbol{\xi}))_{j=1}^4.$$

We test the performance of μ-reversible Metropolis algorithms for computing expectations w.r.t. μ of a function $f : \ell^2(\mathbb{R}) \to \mathbb{R}$. We consider four Metropolis algorithms denoted by RW, pCN, GN-RW and gpCN with different proposal kernels:

- RW: Gaussian random walk proposal $P_1(\boldsymbol{\xi}, \cdot) = N(\boldsymbol{\xi}, s^2 C)$,
- pCN: pCN proposal $P_2(\boldsymbol{\xi}, \cdot) = N(\sqrt{1 - s^2}\boldsymbol{\xi}, s^2 C)$,
- GN-RW: Gauss–Newton random walk proposal $P_3(\boldsymbol{\xi}, \cdot) = N(\boldsymbol{\xi}, s^2 C_\Gamma)$,
- gpCN: gpCN proposal $P_4(\boldsymbol{\xi}, \cdot) = N(A_\Gamma \boldsymbol{\xi}, s^2 C_\Gamma)$.

Here we choose $\Gamma = \sigma_\varepsilon^{-2} L L^\top$ with $L = \nabla G(\boldsymbol{\xi}_{\mathrm{MAP}})$ and

$$\boldsymbol{\xi}_{\mathrm{MAP}} = \underset{\xi \in \mathrm{Im}\, C^{1/2}}{\mathrm{argmin}} \left(\sigma_\varepsilon^{-2} |\boldsymbol{y} - G(\boldsymbol{\xi})|^2 + \|C^{-1/2}\boldsymbol{\xi}\|^2 \right).$$

The solution of (18) is given by $p(x) = 2S_x(\mathrm{e}^{-u})/S_1(\mathrm{e}^{-u})$ with $S_x(f) = \int_0^x f(y)\mathrm{d}y$, and thus, the gradient $\nabla G(\boldsymbol{\xi})$ can be easily computed by differentiating the explicit formula for p w.r.t. $\boldsymbol{\xi}$.[1] Furthermore, we apply the Levenberg–Marquardt algorithm to solve the above optimization problem for the MAP estimator $\boldsymbol{\xi}_{\mathrm{MAP}}$. For all Metropolis algorithms we tune s such that the average acceptance rate is about 0.25.[2] As a metric for comparison we consider and estimate the *effective sample size*

$$\mathrm{ESS} = \mathrm{ESS}(n, f, (\boldsymbol{\xi}_k)_{k \in \mathbb{N}}) = n \left[1 + 2 \sum_{k \geq 0} \gamma_f(k) \right]^{-1}.$$

Here n is the number of samples taken from a Markov chain $(\boldsymbol{\xi}_k)_{k \in \mathbb{N}}$ with, say, a Metropolis transition kernel M and γ_f denotes the autocorrelation function $\gamma_f(k) = \mathrm{Corr}(f(\boldsymbol{\xi}_{n_0}), f(\boldsymbol{\xi}_{n_0+k}))$ for a quantity of interest f.

The value of ESS corresponds to the number of independent samples w.r.t. μ which would approximately yield the same mean squared error as the MCMC estimator $S_{n,n_0}(f)$ for computing $\mathbb{E}_\mu(f)$. This can be justified under the assumption that $\boldsymbol{\xi}_{n_0} \sim \mu$, since then by virtue of [26, Proposition 3.26] we have

$$\lim_{n \to \infty} n \cdot \mathbb{E} \left| S_{n,n_0}(f) - \mathbb{E}_\mu(f) \right|^2 = \sigma_{f,M}^2,$$

$$1 + 2 \sum_{k \geq 0} \gamma_f(k) = \frac{\sigma_{f,M}^2}{\mathbb{E}_\mu(f^2) - \mathbb{E}_\mu(f)^2}$$

where $\sigma_{f,M}^2$ denotes the asymptotic variance of the estimator $S_{n,n_0}(f)$ as in Sect. 2.1.

For numerical simulations we use a uniform discretization of $[0, 1]$ with $\Delta x = 2^{-10}$ and apply the trapezoidal rule for evaluating integrals w.r.t. $\mathrm{d}x$. Furthermore, we truncate the expansion (19) after N terms where we vary N in order to test the Metropolis algorithms for dimension-independent performance. The noise-free observations are generated by $u(x) = 2\sin(2\pi x)$. We also consider different noise levels σ_ε to examine the effect of smaller variances σ_ε^2, leading to more concentrated posterior distributions μ, to the performance of the Metropolis algorithms. In all cases we take $n_0 = 10^5$ as burn-in length and $n = 10^6$ as sample size. We use $f(\boldsymbol{\xi}) := \int_0^1 \mathrm{e}^{u(x,\boldsymbol{\xi})}\,\mathrm{d}x$ as

---

[1] In general elliptic PDEs can be solved in a weak sense by variational methods. Then adjoint methods known from PDE-constrained optimization and parameter identification can be employed to compute $\nabla G(\boldsymbol{\xi})$, see [31, Chapter 6] for details.

[2] The empirical performance of each algorithm was best for this particular tuning.
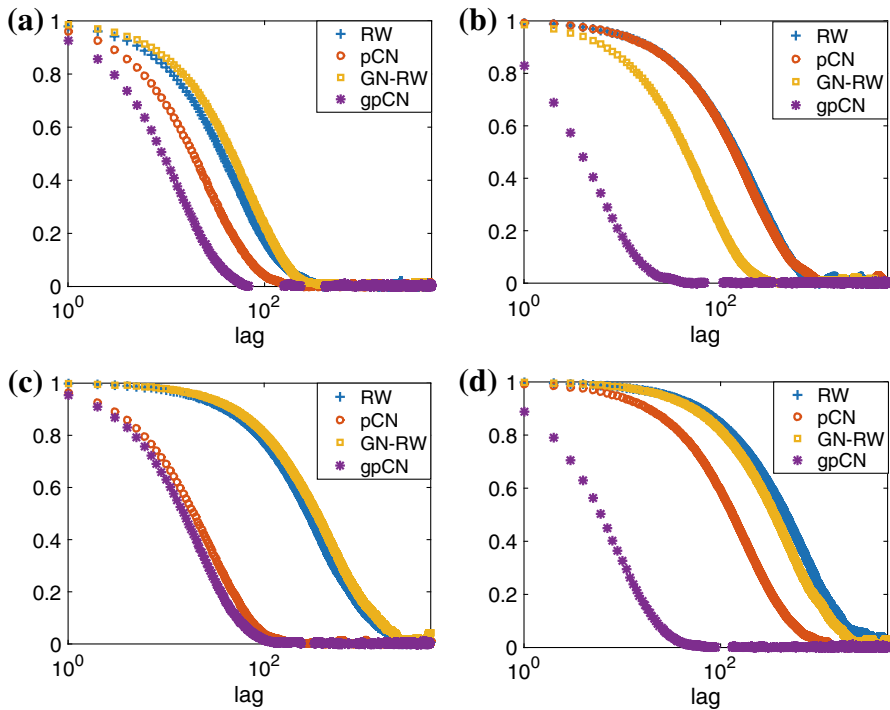
Fig. 2 Autocorrelation of f given samples generated by the four Metropolis algorithms denoted by RW, pCN, GN-RW and gpCN for: **a** state dimension $N = 50$ and noise standard deviation $\sigma_\varepsilon = 0.1$; **b** $N = 50$ and $\sigma_\varepsilon = 0.01$; **c** $N = 400$ and $\sigma_\varepsilon = 0.1$; **d** $N = 400$ and $\sigma_\varepsilon = 0.01$

the quantity of interest.[3] To estimate the ESS we use the initial monotone sequence estimators;[4] for details we refer to [10, Section 3.3].

The results of the simulations are illustrated in Figs. 2 and 3. The former displays the estimated autocorrelation functions γ_f resulting from the four Metropolis algorithms for $N = 50$ and $\sigma_\varepsilon = 0.1$ in (a), for $N = 50$ and $\sigma_\varepsilon = 0.01$ in (b), for $N = 400$ and $\sigma_\varepsilon = 0.1$ in (c) and for $N = 400$ and $\sigma_\varepsilon = 0.01$ in (d). In Fig. 3 we display the estimated ESS for varying $\sigma_\varepsilon = 0.1, 0.05, 0.025, 0.01$ with fixed $N = 100$ in (a) and varying $N = 50, 100, 200, 400, 800$ with fixed $\sigma_\varepsilon = 0.1$ in (b).

We see in both figures that the performance of pCN and gpCN is independent of the dimension and only GN-RW and gpCN perform robustly w.r.t. the noise variance. Thus, the gpCN Metropolis seems to be the only algorithm with both desirable properties. Intuitively, the variance-independent performance might come from the fact that our choice of C_Γ incorporates the noise covariance $\sigma_\varepsilon^2 I$ in a way as the posterior covariance might depend on. Thus, the smaller σ_ε becomes, i.e., the more pronounced

---

[3] We also studied other functions such as $f(\xi) = \xi_1$, $f(\xi) = \max_x e^{u(x,\xi)}$ and $f(\xi) = p(0.5, \xi)$ but the results of the comparison were essentially the same.

[4] We also estimated the ESS by batch means (100 batches of size 10^4) to control our simulations. This led to similar results.
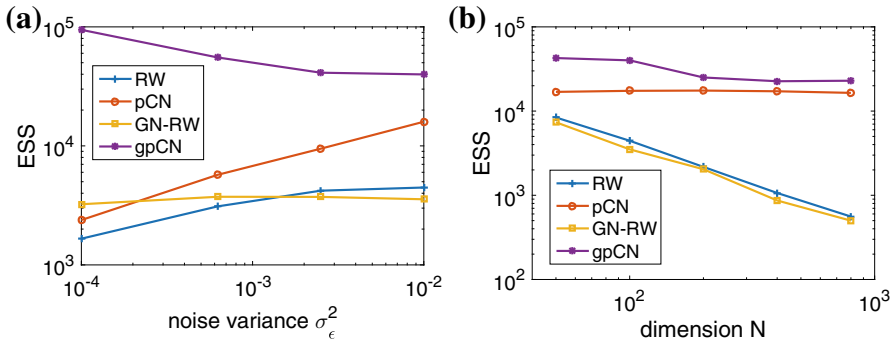
Fig. 3 Dependence of empirical ESS for each Metropolis algorithm RW, pCN, GN-RW and gpCN w.r.t.: **a** noise variance with fixed state dimension $N = 100$; **b** state dimension with fixed noise variance $\sigma_\varepsilon^2 = 0.01$

the change from prior to posterior is, the more pronounced is also the adaptation in the proposal covariance by $C_\Gamma = (C^{-1} + \sigma_\varepsilon^{-2} L L^\top)^{-1}$. Moreover, the gpCN performs best among the four algorithms also in absolute terms of the ESS.

4 Qualitative Comparison of gpCN Metropolis

In this section we develop qualitative comparison arguments for Metropolis algorithms in a general setting and apply these results to the gpCN Metropolis algorithms. In particular, we relate the existence of a spectral gap for the gpCN to the existence of a spectral gap of the pCN Metropolis. Here it is worth mentioning that in [13] sufficient conditions for the latter were proven under additional regularity assumptions on the function Φ in (1). With our approach we do not need to rely on those conditions and will benefit from any improvement of the results stated in [13].

We start with stating a general comparison result for the spectral gaps of Metropolis algorithms with equivalent proposals. We then verify the corresponding assumptions for the gpCN Metropolis: positivity and equivalence to the pCN proposal. In order to derive our main theorem, we consider in Sect. 4.4 restrictions of the target measure μ to arbitrary R-balls in $\mathcal{H}$ and prove convergence of the gpCN Metropolis to these restricted measures.

4.1 Comparison of Spectral Gaps

Let K be a μ-reversible transition kernel on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, i.e., the associated Markov operator $K: L_2(\mu) \to L_2(\mu)$ is self-adjoint. Let the largest element of the spectrum $\operatorname{spec}(K \mid L_2^0(\mu))$ be given by

$$\Lambda(K) := \sup\{\lambda : \lambda \in \operatorname{spec}(K \mid L_2^0(\mu))\}$$

and define the *conductance of K* (w.r.t. μ) by

$$\varphi(K) := \inf_{\mu(A)\in(0,1/2]} \frac{\int_A K(u, A^c)\mu(\mathrm{d}u)}{\mu(A)}.$$

Under the assumptions above the Cheeger inequality for Markov operators, see [18], given by

$$\frac{\varphi(K)^2}{2} \leq 1 - \Lambda(K) \leq 2\varphi(K) \tag{20}$$

provides a useful relation between $\Lambda(K)$ and the conductance $\varphi(K)$.

Let us assume that M_1 and M_2 are μ-reversible transition kernels of Metropolis algorithms with the same acceptance probability α and proposals P_1 and P_2, respectively. Then, we obtain the following result.

Lemma 7 *Let μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and for $i = 1, 2$ let*

$$M_i(u, \mathrm{d}v) = \alpha(u, v) P_i(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \int_{\mathcal{H}} (1 - \alpha(u, w)) P_i(u, \mathrm{d}w)$$

be μ-reversible Metropolis kernels. Assume that for any $u \in \mathcal{H}$ the Radon–Nikodym derivative of $P_1(u, \mathrm{d}v)$ w.r.t. $P_2(u, \mathrm{d}v)$ exists, i.e., the proposal kernels admit a density

$$\rho(u, v) = \frac{\mathrm{d}P_1(u)}{\mathrm{d}P_2(u)}(v), \quad u, v \in \mathcal{H}.$$

If for a number $p > 1$ we have

$$\kappa_p := \sup_{\mu(A)\in(0,1/2]} \frac{\int_A \int_{A^c} \rho(u, v)^p P_2(u, \mathrm{d}v)\, \mu(\mathrm{d}u)}{\mu(A)} < \infty, \tag{21}$$

then

$$\varphi(M_1) \leq \kappa_p^{1/p}\, \varphi(M_2)^{(p-1)/p}.$$

Proof Let $A \in \mathcal{B}(\mathcal{H})$ with $\mu(A) \in (0, 1/2]$. Further, let $q = p/(p-1)$ such that $1/q + 1/p = 1$. Then

$$\int_A M_1(u, A^c)\, \mu(\mathrm{d}u) = \int_{\mathcal{H}} \int_{\mathcal{H}} \mathbf{1}_{A^c}(v)\mathbf{1}_A(u)\, \alpha(u, v)\, P_1(u, \mathrm{d}v)\, \mu(\mathrm{d}u)$$

$$= \int_{\mathcal{H}} \int_{\mathcal{H}} \mathbf{1}_{A^c}(v)\mathbf{1}_A(u)\, \alpha(u, v)\, \rho(u; v)\, P_2(u, \mathrm{d}v)\, \mu(\mathrm{d}u).$$

Note that $P_2(u, \mathrm{d}v)\mu(\mathrm{d}u)$ is a probability measure on $(\mathcal{H} \times \mathcal{H}, \mathcal{B}(\mathcal{H} \times \mathcal{H}))$ and we can apply Hölder's inequality according to this measure with parameters p and q. Thus, by using $\alpha(u, v) = \alpha(u, v)^{1/q}\alpha(u, v)^{1/p}$ we obtain

$$\int_A M_1(u, A^c) \, \mu(\mathrm{d}u)$$

$$\leq \left(\int_A M_2(u, A^c) \, \mu(\mathrm{d}u) \right)^{1/q} \left(\int_A \int_{A^c} \rho(u, v)^p \alpha(u, v) \, P_2(u, \mathrm{d}v) \, \mu(\mathrm{d}u) \right)^{1/p}$$

$$\leq \left(\int_A M_2(u, A^c) \, \mu(\mathrm{d}u) \right)^{1/q} \left(\int_A \int_{A^c} \rho(u, v)^p \, P_2(u, \mathrm{d}v) \, \mu(\mathrm{d}u) \right)^{1/p}$$

Dividing by $\mu(A)$, applying $\mu(A)^{-1} = \mu(A)^{-1/q} \, \mu(A)^{-1/p}$ and taking the infimum yield

$$\varphi(M_1) \leq \varphi(M_2)^{1/q} \kappa_p^{1/p}.$$

$\square$

Employing comparison inequalities in terms of the conductance is not an entirely new idea, see for example [19, Proof of Theorem 4]. There the authors obtained a conductance inequality for transition kernels with bounded Radon–Nikodym derivatives w.r.t. each other. An immediate consequence of Lemma 7 and (20) is the following theorem.

Theorem 8 (Spectral gap comparison) *Let the assumptions of Lemma 7 be satisfied and let the Markov operators associated with M_1 and M_2 be positive and self-adjoint on $L_2(\mu)$. Then*

$$\left(\frac{\mathrm{gap}(M_1)}{2} \right)^p \leq \kappa_p \, (2 \, \mathrm{gap}(M_2))^{(p-1)/2}.$$

We apply Theorem 8 to prove our convergence result for the gpCN Metropolis. We therefore verify in the following section the condition that the corresponding Markov operator is positive.

4.2 Positivity of Metropolis with Gaussian Proposals

Recall that $\langle f, g \rangle_\mu = \int_{\mathcal{H}} fg \, \mathrm{d}\mu$ denotes the inner-product of $L_2(\mu)$ and that a Markov operator $K : L_2(\mu) \to L_2(\mu)$ is positive if $\langle Kf, f \rangle_\mu \geq 0$ for all $f \in L_2(\mu)$.

Lemma 9 (Positivity of proposals) *Let $\mu_0 = N(0, C)$ be a Gaussian measure on a separable Hilbert space $\mathcal{H}$ and let $P(u, \cdot) = N(Au, Q)$ be a μ_0-reversible proposal kernel with a bounded, linear operator $A : \mathcal{H} \to \mathcal{H}$. If there exists a bounded, linear operator $B : \mathcal{H} \to \mathcal{H}$ such that*

$$B^2 = A, \qquad BC = CB^*,$$

and $D := C - BCB^*$ is positive and trace class, then the Markov operator associated with the proposal P is positive on $L_2(\mu_0)$.

Proof Because of the assumptions on B and D we obtain that the proposal kernel $P_1(u, \cdot) = N(Bu, D)$ is well defined. Further, since $BCB^* + D = C$ we derive

$$P_1(u, dv)\mu_0(du) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} C & CB^* \\ BC & C \end{bmatrix}\right),$$

which leads by $BC = CB^*$ to the μ_0-reversibility of P_1 and, thus, to the self-adjointness of its associated Markov operator in $L_2(\mu_0)$. It remains to prove that $P_1^2 = P$ holds for the associated Markov operators which then immediately yields the assertion. The equality of the Markov operators is equivalent to the equality of the measures $P_1^2(u, \cdot)$ and $P(u, \cdot)$ for all $u \in \mathcal{H}$. In order to show that $P_1^2(u, \cdot) = P(u, \cdot)$ for all $u \in \mathcal{H}$, we take $(\xi_n)_{n\in\mathbb{N}}$ to be an i.i.d. sequence with $\xi_1 \sim N(0, D)$ and construct an auxiliary Markov chain by

$$X_{n+1} = BX_n + \xi_n, \quad n \geq 1,$$

where $X_1 = u$ for an arbitrary $u \in \mathcal{H}$. The transition kernel of the chain $(X_n)_{n\in\mathbb{N}}$ is the kernel P_1. In particular, for $G \in \mathcal{B}(\mathcal{H})$ holds $\mathbb{P}[X_3 \in G] = P_1^2(u, G)$. By

$$X_3 = BX_2 + \xi_2 = B^2u + B\xi_1 + \xi_2$$

and $B\xi_1 + \xi_2 \sim N(0, BDB^* + D)$ we obtain $X_3 \sim N(B^2u, BDB^* + D)$. Due to the assumptions we have $B^2 = A$ and

$$BDB^* + D = B(C - BCB^*)B^* + C - BCB^* = C - ACA^*.$$

The last step $C - ACA^* = Q$ follows by the assumed μ_0-reversibility of P, because we know from Sect. 3.2 that P being μ_0-reversible is equivalent to A and Q satisfying $AC = CA^*$ and $ACA^* + Q = C$. We thus arrive at $X_3 \sim N(Au, Q)$ which proves $P_1^2(u, \cdot) = P(u, \cdot)$. □

The next lemma extends the previous result to Markov operators associated with Metropolis algorithms. The proof follows by the same line of arguments as developed in [27, Section 3.4] and is therefore omitted.

Lemma 10 (Positivity of Metropolis kernels) *Let μ be a measure on $\mathcal{H}$ given by (1) and let P be a μ_0-reversible proposal kernel whose associated Markov operator is positive on $L_2(\mu_0)$. Then the Markov operator associated with a μ-reversible Metropolis kernel*

$$M(u, dv) = \alpha(u, v)P(u, dv) + \delta_u(dv)\int_{\mathcal{H}} (1 - \alpha(u, w))P(u, dw)$$

with $\alpha(u, v) = \min\{1, \frac{d\mu}{d\mu_0}(v)\frac{d\mu_0}{d\mu}(u)\}$ is positive on $L_2(\mu)$.

The previous two lemmas lead to the following result about the gpCN Metropolis.

Theorem 11 (Positivity of gpCN Metropolis) *Let $\mu_0 = N(0, C)$ and μ as in* (1) *and let M_Γ denote the gpCN Metropolis kernel as in Corollary 4. Then the associated Markov operator M_Γ is self-adjoint and positive on $L_2(\mu)$.*

Proof It is enough to verify the assumptions of Lemma 9 for the gpCN proposal. Recall that $P_\Gamma(u, \cdot) = N(A_\Gamma u, s^2 C_\Gamma)$ which is μ_0-reversible by construction with bounded $A_\Gamma = C^{1/2}\sqrt{I - s^2(I + H_\Gamma)^{-1}}C^{-1/2}$. By choosing

$$B := C^{1/2}\sqrt[4]{I - s^2(I + H_\Gamma)^{-1}}C^{-1/2},$$

we obtain $B^2 = A_\Gamma$ and $BC = CB^*$. Moreover,

$$D = C - BCB^* = C^{1/2}(I - \sqrt{I - s^2(I + H_\Gamma)^{-1}})C^{1/2}.$$

The eigenvalues of $I - \sqrt{I - s^2(I + H_\Gamma)^{-1}}$ take the form $1 - \sqrt{1 - \frac{s^2}{1+\lambda}} \geq 0$ with $\lambda \geq 0$ being an eigenvalue of H_Γ. Thus, $I - \sqrt{I - s^2(I + H_\Gamma)^{-1}}$ is positive and bounded which yields D being positive and trace class since D is then a product of two Hilbert–Schmidt and one bounded operator. Thus, the conditions of Lemma 9 are satisfied and the assertion follows. □

4.3 Density Between pCN and gpCN Proposal

In this section we show that for any state $u \in \mathcal{H}$ the gpCN proposal is equivalent to the pCN proposal in the sense of measures. Moreover, we will also derive an integrability result for the corresponding density. For proving the equivalence we need the following technical result.

Lemma 12 *Let the assumptions of Corollary 4 be satisfied and define the bounded, linear operator $\Delta_\Gamma : \mathcal{H} \to \mathcal{H}$ by*

$$\Delta_\Gamma := A_0 - A_\Gamma = \sqrt{1 - s^2}I - C^{1/2}\sqrt{I - s^2(I + H_\Gamma)^{-1}}C^{-1/2}. \quad (22)$$

Then $\operatorname{Im} \Delta_\Gamma \subseteq \operatorname{Im} C^{1/2}$, i.e., $C^{-1/2}\Delta_\Gamma$ is a bounded operator on $\mathcal{H}$.

The proof of this lemma can be found in "Proof of Lemma 12" appendix. It is similar to the proof of Lemma 2 and again rather technical. However, Lemma 12 ensures that we can apply the Cameron–Martin theorem, Theorem 21 in "Gaussian Measures" appendix, in the proof of the following result. The other main tool for deriving the next theorem is a variant of the Feldman–Hajek theorem as stated in Theorem 22 in "Gaussian Measures" appendix.

Theorem 13 (Density of pCN w.r.t. gpCN) *With the notation and assumptions of Corollary 4 holds the following.*

1. *The measures $\mu_0 = N(0, C)$ and $\mu_\Gamma = N(0, C_\Gamma)$ are equivalent with*

$$\pi_\Gamma(v) := \frac{d\mu_0}{d\mu_\Gamma}(v) = \frac{\exp\left(\frac{1}{2}\langle \Gamma v, v \rangle\right)}{\sqrt{\det(I + H_\Gamma)}}. \tag{23}$$

2. *For $u \in \mathcal{H}$ the measures $P_0(u, \cdot)$ and $P_\Gamma(u, \cdot)$ are equivalent with*

$$\frac{dP_0(u)}{dP_\Gamma(u)}(v) = \pi_{CM}\left(\Delta_\Gamma u, \frac{1}{s}(v - A_\Gamma u)\right) \pi_\Gamma\left(\frac{1}{s}(v - A_\Gamma u)\right) \tag{24}$$

where Δ_Γ as in (22) and

$$\pi_{CM}(h, v) := \exp\left(-\frac{1}{2}\|C^{-1/2}h\|^2 + \langle C^{-1}h, v \rangle\right). \tag{25}$$

(The subscript in π_{CM} indicates the Cameron–Martin formula.)

Proof We prove (23) by verifying the assumptions of Theorem 22 from "Gaussian Measures" appendix. We observe

$$I - C^{-1/2}C_\Gamma C^{-1/2} = I - (I + H_\Gamma)^{-1}$$

and set $T_\Gamma := I - (I + H_\Gamma)^{-1}$. The eigenvalues $(t_n)_{n\in\mathbb{N}}$ of the self-adjoint operator T_Γ are given by

$$t_n = 1 - \frac{1}{1 + \lambda_n} = \frac{\lambda_n}{1 + \lambda_n} < 1$$

where $(\lambda_n)_{n\in\mathbb{N}}$ are the eigenvalues of the positive trace class operator H_Γ. Thus, T_Γ is also trace class and satisfies $\langle T_\Gamma u, u \rangle < \|u\|^2$ for any $u \in \mathcal{H}$. Then, the assertion follows by Theorem 22 and

$$T_\Gamma(I - T_\Gamma)^{-1} = \left(I - (I + H_\Gamma)^{-1}\right)(I + H_\Gamma) = H_\Gamma$$

as well as

$$\langle H_\Gamma C^{-1/2}v, C^{-1/2}v \rangle = \langle \Gamma v, v \rangle \qquad \forall v \in \mathcal{H}.$$

To show the equivalence of $P_0(u, \cdot)$ and $P_\Gamma(u, \cdot)$ for any $u \in \mathcal{H}$ we introduce the auxiliary kernel $K_\Gamma(u, \cdot) = N(A_\Gamma u, s^2 C)$. The first assertion and a simple change of variables, see Lemma 23 in the appendix, lead to

$$\frac{dK_\Gamma(u)}{dP_\Gamma(u)}(v) = \pi_\Gamma\left(\frac{1}{s}(v - A_\Gamma u)\right), \qquad u, v \in \mathcal{H}.$$

Thus, it remains to prove the equivalence of $K_\Gamma(u, \cdot)$ and $P_0(u, \cdot)$ for any $u \in \mathcal{H}$. By the Cameron–Martin formula, see Theorem 21 in "Gaussian Measures" appendix, this holds iff

$$\mathrm{Im}(A_\Gamma - \sqrt{1 - s^2}I) \subseteq \mathrm{Im}(C^{1/2})$$

which was shown in Lemma 12. Now Theorem 21 combined with a change of variables, see Lemma 23, yields

$$\frac{\mathrm{d}P_0(u)}{\mathrm{d}K_\Gamma(u)}(v) = \pi_{\mathrm{CM}}\left(\left(\sqrt{1 - s^2}I - A_\Gamma\right)u, \frac{1}{s}(v - A_\Gamma u)\right)$$

and the assertion follows by

$$\frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v) = \frac{\mathrm{d}P_0(u)}{\mathrm{d}K_\Gamma(u)}(v)\frac{\mathrm{d}K_\Gamma(u)}{\mathrm{d}P_\Gamma(u)}(v).$$

$\square$

Note that Theorem 13 implies that for any $\Gamma_1, \Gamma_2 \in \mathcal{L}_+(\mathcal{H})$ there exists a density between the two gpCN proposals $P_{\Gamma_1}(u)$ and $P_{\Gamma_2}(u)$. However, for the application of Theorem 8 we still have to verify condition (21). This is partly addressed in the following result.

Theorem 14 (Integrability of gpCN density) *Let the assumptions of Lemma 12 be satisfied and set*

$$\rho_\Gamma(u, v) := \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v), \quad u, v \in \mathcal{H}.$$

Then, for any $0 < p < 1 + \frac{1}{2\|H_\Gamma\|}$ there exist constants $c = c(p, H_\Gamma) < \infty$ and $b = b(p, \|C^{-1/2}\Delta_\Gamma\|) < \infty$ such that

$$\int_{\mathcal{H}} \rho_\Gamma^p(u, v)\, P_\Gamma(u, \mathrm{d}v) \leq c \exp\left(\frac{b}{2}\|u\|^2\right).$$

Proof We employ the same notation as in Theorem 13, i.e., let $\mu_0 = N(0, C)$ and $\mu_\Gamma = N(0, C_\Gamma)$ as well as π_Γ and π_{CM} be as in (23) and (25), respectively. By Theorem 13 we know

$$\rho_\Gamma(u, v) = \pi_{\mathrm{CM}}\left(\Delta_\Gamma u, \frac{1}{s}(v - A_\Gamma u)\right)\pi_\Gamma\left(\frac{1}{s}(v - A_\Gamma u)\right).$$

By first applying a change of variables, see Lemma 23, and then the Cauchy–Schwarz inequality we obtain

$$
\begin{aligned}
\int_{\mathcal{H}} & \rho_\Gamma^p(u, v) \, P_\Gamma(u, \mathrm{d}v) \\
&= \int_{\mathcal{H}} \pi_{\mathrm{CM}}^p(\Delta_\Gamma u, v) \, \pi_\Gamma^p(v) \, \mu_\Gamma(\mathrm{d}v) \\
&= \int_{\mathcal{H}} \pi_{\mathrm{CM}}^p(\Delta_\Gamma u, v) \, \pi_\Gamma^{p-1}(v) \, \mu_0(\mathrm{d}v) \\
&\leq \left(\int_{\mathcal{H}} \pi_{\mathrm{CM}}^{2p}(\Delta_\Gamma u, v) \mu_0(\mathrm{d}v) \right)^{1/2} \left(\int_{\mathcal{H}} \pi_\Gamma^{2p-2}(v) \mu_0(\mathrm{d}v) \right)^{1/2}.
\end{aligned}
$$

Furthermore, we have by applying (35) from "Gaussian Measures" appendix

$$
\begin{aligned}
\int_{\mathcal{H}} \pi_{\mathrm{CM}}^{2p}\left(\Delta_\Gamma u, v\right) \mu_0(\mathrm{d}v) &= \int_{\mathcal{H}} \mathrm{e}^{-\frac{2p}{2}\|C^{-1/2}\Delta_\Gamma u\|^2} \, \mathrm{e}^{2p \langle C^{-1}\Delta_\Gamma u, v\rangle} \, \mu_0(\mathrm{d}v) \\
&= \exp\left((2p^2 - p)\|C^{-1/2}\Delta_\Gamma u\|^2 \right).
\end{aligned}
$$

We apply $\|C^{-1/2}\Delta_\Gamma u\| \leq \|C^{-1/2}\Delta_\Gamma\| \, \|u\|$ and set

$$
b := (2p^2 - p) \, \|C^{-1/2}\Delta_\Gamma\|^2.
$$

Note that $b \leq 0$ for $p \leq \frac{1}{2}$. Due to the assumptions on p we have

$$
\langle (2p - 2)H_\Gamma v, v\rangle < \frac{\langle H_\Gamma v, v\rangle}{\|H_\Gamma\|} \leq \|v\|^2, \qquad v \in \mathcal{H}.
$$

Thus, we can apply (36) from "Gaussian Measures" appendix and get

$$
\begin{aligned}
\int_{\mathcal{H}} \pi_\Gamma^{2p-2}(v)\mu_0(\mathrm{d}v) &= \int_{\mathcal{H}} \frac{\exp\left(\frac{1}{2}\langle (2p-2)H_\Gamma \, C^{-1/2}v, C^{-1/2}v\rangle\right)}{\det(I + H_\Gamma)^{(2p-2)/2}} \mu_0(\mathrm{d}v) \\
&= \left(\det(I - (2p-2)H_\Gamma) \, \det(I + H_\Gamma)^{2p-2} \right)^{-1/2} \\
&=: c^2.
\end{aligned}
$$

Since H_Γ is positive and trace class, $\det(I + H_\Gamma)$ is well defined (see "Gaussian Measures" appendix) and $\det(I + H_\Gamma) \in [1, \infty)$. Furthermore, due to $\langle (2p - 2)H_\Gamma v, v\rangle < \|v\|^2$, the eigenvalues of $(2p-2)H_\Gamma$ lie within $(-\infty, 1)$ which ensures that $\det(I - (2p-2)H_\Gamma) > 0$ and, hence, $0 < c^2 < \infty$. This proves the assertion. □

Thus, the above theorem allows us to estimate the integral in (21). We obtain for $0 < p < 1 + 1/(2\|H_\Gamma\|)$ that

$$
\int_A \int_{A^c} \rho_\Gamma(u; v)^p \, P_\Gamma(u, \mathrm{d}v) \, \mu(\mathrm{d}u) \leq c \int_A \exp\left(\frac{b}{2} \|u\|^2 \right) \mu(\mathrm{d}u).
$$

Unfortunately, if we divide the right-hand side by $\mu(A)$ and take the supremum over all $\{A : 0 < \mu(A) \leq 0.5\}$ this is unbounded. In the next section we introduce restrictions of the target measure for which we can circumvent this problem.

4.4 Restrictions of the Target Measure

In order to show boundedness of κ_p from (21) for the gpCN proposal we consider restrictions of the target measure to bounded sets. For appropriately chosen sets, the restricted measures become arbitrarily close to the target measure. Let $R \in (0, \infty]$ and set

$$\mathcal{H}_R := \{u \in \mathcal{H} : \|u\| < R\}.$$

Definition 15 (*Restricted measure*) Let μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and $R \in (0, \infty]$. We define its restriction to $\mathcal{H}_R$ as the probability measure μ_R on $\mathcal{H}$ given by

$$\mu_R(\mathrm{d}u) := \frac{1}{\mu(\mathcal{H}_R)} \boldsymbol{I}_{\mathcal{H}_R}(u)\mu(\mathrm{d}u). \tag{26}$$

For sufficiently large R the measure μ_R is close to μ, because

$$\|\mu_R - \mu\|_{\mathrm{tv}} = \int_{\mathcal{H}} \left| \frac{\mathrm{d}\mu_R}{\mathrm{d}\mu}(u) - 1 \right| \mathrm{d}\mu(u) = \mu(\mathcal{H}_R^c) + 1 - \mu(\mathcal{H}_R) = 2\mu(\mathcal{H}_R^c)$$

and since μ is a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ there exists for any $\varepsilon > 0$ a number $R > 0$ such that $2\mu(\mathcal{H}_R^c) < \varepsilon$. Let us mention here that restricted measures appear, for example, also in [3, Equation (3.5)] and in the recent work [15], in order to analyze the convergence of Metropolis–Hastings-based algorithms.

We ask now whether good convergence properties of a μ-reversible transition kernel K are inherited on a suitably modified μ_R-reversible transition kernel K_R.

Definition 16 (*Restricted transition kernel*) Let K be a transition kernel on $\mathcal{H}$ and $R \in (0, \infty]$. We define its restriction to $\mathcal{H}_R$ as the following transition kernel $K_R : \mathcal{H} \times \mathcal{B}(\mathcal{H}) \to [0, 1]$ given by

$$K_R(u, \mathrm{d}v) := \boldsymbol{I}_{\mathcal{H}_R}(v)\, K(u, \mathrm{d}v) + K(u, \mathcal{H}_R^c)\, \delta_u(\mathrm{d}v). \tag{27}$$

Note that if K is μ-reversible, then K_R is μ_R-reversible and if K is of Metropolis form (4), then so is K_R.

Proposition 17 *Let μ be a probability measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ and K be a μ-reversible transition kernel. Then for any $R > 0$ the transition kernel K_R given in (27) is μ_R-reversible with μ_R as in (26). Moreover, for a Metropolis kernel M of the form (4) the corresponding restricted kernel M_R is again a Metropolis kernel*

$$M_R(u, \mathrm{d}v) = \alpha_R(u, v) P(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \left(1 - \int_{\mathcal{H}} \alpha_R(u, w) P(u, \mathrm{d}w) \right)$$

with $\alpha_R(u, v) := \mathbf{1}_{\mathcal{H}_R}(v) \alpha(u, v)$.

Proof Recall that K is μ-reversible iff

$$\int_A K(u, B) \, \mu(\mathrm{d}u) = \int_B K(u, A) \, \mu(\mathrm{d}u), \qquad \forall A, B \in \mathcal{B}(\mathcal{H}).$$

Let $A, B \in \mathcal{B}(\mathcal{H})$. We have

$$\begin{aligned}
\int_A K_R(u, B) \, \mu_R(\mathrm{d}u) &= \int_A K(u, B \cap \mathcal{H}_R) \, \mu_R(\mathrm{d}u) + \int_{A \cap B} K(u, \mathcal{H}_R^c) \, \mu_R(\mathrm{d}u) \\
&= \frac{1}{\mu(\mathcal{H}_R)} \int_{A \cap \mathcal{H}_R} K(u, B \cap \mathcal{H}_R) \, \mu(\mathrm{d}u) \\
&\quad + \int_{A \cap B} K(u, \mathcal{H}_R^c) \, \mu_R(\mathrm{d}u).
\end{aligned}$$

Because of the μ-reversibility of K we can interchange A and B which leads to the first assertion. The second statement follows by

$$\begin{aligned}
M_R(u, \mathrm{d}v) &= \mathbf{1}_{\mathcal{H}_R}(v) M(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) M(u, \mathcal{H}_R^c) \\
&= \mathbf{1}_{\mathcal{H}_R}(v) \alpha(u, v) P(u, \mathrm{d}v) \\
&\quad + \delta_u(\mathrm{d}v) \left(1 - \int_{\mathcal{H}} \alpha(u, w) P(u, \mathrm{d}w) + \int_{\mathcal{H}_R^c} \alpha(u, w) P(u, \mathrm{d}w) \right) \\
&= \mathbf{1}_{\mathcal{H}_R}(v) \alpha(u, v) P(u, \mathrm{d}v) + \delta_u(\mathrm{d}v) \left(1 - \int_{\mathcal{H}_R} \alpha(u, w) P(u, \mathrm{d}w) \right).
\end{aligned}$$

$\square$

Now we ask whether a spectral gap of K on $L_2(\mu)$ implies a spectral gap of the Markov operator associated with K_R on $L_2(\mu_R)$. Note that

$$K_R f(u) = \int_{\mathcal{H}} f(v) \, K_R(u, \mathrm{d}v) = \int_{\mathcal{H}_R} f(v) \, K(u, \mathrm{d}v) + f(u) K(u, \mathcal{H}_R^c).$$

We have the following relation between $\|K_R\|_{\mu_R}$ and $\|K\|_\mu$.

Lemma 18 *With the notation and assumptions from above holds*

$$\|K_R\|_{\mu_R} \le \|K\|_\mu + \sup_{u \in \mathcal{H}_R} K(u, \mathcal{H}_R^c). \tag{28}$$

Furthermore, if the Markov operator K is positive on $L_2(\mu)$, then K_R is also positive on $L_2(\mu_R)$.

Proof For $f \in L_2(\mu_R)$ let

$$(Ef)(u) := \mathbf{1}_{\mathcal{H}_R}(u) f(u) \in L_2(\mu).$$

Note that $\|f\|_{2,\mu_R} = \frac{1}{\sqrt{\mu(\mathcal{H}_R)}} \|Ef\|_{2,\mu}$ and for $\int_{\mathcal{H}_R} f \, d\mu_R = 0$ follows $\int_{\mathcal{H}} Ef \, d\mu = 0$. Further, for any $f \in L_2(\mu_R)$ we have

$$\|K_R f\|_{2,\mu_R}^2 = \int_{\mathcal{H}_R} \left| \int_{\mathcal{H}_R} f(v) \, K(u, dv) + f(u) \, K(u, \mathcal{H}_R^c) \right|^2 \mu_R(du)$$

$$= \int_{\mathcal{H}_R} \left| \int_{\mathcal{H}} Ef(v) \, K(u, dv) + Ef(u) \, K(u, \mathcal{H}_R^c) \right|^2 \mu_R(du)$$

$$= \|K(Ef) + g \, Ef\|_{2,\mu_R}^2$$

with $g(u) := \mathbf{1}_{\mathcal{H}_R}(u) \, K(u, \mathcal{H}_R^c)$. Then

$$\frac{\|K_R f\|_{2,\mu_R}}{\|f\|_{2,\mu_R}} = \frac{\|K(Ef) + g \, Ef\|_{2,\mu_R}}{\|Ef\|_{2,\mu_R}} = \frac{\|E(K(Ef)) + g \, Ef\|_{2,\mu}}{\|Ef\|_{2,\mu}}$$

$$\leq \frac{\|K(Ef)\|_{2,\mu} + \|g \, Ef\|_{2,\mu}}{\|Ef\|_{2,\mu}}$$

$$\leq \frac{\|K(Ef)\|_{2,\mu}}{\|Ef\|_{2,\mu}} + \sup_{u \in \mathcal{H}_R} K(u, \mathcal{H}_R^c),$$

where we applied $\|Ef\|_{2,\mu} \leq \|f\|_{2,\mu}$ in the first inequality. By taking the supremum over all $f \in L_2^0(\mu_R)$ and because of $E(L_2^0(\mu_R)) \subseteq L_2^0(\mu)$ the first assertion follows. Moreover, we have for $f \in L_2(\mu_R)$ that

$$\langle K_R f, f \rangle_{\mu_R} = \int_{\mathcal{H}} K_R f(u) \, f(u) \, \mu_R(du)$$

$$= \int_{\mathcal{H}} \left(\int_{\mathcal{H}_R} f(v) \, K(u, dv) + f(u) \, K(u, \mathcal{H}_R^c) \right) f(u) \, \mu_R(du)$$

$$= \int_{\mathcal{H}} \int_{\mathcal{H}} (Ef)(v) \, K(u, dv) \, (Ef)(u) \, \frac{\mu(du)}{\mu(\mathcal{H}_R)}$$

$$+ \int_{\mathcal{H}} f^2(u) \, K(u, \mathcal{H}_R^c) \, \mu_R(du).$$

The second term is always positive since $f^2(u) \, K(u, \mathcal{H}_R^c) \geq 0$ for all $u \in \mathcal{H}$, and the first term coincides with $\langle K(Ef), Ef \rangle_\mu / \mu(\mathcal{H}_R)$. Thus, the second statement is proven. $\square$

Lemma 18 tells us that there exists an absolute spectral gap of K_R if there exists an absolute spectral gap of K and $\sup_{u \in \mathcal{H}_R} K(u, \mathcal{H}_R^c)$ is sufficiently small. Indeed, we can apply this result to the pCN Metropolis algorithm.

Theorem 19 (Spectral gap of restricted pCN Metropolis) *Let μ be as in (1) and let M_0 denote the μ-reversible pCN Metropolis kernel. If there exists a spectral gap of M_0 in $L_2(\mu)$, then for any $\varepsilon > 0$ there exists a number $R \in (0, \infty)$ such that $M_{0,R}$ possesses a spectral gap in $L_2(\mu_R)$, i.e.,*

$$\text{gap}(M_{0,R}) = 1 - \|M_{0,R}\|_{\mu_R} \geq \text{gap}(M_0) - \varepsilon,$$

where μ_R as in (26) and $M_{0,R}$ according to Definition 16.

Proof Given the results of Proposition 17 and Lemma 18 it suffices to prove that for any $\varepsilon > 0$ there exists an $R > 0$ such that $\sup_{u \in \mathcal{H}_R} M_0(u, \mathcal{H}_R^c) \leq \varepsilon$. We recall that the proposal kernel of M_0 is $P_0(u, \cdot) = N(\sqrt{1 - s^2}u, s^2C)$ and obtain with $\mu^s := N(0, s^2C)$ that

$$
\begin{aligned}
\sup_{u \in \mathcal{H}_R} M_0(u, \mathcal{H}_R^c) &\leq \sup_{u \in \mathcal{H}_R} P_0(u, \mathcal{H}_R^c) = \sup_{u \in \mathcal{H}_R} \int_{\|\sqrt{1-s^2}u+v\| \geq R} d\mu^s(v) \\
&\leq \sup_{u \in \mathcal{H}_R} \int_{\|\sqrt{1-s^2}u\|+\|v\| \geq R} d\mu^s(v) \\
&= \sup_{u \in \mathcal{H}_R} \int_{\|v\| \geq R - \sqrt{1-s^2}\|u\|} d\mu^s(v) \\
&\leq \int_{\|v\| \geq (1-\sqrt{1-s^2})R} d\mu^s(v) = \mu_0(\mathcal{H}_{R_s}^c)
\end{aligned}
$$

where $R_s = \frac{1-\sqrt{1-s^2}}{s} R$ and $\mu_0 = N(0, C)$. Again, since μ_0 is a probability measure on $\mathcal{H}$ we know that there exists a number R, such that $\mu_0(\mathcal{H}_{R_s}^c) \leq \varepsilon$. $\qquad\square$

4.5 Spectral Gap of Restricted gpCN Metropolis

Now, we are able to formulate and to prove our main convergence result.

Theorem 20 (Convergence of restricted gpCN Metropolis) *Let μ be as in (1) and assume that the pCN Metropolis kernel possesses a spectral gap in $L_2(\mu)$, i.e., $\text{gap}(M_0) > 0$. Then, for any $\Gamma \in \mathcal{L}_+(\mathcal{H})$ and any $\varepsilon \in (0, \text{gap}(M_0))$ there exists a number $R_0 = R_0(\varepsilon) \in (0, \infty)$ such that for any $R \geq R_0$ holds*

$$\|\mu - \mu_R\|_{tv} < \varepsilon \quad \text{and} \quad \text{gap}(M_{\Gamma,R}) > 0$$

where $\text{gap}(M_{\Gamma,R}) = 1 - \|M_{\Gamma,R}\|_{\mu_R}$ denotes the spectral gap of $M_{\Gamma,R}$ in $L_2(\mu_R)$.

Proof By Theorem 19 we have that for any $\varepsilon \in (0, \text{gap}(M_0))$ there exists a number $R_0 \in (0, \infty)$ such that for any $R \geq R_0$ holds

$$\|\mu - \mu_R\|_{tv} \leq \varepsilon \quad \text{and} \quad \text{gap}(M_{0,R}) > 0.$$

Moreover, Proposition 17, Lemma 18 and Theorem 11 yield that for any $\Gamma \in \mathcal{L}_+(\mathcal{H})$ the Markov operator associated with $M_{\Gamma,R}$ is self-adjoint and positive on $L_2(\mu_R)$. In particular, $M_{\Gamma,R}$ is again a Metropolis kernel with proposal P_Γ and acceptance probability α_R. Thus, in order to apply Theorem 8 to $M_{0,R}$ and $M_{\Gamma,R}$ it remains to verify that there exists a $p > 1$ so that

$$\kappa_{p,R} := \sup_{\mu_R(A) \in (0,1/2]} \frac{\int_A \int_{A^c} \rho_\Gamma(u,v)^p \, P_\Gamma(u,\mathrm{d}v) \, \mathrm{d}\mu_R(u)}{\mu_R(A)} < \infty$$

where $\rho_\Gamma(u,v) = \frac{\mathrm{d}P_0(u)}{\mathrm{d}P_\Gamma(u)}(v)$. By Theorem 14 we have for any $p < 1 + \frac{1}{2\|H_\Gamma\|}$ that

$$\kappa_{p,R} \leq \sup_{\mu_R(A) \in (0,1/2]} \frac{\int_A c \exp\left(\frac{b}{2}\|u\|^2\right) \mathrm{d}\mu_R(u)}{\mu_R(A)} \leq c \exp\left(\frac{b}{2} R^2\right) < \infty.$$

Hence, Theorem 8 leads to

$$\mathrm{gap}(M_{\Gamma,R})^{(p-1)/2} \geq \frac{1}{2^{(3p-1)/2}} \frac{\mathrm{gap}(M_{0,R})^p}{\kappa_{p,R}} > 0$$

which proves the assertion. □

Theorem 20 tells us that the corresponding restricted gpCN Metropolis converges exponentially fast to any, arbitrarily close, restriction μ_R of μ whenever the pCN Metropolis has a spectral gap, e.g., under the conditions of [13, Theorem 2.14]. In particular, Theorem 20 is a statement about the inheritance of geometric convergence from the pCN to the restricted gpCN Metropolis. We emphasize that a quantitative comparison of their spectral gaps is not proven. We provide a lower bound for $\mathrm{gap}(M_{\Gamma,R})$ in nonlinear terms of the pCN Metropolis. Additionally, the stated estimate behaves rather poor in R; more precise, it decays exponentially as $R \to \infty$.

Although we argued in the above theorem with restrictions of μ in order to bound κ_p from Theorem 8, let us mention that in simulations when R is sufficiently large one cannot distinguish between μ and μ_R as well as between Markov chains with transition kernels M_Γ and $M_{\Gamma,R}$.

Moreover, we conjecture that the gpCN Metropolis targeting μ has a strictly positive spectral gap whenever the pCN Metropolis has one. Recalling the results of the numerical simulations in Sect. 3.3 we even conjecture that the spectral gap of the gpCN Metropolis with suitably chosen $\Gamma \in \mathcal{L}_+(\mathcal{H})$ is much larger than the one of the pCN Metropolis.

5 Outlook on gpCN Proposals with State-Dependent Covariances

In this section we comment on state-dependent proposal covariances as they are a natural extension of the idea behind the gpCN proposal. The advantage of such a state-dependent approach is that the resulting Metropolis algorithm might be even better adapted to the target measure by allowing locally different proposal covariances. For

an illustrative motivation of state-dependent proposal covariances we refer to [11,22], and for recent positive and negative theoretic results we refer to [20]. In the Hilbert space setting we are now able to define MH algorithms by means of Theorem 13. Consider the proposal kernel

$$P_{\text{loc}}(u, \cdot) = N(A_{\Gamma(u)}u, s^2 C_{\Gamma(u)}) \tag{29}$$

where we assume that for $u \in \mathcal{H}$ we have $\Gamma(u) \in \mathcal{L}_+(\mathcal{H})$ and that the corresponding mapping $u \mapsto \Gamma(u)$ is measurable. Further, by $A_{\Gamma(u)}$ and $C_{\Gamma(u)}$ we denote the components of the gpCN proposal for $\Gamma = \Gamma(u)$. Following the heuristic presented in Sect. 3.1 for Bayesian inference problems where Φ in (1) is of the form (10), we could choose for instance

$$\Gamma(u) = \nabla G(u)^* \, \Sigma^{-1} \, \nabla G(u). \tag{30}$$

When considering the measure $\eta_{\text{loc}}(du, dv) = P_{\text{loc}}(u, dv)\mu_0(du)$ we notice that η_{loc} is no longer a Gaussian measure due to the dependence of Γ on u. However, to construct a μ-reversible Metropolis kernel with the proposal P_{loc} above, we can apply the same trick as in [1, Theorem 4.1]. Namely, with $\rho_\Gamma(u, v) = \frac{d P_0(u)}{d P_\Gamma(u)}(v)$ as given in Theorem 13 we obtain

$$
\begin{aligned}
P_{\text{loc}}(u, dv)\mu_0(du) &= \frac{1}{\rho_{\Gamma(u)}(u, v)} \, P_0(u, dv)\mu_0(du) \\
&= \frac{1}{\rho_{\Gamma(u)}(u, v)} \, P_0(v, du)\mu_0(dv) \\
&= \frac{\rho_{\Gamma(v)}(v, u)}{\rho_{\Gamma(u)}(u, v)} \, P_{\text{loc}}(v, du)\mu_0(dv),
\end{aligned}
$$

where we used the μ_0-reversibility of the pCN proposal P_0. Hence, according to the general Metropolis kernel construction outlined in Sect. 2.2, we have that a Metropolis kernel M_{loc} with proposal P_{loc} and acceptance probability

$$\alpha_{\text{loc}}(u, v) = \min\left\{1, \exp(\Phi(u) - \Phi(v)) \, \frac{\rho_{\Gamma(u)}(u, v)}{\rho_{\Gamma(v)}(v, u)}\right\} \tag{31}$$

is μ-reversible. Note that the same construction can analogously be applied to proposals of the form

$$P'_{\text{loc}}(u, \cdot) = N(\sqrt{1 - s^2}\,u, s^2 C_{\Gamma(u)}), \tag{32}$$

where the modified acceptance probability is then given by

$$\alpha'_{\text{loc}}(u, v) = \min\left\{1, \exp(\Phi(u) - \Phi(v)) \, \frac{\pi_{\Gamma(u)}(\frac{1}{s}[v - A_0 u])}{\pi_{\Gamma(v)}(\frac{1}{s}[u - A_0 v])}\right\} \tag{33}$$

with π_Γ as stated in Theorem 13. The arguments above show that this type of algorithms are well posed in infinite dimensions. Of course, the question arises if the additional computational costs of evaluating $\Gamma(u)$ and $\rho_{\Gamma(u)}$ or $\pi_{\Gamma(u)}$ in each step pay off in a significantly higher statistical efficiency. Related to this concern, one could think of substituting $\nabla G(u)$ in (30) by a cheaper approximation in order to reduce the computational work. This might help to make MH algorithms with local proposal covariances feasible. Unfortunately, the tools and results developed and presented in Sect. 4 are not sufficient to prove spectral gaps of these MH algorithms with state-dependent proposals. The main reason for this is the missing reversibility of the proposals w.r.t. μ_0. This condition played a key role in Theorem 8 and is the main reason why the analysis of Sect. 4 is not applicable. We leave this open for future research.

Appendix

Gaussian Measures

The following brief introduction to Gaussian measures is based on the presentations given in [6, Section 1] and [12, Section 3]. Another comprehensive reference for this topic is [2].

Let $\mathcal{H}$ be a Hilbert space with norm $\|\cdot\|$ and inner-product $\langle\cdot,\cdot\rangle$ and let $\mathcal{L}^1_+(\mathcal{H})$ denote the set of all linear, bounded, self-adjoint, positive and trace class operators $A : \mathcal{H} \to \mathcal{H}$.

Let μ be a measure on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, and for simplicity let us assume that $\int_{\mathcal{H}} \|v\|^2 \mu(\mathrm{d}v) < \infty$. The *mean* $m \in \mathcal{H}$ of μ is defined as the Bochner integral $m = \int_{\mathcal{H}} v\, \mu(\mathrm{d}v)$ and the *covariance* of μ is the unique operator $C \in \mathcal{L}^1_+(\mathcal{H})$ given by

$$\langle Cu, u' \rangle = \int_{\mathcal{H}} \langle u, v - m \rangle \langle u', v - m \rangle \mu(\mathrm{d}v), \qquad \forall u, u' \in \mathcal{H}.$$

A measure μ on $\mathcal{H}$ is called a *Gaussian measure* with mean $m \in \mathcal{H}$ and covariance operator $C \in \mathcal{L}^1_+(\mathcal{H})$, denoted by $N(m, C)$, iff

$$\int_{\mathcal{H}} \mathrm{e}^{\mathrm{i}\langle u, v \rangle} \mu(\mathrm{d}v) = \mathrm{e}^{\mathrm{i}\langle m, u \rangle - \frac{1}{2}\langle Cu, u \rangle}, \qquad \forall u \in \mathcal{H}.$$

This definition is equivalent to $\langle u \rangle_* \mu = N(\langle u, m \rangle, \langle Cu, u \rangle)$ for all $u \in \mathcal{H}$ where $\langle u \rangle : \mathcal{H} \to \mathbb{R}$ with $\langle u \rangle(v) := \langle u, v \rangle$ and where $\langle u \rangle_* \mu$ denotes the pushforward measure of μ under the mapping $\langle u \rangle$. Gaussian measures are uniquely determined by their mean and covariance, i.e., for any $m \in \mathcal{H}$ and any $C \in \mathcal{L}^1_+(\mathcal{H})$ there exists a unique Gaussian

measure $\mu = N(m, C)$ on $\mathcal{H}$. Moreover, the set of random variables on $\mathcal{H}$ distributed according to a Gaussian measure is closed w.r.t. affine transformations. In detail, let $X \sim N(m, C)$ be a Gaussian random variable on $\mathcal{H}$ and let $b \in \mathcal{H}$ and $T : \mathcal{H} \to \mathcal{H}$ be a bounded, linear operator, then due to [6, Proposition 1.2.3] we have

$$b + TX \sim N(b + Tm, TCT^*). \tag{34}$$

The *Cameron–Martin space* $\mathcal{H}_\mu$ of a Gaussian measure $\mu = N(m, C)$ on $\mathcal{H}$ is defined as the image space $\mathrm{Im}\,C^{1/2}$ which forms equipped with $\langle u, v \rangle_{C^{-1}} := \langle C^{-1/2}u, C^{-1/2}v \rangle$ again a Hilbert space. The space $\mathcal{H}_\mu$ has some surprising properties: It is the intersection of all measurable linear subspaces $\mathcal{X} \subseteq \mathcal{H}$ with $\mu(\mathcal{X}) = 1$; if $\ker C = \{0\}$ then $\mathcal{H}_\mu$ is dense in $\mathcal{H}$ and if $\mathcal{H}$ is infinite dimensional then $\mu(\mathcal{H}_\mu) = 0$. Moreover, the space $\mathcal{H}_\mu$ plays an important role for the equivalence of Gaussian measures as rigorously expressed in the Cameron–Martin theorem below. Before stating the result we need some more notation.

In the following let $\mu = N(0, C)$. For $u \in \mathcal{H}_\mu$ we set

$$W_u(v) := \langle C^{-1/2}u, v \rangle, \qquad \forall v \in \mathcal{H},$$

and understand W_u as an element of $L_2(\mu)$. Since the mapping $\mathcal{H}_\mu \ni u \mapsto W_u \in L_2(\mu)$ is an isometry [6, Section 1.2.4], we can define for any $u \in \mathcal{H}$

$$\langle C^{-1/2}u, \cdot \rangle := L_2(\mu)\text{-}\lim_{n \to \infty} W_{u_n}$$

where $u_n \in \mathcal{H}_\mu$ and $u_n \to u$ in $\mathcal{H}$ as $n \to \infty$. And by [6, Proposition 1.2.7] it holds that

$$\int_{\mathcal{H}} e^{\langle C^{-1/2}u, v \rangle} \mu(\mathrm{d}v) = e^{\frac{1}{2}\|u\|^2}, \qquad \forall u \in \mathcal{H}. \tag{35}$$

Hence, if $h \in \mathcal{H}_\mu$, we understand $\langle C^{-1}h, \cdot \rangle$ as $\langle C^{-1/2}(C^{-1/2}h), \cdot \rangle \in L_2(\mu)$.

Theorem 21 (Cameron–Martin formula, [6, Theorem 1.3.6]) *Let $\mu = N(0, C)$ and $\mu_h = N(h, C)$ be Gaussian measures on a separable Hilbert space $\mathcal{H}$. Then, μ and μ_h are equivalent iff $h \in \mathcal{H}_\mu = \mathrm{Im}\,C^{1/2}$ in which case*

$$\frac{\mathrm{d}\mu_h}{\mathrm{d}\mu}(v) = \exp\left(-\frac{1}{2}\|C^{-1/2}h\|^2 + \langle C^{-1}h, v \rangle\right).$$

Thus, two Gaussian measures $N(m, C)$ and $N(m + h, C)$ are only equivalent if $h \in \mathrm{Im}\,C^{1/2}$. Consider now $\mu = N(0, C)$ and $\nu = N(0, Q)$ with $C \neq Q$. Before stating a theorem about the equivalence of μ and ν, we need some more notations. Let $T : \mathcal{H} \to \mathcal{H}$ be in the following a self-adjoint trace class operator and let $(t_n)_{n \in \mathbb{N}}$ denote the sequence of its eigenvalues. We set

$$\det(I + T) := \prod_{n=1}^{\infty}(1 + t_n)$$

and define

$$\langle TC^{-1/2}u, C^{-1/2}u \rangle := \lim_{N \to \infty} \langle TC^{-1/2}\Pi_N u, C^{-1/2}\Pi_N u \rangle, \qquad \mu\text{-a.e.}$$

where Π_N denotes the projection operator to span$\{e_1, \dots, e_N\}$ with e_n denoting the nth eigenvector of C. The existence of the μ-a.e.-limit above is proven in [6, Proposition 1.2.10], and furthermore, if $\langle Tu, u \rangle < \|u\|^2$ holds for any $u \in \mathcal{H}$, then by [6, Proposition 1.2.11] we have

$$\int_{\mathcal{H}} e^{\frac{1}{2}\langle TC^{-1/2}u, C^{-1/2}u \rangle} \, d\mu(u) = \frac{1}{\sqrt{\det(1 - T)}}. \tag{36}$$

Theorem 22 ([6, Proposition 1.3.11]) *Let $\mu = N(0, C)$ and $\nu = N(0, Q)$ be Gaussian measures on a separable Hilbert space $\mathcal{H}$. If $T := I - C^{-1/2}QC^{-1/2}$ is self-adjoint, trace class and satisfies $\langle Tu, u \rangle < \|u\|^2$ for any $u \in \mathcal{H}$, then μ and ν are equivalent with*

$$\frac{d\nu}{d\mu}(u) = \frac{1}{\sqrt{\det(I - T)}} \exp\left(-\frac{1}{2}\langle T(I - T)^{-1} C^{-1/2}u, C^{-1/2}u \rangle \right), \quad u \in \mathcal{H}.$$

We note that the assumptions of Theorem 22 can be relaxed to $I - C^{-1/2}QC^{-1/2}$ being Hilbert–Schmidt which is known as Feldman–Hajek theorem. Also in this case expression for the Radon–Nikodym derivative can be obtained, see [2, Corollary 6.4.11].

Finally, we recall two simple but useful facts resulting from a change of variables.

Lemma 23 *Let $\mathcal{H}$ be a separable Hilbert space, $0 < s < \infty$ and $h \in \mathcal{H}$.*

- *Assume $\mu = N(m, C)$, $\nu = N(m + h, s^2 C)$ on $\mathcal{H}$ and $f : \mathcal{H} \to \mathbb{R}$. Then*

$$\int_{\mathcal{H}} f(v)\mu(dv) = \int_{\mathcal{H}} f\left(\frac{1}{s}(v - h) \right) \nu(dv).$$

- *Assume $\mu_1 = N(m_1, C_1)$ and $\mu_2 = N(m_2, C_2)$ are equivalent with $\frac{d\mu_2}{d\mu_1}(u) = \pi(u)$. Then the measures $\nu_1 = N(m_1 + h, s^2 C_1)$ and $\nu_2 = N(m_2 + h, s^2 C_2)$ are also equivalent with*

$$\frac{d\nu_2}{d\nu_1}(u) = \pi\left(\frac{u - h}{s} \right).$$

Proofs

The following proofs are rather operator theoretic and rely heavily on the holomorphic functional calculus. We refer to [8, Section VII.3] for a comprehensive introduction.

Proof of Lemma 2

From the proof of Proposition 1 we know that $(I + H_\Gamma)^{-1} : \mathcal{H} \to \mathcal{H}$ is self-adjoint and that $\|(I + H_\Gamma)^{-1}\| \leq 1$. Thus, $I - s^2(I + H_\Gamma)^{-1}$ is also a self-adjoint, bounded and positive operator on $\mathcal{H}$ and its square root operator appearing in (16) exists. This yields the well definedness of $A_\Gamma : \text{Im } C^{1/2} \to \mathcal{H}$. We now prove that A_Γ is a bounded operator on $\text{Im } C^{1/2}$. For $s = 0$ we get $A_\Gamma = I$ and the assertion follows, so that we assume $s \in (0, 1)$. Let us now define $f : \mathbb{C}\setminus\{-1\} \to \mathbb{C}$ by

$$f(z) = \sqrt{1 - s^2(1 + z)^{-1}}.$$

The function f is analytic in the complex half plane $\{z \in \mathbb{C} : \Re(z) > s^2 - 1\}$, since $\Re(1 + z) > s^2$ implies

$$\Re\left((1 + z)^{-1}\right) = \frac{\Re(1 + z)}{|1 + z|^2} \leq \frac{1}{\Re(1 + z)} < \frac{1}{s^2}.$$

Denoting $\gamma := \|H_\Gamma\|$ the spectrum of $H_\Gamma = C^{1/2}\Gamma C^{1/2}$ is contained in $[0, \gamma]$. Then, since $s < 1$ we have that f is analytic in a neighborhood, say, $\mathcal{N}[0, \gamma]$ of $[0, \gamma]$. Hence, by functional calculus we obtain

$$\sqrt{I - s^2(I + H_\Gamma)^{-1}} = f(H_\Gamma) = \frac{1}{2\pi i} \int_{\partial\mathcal{N}[0,\gamma]} f(\zeta)(\zeta I - H_\Gamma)^{-1} \, d\zeta.$$

Due to analyticity we can approximate f by a sequence of polynomials p_n with degree n which converge uniformly on $\mathcal{N}[0, \gamma]$ to f for $n \to \infty$. Then, by [8, Lemma VII.3.13] holds

$$\|p_n(H_\Gamma) - f(H_\Gamma)\|_{\mathcal{H} \to \mathcal{H}} \to 0,$$

for $n \to \infty$. Since the polynomials p_n can be represented as $p_n(z) = \sum_{k=0}^{n} a_k^{(n)} z^k$, we obtain further

$$C^{1/2} p_n(H_\Gamma) = C^{1/2} \sum_{k=0}^{n} a_k^{(n)} (C^{1/2}\Gamma C^{1/2})^k = p_n(C\Gamma) C^{1/2}.$$

By [14, Proposition 1] we have

$$\text{spec}(C\Gamma \mid \mathcal{H}) = \text{spec}(C^{1/2}\Gamma C^{1/2} \mid \mathcal{H}) \subseteq [0, \gamma]$$

where $\text{spec}(\cdot \mid \mathcal{H})$ denotes the spectrum on $\mathcal{H}$, and, thus, we can conclude $\|p_n(C\Gamma) - f(C\Gamma)\|_{\mathcal{H} \to \mathcal{H}} \to 0$ as $n \to \infty$ again by [8, Lemma VII.3.13]. Hence,

$$C^{1/2} f(H_\Gamma) = \lim_{n\to\infty} C^{1/2} p_n(H_\Gamma) = \lim_{n\to\infty} p_n(C\Gamma) C^{1/2} = f(C\Gamma)C^{1/2}$$

and

$$A_\Gamma = C^{1/2} f(H_\Gamma) C^{-1/2} = f(C\Gamma) C^{1/2} C^{-1/2} = f(C\Gamma)$$

where $f(C\Gamma)$ is by construction a bounded operator on $\mathcal{H}$.

Proof of Lemma 12

By [7, Theorem 1] the relation $\mathrm{Im}(\Delta_\Gamma) \subseteq \mathrm{Im}(C^{1/2})$ holds iff there exists a bounded operator $B : \mathcal{H} \to \mathcal{H}$ such that

$$\Delta_\Gamma = C^{1/2} B. \tag{37}$$

Thus, $\mathrm{Im}(\Delta_\Gamma) \subseteq \mathrm{Im}(C^{1/2})$ is equivalent to $C^{-1/2}\Delta_\Gamma$ being bounded on $\mathcal{H}$. In order to construct and analyze the operator B, we define $f : \mathbb{C} \backslash \{-1\} \to \mathbb{C}$ by

$$f(z) := \sqrt{1 - s^2 (1 + z)^{-1}} - \sqrt{1 - s^2},$$

which is analytic in $\{z \in \mathbb{C} : \Re(z) > s^2 - 1\}$, cf. the proof of Lemma 2, and particularly in

$$V = \{z \in \mathbb{C} : \mathrm{dist}(z, [0, \gamma]) \le \varepsilon\}, \qquad 0 < \varepsilon < 1 - s^2,$$

where $\gamma := \|H_\Gamma\|$. We have the following representation

$$
\begin{aligned}
-\Delta_\Gamma &= A_\Gamma - \sqrt{1 - s^2} I \\
&= C^{1/2} \left(\sqrt{I - s^2 (I + H_\Gamma)^{-1}} - \sqrt{1 - s^2} I \right) C^{-1/2} \\
&= C^{1/2} f(H_\Gamma) C^{-1/2}
\end{aligned}
$$

with

$$f(H_\Gamma) = \frac{1}{2\pi i} \int_{\partial V} f(\zeta) (\zeta I - H_\Gamma)^{-1} \, d\zeta$$

see [8, Chapter VII.3]. Hence, if we can prove that $B = -f(H_\Gamma) C^{-1/2}$ is a bounded operator on $\mathcal{H}$, we have shown the assertion.

To this end let $p_n(z) = \sum_{k=0}^{n} a_k^{(n)} z^k$ be polynomials of degree n, with $n \in \mathbb{N}$, which converge uniformly on V to f. Such polynomials exist due to the analyticity of f and by the fact that $f(0) = 0$ we can assume w.l.o.g. that $a_0^{(n)} = 0$ for all $n \in \mathbb{N}$. This leads to

$$p_n(H_\Gamma) = C^{1/2}\Gamma^{1/2} \left(\sum_{k=1}^{n} a_k^{(n)} (\Gamma^{1/2} C \Gamma^{1/2})^{k-1} \right) \Gamma^{1/2} C^{1/2}$$

$$= C^{1/2}\Gamma^{1/2} \, q_{n-1}(\Gamma^{1/2} C \Gamma^{1/2}) \, \Gamma^{1/2} C^{1/2}$$

with $q_{n-1}(z) := \sum_{k=1}^{n} a_k^{(n)} z^{k-1} = p_n(z)/z$. Now, [14, Proposition 1] implies that the operators $C^{1/2}\Gamma C^{1/2}$ and $\Gamma^{1/2} C \Gamma^{1/2}$ share the same spectrum, since C and Γ are positive. Thus, $\mathrm{spec}(\Gamma^{1/2} C \Gamma^{1/2} \mid \mathcal{H}) \subset [0, \gamma]$ and we have

$$q_n(\Gamma^{1/2} C \Gamma^{1/2}) = \frac{1}{2\pi i} \int_{\partial V} q_n(\zeta) \, (\zeta I - \Gamma^{1/2} C \Gamma^{1/2})^{-1} \, d\zeta, \qquad n \in \mathbb{N}.$$

Moreover, the polynomials q_n are a Cauchy sequence in $C(\partial V)$, since

$$\sup_{\zeta \in \partial V} |q_n(\zeta) - q_m(\zeta)| \;\leq\; \sup_{\zeta \in \partial V} \frac{|\zeta|}{\min_{\eta \in \partial V} |\eta|} |q_n(\zeta) - q_m(\zeta)|$$

$$= \frac{1}{\min_{\eta \in \partial V} |\eta|} \sup_{\zeta \in \partial V} |\zeta q_n(\zeta) - \zeta q_m(\zeta)|$$

$$= \frac{1}{\min_{\eta \in \partial V} |\eta|} \sup_{\zeta \in \partial V} |p_{n+1}(\zeta) - p_{m+1}(\zeta))|$$

where $\min_{\eta \in \partial V} |\eta| = \varepsilon > 0$ due to our choice of V. Thus, the polynomials q_n converge uniformly on ∂V to a function g. This implies that the operators $q_n(\Gamma^{1/2} C \Gamma^{1/2})$ converge in the operator norm to a bounded operator

$$g(\Gamma^{1/2} C \Gamma^{1/2}) := \frac{1}{2\pi i} \int_{\partial V} g(\zeta) \, (\zeta I - \Gamma^{1/2} C \Gamma^{1/2})^{-1} \, d\zeta.$$

We arrive at

$$f(H_\Gamma) = \lim_{n \to \infty} p_n(C^{1/2} \Gamma C^{1/2})$$

$$= \lim_{n \to \infty} C^{1/2} \Gamma^{1/2} \, q_{n-1}(\Gamma^{1/2} C \Gamma^{1/2}) \, \Gamma^{1/2} C^{1/2}$$

$$= C^{1/2} \Gamma^{1/2} \, g(\Gamma^{1/2} C \Gamma^{1/2}) \, \Gamma^{1/2} C^{1/2},$$

which yields

$$B = -f(H_\Gamma) C^{-1/2} = -C^{1/2} \Gamma^{1/2} \, g(\Gamma^{1/2} C \Gamma^{1/2}) \Gamma^{1/2}$$

being bounded on $\mathcal{H}$.

References

1. A. Beskos, G. Roberts, A. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stoch. Dynam.*, 8(3):319–350, 2008.

2. V. Bogachev. *Gaussian Measures*. American Mathematical Society, 1998.
3. N. Bou-Rabee and M. Hairer. Nonasymptotic mixing of the MALA algorithm. *IMA J. Numer. Anal.*, 33(1):80–110, 2013.
4. S. Cotter, G. Roberts, A. Stuart, and D. White. MCMC methods for functions: Modifying old algorithms to make them faster. *Stat. Sci.*, 28(3):283–464, 2013.
5. T. Cui, K. Law, and Y. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.
6. G. Da Prato and J. Zabczyk. *Second Order Partial Differential Equations in Hilbert Spaces*. Cambridge University Press, Cambridge, 2004.
7. R. Douglas. On majorization, factorization, and range inclusion of operators on Hilbert space. *Proc. Amer. math. Soc.*, 17:413–415, 1966.
8. N. Dunford and J. Schwartz. *Linear Operators, Part I: General Theory*. Wiley-Interscience, New York, 1958.
9. O. Ernst, B. Sprungk, and H.-J. Starkloff. Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems. *SIAM/ASA J. Uncertainty Quantification*, 3(1):823–851, 2015.
10. C. Geyer. Practical Markov chain Monte Carlo. *Stat. Sci.*, 7(4):473–483, 1992.
11. M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B*, 73(2):123–214, 2010.
12. M. Hairer. An introduction to stochastic PDEs. Lecture notes, 2009.
13. M. Hairer, A. Stuart, and S. Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.
14. M. Hladnik and M. Omladič. Spectrum of the product of operators. *Proc. Amer. math. Soc.*, 102(2):300–302, 1988.
15. Z. Hu, Z. Yao, and J. Li. On an adaptive preconditioned Crank–Nicolson algorithm for infinite dimensional Bayesian inferences. arXiv:1511.05838, 2015.
16. C. Kipnis and S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
17. K. Law. Proposals which speed up function-space MCMC. *J. Comput. Appl. Math.*, 262:127–138, 2014.
18. G. Lawler and A. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988.
19. A. Lee and K. Łatuszyński. Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika*, 101(3):655–671, 2014.
20. S. Livingstone. Geometric ergodicity of the Random Walk Metropolis with position-dependent proposal covariance. arXiv:1507.05780, 2015.
21. A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 65:385–397, 1984.
22. J. Martin, L. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.*, 34(3):A1460–A1487, 2012.
23. F. Pinski, G. Simpson, A. Stuart, and H. Weber. Algorithms for Kullback–Leibler approximation of probability measures in infinite dimensions. *SIAM J. Sci. Comput.*, 37(6):A2733–A2757, 2015.
24. G. Roberts and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2(2):13–25, 1997.
25. G. Roberts and J. Rosenthal. Optimal scaling for various Metropolis–Hastings algorithms. *Stat. Sci.*, 16(4):351–367, 2001.
26. D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math. (Rozprawy Mat.)*, 485:1–93, 2012.
27. D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. *Electron. Commun. Probab.*, 18:1–8, 2013.
28. A. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.
29. L. Tierney. Markov chains for exploring posterior distributions. *Ann. Stat.*, 22(4):1701–1762, 1994.
30. L. Tierney. A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.
31. C. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.