

MCMC Algorithms for Posteriors on Matrix Spaces

Alexandros Beskos*

Department of Statistical Science, University College London
and

Kengo Kamatani†

Department of Engineering Science, Osaka University

November 8, 2021

Abstract

We study Markov chain Monte Carlo (MCMC) algorithms for target distributions defined on matrix spaces. Such an important sampling problem has yet to be analytically explored. We carry out a major step in covering this gap by developing the proper theoretical framework that allows for the identification of ergodicity properties of typical MCMC algorithms, relevant in such a context. Beyond the standard Random-Walk Metropolis (RWM) and preconditioned Crank–Nicolson (pCN), a contribution of this paper in the development of a novel algorithm, termed the ‘Mixed’ pCN (MpCN). RWM and pCN are shown *not* to be geometrically ergodic for an important class of matrix distributions with heavy tails. In contrast, MpCN is robust across targets with different tail behaviour and has very good empirical performance within the class of heavy-tailed distributions. Geometric ergodicity for MpCN is not fully proven in this work, as some remaining drift conditions are quite challenging to obtain owing to the complexity of the state space. We do, however, make a lot of progress towards a proof, and show in detail the last steps left for future work. We illustrate the computational performance of the various algorithms through numerical applications, including calibration on real data of a challenging model arising in financial statistics.

Keywords: Preconditioned Crank–Nicolson; Drift Condition; Matrix-Valued Stochastic Differential Equation.

*Alexandros Beskos acknowledges support from a Leverhulme Trust Prize.

†Kengo Kamatani acknowledges support from JSPS KAKENHI Grant Numbers 16K00046, 20H04149 and JST CREST Grant Number JPMJCR14D7.

1 Introduction

Statistical models with parameters defined on matrix spaces arise naturally in many applications, with maybe most typical the case of covariance matrices. Related measures have thus been developed, with most prominent the Matrix-Normal, Wishart and Inverse-Wishart distributions (Barnard et al., 2000). Numerous extensions have appeared, see e.g. O’Malley and Zaslavsky (2008); Huang and Wand (2013) for a scaled and an hierarchical Inverse-Wishart, Barnard et al. (2000) for a strategy that extracts the correlation matrix, Roverato (2002) and Dobra et al. (2011) for the Hyper-Inverse-Wishart and G-Wishart distributions, respectively.

Beyond standard conjugate settings – e.g. Inverse-Wishart prior for the covariance of Gaussian observations – more involved hierarchical models have generated a need for developing a suite of accompanying MCMC methods. This work reviews standard algorithms and introduces a novel one (MpCN), motivated by an MCMC method on vector-spaces used in Kamatani (2017). The paper invokes a theoretical framework that permit the analysis of the ergodicity properties of some of the presented algorithms or – in the case of MpCN – makes a lot of progress towards a (quite challenging) proof, and describes the last remaining steps left for future research. We show that RWM and pCN do not work well even for an Inverse-Wishart target as they are not geometrically ergodic for heavy-tailed distributions. In contrast, MpCN has much better empirical performance on such targets. Our main contributions are summarised as follows.

(i) We develop a new MCMC method – MpCN –, and provide a motivation for its underpinnings. MpCN is characterised by better empirical performance against RWM or pCN, in a number of numerical studies.

(ii) We prove that targets on the space of positive definite matrices can be ‘upcasted’ onto corresponding laws on unrestricted matrices, the latter space permitting a direct path for the development of MCMC methodology.

(iii) We prove that RWM, pCN are not geometrically ergodic for a wide class of matrix-valued targets. We make a lot of progress into demonstrating geometric ergodicity for MpCN, and highlight the remaining steps for the completed proof.

(iv) We run MpCN on a challenging hierarchical model providing a matrix-extension of the influential scalar Stochastic Volatility (SV) dynamic jump-model by Barndorff-Nielsen and Shephard (2001). We stress that the paper focuses on MCMC methods with ‘blind’ proposals. This is in agreement with the selected SV application – and, more generally, modern pseudo-marginal methods (Andrieu and Roberts, 2009) for complex models – where guided proposals or Gibbs sampler schemes are typically cumbersome and impractical.

We note that there are at least two main difficulties in constructing MCMC on a matrix space. First, matrix calculations such as multiplication, inversion and eigenvalue decomposition can be expensive. So, attempts to use gradient-based methods can lead to prohibitively high computational costs. In this work, consideration of derivatives is completely avoided. Second, involved matrices may contain specific structure, thus careful selection of the proposal kernel is required for such structure to be preserved. For example, we mainly consider symmetric positive definite matrix spaces representing a cone in the space of diagonal matrices. Thus, the naïve random-walk Metropolis cannot be used as it will not preserve positive definiteness.

The paper develops as follows. Section 2 introduces relevant measures on matrix spaces. Section 3 presents MCMC algorithms on such spaces. Section 4 develops ergodicity results for the RWM and pCN algorithms. Section 5 investigates ergodicity properties of MpCN. Section 6 shows a collection of numerical results. Section 7 provides conclusions and points at future work.

Notation: $\mathbb{R}_+ := (0, \infty)$. We write $X =_d Y$ if variables X, Y have the same law. $M(p, q)$ is the set of $p \times q$ (real-valued) matrices, $GL(p)$ the group of $p \times p$ invertible matrices, $Sym(p)$ the set of $p \times p$ symmetric matrices, $P^+(p)$ the set of $p \times p$ symmetric, positive-definite matrices, $\mathcal{O}(p)$ the set of orthogonal matrices, $\mathcal{O}(p, q)$, $p \geq q$, the space of $p \times q$ matrices U such that $U^\top U = I_q$. We use the notation $A = (A_{ij})$ to indicate individual matrix elements. Let $\|A\|_F$ be the Frobenius norm, with inner product $\langle A, B \rangle_F = \text{tr}(A^\top B)$, where $\text{tr}(\cdot)$ is the trace of a matrix. A^\top is the transpose of A , and $\det(A)$ its determinant. The derivative of $h : M(p, q) \rightarrow \mathbb{R}$ is the $p \times q$ matrix $Dh(A) := (\partial h(A)/\partial A_{ij})$. For $A \in P^+(p)$, we denote by $A^{1/2}$ the matrix $B \in P^+(p)$ such that $BB = A$.

2 Measures on Matrix Spaces

2.1 Reference Measures

For $\mathcal{U} \in P^+(p)$, $p \geq q$, we define $\nu_{\mathcal{U}}$ on $M(p, q)$,

$$\nu_{\mathcal{U}}(dX) := \frac{\Gamma_q(p/2)}{\pi^{pq/2}} \left(\frac{(\det \mathcal{U})^{-q/2}}{\det(X^{\top} \mathcal{U}^{-1} X)^{p/2}} \right) \text{Leb}(dX),$$

$$\text{Leb}(dX) = \prod_{i=1}^p \prod_{j=1}^q dX_{ij},$$

$\Gamma_p(\cdot)$ is the multivariate Gamma function. We denote $\nu_{\mathcal{U}}$ simply as ν when $\mathcal{U} = I_p$. Leb denotes the Lebesgue measure, on a space of dimensions implied by the context. The Lebesgue measure and $\nu_{\mathcal{U}}$ will be used as reference measures on $M(p, q)$. Note that $\nu_{\mathcal{U}}$ is invariant on $M(p, q)$ under right multiplication, $X \mapsto Xa$, $a \in \text{GL}(q)$. If $p = q$ then $\nu \equiv \nu_{\mathcal{U}}$ is a unimodular Haar measure on the locally compact topological group $\text{GL}(q)$ and $\nu \equiv \nu_{-1}$, where $\nu_{-1}(A) := \nu(A^{-1})$ for any Borel set $A \subseteq \text{GL}(q)$; see e.g. Sections 5, 7 of Farrell (1985), Section 60 of Halmos (1950).

On $P^+(q)$, we define the measure,

$$\mu(dS) := (\det S)^{-(q+1)/2} \prod_{1 \leq i \leq j \leq q} dS_{ij}. \quad (1)$$

We note that μ is invariant under the transform $S \mapsto aSa^{\top}$, $a \in \text{GL}(q)$; see Section 5 of Farrell (1985). Also, as in Problem 7.10.6 of Farrell (1985), $\mu \equiv \mu_{-1}$.

For the compact topological space $\mathcal{O}(p, q)$, there is a uniform probability distribution that will be denoted $d\mathbf{U}$ in this work. Note that the uniform distribution on $\mathcal{O}(p, q)$, for $p \geq q$, is the marginal on $\mathcal{O}(p, q)$ of the uniform distribution on $\mathcal{O}(p)$, see Theorem 3.3.1 of Chikuse (2003). The reference measures satisfy,

$$\int_{M(p, q)} f(x) \nu(dx) = \int_{M(p, q)} f(\mathcal{U}^{-1/2} x) \nu_{\mathcal{U}}(dx) = \int_{P^+(q)} \mu(dS) \int_{\mathcal{O}(p, q)} f(\mathbf{U} S^{1/2}) d\mathbf{U}, \quad (2)$$

for $f : M(p, q) \rightarrow \mathbb{R}$.

2.2 Probability Measures

Example 2.1. For parameters $M \in M(p, q)$, $\Sigma \in P^+(p)$, $T \in P^+(q)$, the Matrix-Normal distribution $N_{p,q}(M, \Sigma, T)$ has density with respect to Leb,

$$\phi_{p,q}(X; M, \Sigma, T) = \frac{\exp\left(-\text{tr}[T^{-1}(X - M)^\top \Sigma^{-1}(X - M)]/2\right)}{(2\pi)^{pq/2}(\det \Sigma)^{q/2}(\det T)^{p/2}}.$$

Example 2.2. Let $r \in \mathbb{R}$, $r > q - 1$, and $T \in P^+(q)$. The Wishart distribution $W_q(r, T)$ is a probability measure on $P^+(q)$ with density with respect to μ ,

$$\pi(S) = \frac{\det(S)^{r/2} \exp(-\text{tr}[T^{-1}S]/2)}{2^{rq/2}(\det T)^{r/2}\Gamma_q(r/2)}. \quad (3)$$

Example 2.3. Let $r \in \mathbb{R}$, $r > q - 1$, and $T \in P^+(q)$. The Inverse-Wishart distribution $W_q^{-1}(r, T)$ is a probability measure on $P^+(q)$ with density with respect to μ ,

$$\pi(S) = \frac{(\det T)^{r/2} \exp(-\text{tr}[TS^{-1}]/2)}{2^{rq/2} \det(S)^{r/2} \Gamma_q(r/2)}.$$

Remark 1. We summarize some relevant distribution properties.

(i) If $X \sim N_{p,q}(M, \Sigma, T)$ then,

$$\mathbb{E}[(X - M)(X - M)^\top] = \Sigma \text{tr}(T), \quad \mathbb{E}[(X - M)^\top(X - M)] = T \text{tr}(\Sigma).$$

Also, for $A \in M(r, p)$, $B \in M(q, s)$ full rank matrices with $r \leq p$ and $q \geq s$,

$$AXB \sim N_{r,s}(AMB, A\Sigma A^\top, B^\top T B).$$

(ii) If $X_i \sim N_{p,q}(M_i, \Sigma_i, T)$, $i = 1, 2$, are independent then,

$$X_1 + X_2 \sim N_{p,q}(M_1 + M_2, \Sigma_1 + \Sigma_2, T).$$

(iii) If $X \sim N_{p,q}(0, I_p, T)$ then $S = X^\top X \sim W_q(p, T)$.

(iv) If $S \sim W_q(r, T)$ then $S^{-1} \sim W_q^{-1}(r, T^{-1})$.

2.3 Upcasting $P^+(q)$ Onto $M(p, q)$

Let $p \geq q$ and $\mathcal{U} \in P^+(p)$. We show that a distribution on $P^+(q)$ can be expressed as transform of one on the larger space $M(p, q)$ via the surjective mapping $x \mapsto x^\top \mathcal{U}^{-1} x$.

Theorem 2.4. *Let $\tilde{\Pi}(dS) = \tilde{\pi}(S)\mu(dS)$ be a probability measure on $P^+(q)$. Consider the distribution $\Pi(dx)$ on $M(p, q)$, $p \geq q$, defined as,*

$$\Pi(dx) := \tilde{\pi}(x^\top \mathcal{U}^{-1} x) \nu_{\mathcal{U}}(dx). \quad (4)$$

If $X \sim \Pi(dx)$ then $S = X^\top \mathcal{U}^{-1} X \sim \tilde{\Pi}(dS)$.

Proof. By Equation (2), if $f(x) = g(x^\top \mathcal{U}^{-1} x)$, for $g : P^+(q) \rightarrow \mathbb{R}$, we have

$$\begin{aligned} \int_{M(p,q)} g(x^\top \mathcal{U}^{-1} x) \nu_{\mathcal{U}}(dx) &= \int_{M(p,q)} g(x^\top x) \nu(dx) \\ &= \int_{P^+(q)} g(S) \mu(dS). \end{aligned}$$

We have obtained a change of variables formula for map $x \mapsto x^\top \mathcal{U}^{-1} x = S$. Replacing $g(\cdot)$ with $(g \cdot \tilde{\pi})(\cdot)$ gives,

$$\int_{P^+(q)} g(S) \tilde{\Pi}(dS) = \int_{M(p,q)} g(x^\top \mathcal{U}^{-1} x) \Pi(dx).$$

This completes the proof. □

We will consider MCMC methods on $M(p, q)$. In light of Theorem 2.4, such algorithms are directly relevant for distributions $\tilde{\Pi}(dS) = \tilde{\pi}(S)\mu(dS)$ on $P^+(q)$ since we can execute the MCMC algorithm on $M(p, q)$, with target law $\Pi(dx) = \tilde{\pi}(x^\top \mathcal{U}^{-1} x) \nu_{\mathcal{U}}(dx)$, and apply the transform $x \mapsto x^\top \mathcal{U}^{-1} x = S$ on the collected x -samples.

3 Matrix-Valued MCMC Methods

We provide some MCMC methods on $M(p, q)$, $p \geq q$. As noted in Section 1, we focus on MCMC algorithms involving blind proposals, i.e. containing no information about the target. Let $\Pi(dx)$ be a target distribution on $M(p, q)$, and – in a Metropolis setting – $Q(x, dy)$ a proposal Markov kernel. Assume that ξ is a σ -finite measure such that Π is

absolutely continuous with respect to ξ , and $Q(x, dy)$ is ξ -reversible. Then, the triplet (Π, Q, ξ) give rise to a Metropolis–Hastings kernel, so that, for Borel sets $A \subseteq M(p, q)$,

$$P(x, A) = \int_A Q(x, dy)\alpha(x, y) + R(x) \cdot \delta_x(A),$$

where we have defined $R(x) = 1 - \int Q(x, dy)\alpha(x, y)$, for acceptance probability function that has the simple form,

$$\alpha(x, y) = \min\{1, \pi(y)/\pi(x)\}, \quad \pi(x) := (d\Pi/d\xi)(x). \quad (5)$$

All Metropolis–Hastings kernels – on $M(p, q)$ – in this paper will correspond to instances of such triplets (Π, Q, ξ) . A similar notation is adopted for measures restricted on $P^+(q)$.

3.1 RWM and pCN

The proposal kernel $Q(x, dy)$ of the RWM algorithm on $M(p, q)$ is defined by the update

$$\text{RWM} : \quad y = x + w, \quad (6)$$

with $w \sim N_{p,q}(0, U, V)$, $U \in P^+(p)$, $V \in P^+(q)$. The proposal kernel $Q(x, \cdot) = N_{p,q}(x, U, V)$ is reversible with respect to the Lebesgue measure on $M(p, q)$. Thus, following the notation we established above, we now have the triplet (Π, Q, Leb) , and the acceptance probability $\alpha(x, y)$ is as in (5), with $\pi(x)$ the density of Π with respect to Leb . Let $\rho \in [0, 1)$. The pCN method on $M(p, q)$ is determined via the proposal,

$$\text{pCN} : \quad y = \rho^{1/2} x + (1 - \rho)^{1/2} w, \quad (7)$$

with $w \sim N_{p,q}(0, U, V)$. The proposal kernel $Q(x, \cdot) = N_{p,q}(\rho^{1/2}x, U, (1 - \rho)V)$ is reversible with respect to $N_{p,q}(0, U, V)$. We have the triplet $(\Pi, Q, N_{p,q}(0, U, V))$, so the acceptance probability is as in (5), where $\pi(x)$ is density of Π with respect to $N_{p,q}(0, U, V)$. The pCN algorithm was introduced in Section 4.2 of Neal (1999). As shown there, and in several more recent works (see e.g. Beskos et al., 2008; Cotter et al., 2013) pCN can be very effective in high dimensions in the context of Gaussian priors and not highly informative observations. The algorithms are well-defined even on infinite-dimensional Hilbert spaces; this has sparked the use of pCN in the area of Bayesian Inverse Problems, see e.g. the overview in Stuart (2010).

3.2 MpCN on $M(p, q)$

3.2.1 Derivation via Bayesian Paradigm

We return briefly at the scenario of Section 2.3, i.e. adopt the viewpoint that the original target distribution is given as $\tilde{\Pi}(dS) = \tilde{\pi}(S)\mu(dS)$ on $P^+(q)$, and one aims to generate $X \sim \Pi(dx) = \tilde{\pi}(x^\top \mathcal{U}^{-1}x)\nu_{\mathcal{U}}(dx)$ on $M(p, q)$, $p \geq q$, and return $S = X^\top \mathcal{U}^{-1}X$. In such an upcasted setting, one can provide a motivation for the derivation of MpCN. Let x_i denote the i th row of $x \in M(p, q)$. Since the density of $\Pi(dx)$ with respect to Leb writes as,

$$f(z_1^\top z_1 + \cdots + z_p^\top z_p); \quad z := \mathcal{U}^{-1/2}x,$$

where $f(S) = \tilde{\pi}(S) \det(S)^{-p/2}$, $S \in P^+(q)$, it is clear that, under $X \sim \Pi(dx)$,

$$Z_{1\cdot} =_d Z_{2\cdot} =_d \cdots =_d Z_{p\cdot},$$

so all rows of $Z := \mathcal{U}^{-1/2}X$ have the same marginal law. Looking back at the pCN proposal (7), as adjusted under the linear transform $x \mapsto \mathcal{U}^{-1/2}x =: z$, the above understanding can provide guidance for tuning the algorithmic parameter $V \in P^+(q)$ – corresponding to the variance of each row-vector of the noise $\mathcal{U}^{-1/2}w$ – given information about the current position $z = \mathcal{U}^{-1/2}x$ of the MCMC chain. In a ‘classical’ approach, a likelihood-based choice would simply be the sample variance over the rows, $\hat{V} := \sum_{i=1}^p z_i^\top z_i / p$. A Bayesian approach seems preferable as it will ultimately provide an algorithm with a heavier-tailed proposal. A choice of *Jeffrey’s prior* $\mu(dV)$ (see e.g. Geisser and Cornfield, 1963), with μ as given in (1), combined with a likelihood $\phi_{p,q}(z; 0, \mathcal{U}^{-1/2}U\mathcal{U}^{-1/2}, V)$, $U \in P^+(p)$, is easily shown to provide the posterior $W_q^{-1}(p, x^\top U^{-1}x)$ for V .

3.2.2 MpCN Proposal and Acceptance Probability

The above thinking gives rise to the following proposal, applicable for a general law $\Pi(dx)$ on $M(p, q)$ – not only in the upcasted scenario adapted above for purposes of providing some rationale under a particular viewpoint –,

$$\begin{aligned} \text{MpCN : } \quad y &= \rho^{1/2}x + (1 - \rho)^{1/2}w; \\ w &\sim N_{p,q}(0, U, V); \quad V \sim W_q^{-1}(p, x^\top U^{-1}x). \end{aligned} \tag{8}$$

Remark 2. *i. In the scalar case, it is rather common to choose an Inverse-Gamma distribution for a variance parameter $V \in \mathbb{R}_+$, thus an Inverse-Wishart distribution in the matrix case, $V \in P^+(q)$, appears as a fairly natural choice.*

ii. The law of $[w|x]$ is that of a Matrix-Student- t distribution with Lebesgue density proportional to $z \mapsto \det(x^\top U^{-1}x + z^\top U^{-1}z)^{-p}$; see e.g. Dickey (1967).

MpCN is a Metropolis method with target $\Pi(dx)$ on $M(p, q)$, and proposal as determined in (8). Let Q denote the related proposal transition kernel. The lemma that follows shows that Q is ν_U -reversible, so we have the triplet (Π, Q, ν_U) and the acceptance probability is as in (5), where $\pi(x)$ is the density of Π with respect to ν_U .

Lemma 3.1. *The MpCN proposal kernel Q on $M(p, q)$ has density $\mathbf{q}(x, y)$ with respect to $\nu_U(dy)$, that writes as,*

$$\mathbf{q}(x, y) = (1 - \rho)^{pq/2} c_{p,q} \frac{\det(x^\top U^{-1}x)^{p/2} \times \det(y^\top U^{-1}y)^{p/2}}{\det(R(x, y))^p}, \quad (9)$$

where, $c_{p,q} = \frac{\Gamma_q(p)}{\Gamma_q(p/2)^2}$, and,

$$R(x, y) = x^\top U^{-1}x + y^\top U^{-1}y - \rho^{1/2} x^\top U^{-1}y - \rho^{1/2} y^\top U^{-1}x.$$

Therefore, Q is ν_U -reversible.

Proof. Following the definition of MpCN, the joint distribution of $[V, y|x]$ writes as,

$$W_q^{-1}(dV; p, x^\top U^{-1}x) N_{p,q}(dy; \rho^{1/2}x, U, (1 - \rho)V). \quad (10)$$

It remains to integrate out V in (10), and take the density of the resulted distribution of $[y|x]$ with respect to $\nu_U(dy)$. All such calculations can be carried out analytically due to $\mu(dV)$ being conjugate with respect to Matrix-Normal distribution with right-covariance matrix V . Thus, tedious but otherwise straightforward calculations give the density $\mathbf{q}(x, y)$ in expression (9). Since $\mathbf{q}(x, y) = \mathbf{q}(y, x)$, we have $\nu_U(dx)Q(x, dy) = \nu_U(dy)Q(y, dx)$. \square

3.3 Random-Walk Property of MpCN on $P^+(q)$

Returning to the context of an initial target $\tilde{\Pi}(dS)$ on $P^+(q)$, we show here that, when the operators \mathcal{U} and U – used, respectively, when upcasting the target $\tilde{\Pi}$ onto $M(p, q)$ and as a

parameter in the MpCN proposal – coincide (i.e., $\mathcal{U} = U$), then MpCN on $M(p, q)$ induces a Markovian kernel on $P^+(q)$ under the transform $x \mapsto x^\top U^{-1}x$. Such kernel is easier to analyse as it exhibits random-walk-type behavior, as shown in the proposition and theorem that follow. We define the operation $A \circ B = B^{1/2}AB^{1/2}$, $A, B \in P^+(q)$; it follows that $(A \circ B)^{-1} = A^{-1} \circ B^{-1}$ and $\text{tr}(A \circ B) = \text{tr}(AB)$. Recall also the definition of the uniform measure $d\mathbf{U}$ in Section 2.1.

Proposition 3.2. *Let $\tilde{\Pi}(dS) = \tilde{\pi}(S)\mu(dS)$ and $\Pi(dx) = \tilde{\pi}(x^\top U^{-1}x)\nu_U(dx)$, be probability measures on $P^+(q)$ and $M(p, q)$ respectively, with $U \in P^+(q)$, $p \geq q$. Given $x \in M(p, q)$, set $y \sim Q(x, dy)$, where $Q = Q(\cdot; \rho, U)$ is the proposal kernel of MpCN, with parameters $\rho \in [0, 1)$ and U .*

(i) *We have the representation, for $\det(x^\top U^{-1}x) \neq 0$,*

$$y^\top U^{-1}y = \epsilon \circ (x^\top U^{-1}x),$$

for some $P^+(q)$ -valued random variable ϵ with law that does not depend on x, U .

(ii) *The law of ϵ , $\mathcal{L}(d\epsilon)$, writes as,*

$$\mathcal{L}(d\epsilon) = (1 - \rho)^{pq/2} c_{p,q} \mu(d\epsilon) \int_{\mathcal{O}(p,q)} \frac{\det(\epsilon)^{p/2}}{\det(I_q + \epsilon - \rho^{1/2}(\mathbf{U}_1 \epsilon^{1/2} + \epsilon^{1/2} \mathbf{U}_1^\top))^p} d\mathbf{U}, \quad (11)$$

with $c_{p,q}$ as given in Lemma 3.1, where \mathbf{U}_1 denotes the first q rows of \mathbf{U} . In particular, $\mathcal{L}(d\epsilon)$ depends on ρ , but not on U .

(iii) *We have $\epsilon =_d \epsilon^{-1}$.*

Proof. Let $u_1(x) = U^{-1/2}x(x^\top U^{-1}x)^{-1/2} \in \mathcal{O}(p, q)$. There exists an orthogonal complement, $u_2(x) \in M(p, p - q)$ as an analytic function, such that $u(x) = (u_1(x), u_2(x)) \in \mathcal{O}(p)$. We consider linear transformations of y ,

$$z_i = u_i(x)^\top U^{-1/2}y(x^\top U^{-1}x)^{-1/2}, \quad i = 1, 2,$$

and set $z = (z_1, z_2) \in M(p, q)$. Then $\epsilon = z^\top z$. By Lemma 3.1, and the change of variables formula (Lemma 1.5.1 of Chikuse, 2003), the law of z is,

$$(1 - \rho)^{pq/2} c_{p,q} \frac{\det(z^\top z)^{p/2}}{\det(I_q + z^\top z - \rho^{1/2}(z_1 + z_1^\top))^p} \nu(dz),$$

which does not involve x or U . Also, by the change of variables formula (2), the law of $\epsilon = z^\top z$ is as in (11). Finally, the law, $\mathcal{L}_{-1}(d\epsilon)$, of the inverse ϵ^{-1} writes as (upon recalling that $\mu_{-1} \equiv \mu$),

$$\begin{aligned}\mathcal{L}_{-1}(d\epsilon) &= (1 - \rho)^{pq/2} c_{p,q} \mu(d\epsilon) \times \int_{\mathcal{O}(p,q)} \frac{\det(\epsilon^{-1})^{p/2}}{\det(I_q + \epsilon^{-1} - \rho^{1/2}(\mathbf{U}_1 \epsilon^{-1/2} + \epsilon^{-1/2} \mathbf{U}_1^\top))^p} d\mathbf{U} \\ &= (1 - \rho)^{pq/2} c_{p,q} \mu(d\epsilon) \times \int_{\mathcal{O}(p,q)} \frac{\det(\epsilon)^{p/2}}{\det(I_q + \epsilon - \rho^{1/2}(\mathbf{U}_1^\top \epsilon^{1/2} + \epsilon^{1/2} \mathbf{U}_1))^p} d\mathbf{U}.\end{aligned}$$

From Theorem 3.3.1 of Chikuse (2003), the uniform distribution on $\mathcal{O}(p, q)$ is the marginal (on $\mathcal{O}(p, q)$) of the uniform distribution on $\mathcal{O}(p)$. Also, by Section 1.4.1 of Chikuse (2003), the uniform distribution on $\mathcal{O}(p)$ is invariant under the matrix transpose operation. Thus, the distribution of \mathbf{U}_1 is also invariant under the matrix transpose operation. The proof is now complete. \square

Proposition 3.2 leads to the theorem below.

Theorem 3.3. *Let $\tilde{\Pi}(dS) = \tilde{\pi}(S)\mu(dS)$ be a target on $P^+(q)$. Define the corresponding upcasted law on $M(p, q)$, $\Pi(dx) = \tilde{\pi}(x^\top U^{-1}x)\nu_U(dx)$, $U \in P^+(q)$, and let $Q = Q(\cdot; \rho, U)$ be the MpCN proposal kernel. Consider also the kernel $\tilde{Q}(S, \cdot) :=_d \epsilon \circ S$, $S \in P^+(q)$, with the law $\mathcal{L}(d\epsilon)$ of $\epsilon \in P^+(q)$ as determined in Proposition 3.2(ii), for parameters $\rho \in [0, 1)$ and $p, p \geq q$.*

If $\{X_n\}_{n \geq 0}$ is the MpCN Markov chain with target Π and proposal Q , then the process $\{S_n\}_{n \geq 0}$ with,

$$S_n := X_n^\top U^{-1} X_n,$$

is a Metropolis–Hastings Markov chain, with respect to its own filtration, with target $\tilde{\Pi}$ and proposal \tilde{Q} (and initial position $S_0 =_d X_0^\top U^{-1} X_0$). Moreover, it is a Random-Walk Markov chain, in the sense that \tilde{Q} is the transition kernel of a random walk $T_n = \epsilon_n \circ T_{n-1}$, $n \geq 1$, where ϵ_n follows a probability distribution \mathcal{L} satisfying $\mathcal{L} = \mathcal{L}_{-1}$.

Given Theorem 3.3, we will refer – without confusion – to the Markov kernel, denoted \tilde{P} , with target $\tilde{\Pi}$ and proposal $\tilde{Q} = \tilde{Q}(\cdot; \rho, p)$ as “an MpCN kernel on $P^+(q)$ ”. Notice that both $\tilde{\Pi}$ and \tilde{Q} do not depend on the choice of parameter U . The MpCN kernel on $P^+(q)$ is a Random-Walk Metropolis kernel on this space.

Proof of Theorem 3.3. We show that $\{S_n\}_{n \geq 0}$ is in the class of Metropolis–Hastings Markov chains introduced in the beginning of Section 3, with triplet $(\tilde{\Pi}, \tilde{Q}, \mu)$. First, we prove μ -reversibility of \tilde{Q} , by making use of the ν_U -reversibility of Q itself. Indeed, for any given Borel sets $A, B \subseteq P^+(q)$, upon defining the sets $A^* = \{x \in M(p, q) : x^\top U^{-1}x \in A\}$ and $B^* = \{x \in M(p, q) : x^\top U^{-1}x \in B\}$, we have,

$$\begin{aligned} \int_A \tilde{Q}(S, B) \mu(dS) &= \int_{A^*} Q(x, B^*) \nu_U(dx) \\ &= \int_{B^*} Q(x, A^*) \nu_U(dx) = \int_B \tilde{Q}(S, A) \mu(dS). \end{aligned}$$

The acceptance probability of $\{X_n\}_{n \geq 0}$ coincides with that of the triplet $(\tilde{\Pi}, \tilde{Q}, \mu)$ via the surjective mapping $x \mapsto x^\top U^{-1}x$. The proof is complete. \square

Proposition 3.4. *Let $\{S_n\}_{n \geq 0}$ be a Markov process on $P^+(q)$ with transition kernel corresponding to that of an MpCN chain with target $\tilde{\Pi}$ and proposal $\tilde{Q}(\cdot; \rho, p)$. Then, $\{S_n^{-1}\}_{n \geq 0}$ is a Markov process on $P^+(q)$ with transition kernel corresponding to that of an MpCN transition kernel with target $\tilde{\Pi}_{-1}$ and identical proposal $\tilde{Q}(\cdot; \rho, p)$.*

Proof. Recall that $\{S_n\}_{n \geq 0}$ corresponds to the triplet $(\tilde{\Pi}, \tilde{Q}, \mu)$. We show that $\{S_n^{-1}\}_{n \geq 0}$ corresponds to triplet $(\tilde{\Pi}_{-1}, \tilde{Q}, \mu)$. Since $S \mapsto S^{-1}$ is a bijection in $P^+(q)$, $\{S_n^{-1}\}_{n \geq 0}$ is also a Markov chain. The third component of the triplet is invariant under this transform, since $\mu_{-1} = \mu$. The second component is also invariant since $\tilde{Q}(S, A) = \tilde{Q}(S^{-1}, A^{-1})$, for $A \subseteq P^+(q)$, because $\epsilon =_d \epsilon^{-1}$ by Proposition 3.2(iii). The first component of the triplet $\tilde{\Pi}$ becomes $\tilde{\Pi}_{-1}$. Finally, the acceptance probability $\alpha(S, S^*)$ of the triplet $(\tilde{\Pi}, \tilde{Q}, \mu)$ is the same as that of $(\tilde{\Pi}_{-1}, \tilde{Q}, \mu)$ via the projection $S \mapsto S^{-1}$. Therefore $\{S_n^{-1}\}_{n \geq 0}$ is the Metropolis chain determined by $(\tilde{\Pi}_{-1}, \tilde{Q}, \mu)$. The proof is complete. \square

Notice that if a Markov chain $\{S_n\}_{n \geq 0}$ is geometrically ergodic, and $T_n = f(S_n)$ forms a Markov chain – for a map f , on appropriate domains – then $\{T_n\}_{n \geq 0}$ is also geometrically ergodic, since the σ -algebra generated by $\{S_n\}_{n \geq 0}$ contains that of $\{T_n\}_{n \geq 0}$. This observation is used in the following two statements.

- (i) If the upcasted MpCN kernel on $M(p, q)$ is geometrically ergodic, then the deduced MpCN kernel on $P^+(q)$ is also geometrically ergodic.

- (ii) Proposition 3.4 implies that, if the MpCN kernel on $P^+(q)$ targeting $\tilde{\Pi}$ is geometrically ergodic, then the MpCN kernel targeting $\tilde{\Pi}_{-1}$ (for the same ρ, p) is also geometrically ergodic. Indicatively, geometric ergodicity of MpCN with Wishart target $\tilde{\Pi} \equiv W_q(r, T)$ is equivalent to geometric ergodicity of MpCN with Inverse-Wishart target $\tilde{\Pi}_{-1} \equiv W_q^{-1}(r, T^{-1})$, for $r \in \mathbb{R}$, $r > q - 1$, and $T \in P^+(q)$.

4 Ergodicity Results for RWM and pCN

The ergodic properties of the RWM and pCN kernels on a vector space have been studied in Mengersen and Tweedie (1996); Roberts and Tweedie (1996); Jarner and Hansen (2000); Rudolf and Sprungk (2016); Kamatani (2017). In this section, we investigate ergodicity on a matrix space. First, we note that RWM and pCN kernels are ergodic under fairly general assumptions (e.g., Meyn and Tweedie, 1994; Kulik and Scheutzow, 2015). Therefore, in this section we concentrate on geometric ergodicity.

Throughout this section, and unless specified otherwise, $U \in P^+(p)$, $V \in P^+(q)$, $\rho \in [0, 1)$ are the parameters appearing in the RWM and pCN proposal kernels on $M(p, q)$, $p \geq q$. In this paper, a Markov kernel P on (E, \mathcal{E}) is said to be geometrically ergodic if there is a probability measure Π and $r \in (0, 1)$ such that

$$V(x) = r^{-n} \|P^n(x, \cdot) - \Pi\|_{\text{TV}}$$

is Π -integrable, where for general measures μ, ν , $\|\mu - \nu\|_{\text{TV}} = \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|$.

4.1 Ergodicity for RWM

We provide a sufficient and a necessary condition for the geometric ergodicity of the RWM kernel. Many MCMC methods do not work well for target distributions with contour manifolds that degenerate – in a proper sense – in the tails; see, e.g., Section 5 of Jarner and Hansen (2000). To exclude such cases, we consider the following class of functions.

Definition 4.1. *A continuously differentiable function $h : M(p, q) \mapsto \mathbb{R}_+$ satisfies the*

contour condition if,

$$\lim_{\|x\|_F \rightarrow +\infty} \left\langle \frac{Dh(x)}{\|Dh(x)\|_F}, \frac{x}{\|x\|_F} \right\rangle_F < 0. \quad (12)$$

A sufficient condition for geometric ergodicity is formulated by using the above contour condition together with the following exponentially light tail condition (13). Though the theory in Jarner and Hansen (2000) concerns only vector-valued processes, we can apply their result if we consider the $M(p, q)$ space as a vector space of length pq .

Proposition 4.2 (Jarner and Hansen 2000). *Consider the law $\Pi(dx) = \pi(x) \text{Leb}(dx)$ on $M(p, q)$. Assume that $\log \pi$ is continuously differentiable, satisfies the contour condition, and,*

$$\lim_{\|x\|_F \rightarrow +\infty} \left\langle D \log \pi(x), \frac{x}{\|x\|_F} \right\rangle = -\infty. \quad (13)$$

Then, RWM with target Π is geometrically ergodic.

A necessary condition can be formulated via a moment requirement, that first appeared in Jarner and Tweedie (2003) for Euclidean spaces. In Kamatani (2017) such results were generalized to metric spaces including the matrix space $M(p, q)$ with Frobenius norm $\|\cdot\|_F$. The result can also be applied to probability measures on $P^+(q)$, upcasted onto $M(p, q)$.

Proposition 4.3 (Jarner and Tweedie 2003; Kamatani 2017).

(i) *If the RWM kernel with target $\Pi(dx)$ on $M(p, q)$ is geometrically ergodic, then, for some $s > 0$,*

$$\int_{M(p, q)} \exp(s\|x\|_F) \Pi(dx) < \infty.$$

(ii) *Let $\tilde{\Pi}(dS) = \tilde{\pi}(S) \mu(dS)$ be a probability distribution on $P^+(q)$, with corresponding upcasted law on $M(p, q)$, $\Pi(dx) = \tilde{\pi}(x^\top \mathcal{U}^{-1} x) \nu_{\mathcal{U}}(dx)$. If the RWM chain with target Π is geometrically ergodic, then, for some $s > 0$,*

$$\int_{P^+(q)} \exp\left(s \sqrt{\text{tr}(S)}\right) \tilde{\Pi}(dS) < \infty.$$

For example, $\alpha \geq 1$ is necessary and $\alpha > 1$ sufficient for the Lebesgue density $\pi(x) \propto \exp(-\beta\|x\|_F^\alpha)$, $\beta > 0$. In contrast, RWM is expected not to be geometrically ergodic when $\pi(x)$ has heavy tails.

Regular variation on \mathbb{R} is a concept met in a large literature – see, e.g., Resnick (2007); Bingham et al. (1989) – and can be used to characterise popular classes of heavy-tailed functions. Regular variation on $P^+(q)$ is a natural extension from \mathbb{R} . We first define an appropriate metric. A matrix $A \in P^+(q)$ writes as $A = \mathbf{U}\Lambda\mathbf{U}^\top$, for $\mathbf{U} \in \mathcal{O}(q)$ and positive diagonal matrix Λ . The logarithmic map, $\log : P^+(q) \rightarrow \text{Sym}(q)$, is given by $\log A = \mathbf{U}(\log \Lambda)\mathbf{U}^\top$ where $\log \Lambda := \text{diag}\{\log \lambda_1, \dots, \log \lambda_q\}$. We consider a metric $\mathbf{d}(\cdot, \cdot)$ on $P^+(q)$, such that,

$$\mathbf{d}(A, B) := \|\log(A \circ B^{-1})\|_F = \left\{ \sum_{i=1}^q \log^2 \lambda_i \right\}^{1/2}, \quad (14)$$

where $\lambda_1, \dots, \lambda_q$ are eigenvalues of $A \circ B^{-1}$, $A, B \in P^+(q)$. This is a natural distance induced by the logarithmic map; see Theorem XII.1.3 of Lang (1999). Note that a matrix can have several square roots, in general, but the value of $\mathbf{d}(A, B)$ is unique. The topology induced by the metric \mathbf{d} and that by the Frobenius norm are different; the former fits naturally to $P^+(q)$. $P^+(q)$ is a complete metric space under \mathbf{d} , but not under the Frobenius norm: e.g., observe that $A_n = \text{diag}\{n^{-1}, 1, \dots, 1\}$ forms a Cauchy sequence under the Frobenius norm, with a limit $A = \text{diag}\{0, 1, \dots, 1\} \notin P^+(q)$.

Definition 4.4. *A function $h : P^+(q) \rightarrow \mathbb{R}_+$ is regularly varying if there exists $r \in \mathbb{R}$, with $r > q - 1$, such that,*

$$\frac{h(zx)}{h(zI_q)} \xrightarrow{z \rightarrow +\infty} \det(x)^{-r/2},$$

locally uniformly in $x \in P^+(q)$ under the topology induced by the metric $\mathbf{d}(\cdot, \cdot)$ in (14).

Remark 3. *The probability density function of the Inverse-Wishart law $W_q^{-1}(r, T)$ is regularly varying since,*

$$\frac{\pi(zx)}{\pi(zI_q)} = \det(x)^{-r/2} \exp(-z^{-1} \text{tr}(T(x^{-1} - I_q))) \xrightarrow{z \rightarrow +\infty} \det(x)^{-r/2}.$$

Corollary 4.5. *Let $\tilde{\Pi}(\text{d}S) = \tilde{\pi}(S)\mu(\text{d}S)$ be a probability law on $P^+(q)$. If $\tilde{\pi}$ is continuous, regularly varying, then the RWM chain with target $\Pi(\text{d}x) = \tilde{\pi}(x^\top \mathcal{U}^{-1}x)\nu_{\mathcal{U}}(\text{d}x)$ on $M(p, q)$ is not geometrically ergodic.*

Proof. Any $S \in P^+(q)$ writes as $S = \mathbf{U}^\top \Lambda \mathbf{U} = (\sum_i \lambda_i) \mathbf{U}^\top \Lambda^* \mathbf{U} = z \mathbf{U}^\top \Lambda^* \mathbf{U}$, with $z = \sum_i \lambda_i$, $\mathbf{U} \in \mathcal{O}(q)$, $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_q\}$, $\Lambda^* = \text{diag}\{\lambda_1^*, \dots, \lambda_q^*\}$, $0 < \lambda_1 < \dots < \lambda_q$, where we have set $\lambda_i^* = \lambda_i / \sum_j \lambda_j$. We define a bounded set, for $\varepsilon \geq 0$,

$$\Gamma_\varepsilon := \left\{ \lambda^* = (\lambda_1^*, \dots, \lambda_{q-1}^*) \in \mathbb{R}_+^{q-1} : \varepsilon < \lambda_1^* < \dots < \lambda_q^*, \lambda_q^* = 1 - \sum_{i=1}^{q-1} \lambda_i^* \right\}.$$

We introduced ε here so that we can use Lebesgue's dominated convergence theorem in the following inequality. By Theorem 5.3.1 of Farrell (1985), for $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, we have,

$$\begin{aligned} \int_{S \in P^+(q)} f(\text{tr}(S)) \tilde{\Pi}(dS) &= c \int f(\text{tr} \Lambda) \prod_{i < j} (\lambda_j - \lambda_i) \prod_{i=1}^q \lambda_i^{-(q+1)/2} d\lambda_i \int_{\mathcal{O}(q)} \tilde{\pi}(\mathbf{U}^\top \Lambda \mathbf{U}) d\mathbf{U} \\ &= c \int_{\mathbb{R}_+} f(z) \left\{ \int_{\lambda^* \in \Gamma_0} \prod_{i=1}^q \prod_{j=i+1}^q (\lambda_j^* - \lambda_i^*) (\lambda_i^*)^{-(q+1)/2} d\lambda^* \int_{\mathcal{O}(q)} \tilde{\pi}(z \mathbf{U}^\top \Lambda^* \mathbf{U}) d\mathbf{U} \right\} z^{-1} dz \\ &\geq c \int_{\mathbb{R}_+} f(z) \left\{ \int_{\lambda^* \in \Gamma_\varepsilon} \prod_{i=1}^q \prod_{j=i+1}^q (\lambda_j^* - \lambda_i^*) (\lambda_i^*)^{-(q+1)/2} d\lambda^* \int_{\mathcal{O}(q)} \tilde{\pi}(z \mathbf{U}^\top \Lambda^* \mathbf{U}) d\mathbf{U} \right\} z^{-1} dz \\ &= c \int_{\mathbb{R}_+} f(z) h_\varepsilon(z) z^{-1} dz, \end{aligned}$$

with $h_\varepsilon(z)$ defined above in an obvious way. Given that $\tilde{\pi}$ is assumed to be regularly varying, $h_\varepsilon(z)$ can also be shown to be regularly varying, since for any $x > 0$ we have,

$$\frac{\tilde{\pi}(zx \mathbf{U}^\top \Lambda^* \mathbf{U})}{\tilde{\pi}(zI_q)} \xrightarrow{z \rightarrow \infty} \det(x \mathbf{U}^\top \Lambda^* \mathbf{U})^{-r/2} = x^{-rq/2} \det(\mathbf{U}^\top \Lambda^* \mathbf{U})^{-r/2},$$

for some $r \in \mathbb{R}$, with $r > q - 1$; hence, from Lebesgue's dominated convergence theorem,

$$\frac{h_\varepsilon(zx)}{h_\varepsilon(z)} = \frac{h_\varepsilon(zx)/\tilde{\pi}(zI_q)}{h_\varepsilon(z)/\tilde{\pi}(zI_q)} \xrightarrow{z \rightarrow \infty} x^{-rq/2}.$$

Thus, for $f(x) = \exp(s\sqrt{x})$, $s, x > 0$, we have obtained,

$$\int_{S \in P^+(q)} \exp(s\sqrt{\text{tr}(S)}) \tilde{\Pi}(dS) \geq c \int_0^\infty \exp(s\sqrt{z}) h_\varepsilon(z) z^{-1} dz.$$

By Theorem 1.5.6(iii) of Bingham et al. (1989), $h_\varepsilon(z) z^{-1} \geq Cz^{-rq/2-1-\delta}$, for any $\delta > 0$, for some $C > 0$, and sufficiently large $z > 0$. The claim of Corollary 4.5 follows, since,

$$\lim_{z \rightarrow \infty} \exp(s\sqrt{z}) h_\varepsilon(z) z^{-1} = +\infty,$$

for any $s > 0$, that violates the integrability condition in Proposition 4.3(ii). \square

4.1.1 Example Cases

Using the above results, we check geometric ergodicity for RWM for the standard probability measures written down in Section 2.2.

1. Let $\pi(x)$ be the density function of $N_{p,q}(M, \Sigma, T)$ under Leb. Then, we have the derivative $D \log \pi(x) = -\Sigma^{-1}(x - M)T^{-1}$. Thus, as $\|x\|_F \rightarrow \infty$, the inner product term at the contour condition (12) is dominated from above by,

$$-\inf_{e:\|e\|_F=1} \frac{\langle \Sigma^{-1}eT^{-1}, e \rangle_F}{\|\Sigma^{-1}eT^{-1}\|_F} \leq -\frac{\inf_{e:\|e\|_F=1} \langle \Sigma^{-1}eT^{-1}, e \rangle_F}{\sup_{e:\|e\|_F=1} \|\Sigma^{-1}eT^{-1}\|_F}.$$

The denominator is bounded from above, and the numerator is bounded away from 0 since Σ and T are positive definite. Thus, the density satisfies the contour condition. We can also check that π satisfies (13). Therefore, RWM is geometrically ergodic.

2. Consider the Wishart distribution $W_q(r, T)$, $r \in \mathbb{R}$, $r > q - 1$. Since the probability measure is defined on $P^+(q)$, we use the upcasting strategy to apply RWM. The density function of the upcasted law of $W_q(r, T)$ under Leb is,

$$\pi(x) = \text{const.} \times \det(x^\top \mathcal{U}^{-1}x)^{r/2} \exp(-\text{tr}[T^{-1}(x^\top \mathcal{U}^{-1}x)]/2) \det(x^\top \mathcal{U}^{-1}x)^{-p/2}.$$

Thus,

$$D \log \pi(x) = \frac{r-p}{2} D \log(\det(x^\top \mathcal{U}^{-1}x)) - \mathcal{U}^{-1}xT^{-1}.$$

By the triangle inequality, inequalities (12) and (13) will follow once we show,

$$\lim_{\|x\|_F \rightarrow \infty} \frac{\|D \log(\det(x^\top \mathcal{U}^{-1}x))\|_F}{\|x\|_F} = 0.$$

By Equation (15.8.6) of Harville (1997),

$$\begin{aligned} \frac{d}{dx_{ij}} \log(\det(x^\top \mathcal{U}^{-1}x)) &= \text{tr} \left[(x^\top \mathcal{U}^{-1}x)^{-1} \frac{d}{dx_{ij}} (x^\top \mathcal{U}^{-1}x) \right] \\ &= 2 [\mathcal{U}^{-1}x(x^\top \mathcal{U}^{-1}x)^{-1}]_{ij}. \end{aligned}$$

Therefore,

$$\begin{aligned} \|D \log(\det(x^\top \mathcal{U}^{-1}x))\|_F^2 &= 4 \sum_{ij} [\mathcal{U}^{-1}x(x^\top \mathcal{U}^{-1}x)^{-1}]_{ij}^2 \\ &\leq 4 \|x\|_F^{-1} \sup_{e:\|e\|_F=1} \sum_{ij} [\mathcal{U}^{-1}e(e^\top \mathcal{U}^{-1}e)^{-1}]_{ij}^2, \end{aligned}$$

which converges to 0 as $\|x\|_F \rightarrow \infty$. So inequalities (12) and (13) hold. This implies that RWM is geometrically ergodic for the Wishart distribution.

3. By Corollary 4.5, since the Inverse-Wishart distribution has regularly varying density function (see Remark 3), RWM is not geometrically ergodic for $W_q^{-1}(r, T)$.

4.2 Ergodicity of pCN kernel

The pCN algorithm is more sensitive as to the choice of target distribution. We establish a necessary condition for geometric ergodicity by using the decay of deviation of $\log \pi$, with π the Lebesgue density of the target. By the following result, if $\pi(x) \propto \exp(-\beta\|x\|_F^\alpha)$, $x \in M(p, q)$, then it is necessary that $\alpha \geq 2$.

Proposition 4.6.

- (i) Consider $\Pi(dx) = \pi(x) \text{Leb}(dx)$. Let λ be the largest amongst all eigenvalues of U , V (recall these are parameters appearing at the pCN proposal kernel). Set,

$$C_{r,\varepsilon}(s) := \sup_{s': \|s-s'\|_F \leq \varepsilon} r^{-2} \left| \log \pi(rs) - \log \pi(rs') \right|.$$

If, for some $\varepsilon > 1 - \rho^{1/2}$,

$$\inf_{s: \|s\|_F=1} \liminf_{r \rightarrow \infty} C_{r,\varepsilon}(s) < \frac{1-\rho}{2} \lambda^{-2},$$

then the pCN kernel with target Π is not geometrically ergodic.

- (ii) Consider $\tilde{\Pi}(dS) = \tilde{\pi}(S) \mu(dS)$. Let $\tilde{\lambda}$ be the largest amongst all eigenvalues of U and V (recall U is the matrix appearing in the definition of the upcasted target Π). Set,

$$C_{r,\varepsilon}(S) := \sup_{S': d(S,S') < \varepsilon} r^{-2} \left| \log \tilde{\pi}(r^2 S) - \log \tilde{\pi}(r^2 S') \right|.$$

If, for some $\varepsilon > q^{1/2} 2 \log(1 - \rho^{1/2})$,

$$\inf_{S \in P^+(q), \text{tr}(S)=1} \liminf_{r \rightarrow \infty} C_{r,\varepsilon}(S) < \frac{1-\rho}{2} \tilde{\lambda}^{-2},$$

then the pCN chain with target Π is not geometrically ergodic.

See Appendix A for the proof. We state an immediate consequence of Proposition 4.6 for probability measures with regularly varying densities on $P^+(q)$.

Corollary 4.7. *Consider the target $\tilde{\Pi}(dS) = \tilde{\pi}(S)\mu(dS)$ on $P^+(q)$. If $\tilde{\pi}$ is continuous, regularly varying, then the pCN kernel with the corresponding upcasted target $\Pi(dx)$ on $M(p, q)$ is not geometrically ergodic.*

Proof. By the regularly varying property, for some $r \in \mathbb{R}$, $r > q - 1$,

$$\begin{aligned} \sup_{S':d(S,S')<\varepsilon} \left| \log \tilde{\pi}(rS) - \log \tilde{\pi}(rS') \right| &= \sup_{S':d(S,S')<\varepsilon} \left| \log \frac{\tilde{\pi}(rS)/\tilde{\pi}(rI_q)}{\tilde{\pi}(rS')/\tilde{\pi}(rI_q)} \right| \\ &\xrightarrow{r \rightarrow \infty} \sup_{S':d(S,S')<\varepsilon} \left| \log \frac{(\det S)^{-r/2}}{(\det S')^{-r/2}} \right| < \infty. \end{aligned}$$

In particular,

$$\sup_{S':d(S,S')<\varepsilon} r^{-1} \left| \log \tilde{\pi}(rS) - \log \tilde{\pi}(rS') \right| \xrightarrow{r \rightarrow \infty} 0.$$

Thus, the pCN kernel is not geometrically ergodic. \square

4.2.1 Example Cases

We check geometric ergodicity for pCN for the standard targets in Section 2.2. We only state negative results as we did not have positive results in this paper.

1. Let $\pi(x)$ be the density function of $N_{p,q}(M, \Sigma, T)$. Let λ and λ_* be the largest and smallest amongst all eigenvalues of Σ and T . Then,

$$\begin{aligned} C_{r,\varepsilon}(s) &\xrightarrow{r \rightarrow \infty} \sup_{s':\|s-s'\|_F < \varepsilon} \left| \text{tr} \left[T^{-1}s^\top \Sigma^{-1}s - T^{-1}s'^\top \Sigma^{-1}s' \right] / 2 \right| \\ &= \sup_{s':\|s-s'\|_F < \varepsilon} \left| \langle (s - s')T^{-1}, \Sigma^{-1}(s + s') \rangle_F / 2 \right| \\ &\leq \lambda_*^{-2} \sup_{s':\|s-s'\|_F < \varepsilon} \left\{ \|s - s'\|_F \|s + s'\|_F / 2 \right\} \\ &\leq \lambda_*^{-2} \varepsilon (\|s\|_F + \varepsilon / 2). \end{aligned}$$

Thus, if,

$$\lambda_*^{-2} (1 - \rho^{1/2})(1 + (1 - \rho^{1/2})/2) < \frac{1-\rho}{2} \lambda^{-2},$$

which is simplified to,

$$\frac{3 - \rho^{1/2}}{1 + \rho^{1/2}} < \frac{\lambda_*^2}{\lambda^2},$$

then pCN is not geometrically ergodic. Therefore, pCN is not *always* geometrically ergodic for the matrix-normal distribution.

2. Similar calculations yield that, for a Wishart distribution $W_q(r, T)$, we have,

$$C_{r,\varepsilon}(S) \xrightarrow{r \rightarrow \infty} \sup_{S': d(S, S') < \varepsilon} |\text{tr}[T^{-1}(S - S')]| / 2.$$

Therefore, if T is large enough, then the condition in Proposition 4.6(ii) is satisfied, and pCN is not geometrically ergodic.

3. pCN is not geometrically ergodic for the upcasted Inverse-Wishart distribution by Corollary 4.7, since the relevant probability density function is a regularly varying function (see Remark 3).

5 MpCN Ergodicity Investigation

Consider a target law $\Pi(dx) = \pi(x)\nu_{\mathcal{U}}(dx)$. If $\pi(x) > 0$ for any full-rank $x \in M(p, q)$ then MpCN is ergodic. The MpCN algorithm is expected to have good convergence properties even for heavy-tailed target distributions. We have made a lot of progress towards proving *geometric* ergodicity for MpCN, via use of Foster–Lyapunov drift criterion (Meyn and Tweedie, 1993) and Dirichlet form. The last remaining steps for a fully completed proof remain subject of future research.

5.1 Our Results

Our investigation focuses mainly at the setting of Section 3.3, when: the initial target is $\tilde{\Pi}(dS) = \tilde{\pi}(dS)\mu(dS)$, $S \in P^+(q)$; the upcasted distribution is $\Pi(dx) = \tilde{\pi}(x^\top \mathcal{U}^{-1}x)\nu_{\mathcal{U}}(dx)$, $x \in M(p, q)$; the choice $\mathcal{U} = U$ gives rise to a Markov chain on $P^+(q)$ defined via what we have called the MpCN kernel \tilde{P} on $P^+(q)$, with target $\tilde{\Pi}$ and proposal kernel $\tilde{Q} = \tilde{Q}(\cdot; \rho, p)$

both of which do not depend on U – see also the comment after Theorem 3.3. To prove geometric ergodicity, it suffices to show, that for drift function $\mathbf{V} : P^+(q) \rightarrow \mathbb{R}_+$,

$$\limsup_{\text{tr}(S) \rightarrow \infty} \frac{\tilde{P}\mathbf{V}(S) - \mathbf{V}(S)}{\mathbf{V}(S)} < 0; \quad (15)$$

$$\limsup_{\det(S) \rightarrow 0} \frac{\tilde{P}\mathbf{V}(S) - \mathbf{V}(S)}{\mathbf{V}(S)} < 0. \quad (16)$$

We provide here a brief explanation for the requirement to consider the limits $\text{tr}(S) \rightarrow \infty$, $\det(S) \rightarrow 0$, and not the ones $\|S\|_F \rightarrow \infty$, $\|S\|_F \rightarrow 0$ that might appear as more natural candidates. Since $S \mapsto \det(S)$, $S \mapsto \text{tr}(S)$ are continuous in the metric $\mathbf{d}(A, B)$ in (14), $\{S \in P^+(q) : \epsilon \leq \det(S), \text{tr}(S) \leq \epsilon^{-1}\}$ is a compact set for $\epsilon \in (0, 1)$. On the other hand, every compact set is a small set for \tilde{P} . Thus, (15) and (16) give that $(\tilde{P}\mathbf{V} - \mathbf{V})/\mathbf{V}$ is smaller than a negative constant outside a small set, implying that the geometric drift condition is satisfied. In contrast, $\{S \in P^+(q) : \epsilon \leq \|S\|_F \leq \epsilon^{-1}\}$ is *not* a compact set in the metric $\mathbf{d}(A, B)$. Thus, indeed one needs to work with the limits $\text{tr}(S) \rightarrow \infty$, $\det(S) \rightarrow 0$, and not with $\|S\|_F \rightarrow \infty$, $\|S\|_F \rightarrow 0$. We prove (15), for an appropriate drift function. We start with a definition (for matrices $S, T \in P^+(q)$, we write $S < T$ if $T - S \in P^+(q)$).

Definition 5.1. (i) For $T \in P^+(p)$, $h : M(p, q) \rightarrow \mathbb{R}$ is rapidly varying if, for full rank $x, y \in M(p, q)$, and $z \in P^+(q)$,

$$\frac{h(yz)}{h(xz)} \xrightarrow{\text{tr}(z) \rightarrow \infty} \begin{cases} \infty, & \text{if } x^\top T^{-1}x > y^\top T^{-1}y; \\ 0, & \text{if } x^\top T^{-1}x < y^\top T^{-1}y. \end{cases}$$

(ii) For target distribution $\tilde{\Pi}(\mathrm{d}S) = \tilde{\pi}(S)\mu(\mathrm{d}S)$ we will call $\tilde{\pi}$ rapidly varying when $\pi(x) = \tilde{\pi}(x^\top U^{-1}x)$ is rapidly varying.

The above is a natural extension of rapid variation for scalar-valued functions; see, e.g., Section 2.4 of Resnick (2007) for details on the scalar case. Rapid variation is relevant for several light-tailed distributions. The Lebesgue density of Matrix-Normal distribution is rapidly varying, as is the density $\tilde{\pi} = \tilde{\pi}(S)$ of the Wishart distribution in (3).

Proposition 5.2. If $\tilde{\pi}(S)$ is strictly positive, continuous, rapidly varying function, then (15) holds for drift function $\mathbf{V}(S) = \tilde{\pi}(S)^{-\alpha}$, with any $\alpha \in (0, 1)$.

Proof. The proof is given in Appendix B. \square

It remains to establish (16), where $\det(S) \rightarrow 0$, or equivalently, $S \rightarrow S_0$, with S_0 degenerate, positive semi-definite, symmetric matrix. We prove the drift inequality for the special case $S_0 = 0$; this also provides a proof of (16) in the trivial scalar case $q = 1$.

Proposition 5.3. *Assume that $\tilde{\pi}$ is strictly positive, continuous. Suppose that for some $\zeta \neq 0$, for any $\epsilon \in P^+(q)$,*

$$\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \xrightarrow{\text{tr}(S) \rightarrow 0} \det(\epsilon)^{\zeta/2}.$$

For $\mathbf{V}(S) = \tilde{\pi}(S)^{-\alpha}$, $\alpha > 0$, the MpCN kernel \tilde{P} on $P^+(q)$, with target $\tilde{\Pi}$, satisfies,

$$\limsup_{\text{tr}(S) \rightarrow 0} \frac{\tilde{P}\mathbf{V}(S) - \mathbf{V}(S)}{\mathbf{V}(S)} < 0.$$

Proof. Recall that the proposal writes as $\epsilon \circ S$ with $\epsilon \sim \mathcal{L}(\text{d}\epsilon)$. By the dominated convergence theorem,

$$\begin{aligned} \frac{\tilde{P}\mathbf{V}(S) - \mathbf{V}(S)}{\mathbf{V}(S)} &= \mathbb{E} \left[\left\{ \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right)^{-\alpha} - 1 \right\} \min \left\{ 1, \frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right\} \right] \\ &\rightarrow \mathbb{E} \left[\left\{ \det(\epsilon)^{-\alpha\zeta/2} - 1 \right\} \min \left\{ 1, \det(\epsilon)^{\zeta/2} \right\} \right]. \end{aligned}$$

Since ϵ and ϵ^{-1} have the same law, we obtain,

$$\begin{aligned} &\mathbb{E} \left[\left\{ \det(\epsilon)^{-\alpha\zeta/2} - 1 \right\} \min \left\{ 1, \det(\epsilon)^{\zeta/2} \right\} \right] \\ &= \mathbb{E} \left[\left\{ \det(\epsilon)^{-\alpha\zeta/2} - 1 \right\}, \det \epsilon \geq 1 \right] + \mathbb{E} \left[\left\{ \det(\epsilon)^{-\alpha\zeta/2} - 1 \right\} \det(\epsilon)^{\zeta/2}, \det \epsilon < 1 \right] \\ &= \mathbb{E} \left[\left\{ \det(\epsilon)^{-\alpha\zeta/2} - 1 \right\}, \det \epsilon \geq 1 \right] + \mathbb{E} \left[\left\{ \det(\epsilon)^{\alpha\zeta/2} - 1 \right\} \det(\epsilon)^{-\zeta/2}, \det \epsilon > 1 \right] \\ &= \mathbb{E} \left[\left\{ \det(\epsilon)^{-\alpha\zeta/2} - 1 \right\} \left\{ 1 - \det(\epsilon)^{-(1-\alpha)\zeta/2} \right\}, \det \epsilon \geq 1 \right], \end{aligned}$$

with the last expression being negative. \square

In the proof above, the negativity of the limiting integral follows directly from the symmetry $\mathcal{L}(\epsilon) = \mathcal{L}(\epsilon^{-1})$ when $S \rightarrow 0$. The general scenario $S \rightarrow S_0$ is more complicated. In this case, the limiting integral depends on S_0 , and the symmetry does not simplify the integral. Thus, the above proof cannot be applied to the case $S_0 \neq 0$.

To stress the effect of algorithmic parameter ρ , we write the MpCN kernel on $M(p, q)$ (resp. $P^+(q)$) as P_ρ (resp. \tilde{P}_ρ) and the corresponding proposal as \tilde{Q}_ρ . In all cases, $\rho \in [0, 1)$.

Proposition 5.4. (i) If MpCN kernel P_0 on $M(p, q)$ has a spectral gap, then so does P_ρ for any $\rho \in [0, 1)$.

(ii) If MpCN kernel \tilde{P}_0 on $P^+(q)$ has a spectral gap, then so does \tilde{P}_ρ for any $\rho \in [0, 1)$.

Proof. The proof is given in Appendix C. □

Remark 4. See Appendix C for details on the concept of spectral gap. By Proposition 5.4, geometric ergodicity of the MpCN kernel with target Π on $M(p, q)$ (or $\tilde{\Pi}$ on $P^+(q)$) and parameter $\rho \in [0, 1)$ is implied by that of the MpCN kernel for $\rho = 0$. The result is important, as working with \tilde{P}_0 simplifies a lot the involved matrix calculations for deriving drift conditions. From a practical point of view, the result allows for numerical evidence over inequality (16) – as in Section 5.2 that follows – by taking advantage of the fact that the choice $\rho = 0$ provides a much more manageable expression for the distribution of the noise ϵ in (11), involved in the MpCN proposal (Proposition 5.5 below exploits this fact).

Proposition 5.5. If $\rho = 0$, the eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_q$ of $\epsilon \sim \mathcal{L}(d\epsilon)$ have the following joint Lebesgue density function, up to a normalising constant,

$$\prod_{i=1}^q \frac{\lambda_i^{(p-q-1)/2}}{(1 + \lambda_i)^p} \prod_{i < j} (\lambda_j - \lambda_i).$$

Proof. This follows from Theorem 5.3.1 of Farrell (1985) together with the analytical expression of $\mathcal{L}(d\epsilon)$, since, for any integrable f ,

$$\begin{aligned} \int f(\epsilon) \mathcal{L}(d\epsilon) &\propto \int f(\epsilon) \frac{\det(\epsilon)^{p/2}}{\det(I_q + \epsilon)^p} \mu(d\epsilon) \\ &= \int f(\epsilon) \frac{\det(\epsilon)^{(p-q-1)/2}}{\det(I_q + \epsilon)^p} \text{Leb}^{q(q+1)/2}(d\epsilon) \\ &= \int \prod_{i=1}^q \frac{\lambda_i^{(p-q-1)/2}}{(1 + \lambda_i)^p} \prod_{i < j} (\lambda_j - \lambda_i) d\lambda_1 \cdots d\lambda_q \int_{\mathcal{O}(q)} f(\mathbf{U}\Lambda\mathbf{U}^\top) d\mathbf{U}, \end{aligned}$$

where Λ is diagonal with elements $\lambda_1, \dots, \lambda_q$. □

5.2 Numerical Evidence over Drift Condition (16)

We consider the case of a Wishart target $\tilde{\Pi} = W_q(r, I_q)$ with $r \in \mathbb{R}$, $r > q - 1$ for $V(S) = \tilde{\pi}(S)^{-\alpha}$. In this case,

$$\frac{\tilde{P}_0 V(S) - V(S)}{V(S)} = \int_{P^+(q)} \{\eta(\epsilon, S)^{-\alpha} - 1\} \min\{1, \eta(\epsilon, S)\} \mathcal{L}(d\epsilon), \quad (17)$$

where,

$$\eta(\epsilon, S) = \frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} = (\det \epsilon)^{r/2} \exp\left(-\frac{1}{2} \text{tr}[(\epsilon - I_q)S]\right).$$

(Note that the right-hand side of (17) can be defined even if S is degenerate.) We want to numerically investigate (16). By continuity, it suffices to show that (17) is always negative if S is degenerate. Furthermore, we can assume that S is diagonal. To see this, first observe that for any $S \in P^+(q)$, there is $U \in \mathcal{O}(q)$ such that $U^\top S U$ is diagonal. We also have $\eta(U \epsilon U^\top, S) = \eta(\epsilon, U^\top S U)$, and the law of $U \epsilon U^\top$ is the same as that of ϵ when $\rho = 0$. Therefore, it is enough to show that (17) is always negative if S is a degenerate diagonal matrix. The law of the eigenvalues $\lambda_1, \dots, \lambda_q$ is determined in Proposition 5.5, and ϵ is decomposed as $\epsilon = U \Lambda U^\top$ where U is uniformly distributed in $\mathcal{O}(q)$ and Λ is a diagonal matrix with diagonal elements $\lambda_1, \dots, \lambda_q$. Thus, we can now evaluate (17) via numerical integration. To evaluate this integral, we used importance sampling with the Pareto distribution as a reference measure. In Fig. 1 we use 10^5 random samples for each given of S .

We fix $q = 2$, so we can assume that $S = \text{diag}(s, 0)$, $s > 0$. As we see in Fig. 1, left panel, for small enough $\alpha \in (0, 1)$, the value of (17) is (numerically found to be) negative. Thus, we can numerically confirm the drift condition for $\tilde{\Pi} = W_q(r, I_q)$ for $r = p = 2$ from Fig. 1, left panel, and for $r = 4$, $p = 5$ from Fig. 1 right (recall the law of ϵ depends on p). Other choices of r and p yield similar figures. A Dirichlet form argument can be used to show that geometric ergodicity for $\tilde{\Pi} = W_q(r, I_q)$ implies geometric ergodicity for $\tilde{\Pi} = W_q(r, T)$, for general $T \in P^+(q)$. Recall also – see the comment after Proposition 3.4 – that geometric ergodicity of the MpCN kernel for a Wishart target implies geometric ergodicity for an Inverse-Wishart target.

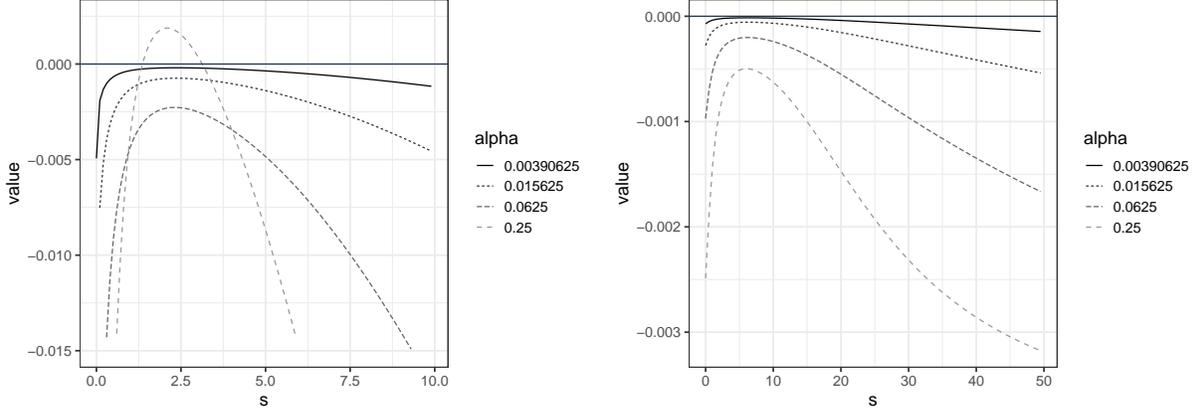


Figure 1: Numerical evaluation of the relative expected drift change $(\tilde{P}_0\mathbf{V}(S) - \mathbf{V}(S))/\mathbf{V}(S)$ for target $\tilde{\Pi} = W_q(r, I_q)$, at $S = \text{diag}(s, 0)$, for $r = p = 2$ (left) and $r = 4, p = 5$ (right).

6 Simulation Experiments

In this section, we discuss and analyse the difference in performance among the algorithms considered in this work, i.e. RWM, pCN and MpCN.

6.1 Tuning Parameters and Performance on Simple Targets

First, the choice of tuning parameters is discussed. Consider a target distribution defined on $M(p, q)$. Matrices U and V , appearing in RWM and pCN, are scaling parameters that can be learned from a initial phase of the MCMC algorithms. Upon recalling the second moment properties of the Matrix-Normal law in Remark 1(i), we can have the estimates,

$$\hat{U} = \hat{E}_1, \quad \hat{V} = \hat{E}_2 / \text{tr}(\hat{E}_2),$$

having defined,

$$\hat{E}_1 = (L - 1)^{-1} \sum_{l=1}^L (X_l - \bar{X})(X_l - \bar{X})^\top, \quad \hat{E}_2 = (L - 1)^{-1} \sum_{l=1}^L (X_l - \bar{X})^\top (X_l - \bar{X}),$$

where X_1, \dots, X_L , $L \geq 1$, are samples from the MCMC method and \bar{X} is the sample mean. In the case of RMW, one should introduce a scalar, $\sigma^2 > 0$, to allow for controlling the acceptance probability, so that the proposal writes as $y = x + \sigma \xi$, $\xi \sim N(0, \hat{U}, \hat{V})$ – this is not needed in the case of pCN and MpCN. Recall that MpCN does not involve V as above,

but simulates from an Inverse-Wishart law. As ρ decreases and σ^2 increases, the acceptance probability decreases. Empirically, a good choice of average acceptance probability for all algorithms is around 20% to 40%.

Assume that the target distribution is defined on $P^+(q)$, and that $\mathcal{U} = U$, in which case the MCMC algorithms give rise to Markov chains on $P^+(q)$ with dynamics that do not depend on U . Without loss of generality, let $U = I_p$. The dimension p is a tuning parameter. Let $S \in P^+(q)$ and $S = x_p^\top x_p$ for some $x_p \in M(p, q)$. For RWM, the proposed value from x_p is $y_p = x_p + w_p$, $w_p \sim N_{p,q}(0, I_p, \sigma^2 V)$. We set $S^* = y_p^\top y_p$. Then $\sigma^{-2}S^*$ follows the noncentral Wishart distribution with noncentral matrix $\sigma^{-2}S$, covariance V , and p degrees of freedom (see Section 1.5.4 of Chikuse, 2003). By the properties of non-central Wishart distribution, S^* has the same law as that of the matrix sum,

$$u^\top u + v^\top v = \begin{pmatrix} u \\ v \end{pmatrix}^\top \begin{pmatrix} u \\ v \end{pmatrix}, \quad u \sim N_{q,q}(S^{1/2}, I_q, \sigma^2 V), \quad v \sim N_{p-q,q}(0, I_{p-q}, \sigma^2 V).$$

The first term on the left side does not depend on p , while the second term is $\sum_{i=1}^{p-q} v_i v_i^\top$, where $v_i \sim N_q(0, V)$, and it is monotonically increasing with p . Thus S^* increases monotonically with p , an effect that is illustrated in Fig. 2, top-left panel. In Fig. 2, the x -axis and the y -axis are the logarithms of two out of q eigenvalues of the proposed position S^* , when the current position S is the identity. A similar calculation shows that the above observation is also true for pCN (Fig. 2, top-centre panel). In contrast, for MpCN the behavior of S for varying p is different. Recall that the matrix V is random in this case. By the law of large numbers, as $p \rightarrow \infty$,

$$\frac{V^{-1}}{p} \longrightarrow (x^\top x)^{-1} = S^{-1}, \quad (w^\top w) \longrightarrow S,$$

where V and w are as in (8). Also, $x^\top w \sim N_{q,q}(0, S, V)$ converges to 0 since $V \rightarrow 0$. One thus has $(y^\top y) \longrightarrow (x^\top x)$, and the acceptance probability converges to 1, where y is the proposed value of x . Therefore, the kernel degenerates as $p \rightarrow \infty$ (see Fig. 2, top-right panel). In general, in addition to the choice of σ and ρ (Fig. 2, bottom panel), the choice of p provides additional flexibility in the design of the MCMC kernels.

We now check the performance of the three algorithms for a Wishart target, $W_q(r, T)$, and an Inverse-Wishart one, $W_q^{-1}(r, T)$. We set $q = 8$, $r = 16$, and produce T via a sample

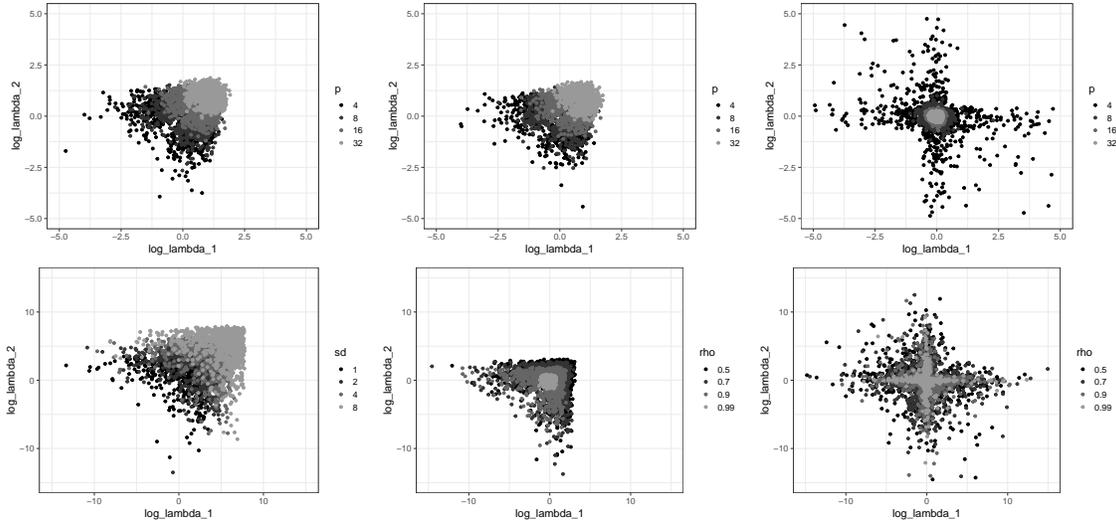


Figure 2: The effect of increasing p (top) and the effect of σ and ρ (bottom) for the proposed position S^* when the current position is $S = I_q$, for RWM (left), pCN (center) and MpCN (right). In all cases, $q = 4$, for RWM and pCN we have $V = I_q$, and 10^3 realisations of S^* are used for each figure.

$T \sim W_q(q, I_q)$. In the following experiment, we set $V = I_q$ for RWM, pCN, and set $p = q$. Parameters σ , ρ are chosen so that the acceptance probability is between 20% and 40%. We measure the distance between the empirical means of the MCMC algorithms and the mean of the target distribution via metric $d(\cdot, \cdot)$ defined in (14). The computational cost for each iteration of MpCN is two to three times that of RWM and pCN since MpCN uses a simulation of the inverse Wishart distribution, that requires the evaluation of an inverse matrix and the eigencomposition of a $q \times q$ matrix per iteration. Even taking this into account, MpCN performs much better for both the Wishart and Inverse-Wishart target (Fig. 3). For the Wishart target, pCN performs much worse than the other two methods, whereas pCN and RWM show similar weak performance for the Inverse-Wishart target. These findings align with the theory since pCN is not geometrically ergodic for the Wishart target, whereas both pCN and RWM are not geometrically ergodic for the Inverse-Wishart. For the heavy-tail scenario, the choice of an incremental distribution with heavy tails sometimes improves the performance of RWM (Jarner and Tweedie, 2003; Jarner and Roberts, 2007). However, attempting some choices in this direction did not improve the

results in our setting, thus numerics from such methods have not been included in the above plots.

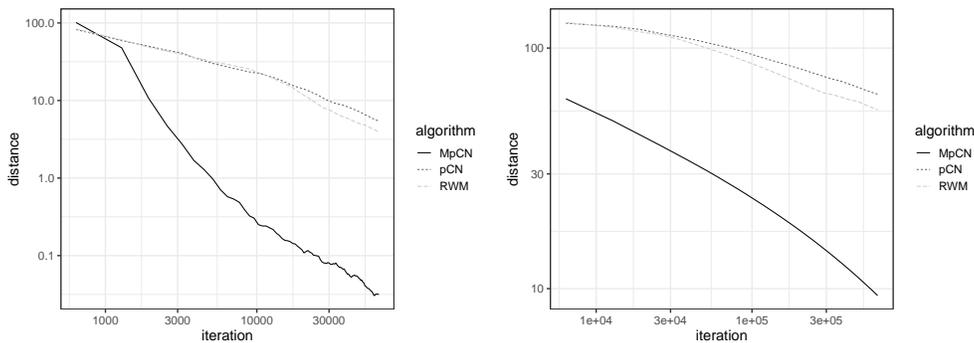


Figure 3: The distance between the empirical average and the true mean for a Wishart target (left) and an Inverse-Wishart target (right) in log – log scale.

6.2 SDE Models for Covariance Matrix

The prominent work of Barndorff-Nielsen and Shephard (2001) introduced the scalar, non-Gaussian, Ornstein–Uhlenbeck (OU) stochastic volatility (SV) class of models of the form,

$$d\sigma_t^2 = -\omega \sigma_t^2 dt + d\ell_t, \quad (18)$$

for a decay-rate parameter $\omega > 0$, where the driving noise $\{\ell_t\}$ is a Lévy process (Sato, 1999) of positive increments and no drift – such process is termed a ‘subordinator’ process. Under conditions, the differential dynamics give rise to a stationary OU process $\{\sigma_t^2\}$, with values in \mathbb{R}_+ . The analysis in Barndorff-Nielsen and Shephard (2001) illustrated that such class of models offers a great degree of flexibility in the specification of both the marginal distribution of σ_t^2 and various dynamical properties of the process, so that the model can match stylised empirical properties of observed time-series in financial economics. Parameter estimation in Barndorff-Nielsen and Shephard (2001) is carried out using method of moments. Later, Dellaportas et al. (2015) consider the important instance in this above class of models where $\{\ell_t\}$ is a compound Poisson process – in which case the marginal law of $\{\sigma_t^2\}$ is that of a Gamma distribution – and develop a complex, sophisticated MCMC algorithm for carrying out full Bayesian inference.

Barndorff-Nielsen and Shephard (2001) briefly discuss multivariate extensions of the proposed modelling framework. This direction is explored in detail in Barndorff-Nielsen and Stelzer (2007), where a non-Gaussian OU on $P^+(q)$ is carefully constructed via the differential equation,

$$d\Sigma_t = -(\Omega\Sigma_{t-} + \Sigma_{t-}\Omega^\top)dt + dL_t, \quad (19)$$

for $\Sigma_0 = \sigma_0 \in P^+(q)$, $\Omega \in M(q, q)$, and matrix subordinator Lévy process $\{L_t\}$, i.e., for $0 < s < t$, $L_t - L_s$ is positive semi-definite. Barndorff-Nielsen and Stelzer (2007) show that the conditions $\mathbb{E} \log^+ \|L_t\|_F < \infty$ and the spectrum of Ω being $\sigma(\Omega) \subset (0, \infty) + \mathbf{i} \mathbb{R}$, imply that SDE (19) has a stationary solution. The solution of (19) in-between jump times of the Lévy process, writes as,

$$\Sigma_t = \exp \{ -\Omega(t-s) \} \Sigma_s \exp \{ -\Omega^\top(t-s) \}.$$

An inferential objective arising within this important class of multivariate non-Gaussian SV models is the estimation of Ω and of parameters involved in the specification of $\{L_t\}$ based on observations related to Σ_t .

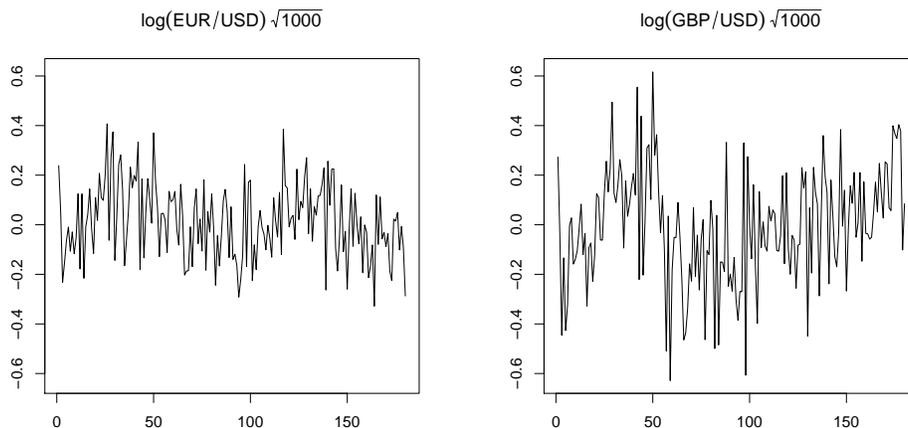


Figure 4: Observed linearly detrended $n = 180$ daily log-returns – rescaled by $\sqrt{1000}$ – for the period 23/06/20 to 01/03/21.

This design gives rise to a flexible, intuitive model, building upon mean-reverting-type dynamics imposed directly on the covariance matrix Σ_t . Such direction differs from typical

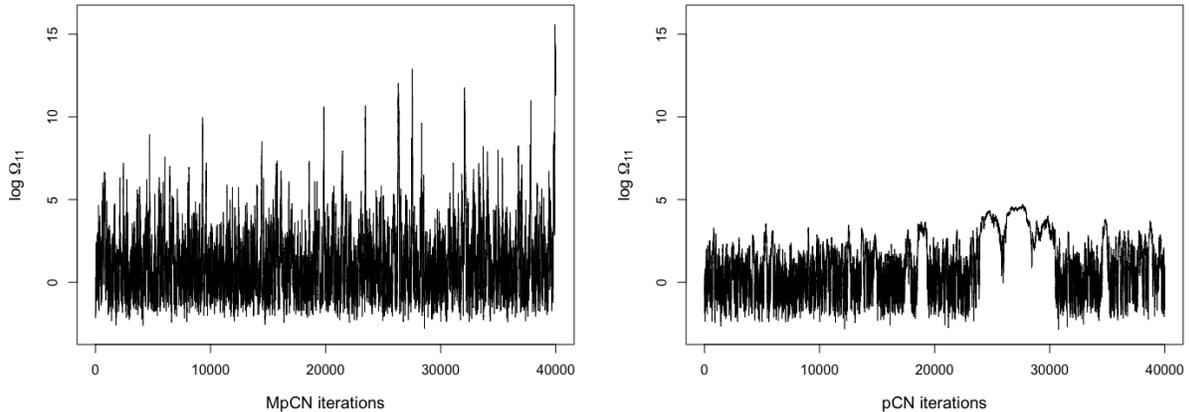


Figure 5: MpCN and pCN traceplots – in log-scale – for the prior of Ω_{11} .

approaches in the literature that work with various covariance decompositions (e.g. Dellaportas et al., 2015), thus lacking direct interpretation of the dynamical behaviour for Σ_t itself. To simplify the inferential setting, we consider a case where: i) the unknown parameter is Ω , and $\Omega \in P^+(q)$; ii) the Lévy process corresponds to a compound Poisson one, with iid jumps comprised of scalar, independent Exponential variables, with all involved parameters assumed known. Data $\{y_i\}$ are obtained at times $0 < t_1 < \dots < t_n$, $n \geq 1$, with $[y_i | \Sigma_{t_i}] \sim N(0, \Sigma_{t_i})$, $1 \leq i \leq n$.

We use data corresponding to $n = 180$ daily log-returns of the Euro-Dollar and Sterling-Dollar exchange rates, as observed in the period from 23/06/20 to 01/03/21. Thus, we look at a bivariate scenario, $q = 2$, with $t_i = i$. The data – after linearly detrending and rescaling by $\sqrt{1000}$ – are shown in Fig. 4. In contrast to the complex methodology used in Dellaportas et al. (2015) for performing Bayesian inference in the scalar case, we follow the direction of a pseudo-marginal algorithm (e.g. Andrieu and Roberts, 2009; Andrieu et al., 2010) to treat the latent paths of the covariance process $\{\Sigma_t\}$, thereby replacing the intractable likelihood with unbiased estimators obtained by a particle filter. For the driving compound Poisson process, we use intensity 0.4 and Exponential jumps with mean equal to 1/60. Such values were selected via preliminary runs of MCMC algorithms with these parameters assumed unknown. We fix $\sigma_0 = \text{diag}\{0.05, 0.05\}$, based again on preliminary runs. The prior for Ω is the inverse Wishart distribution $W_2^{-1}(2, I_2)$, we stress that pursuing joint estimation of

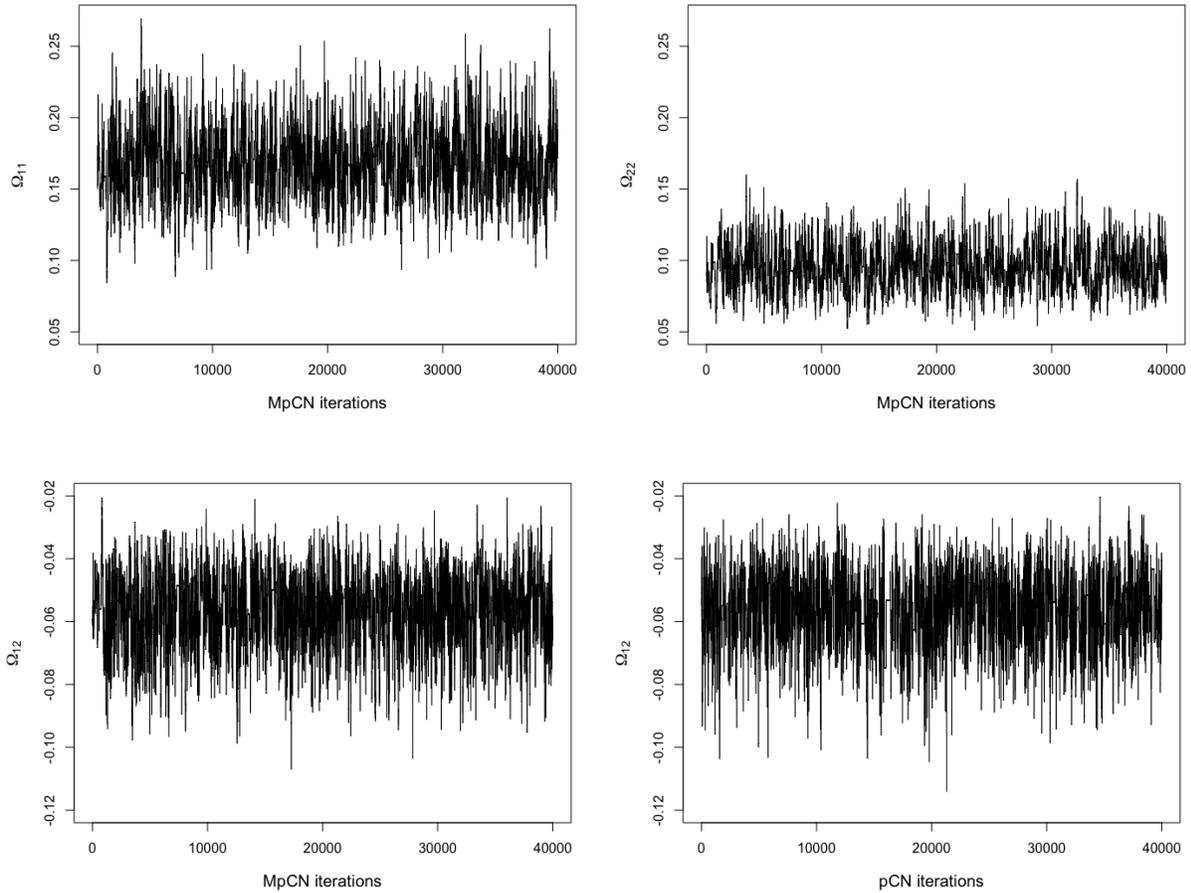


Figure 6: MpCN traceplots (first 3 plots) and pCN traceplots (last plot, bottom right) for the posteriors of Ω_{11} , Ω_{22} , Ω_{12} given the data in Fig. 4.

Ω with some of the above parameters in our bivariate setting is quite a challenging task, that is beyond the context of our work, and one that would deviate from our main interest of investigating the performance of algorithms we have defined on the previous parts of the paper and treat posterior sampling on $P^+(q)$.

We applied MpCN and pCN, with $p = 2$, both on the prior and the posterior distribution, on an Intel Xeon E5-2690, 2.9GHz, Memory 80GB (with C code). Our theory has shown that pCN is not geometrically ergodic when applied on the Inverse-Wishart prior, and this is manifested in the contrasting behaviour of the MpCN and pCN traceplots in Fig. 5 – for both algorithms we use $\rho = 0.9$ giving an average acceptance probability of 46%

and 48%, for MpCN and pCN respectively. When targeting the posterior, for the particle filter we use 1,000 particles and dynamic resampling with ESS threshold 25% (we made no use of parallelisation and applied the vanilla bootstrap filter) – see Andrieu et al. (2010) for details on the use of a particle filter as means of obtaining an unbiased estimate of the likelihood of data to be used within the pseudo-marginal MCMC. Fig. 6 shows MpCN and pCN traceplots generated using $\rho = 0.999$, giving an average acceptance probability of 22% and 14%, for MpCN and pCN respectively. Both algorithms required approximately 1 hour per 10,000 iterations. In this case, the consideration of $n = 180$ data points under a Gaussian likelihood has flattened the tails of the posterior, so MpCN and pCN have similar performance when applied on the posterior given all $n = 180$ observations. Note that the likelihood term of even a single data point in this example will change the mathematical nature of the decay of the tails of the target.

A key message we aim to convey is the *robustness* of MpCN in all settings considered in this application vs the poor performance of pCN when applied on the prior distribution.

7 Conclusions and Future Work

Our work presents one of the first contributions towards a systematic analysis (including derivation and ergodicity properties) of MCMC algorithms on matrix spaces. A number of interesting directions have now opened up to future research. We summarise some of them here.

- (i) As a starting point, we focused on blind proposals (this matched the requirements of the SV model). It is natural to move to the study of derivative-driven methods, e.g. MALA or HMC.
- (ii) There are some last steps remaining to obtain of complete proof of geometric ergodicity of the newly developed MpCN algorithm – this is left for future research, building upon the progress made here.
- (iii) The SV model sets up a new research direction in financial statistics. It leads to further investigations on differential models on matrix spaces and Monte Carlo meth-

ods that can be effective therein. This is a challenging task, requiring calibration of high-dimensional matrix-parameters for latent dynamical models (much beyond the 2×2 space, and 180 data points used here).

- (iv) There is a need for development and analysis of MCMC methods on spaces of symmetric positive-definite matrices restricted in sub-domains of $P^+(q)$ arising in the field of Gaussian graphical models. Here interest lies in exploring, e.g., the space of precision matrices, given zeros for a number of partial auto-correlations determined by graphs. For relevant references, see Section 3 of Lenkoski (2013) or Wang and Li (2012). Indicatively, application of the upcasting approach in this context will be important, as it will open up directions for the development of effective MCMC algorithms that respect the space restrictions. Scalability with respect to dimensionality q is also of high significance in this setting.

Acknowledgements

We thank an anonymous referee and the Associate Editor for suggestions that have greatly improved the contents of this paper.

A Proof of Proposition 4.6

First we prove (i). From Proposition 5.1 of Roberts and Tweedie (1996) and continuity of $x \mapsto P(x, \{x\})$, it suffices to show that $\sup_{x \in M(p,q)} P(x, \{x\}) = 1$. By definition,

$$1 - P(x, \{x\}) = \int_{M(p,q)} \alpha(x, y(w)) \phi_{p,q}(w; 0, U, V) dw, \quad (20)$$

for $y = y(w) = \rho^{1/2}x + (1 - \rho)^{1/2}w$; notice also that the acceptance probability writes as,

$$\alpha(x, y(w)) = \min \left\{ 1, \frac{\pi(y(w)) \phi_{p,q}(x; 0, U, V)}{\pi(x) \phi_{p,q}(y(w); 0, U, V)} \right\}.$$

Let $x = rs$, for $r > 0$ and $\|s\|_F = 1$, so that $\|x\|_F = r\|s\|_F = r$. For any $w \in M(p, q)$,

$$\begin{aligned} \left\| \frac{x}{r} - \frac{y(w)}{r} \right\|_F &= \|s - \rho^{1/2}s + r^{-1} (1 - \rho)^{1/2}w\|_F \\ &\xrightarrow{r \rightarrow \infty} 1 - \rho^{1/2}. \end{aligned}$$

For each $w \in M(p, q)$ choose $r_0 = r_0(w) > 0$ so that $\|\frac{x}{r} - \frac{y(w)}{r}\|_F < \varepsilon$ for any $r > r_0$. Then, the definition of $C_{r,\varepsilon}(s)$ in the statement of the Proposition implies that for $r = \|x\|_F > r_0$,

$$|\log \pi(y(w)) - \log \pi(x)| = |\log \pi(r\frac{y(w)}{r}) - \log \pi(r\frac{x}{r})| \leq C_{r,\varepsilon}(r^{-1}x) r^2. \quad (21)$$

We have that, for fixed w , $y(w) = \rho^{1/2}x + \mathcal{O}(1)$ and,

$$\begin{aligned} & -\log \phi_{p,q}(y(w); 0, U, V) + \log \phi_{p,q}(x; 0, U, V) \\ &= \frac{1}{2} \{ \operatorname{tr} [V^{-1}y(w)^\top U^{-1}y(w)] - \operatorname{tr} [V^{-1}x^\top U^{-1}x] \} \\ &= -\frac{1-\rho}{2} \operatorname{tr} [V^{-1}x^\top U^{-1}x] + \mathcal{O}(\|x\|_F) \\ &\leq -\frac{1-\rho}{2} \lambda^{-2} \|x\|_F^2 + \mathcal{O}(\|x\|_F). \end{aligned} \quad (22)$$

Choose a sequence $x_n \in M(p, q)$, for $n = 1, 2, \dots$, with $r_n = \|x_n\|_F \geq r_0$, such that,

$$\limsup_{n \rightarrow \infty} C_{r_n, \varepsilon}(s_n) < \frac{1-\rho}{2} \lambda^{-2},$$

where $s_n = x_n / \|x_n\|_F$. Then, for $y_n(w) = \rho^{1/2}x_n + (1 - \rho)^{1/2}w$, we have,

$$\begin{aligned} \alpha(x_n, y_n(w)) &= \min \left\{ 1, \frac{\pi(y_n(w)) \phi_{p,q}(x_n; 0, U, V)}{\pi(x_n) \phi_{p,q}(y_n(w); 0, U, V)} \right\} \\ &\leq \frac{\pi(y_n(w)) \phi_{p,q}(x_n; 0, U, V)}{\pi(x_n) \phi_{p,q}(y_n(w); 0, U, V)} \\ &= \exp(\log \pi(y_n(w)) - \log \pi(x_n)) \\ &\quad \times \exp(-\log \phi_{p,q}(y_n(w); 0, U, V) + \log \phi_{p,q}(x_n; 0, U, V)) \\ &\leq \exp\left(\left(C_{r_n, \varepsilon}(s_n) - \frac{1-\rho}{2} \lambda^{-2}\right) \|x_n\|_F^2 + \mathcal{O}(\|x_n\|_F)\right) \\ &\longrightarrow_{n \rightarrow \infty} 0, \end{aligned}$$

for any fixed $w \in M(p, q)$. The proof of (i) is completed via the dominated convergence theorem, as from (20) we obtain that $\sup P(x, \{x\}) = 1$.

The proof for (ii) is similar. Let $x = rs$, for $r > 0$ and $\|U^{-1/2}s\|_F = 1$ as above. Then,

for any $w \in M(p, q)$,

$$\begin{aligned}
\mathbf{d}(x^\top U^{-1}x, y(w)^\top U^{-1}y(w)) &= \mathbf{d}\left(\frac{x}{r}^\top \mathcal{U}^{-1} \frac{x}{r}, \frac{y(w)}{r}^\top \mathcal{U}^{-1} \frac{y(w)}{r}\right) \\
&= \mathbf{d}(s^\top \mathcal{U}^{-1}s, (s - \rho^{1/2}s + r^{-1}(1 - \rho)^{1/2}w)^\top \mathcal{U}^{-1}(s - \rho^{1/2}s + r^{-1}(1 - \rho)^{1/2}w)) \\
&\rightarrow \mathbf{d}(s^\top \mathcal{U}^{-1}s, (s - \rho^{1/2}s)^\top \mathcal{U}^{-1}(s - \rho^{1/2}s)) \\
&= \mathbf{d}(I_q, (1 - \rho^{1/2})^2 I_q) = q^{1/2} 2 \log(1 - \rho^{1/2}).
\end{aligned}$$

Let $\pi(x) = \tilde{\pi}(x^\top \mathcal{U}^{-1}x)$. For each $w \in M(p, q)$, there exists $z_0 = z_0(w) > 0$ such that, if $\|\mathcal{U}^{-1/2}x\|_F = z > z_0$ then,

$$|\log \pi(y(w)) - \log \pi(x)| \leq C_{z,\varepsilon} (z^{-2} x^\top \mathcal{U}^{-1}x) z^2.$$

Observe that the trace of $z^{-2}x^\top \mathcal{U}^{-1}x$ is 1. As in the proof of (i),

$$\begin{aligned}
-\log \phi_{p,q}(y(w); 0, U, V) + \log \phi_{p,q}(x; 0, U, V) &= -\frac{1-\rho}{2} \operatorname{tr}[V^{-1}x^\top U^{-1}x] + \mathcal{O}(\|\mathcal{U}^{-1/2}x\|_F) \\
&\leq -\frac{1-\rho}{2} \tilde{\lambda}^{-2} \|\mathcal{U}^{-1/2}x\|_F^2 + \mathcal{O}(\|\mathcal{U}^{-1/2}x\|_F).
\end{aligned}$$

The rest of the proof is the same as above.

B Proof of Proposition 5.2

Recall the concept of rapid variation in Definition 5.1. In the next lemma, we show a key property of probability measures on $P^+(q)$ with rapidly varying densities.

Lemma B.1. *Suppose that the target $\tilde{\Pi}(\mathrm{d}S)$ on $P^+(q)$ has a strictly positive, continuous, rapidly varying density $\tilde{\pi}(S)$ with respect to $\mu(\mathrm{d}S)$. Then, w.p.1,*

$$\lim_{\operatorname{tr}(S) \rightarrow +\infty} \left| \log \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right) \right| = +\infty,$$

where $\epsilon \sim \mathcal{L}(\mathrm{d}\epsilon)$ – see Proposition 3.2 for the definition of $\mathcal{L}(\mathrm{d}\epsilon)$.

Proof. By tightness of $\mathcal{L}(\mathrm{d}\epsilon)$, for any $\delta \in (0, 1)$, there exists $r \in (0, 1)$ such that,

$$\mathbb{P}[r \leq \lambda_{\min}(\epsilon) \leq \lambda_{\max}(\epsilon) \leq r^{-1}] > 1 - \delta,$$

where $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$ are the smallest and largest eigenvalues of $S \in P^+(q)$. Thus, for any $C > 0$, $\delta \in (0, 1)$, we have,

$$\mathbb{P} \left[\left| \log \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right) \right| \leq C \right] \leq \delta + \mathbb{P}[\epsilon \in A(S)],$$

where,

$$A(S) := \left\{ \epsilon \in P^+(q) : \left| \log \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right) \right| \leq C, r \leq \lambda_{\min}(\epsilon) \leq \lambda_{\max}(\epsilon) \leq r^{-1} \right\}.$$

We write $\text{Leb}^{q(q+1)/2}$ for $\prod_{1 \leq i \leq j \leq q} dS_{ij}$. Recall that reference measure $\mu(dS)$ has density $(\det(S))^{-(q+1)/2}$ with respect to $\text{Leb}^{q(q+1)/2}$. Thus, $\mathcal{L}(d\epsilon)$ has bounded density (with respect to $\text{Leb}^{q(q+1)/2}$) on,

$$\{\epsilon \in P^+(q) : r \leq \lambda_{\min}(\epsilon) \leq \lambda_{\max}(\epsilon) \leq r^{-1}\}.$$

So, there exists a constant $c > 0$ such that,

$$\mathbb{P}[\epsilon \in A(S)] \leq c \text{Leb}^{q(q+1)/2}[\epsilon \in A(S)].$$

Now, we consider the variable transformation $\epsilon = \mathbf{U}^\top \Lambda \mathbf{U}$ where $\mathbf{U} \in \mathcal{O}(q)$ and Λ is a diagonal matrix with positive diagonal elements $r \leq \lambda_1 < \lambda_2 < \dots < \lambda_q \leq r^{-1}$. From Theorem 5.3.1 of Farrell (1985), for some constant $c > 0$, we have,

$$\begin{aligned} \text{Leb}^{q(q+1)/2}[\epsilon \in A(S)] &= c \int_{\mathbf{U} \in \mathcal{O}(q)} d\mathbf{U} \int 1_{\{\mathbf{U}^\top \Lambda \mathbf{U} \in A(S)\}} \prod_{i < j} (\lambda_i - \lambda_j) d\lambda_1 \cdots d\lambda_q \\ &= c \int_{\mathbf{U} \in \mathcal{O}(q)} d\mathbf{U} \int 1_{\{\mathbf{U}^\top \Lambda(\lambda_1, \delta) \mathbf{U} \in A(S)\}} \prod_{i < j} (\delta_i - \delta_j) d\lambda_1 d\delta_2 \cdots d\delta_q \end{aligned}$$

where $\delta_1 = 0$, $\delta = \{\delta_i = \lambda_i - \lambda_1; i = 2, \dots, q\}$ and $\Lambda(\lambda_1, \delta)$ is diagonal matrix with values $\lambda_1, \lambda_1 + \delta_2, \dots, \lambda_1 + \delta_q$. Then,

$$\begin{aligned} \text{Leb}^{q(q+1)/2}[\epsilon \in A(S)] &\leq c r^{-q(q-1)/2} \int_{\mathbf{U} \in \mathcal{O}(q)} \int_{0 < \delta_2 < \dots < \delta_q < r^{-1}} \text{Leb}[A_r(S, \mathbf{U}, \delta)] d\mathbf{U} d\delta_2 \cdots d\delta_q, \end{aligned}$$

where,

$$A_r(S, \mathbf{U}, \delta) = \{r < \lambda_1 < r^{-1} - \delta_q : \mathbf{U}^\top \Lambda(\lambda_1, \delta) \mathbf{U} \in A(S)\} \subset \mathbb{R}.$$

The domain of the integral is always contained in a compact set $\mathcal{O}(q) \times [0, r^{-1}]^{d-1}$. Moreover, $A_r(S, \mathbf{U}, \delta)$ is a subset of an open interval (r, r^{-1}) . Thus, we can use the dominated

convergence theorem for $\text{tr}(S) \rightarrow \infty$ for the left-hand side of the above inequality. By the dominated convergence theorem, it suffices to show $\text{Leb}(A_r(S, \mathbf{U}, \delta)) \xrightarrow{\text{tr}(S) \uparrow +\infty} 0$ for each $r > 0$, $\mathbf{U} \in \mathcal{O}(q)$ and δ . Observe that,

$$\begin{aligned} & A_r(S, \mathbf{U}, \delta) \otimes A_r(S, \mathbf{U}, \delta) \\ & \subset \left\{ (\lambda_1, \lambda'_1) : r \leq \lambda_1, \lambda'_1 \leq r^{-1} - \delta_q, \left| \log \left(\frac{\tilde{\pi}((\mathbf{U}^\top \Lambda(\lambda_1, \delta) \mathbf{U}) \circ S)}{\tilde{\pi}((\mathbf{U}^\top \Lambda(\lambda'_1, \delta) \mathbf{U}) \circ S)} \right) \right| \leq 2C \right\}. \end{aligned}$$

Since $\pi(x) = \tilde{\pi}(x^\top \mathbf{U}^{-1} x)$ is a rapidly varying function, we have,

$$\lambda_1 \neq \lambda'_1 \implies \left| \log \left(\frac{\tilde{\pi}((\mathbf{U}^\top \Lambda(\lambda_1, \delta) \mathbf{U}) \circ S)}{\tilde{\pi}((\mathbf{U}^\top \Lambda(\lambda'_1, \delta) \mathbf{U}) \circ S)} \right) \right| \xrightarrow{\text{tr}(S) \uparrow +\infty} +\infty;$$

we used the fact that if $a > b$ then $\mathbf{U}^\top \Lambda(a, \delta) \mathbf{U} - \mathbf{U}^\top \Lambda(b, \delta) \mathbf{U} \in P^+(q)$. The probability of $\lambda_1 = \lambda'_1$ is 0, so we can complete the claim by the dominated convergence theorem. \square

Let \tilde{P}_ρ be the MpCN kernel on $P^+(q)$ – see the statement of Proposition 5.2.

Proof or Proposition 5.2. The MpCN proposal in $P^+(q)$ can be written as $\epsilon \circ S$, with $\epsilon \sim \mathcal{L}(\text{d}\epsilon)$, so we have, for $\mathbf{V}(S) = \tilde{\pi}(S)^{-\alpha}$, $\alpha \in (0, 1)$,

$$\begin{aligned} \frac{\tilde{P}_\rho \mathbf{V}(S) - \mathbf{V}(S)}{\mathbf{V}(S)} &= \mathbb{E} \left[\left\{ \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right)^{-\alpha} - 1 \right\} \min \left\{ 1, \frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right\} \right] \\ &= \mathbb{E} \left[\left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right)^{-\alpha} \min \left\{ 1, \frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right\} \right] - \mathbb{E} \left[\min \left\{ 1, \frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right\} \right]. \end{aligned}$$

By dominated convergence theorem and Lemma B.1, the first term in the right-hand side converges to 0 as $\text{tr}(S) \rightarrow \infty$. Also, by Lemma B.1, the limit of the second term is,

$$\liminf_{\text{tr}(S) \rightarrow +\infty} \mathbb{E} \left[\min \left\{ 1, \frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right\} \right] = \liminf_{\text{tr}(S) \rightarrow +\infty} \mathbb{P} \left[\log \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right) > 0 \right].$$

The proof will be completed if we show that the right-hand side is strictly greater than 0. Thanks to the rapidly varying property,

$$\begin{aligned} \liminf_{\text{tr}(S) \rightarrow +\infty} \mathbb{P} \left[\log \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right) > 0 \right] &\geq \liminf_{\text{tr}(S) \rightarrow +\infty} \mathbb{P} \left[\log \left(\frac{\tilde{\pi}(\epsilon \circ S)}{\tilde{\pi}(S)} \right) > 0, I - \epsilon \in P^+(q) \right] \\ &= \mathbb{P}[I - \epsilon \in P^+(q)]. \end{aligned}$$

Since the law of $\mathcal{L}(\text{d}\epsilon)$ is absolutely continuous with respect to μ , and $\mu[\{\epsilon; I - \epsilon \in P^+(q)\}]$ is positive, the probability is positive. Thus, the drift inequality for $\text{tr}(S) \rightarrow \infty$ follows. \square

C Proof of Proposition 5.4

Consider the standard Hilbert space $L^2(\Pi) = \{f : M(p, q) \rightarrow \mathbb{R}; \|f\|_{L^2(\Pi)}^2 < \infty\}$ where $\|f\|_{L^2(\Pi)} = \sqrt{\langle f, f \rangle}$ with inner product,

$$\langle f, g \rangle = \langle f, g \rangle_{L^2(\Pi)} = \int_{x \in M(p, q)} f(x)g(x)\Pi(dx).$$

A Markov kernel P on $M(p, q)$ is a linear operator on this space via the action,

$$(Pf)(x) = \int_{y \in E} P(x, dy)f(y), \quad f \in L^2(\Pi).$$

A linear operator P is self-adjoint if $\langle Pf, g \rangle = \langle f, Pg \rangle$, $f, g \in L^2(\Pi)$; this is equivalent to Π -reversibility of P . Also, a linear operator is positive if $\langle f, Pf \rangle \geq 0$, $f \in L^2(\Pi)$. By Section XI.8 of Yosida (1995), if P is self-adjoint, its spectrum lies on the real line; if it is also a positive operator, its spectrum lies on \mathbb{R}_+ . First we will show positivity of the pCN kernel. Using this, we will show positivity of the MpCN kernel.

Lemma C.1. *The pCN kernel $P(x, dy)$ on $M(p, q)$ is a positive, self-adjoint linear operator on $L^2(\Pi)$.*

Proof. First, we show that the proposal kernel Q of pCN is positive, self-adjoint on $L^2(\Pi')$, where,

$$Q(x, \cdot) = N_{p, q}(\rho^{1/2}x, (1 - \rho)U, V), \quad \Pi' = N_{p, q}(0, U, V).$$

The self-adjointness property is immediate upon observing,

$$\phi_{p, q}(y; \rho^{1/2}x, (1 - \rho)U, V) \phi_{p, q}(x; 0, U, V) = \phi_{p, q}(x; \rho^{1/2}y, (1 - \rho)U, V) \phi_{p, q}(y; 0, U, V).$$

By the reproductive property of the Gaussian distribution (Remark 1(i, ii)), we have

$$Q^{1/2}(x, \cdot) = N_{p, q}(\rho^{1/4}x, (1 - \rho^{1/2})U, V),$$

where the linear operator $Q^{1/2}$ is defined by $Q^{1/2}(Q^{1/2}f) = Qf$, $f \in L^2(\Pi')$. Observe that $Q^{1/2}$ is the same as Q with $\rho^{1/2}$ in the place of ρ . By this fact, $Q^{1/2}$ is also self-adjoint in $L^2(\Pi')$. Then,

$$\langle f, Qf \rangle_{L^2(\Pi')} = \langle f, Q^{1/2}Q^{1/2}f \rangle_{L^2(\Pi')} = \|Q^{1/2}f\|_{L^2(\Pi')}^2 \geq 0.$$

We now show that P itself is positive, self-adjoint on $L^2(\Pi)$. Since P is a Π -reversible Metropolis kernel, we proceed to the proof of positiveness. We use the decomposition approach introduced in Lemma 3.1 of Rudolf and Ullrich (2013). Recall that the acceptance probability is $\alpha(x, y) = \min\{1, \frac{\pi(y)}{\pi(x)}\}$, where $\pi(x)$ is the density of Π with respect to Π' . We can write,

$$\pi(x) \alpha(x, y) = \min\{\pi(x), \pi(y)\} = \int_0^\infty 1_{K(t)}(x) 1_{K(t)}(y) dt$$

where $K(t) = \{x \in M(p, q) : \pi(x) \geq t\}$. Using the above, we have,

$$\begin{aligned} \langle f, Pf \rangle_{L^2(\Pi)} &= \int_{x, y \in M(p, q)} f(x) f(y) \Pi(dx) P(x, dy) \\ &\geq \int_{x, y \in M(p, q)} f(x) f(y) \alpha(x, y) \Pi(dx) Q(x, dy) \\ &= \int_{x, y \in M(p, q)} f(x) f(y) \alpha(x, y) \pi(x) \Pi'(dx) Q(x, dy) \\ &= \int_0^\infty \int_{x, y \in M(p, q)} [1_{K(t)} f](x) [1_{K(t)} f](y) \Pi(dx) Q(x, dy) dt \\ &= \int_0^\infty \langle 1_{K(t)} f, Q[1_{K(t)} f] \rangle_{L^2(\Pi')} dt \geq 0. \end{aligned}$$

We used the fact that $1_{K(t)} f \in L^2(\Pi')$ by Markov's inequality. Thus, we have that P is a positive operator. \square

Next, we prove positivity and self-adjointness of the MpCN kernel. Such properties are inherited by the pCN kernel. Let,

$$Q_V(x, \cdot) = N_{p, q}(\rho^{1/2} x, U, (1 - \rho)V), \quad \Pi'_V = N_{p, q}(0, U, V), \quad (23)$$

be the proposal kernel of pCN and its invariant distribution, respectively, for matrix parameter V . Observe that

$$\nu_U(dx) W_q^{-1}(dV; p, x^\top U^{-1} x) = \Pi'_V(dx) \mu(dV).$$

As in the proof of Lemma 3.1 and equation (10), following the Bayesian paradigm construction of the MpCN kernel, we have,

$$\begin{aligned} \nu_U(dx) Q(x, dy) &= \int_{V \in P^+(q)} \nu_U(dx) W_q^{-1}(dV; p, x^\top U^{-1} x) Q_V(x, dy) \\ &= \int_{V \in P^+(q)} \Pi'_V(dx) Q_V(x, dy) \mu(dV), \end{aligned}$$

where Q is the MpCN kernel.

Proposition C.2. *The MpCN kernel $P(x, dy)$ is a positive, self-adjoint linear operator on $L^2(\Pi)$.*

Proof. Self-adjointness follows from reversibility. We will prove the positivity part. As in the pCN kernel case, for the target probability distribution $\Pi(dx) = \pi(x)\nu_U(dx)$ we have,

$$\begin{aligned} \langle f, Pf \rangle_{L^2(\Pi)} &= \int f(x)f(y)\Pi(dx)P(x, dy) \\ &\geq \int_{x,y \in M(p,q)} f(x)f(y)\alpha(x, y)\Pi(dx)Q(x, dy) \\ &= \int_0^\infty \int_{x,y \in M(p,q)} [1_{K(t)}f](x)[1_{K(t)}f](y)\nu_U(dx)Q(x, dy)dt. \end{aligned}$$

By (23), together with positivity and self-adjointness of the pCN proposal kernel, we obtain,

$$\begin{aligned} \langle f, Pf \rangle_{L^2(\Pi)} &\geq c_{p,q}^* \int_0^\infty \int_{V \in P^+(q)} \int_{x,y \in M(p,q)} [1_{K(t)}f](x)[1_{K(t)}f](y)\Pi'_V(dx)Q_V(x, dy)\mu(dV)dt \\ &= c_{p,q}^* \int_0^\infty \int_{V \in P^+(q)} \langle 1_{K(t)}f, Q_V[1_{K(t)}f] \rangle_{L^2(\Pi'_V)}\mu(dV)dt \geq 0. \end{aligned}$$

Here, we used the fact that $1_{K(t)}f \in L^2(\Pi'_V)$ by Markov's inequality. \square

Thus, all eigenvalues of the MpCN kernel are in \mathbb{R}_+ . Let $\text{spec}(P)$ be the eigenvalues of a self-adjoint Markov kernel P on $L^2(\Pi)$ – excluding the constant function. A transition kernel P is said to *have a spectral gap* if $1 - |\sup \text{spec}(P)| > 0$ – the latter being equivalent to $1 - \sup \text{spec}(P) > 0$ when P is a positive operator. It is known that, if P is positive and self-adjoint, it is geometrically ergodic if and only if it has a spectral gap (see, e.g., Roberts and Rosenthal, 1997; Roberts and Tweedie, 2001). Note that, by Theorem XI.8.2 of Yosida (1995), the spectral gap $1 - \sup \text{spec}(P)$ writes as,

$$\inf_{f \in L^2(\Pi); \Pi(f)=0} \frac{\mathcal{E}_P(f, f)}{\|f\|^2} = 1 - \sup_{f \in L^2(\Pi); \Pi(f)=0} \frac{\langle f, Pf \rangle}{\|f\|^2}, \quad (24)$$

where,

$$\mathcal{E}_P(f, f) = \frac{1}{2} \int_{x,y \in M(p,q)} (f(x) - f(y))^2 \Pi(dx)P(x, dy).$$

We study the spectral gap of the MpCN kernel. Recall from the main text, that to stress the involvement of parameter ρ , we write the MpCN kernel as P_ρ . We denote by Q_ρ the proposal kernel for P_ρ .

Proof of Proposition 5.4. Part (i):

First we note that,

$$(1 - \rho)x^\top U^{-1}x + (y - \rho^{1/2}x)^\top U^{-1}(y - \rho^{1/2}x) \leq 2x^\top U^{-1}x + 2y^\top U^{-1}y,$$

where the left-hand side is equal to $R(x, y)$ defined in Lemma 3.1; the above inequality follows from the right-hand side being equal to,

$$[(1 - \rho)x^\top U^{-1}x + (y - \rho^{1/2}x)^\top U^{-1}(y - \rho^{1/2}x)] + [(1 - \rho)x^\top U^{-1}x + (y + \rho^{1/2}x)^\top U^{-1}(y + \rho^{1/2}x)].$$

Let $\mathbf{q}_\rho(x, y)$ be the probability density function of $Q_\rho(x, \cdot)$ with respect to ν . Via the explicit form of $\mathbf{q}_\rho(x, y)$ (see Lemma 3.1), the above inequality implies,

$$\mathbf{q}_\rho(x, y) \geq 2^{-pq}\mathbf{q}_0(x, y).$$

Using this inequality,

$$\begin{aligned} \mathcal{E}_{P_\rho}(f, f) &= \frac{1}{2} \int_{x, y \in M(p, q)} (f(x) - f(y))^2 \Pi(dx) P_\rho(x, dy) \\ &= \frac{1}{2} \int_{x, y \in M(p, q)} (f(x) - f(y))^2 \Pi(dx) \alpha(x, y) Q_\rho(x, dy) \\ &\geq 2^{-pq} \frac{1}{2} \int_{x, y \in M(p, q)} (f(x) - f(y))^2 \Pi(dx) \alpha(x, y) Q_0(x, dy) \\ &\geq 2^{-pq} \mathcal{E}_{P_0}(f, f). \end{aligned}$$

Thus, existence of a spectral gap of P_0 implies that P_ρ also has one.

Part (ii):

We define the set,

$$L_\star^2(\Pi) := \{f : M(p, q) \rightarrow \mathbb{R} : f(x) = g(x^\top U^{-1}x) \text{ for a } g \in L^2(\tilde{\Pi})\} \subseteq L^2(\Pi).$$

Spaces $L_\star^2(\Pi)$, $L^2(\tilde{\Pi})$ are isomorphic via the mapping $f \leftrightarrow g$. For $f \in L_\star^2(\Pi)$ we have that $Pf(x) \in L_\star^2(\Pi)$, thus operator P restricted on $L_\star^2(\Pi)$ is positive, self-adjoint, with $Pf(x) \equiv \tilde{P}g(s)$, where $s = x^\top U^{-1}x$. It follows trivially, that we also have $\mathcal{E}_{P_\rho}(f, f) \geq 2^{-pq}\mathcal{E}_{P_0}(f, f)$ – as obtained in Part (i) – for $f \in L_\star^2(\Pi)$, or equivalently $\mathcal{E}_{\tilde{P}_\rho}(g, g) \geq 2^{-pq}\mathcal{E}_{\tilde{P}_0}(g, g)$, for $\mathcal{E}_{\tilde{P}_\rho}(g, g)$ defined in an obvious way. The proof is now complete. \square

References

- Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72(3), 269–342.
- Andrieu, C. and G. O. Roberts (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* 37(2), 697–725.
- Barnard, J., R. McCulloch, and X.-L. Meng (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica* 10(4), 1281–1311.
- Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2), 167–241.
- Barndorff-Nielsen, O. E. and R. Stelzer (2007). Positive-definite matrix processes of finite variation. *Probability and Mathematical Statistics-Wroclaw University* 27(1), 3.
- Beskos, A., G. Roberts, A. Stuart, and J. Voss (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.* 8(3), 319–350.
- Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular variation*, Volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.
- Chikuse, Y. (2003). *Statistics on special manifolds*, Volume 174 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- Cotter, S. L., G. O. Roberts, A. M. Stuart, and D. White (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.* 28(3), 424–446.
- Dellaportas, P., A. Plataniotis, and M. K. Titsias (2015). Scalable inference for a full multivariate stochastic volatility model. *arXiv preprint arXiv:1510.05257*.

- Dickey, J. M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *The Annals of Mathematical Statistics* 38(2), 511–518.
- Dobra, A., A. Lenkoski, and A. Rodriguez (2011). Bayesian inference for general gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association* 106(496), 1418–1433.
- Farrell, R. H. (1985). *Multivariate calculation*. Springer Series in Statistics. Springer-Verlag, New York.
- Geisser, S. and J. Cornfield (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. Ser. B* 25, 368–376.
- Halmos, P. R. (1950). *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y.
- Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer-Verlag, New York.
- Huang, A. and M. P. Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis* 8(2), 439–452.
- Jarner, S. and E. Hansen (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* 85(2), 341–361.
- Jarner, S. F. and G. O. Roberts (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.* 34(4), 781–815.
- Jarner, S. F. and R. L. Tweedie (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* 9(4), 559–578.
- Kamatani, K. (2017). Ergodicity of Markov chain Monte Carlo with reversible proposal. *Journal of Applied Probability* 54(2), 638–654.
- Kulik, A. and M. Scheutzow (2015). A coupling approach to doob's theorem. *Rendiconti Lincei - Matematica e Applicazioni* 26(1), 83–92.

- Lang, S. (1999). *Fundamentals of differential geometry*, Volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.
- Lenkoski, A. (2013). A direct sampler for G-Wishart variates. *Stat* 2(1), 119–128.
- Mengersen, K. L. and R. L. Tweedie (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* 24(1), 101–121.
- Meyn, S. P. and R. L. Tweedie (1993). *Markov Chains and Stochastic Stability*. Springer.
- Meyn, S. P. and R. L. Tweedie (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* 4(4), 981–1011.
- Neal, R. M. (1999). Regression and classification using Gaussian process priors. In *Bayesian statistics, 6 (Alcoceber, 1998)*, pp. 475–501. Oxford Univ. Press, New York.
- O’Malley, A. J. and A. M. Zaslavsky (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association* 103(484), 1405–1418.
- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer series in operations research and financial engineering. New York: Springer.
- Roberts, G. O. and J. S. Rosenthal (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.* 2, no. 2, 13–25 (electronic).
- Roberts, G. O. and R. L. Tweedie (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* 83(1), 95–110.
- Roberts, G. O. and R. L. Tweedie (2001). Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. *J. Appl. Probab.* 38A, 37–41. Probability, statistics and seismology.
- Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scandinavian Journal of Statistics* 29(3), 391–411.

- Rudolf, D. and B. Sprungk (2016, nov). On a generalization of the preconditioned crank–nicolson metropolis algorithm. *Foundations of Computational Mathematics* 18(2), 309–343.
- Rudolf, D. and M. Ullrich (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Probab.* 18, no. 49, 8.
- Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge university press.
- Stuart, A. M. (2010). Inverse problems: a Bayesian perspective. *Acta numerica* 19, 451–559.
- Wang, H. and S. Z. Li (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electronic Journal of Statistics* 6, 168–198.
- Yosida, K. (1995). *Functional analysis*. Classics in Mathematics. Berlin: Springer-Verlag. Reprint of the sixth (1980) edition.