Efficient simulation of high dimensional Gaussian vectors

Nabil Kahalé *

October 25, 2017

Abstract

We describe a Markov chain Monte Carlo method to approximately simulate a centered d-dimensional Gaussian vector X with given covariance matrix. The standard Monte Carlo method is based on the Cholesky decomposition, which takes cubic time and has quadratic storage cost in d. In contrast, the additional storage cost of our algorithm is linear in d. We give a bound on the quadractic Wasserstein distance between the distribution of our sample and the target distribution. Our method can be used to estimate the expectation of $h(X)$, where h is a real-valued function of d variables. Under certain conditions, we show that the mean square error of our method is inversely proportional to its running time. We also prove that, under suitable conditions, the total time needed by our method to obtain a given standardized mean square error is quadratic or nearly quadratic in d. A numerical example is given.

Keywords: Cholesky factorisation, Gaussian vectors, Markov chains, Monte Carlo simulation

1 Introduction

Monte Carlo simulation of Gaussian vectors is commonly used in a variety of fields such as weather and spatial prediction ((Gel, Raftery and Gneiting 2004) and (Diggle and Ribeiro 2007, Chap. 6)), finance (Hull 2012, Chap. 13), and machine learning (Russo and Van Roy 2014, Russo and Van Roy 2016). This paper considers the problem of efficiently sampling a d-dimensional Gaussian vector X with a given mean and a given $d \times d$ covariance matrix V. Since any Gaussian random variable is an affine function of a standard Gaussian random variable, we assume throughout the paper that the components of X are standard Gaussian random variables, and so the diagonal elements of V are 1. Then X can be simulated (Glasserman 2004, Subsection 2.3.3) as follows. Let Z be a d-dimensional vector of independent standard Gaussian random variables, and let A be a $d \times d$ matrix such that

$$AA^T = V. \tag{1.1}$$

Then $AZ \sim N(0, V)$, i.e. AZ is a d-dimensional Gaussian vector with covariance matrix V.

Such a matrix A can be computed in $O(d^3)$ time and $O(d^2)$ space using Cholesky factorization or one of its variants (Golub and Van Loan 2013, Subsections 4.2.5 and 4.2.8). Once A is calculated, AZ can be computed in $O(d^2)$ time. But, in several applications (see e.g. (Gel, Raftery and Gneiting 2004)), d is in the tens of thousands or more, and so the calculation of a Cholesky factorization on a standard computer may not be possible in practice, due to the high running time and/or storage cost. Alternative methods for generating Gaussian vectors have been developed in special cases. For instance, exact and efficient simulation of Gaussian processes on a regular grid in $\mathbb{R}^q$, $q \geq 1$, can be performed (Wood and Chan 1994, Dietrich

* *ESCP Europe, Labex Refi and Big Data Research Center, 75011 Paris, France; e-mail: nkahale@escpeurope.eu.*

and Newsam 1997) using Fast Fourier transforms if the covariance matrix is stationary with respect to translations. Similar Fast Fourier Transform methods can be used for exact simulation of fractional Brownian surfaces on a regular mesh (Stein 2002). Sparse Cholesky decomposition (Rue 2001) and iterative methods (Aune, Eidsvik and Pokern 2013) have been proposed to efficiently generate Gaussian vectors when the precision matrix V^{-1} is sparse.

This paper develops a new Markov Chain Monte Carlo method for approximate generation of a Gaussian vector X with correlation matrix V. Our method is straightforward to implement and can be applied to any correlation matrix V whose elements are known or easy to compute. This condition is verified in many practical applications, as covariance matrices are often specified through a functional form. Our method has a total storage cost of $O(d)$. At iteration n, it produces a d-dimensional vector X_n whose distribution converges (according to the quadratic Wasserstein distance) to $N(0, V)$ as n goes to infinity. Assuming each element of V can be computed in $O(1)$ time, the running time of each iteration is $O(d)$. Our method can for instance be used to approximately simulate spatial Gaussian processes of various types such as Matérn, powered exponential, and spherical on *any subset* of size d of $\mathbb{R}^2$ with $O(d)$ storage cost (background on spatial statistics can be found in (Diggle, Ribeiro Jr and Christensen 2003)). While FFT methods can simulate such processes on regular grids, certain applications (e.g. (Gel, Raftery and Gneiting 2004)) require the simulation of spatial Gaussian processes on non-regular subsets of $\mathbb{R}^2$.

We now describe our method. Let (i_n), $n \geq 0$, be a deterministic or a random sequence in $\{1, \ldots, d\}$, and let (g_n), $n \geq 0$, be a sequence of independent standard Gaussian random variables, independent of (i_n), $n \geq 0$. Define the Markov chain of d-dimensional column vectors X_n, $n \geq 0$, as follows. Let $X_0 = 0$ and, for $n \geq 0$, let

$$X_{n+1} = X_n + (g_n - e_n^T X_n)(V e_n), \tag{1.2}$$

where e_n is the d-dimensional column vector whose i_n-th coordinate is 1 and remaining coordinates are 0 (if $t \in \mathbb{R}$ and u is a vector, tu is the scalar product of t and u). Since Ve_n is the i_n-th column of V, the vector X_{n+1} can be calculated from X_n in $O(d)$ time. Note that X_n can be calculated iteratively with $O(d)$ total storage cost, since only the vector X_j needs to be stored in order to calculate X_{j+1}, for $0 \leq j \leq n-1$. The motivation behind (1.2) is explained in Section 2, and it can also be shown that (1.2) is a variant of the hit-and-run algorithm. A general description of the hit-and-run algorithm can be found in (Smith 1984).

Section 3 shows that, if i_n are independent random variables uniformly distributed over $\{1, \ldots, d\}$, then the quadratic Wasserstein distance between the distribution of X_n and $N(0, V)$ is at most $d/\sqrt{n}$. The quadratic Wasserstein distance between two probability distributions μ and μ' over $\mathbb{R}^d$ (Villani 2009, Definition 6.1) is equal to

$$\mathcal{W}_2(\mu, \mu') = (\inf_{Y \sim \mu, Y' \sim \mu'} \mathrm{E}(||Y - Y'||^2))^{1/2}. \tag{1.3}$$

To put this result into perspective, denote by μ_ϵ the distribution of $N(0, (1-\epsilon)V)$, for $0 \leq \epsilon \leq 1$. Then $\mathcal{W}_2(\mu_\epsilon, \mu_0) = (1 - \sqrt{1-\epsilon})\sqrt{d}$, by (Dowson and Landau 1982, Eq. 16). Thus, after $n = O(d\epsilon^{-2})$ steps, which can be performed in $O(d^2\epsilon^{-2})$ total time, the quadratic Wasserstein distance between the distribution of X_n and μ_0 is at most $\mathcal{W}_2(\mu_\epsilon, \mu_0)$. Section 4 shows that if h is a real-valued function on $\mathbb{R}^d$ satisfying certain conditions, and $i_0, \ldots, i_{n-1}$ are independent random variables uniformly distributed over $\{1, \ldots, d\}$, then $m = \mathrm{E}(h(X))$, where $X \sim N(0, V)$, is well approximated by $n^{-1} \sum_{j=0}^{n-1} h(X_j)$. More precisely, Theorem 4.1 gives explicit bounds on the mean square error

$$\mathrm{MSE}(n) = \mathrm{E}((\frac{\sum_{j=0}^{n-1} h(X_j)}{n} - m)^2)$$

of this estimator. For instance, if h is κ-Lipschitz, Theorem 4.1 implies that $n\mathrm{MSE}(n) \leq 18\kappa^2 d^2$. We give an example with $n = \Theta(d)$ where this bound is tight, up to a constant. To our

knowledge, for general V, no previous methods achieve a similar tradeoff between the running time and the Wasserstein distance, or between the running time and the mean square error, when $n = \Theta(d)$. Section 5 assumes that V is positive definite and shows that, under suitable conditions, $\mathrm{MSE}(n) \sim cn^{-1}$ as n goes to infinity, where c is a constant. It also gives an explicit geometric bound on the Wasserstein distance between the distribution of X_n and $N(0, V)$, and an explicit bound on the mean square error of a related estimator of m. Section 6 gives examples and numerical simulations, and shows that under certain conditions, the total time needed by our method to achieve a given standardized mean square error is $O^*(d^2)$. In particular, under certain conditions on h and if the smallest eigenvalue of V is bounded away from 0, in order to achieve a given standardized mean square error, our algorithm takes $O(d^2 \ln(d))$ time. Concluding remarks are given in a closing section.

An introduction to MCMC methods can be found in (Dellaportas and Roberts 2003). Our proof-techniques are based on coupling arguments. Conductance techniques can also be used to analyse mixing properties of Markov chains (see e.g. (Sinclair 1992, Kahale 1997b, Diaconis 2009)). Chernoff bounds for reversible discrete Markov chains in terms of the spectral gap have been established in (Kahale 1997a, Gillman 1998). Previous theoretical results on the performance of hit-and-run algorithms have focused on their mixing properties (see (Cousins and Vempala 2016, Bélisle, Romeijn and Smith 1993) and references therein). For instance, after appropriate preprocessing, the hit-and-run algorithm for sampling from a convex body (Lovász 1999) produces an approximately uniformly distributed sample point after $O^*(d^3)$ steps. Also, for general log-concave functions, after appropriate preprocessing (Lovász and Vempala 2006), the hit-and-run algorithm mixes in $O^*(d^3)$ steps. Note that the algorithms in (Lovász 1999, Lovász and Vempala 2006) require a pre-processing phase to make the target distribution "well-rounded". When the target distribution is $N(0, V)$, exact rounding can be achieved via a linear transformation induced by a matrix A satisfying (1.1). Rather than using such a matrix A, our Markov chain described in (1.2) performs a random walk along the columns of V. When V is positive definite, the Metropolis and Gibbs algorithms, and an algorithm for sampling from general log-concave functions using a Langevin stochastic differential equation (Durmus and Moulines 2016), could be used to approximately sample from $N(0, V)$, but these algorithms require the calculation of V^{-1}. However, standard algorithms for inverting a matrix take $\Theta(d^3)$ time and $\Theta(d^2)$ space, and so the pre-processing cost of these algorithms is as high as the Cholesky decomposition cost. Omitted proofs are in the appendix.

2 Motivation, notation, and general properties

If x is a d-dimensional vector, denote by $||x||$ the l_2-norm of x. If Z is a centered d-dimensional random vector such that $\mathrm{E}(||Z||^2)$ is finite, let $\mathrm{Cov}(Z) = \mathrm{E}(ZZ^T)$ denote the covariance matrix of Z. Recall that if Z and Z' are independent centered d-dimensional random vectors such that $\mathrm{E}(||Z||^2)$ and $\mathrm{E}(||Z'||^2)$ are finite, and A is a $d \times d$ matrix, then $\mathrm{Cov}(Z+Z') = \mathrm{Cov}(Z)+\mathrm{Cov}(Z')$, and $\mathrm{Cov}(AZ) = A\,\mathrm{Cov}(Z)A^T$.

To motivate (1.2), assume that i_0 is deterministic, let $X \sim N(0, V)$, and let g be a standard Gaussian random variable independent of X. Set

$$X' = X + (g - e_0^T X)(Ve_0). \tag{2.1}$$

Let I be the $d \times d$ identity matrix. As

$$X' = (I - Ve_0e_0^T)X + g(Ve_0),$$

X' is a centered Gaussian vector since it is the sum of two independent centered Gaussian vectors. Furthermore,

$$\mathrm{Cov}(X') = (I - Ve_0e_0^T)V(I - e_0e_0^TV) + (Ve_0)(Ve_0)^T,$$

which after some simplifications, shows that $\mathrm{Cov}(X') = V$. Hence $X' \sim N(0, V)$. (1.2) is obtained from (2.1) by replacing e_0, g, X and X' with e_n, g_n, X_n and X_{n+1}, respectively. Several methods (Glasserman 2004, Subsection 2.3.2) can be used to simulate g_n.

For any $d \times d$ matrix A, the matrix $A^T V A$ is positive semi-definite. Let

$$||A|| = ||A||_V = \sqrt{\mathrm{tr}(A^T V A)}$$

be the Frobenius norm of the matrix $\sqrt{V}A$. If A and B are symmetric $d \times d$ matrices, we say that $A \leq B$ if $B - A$ is positive semi-definite, and we denote by $\lambda_{\max}(A)$ the largest eigenvalue of A.

For $n \geq 0$, let $f_n = \sqrt{V} e_n$ and $P_n = I - f_n f_n^T$. Note that $||f_n||^2 = e_n^T V e_n = 1$. Thus f_n is a (deterministic or random) unit vector and P_n is a (deterministic or random) projection matrix, i.e. $P_n^2 = P_n$. Define the random sequence of d-dimensional vectors Y_n, $n \geq 0$, as follows: $Y_0 = 0$ and

$$Y_{n+1} = P_n Y_n + g_n f_n.$$

By rewriting (1.2) as

$$X_{n+1} = (I - V e_n e_n^T) X_n + g_n (V e_n),$$

it can be shown by induction that $X_n = \sqrt{V} Y_n$.

For $0 \leq j \leq n$, let $M_{j,n} = P_{n-1} P_{n-2} \cdots P_j$, with $M_{n,n} = I$, and let $M_n = M_{0,n}$. Let Z_0 be a d-dimensional vector of independent standard Gaussian random variables which is independent of the sequence (g_n, i_n), $n \geq 0$. For $n \geq 1$, let

$$Z_n = Y_n + M_n Z_0. \tag{2.2}$$

Since $\lambda_{\max}(A) \leq \mathrm{tr}(A)$ for a positive semi-definite matrix A, the following lemma implies that, if the sequence (i_k), $k \geq 0$, is deterministic, then X_n is a centered Gaussian vector, and

$$\lambda_{\max}(V - \mathrm{Cov}(X_n)) \leq ||M_n||^2.$$

As a consequence, any entry of $V - \mathrm{Cov}(X_n)$ is upper-bounded, in absolute value, by $||M_n||^2$.

Lemma 2.1. *If the sequence (i_k), $0 \leq k \leq n-1$, is deterministic, then, for $0 \leq j \leq n$, X_n and Y_n are centered Gaussian vectors, $Z_n \sim N(0, I)$, and*

$$\mathrm{E}(Z_n Z_j^T) = M_{j,n}. \tag{2.3}$$

Furthermore,

$$\mathrm{Cov}(X_n) = V - \sqrt{V} M_n M_n^T \sqrt{V}, \tag{2.4}$$

$\mathrm{Cov}(X_n) \leq V$, *and*

$$\mathrm{tr}(V - \mathrm{Cov}(X_n)) = \mathrm{E}(||X_n - \sqrt{V} Z_n||^2) = ||M_n||^2. \tag{2.5}$$

Lemma 2.1 forms the basis for the proofs of our main results. Indeed, if $M_{j,n}$ goes to 0 as $n - j$ goes to infinity then, by (2.4), $\mathrm{Cov}(X_n)$ converges to V as n goes to infinity. Furthermore, if both j and $n - j$ are sufficiently large, then by (2.3), Z_n and Z_j are nearly independent and, by (2.2), Y_j (resp. Y_n) is close to Z_j (resp. Z_n). Thus, Y_j and Y_n are nearly independent as well, and so are X_j and X_n. These arguments are informal since we have not defined the terms "nearly independent" and "close", but give intuition behind the proofs of Theorems 3.1 and 4.1.

Lemma 2.2 below generalizes some results of Lemma 2.1 when the sequence (i_k), $k \geq 0$, is random.

Lemma 2.2. *If the sequence (i_k), $k \geq 0$, is deterministic or random, the quadratic Wasserstein distance between the distribution of X_n and $N(0, V)$ is at most $\sqrt{\mathrm{E}(||M_n||^2)}$. Furthermore, $Z_n \sim N(0, I)$, X_n is centered, $\mathrm{Cov}(X_n) \leq V$, and*

$$\mathrm{tr}(V - \mathrm{Cov}(X_n)) = \mathrm{E}(||M_n||^2). \tag{2.6}$$

Proof. Since Z_0 and (g_j), $j \geq 0$, and (i_k), $k \geq 0$, are independent, conditioning on $i_0, \ldots, i_n$, the random variables Z_0 and g_j, $j \geq 0$, are independent standard Gaussian. Thus, by Lemma 2.1, $Z_n \sim N(0, I)$, conditioning on $i_0, \ldots, i_n$. Hence, Z_n is independent of $i_0, \ldots, i_n$. Thus, the unconditional distribution of Z_n is $N(0, I)$, and $\sqrt{V} Z_n \sim N(0, V)$. Furthermore, by (2.5),

$$\mathrm{E}(||X_n - \sqrt{V} Z_n||^2 | i_0, \ldots, i_n) = ||M_n||^2,$$

and so, by the tower law,

$$\mathrm{E}(||X_n - \sqrt{V} Z_n||^2) = \mathrm{E}(||M_n||^2).$$

By (1.3), it follows that the quadratic Wasserstein distance between the distribution of X_n and $N(0, V)$ is at most $\sqrt{\mathrm{E}(||M_n||^2)}$. Moreover, it follows from Lemma 2.1 that $\mathrm{E}(X_n | i_0, \ldots, i_n) = 0$. By the tower law, we infer that X_n is centered. Similarly, by Lemma 2.1,

$$\mathrm{E}(X_n X_n^T | i_0, \ldots, i_n) \leq V.$$

Hence, by the tower law, $\mathrm{E}(X_n X_n^T) \leq V$, and so $\mathrm{Cov}(X_n) \leq V$. Once again, (2.6) follows from (2.5) by the tower law. □

3 Upper bound on the Wasserstein distance

We first show the following lemma.

Lemma 3.1. *If P is a $d \times d$ projection matrix and A is a $d \times d$ matrix, then $||AP|| \leq ||A||$.*

Proof. Let $H = A^T V A$. Since $\mathrm{tr}(BC) = \mathrm{tr}(CB)$, $\mathrm{tr}(PHP) = \mathrm{tr}(HP) = \mathrm{tr}(PH)$, and so $\mathrm{tr}(H) - \mathrm{tr}(PHP) = \mathrm{tr}((I-P)H(I-P))$. Since H is positive semi-definite, so is $(I-P)H(I-P)$, and so $\mathrm{tr}(PHP) \leq \mathrm{tr}(H)$. Equivalently, $||AP||^2 \leq ||A||^2$. □

Under the conditions stated in Theorem 3.1 below, by an argument similar to that surrounding Lemma 2.1, it follows from (3.2) that each entry of the matrix $V - \mathrm{Cov}(X_n)$ is at most d^2/n in absolute value.

Theorem 3.1. *Assume that i_n, $n \geq 0$, are independent random variables uniformly distributed over $\{1, \ldots, d\}$. For $n \geq 1$, the quadratic Wasserstein distance between the distribution of X_n and $N(0, V)$ is at most $d/\sqrt{n}$,*

$$\sum_{j=0}^{n} \mathrm{E}(||M_j||^2) \leq d^2, \tag{3.1}$$

and the sequence $\mathrm{E}(||M_j||^2)$ is decreasing. Furthermore, for $n \geq 1$, X_n is centered, $\mathrm{Cov}(X_n) \leq V$ and

$$\mathrm{tr}(V - \mathrm{Cov}(X_n)) \leq \frac{d^2}{n}. \tag{3.2}$$

Proof. For any non-negative integer j, the matrix $e_j e_j^T$ has 1 on its (i_j, i_j) entry, and has 0 on its remaining entries. Hence $\mathrm{E}(e_j e_j^T) = d^{-1} I$. Thus,

$$\begin{aligned} \mathrm{E}(f_j f_j^T) &= \sqrt{V} \mathrm{E}(e_j e_j^T) \sqrt{V} \\ &= d^{-1} V, \end{aligned}$$

and so

$$\mathrm{E}(P_j) = I - d^{-1}V.$$

Let v be a unit d-dimensional vector. For $j \geq 0$, set $v_j = M_j v$. Since P_j is a projection and $v_{j+1} = P_j v_j$ for $j \geq 0$, it follows that

$$\begin{aligned} \mathrm{E}(||v_{j+1}||^2) &= \mathrm{E}(v_j^T P_j v_j) \\ &= \mathrm{E}(||v_j||^2) - d^{-1}\mathrm{E}(v_j^T V v_j). \end{aligned}$$

Hence

$$d^{-1}\mathrm{E}(v_j^T V v_j) = \mathrm{E}(||v_j||^2) - \mathrm{E}(||v_{j+1}||^2). \tag{3.3}$$

As $v_0 = v$, we conclude that

$$\sum_{j=0}^{n} \mathrm{E}(v_j^T V v_j) \leq d.$$

But

$$v_j^T V v_j = v^T M_j^T V M_j v, \tag{3.4}$$

and so, for any unit vector v,

$$v^T(\mathrm{E}(\sum_{j=0}^{n} M_j^T V M_j))v \leq d.$$

Thus, any diagonal entry of the matrix $\mathrm{E}(\sum_{j=0}^{n} M_j^T V M_j)$ is at most d. Hence,

$$\mathrm{tr}(\mathrm{E}(\sum_{j=0}^{n} M_j^T V M_j)) \leq d^2,$$

which implies (3.1). As $M_{j+1}^T = M_j^T P_j$, Lemma 3.1 shows that $||M_{j+1}^T|| \leq ||M_j^T||$. Hence the sequence $\mathrm{E}(||M_j^T||^2)$ is decreasing. Since M_j and M_j^T have the same distribution, $\mathrm{E}(||M_j^T||^2) = \mathrm{E}(||M_j||^2)$. Thus, the sequence $\mathrm{E}(||M_j||^2)$ is decreasing as well. By (3.1), $nE(||M_n||^2) \leq d^2$. We conclude the proof using Lemma 2.2. □

4 Bounding the mean square error

We now define the class of (κ, γ, W)-Lipschitz functions, with $\kappa > 0$ and $\gamma \in (0, 1]$.

Definition 4.1. *Let W be a $d \times d$ positive semi-definite matrix. We say that a real-valued Borel function h of d variables is (κ, γ, W)-Lipschitz if*

$$\mathrm{E}((h(X) - h(X'))^2) \leq \kappa^2(\mathrm{E}(||X - X'||^2))^\gamma \tag{4.1}$$

for any centered Gaussian column vector $\begin{pmatrix} X \\ X' \end{pmatrix}$ with $\mathrm{Cov}(X) \leq W$ and $\mathrm{Cov}(X') \leq W$, where X and X' are d-dimensional.

We say that a function h is (κ, W)-Lipschitz if it is $(\kappa, 1, W)$-Lipschitz. For instance, if h a real-valued κ-Lipschitz function on $\mathbb{R}^d$, i.e. $|h(x) - h(x')| \leq \kappa||x - x'||$ for x, x' in $\mathbb{R}^d$, then h is (κ, W)-Lipschitz for any $d \times d$ positive semi-definite matrix W. The following lemma gives an example of a (κ, W)-Lipschitz function in $\mathbb{R}$ which is not κ'-Lipschitz for any $\kappa' > 0$.

Lemma 4.1. *Let $f(z) = e^z$. Then f is $(e^\nu\sqrt{4\nu + 1}, \nu)$-Lipschitz for $\nu \geq 0$.*

Let h be a (κ, γ, V)-Lipschitz function. Set $m = \mathrm{E}(h(X))$ and $\Sigma^2 = \mathrm{Var}(h(X))$, where $X \sim N(0, V)$, and denote by $\hat{h}$ the real-valued function on $\mathbb{R}^d$ defined by $\hat{h}(x) = h(\sqrt{V}x) - m$. Note that $\mathrm{E}(\hat{h}(Z)) = 0$ if $Z \sim N(0, I)$, since $\sqrt{V}Z \sim N(0, V)$. In particular, by Lemma 2.2, $\mathrm{E}(\hat{h}(Z_j)) = 0$ for $j \geq 0$. In other words, $\mathrm{E}(h(\sqrt{V}Z_j)) = m$, and so the average of $h(\sqrt{V}Z_j)$, $b \leq j \leq n-1$, where b is a burn-in period, is an unbiased estimator of m. The variance of this estimator equals $(n-b)^{-2}\mathrm{E}((\sum_{j=b}^{n-1} \hat{h}(Z_j))^2)$, which we bound using Lemma 4.2 below. Choices for the parameters b and δ will be given in the sequel.

Lemma 4.2. *Let b, n, and δ be integers, with $0 \leq \delta \leq b < n$. Then*

$$\mathrm{E}((\sum_{j=b}^{n-1} \hat{h}(Z_j))^2) \leq 4(n-b)\delta\Sigma^2 + 4\kappa^2 \sum_{b \leq j,\, j+\delta \leq l \leq n-1} \mathrm{E}(||M_{j,l}||^{2\gamma} + ||M_{j,l}^T||^{2\gamma}).$$

For $j \geq 0$, let

$$\beta_j = \hat{h}(Z_j) - \hat{h}(Y_j), \text{ and } \beta = \sum_{j=b}^{n-1} \beta_j. \tag{4.2}$$

The second moments of β_j and of β can be bounded as follows.

Lemma 4.3. *Let b and n be integers, with $0 \leq b < n$. Then*

$$\mathrm{E}({\beta_j}^2) \leq \kappa^2 E(||M_j||^{2\gamma}),$$

and

$$\mathrm{E}(\beta^2) \leq (n-b)\kappa^2 \sum_{j=b}^{n-1} \mathrm{E}(||M_j||^{2\gamma}).$$

Proof. Assume first that the sequence $i_0, \ldots, i_{n-1}$, is deterministic. For $0 \leq j \leq n-1$,

$$\begin{aligned} \mathrm{E}({\beta_j}^2) &= \mathrm{E}(||h(\sqrt{V}Z_j) - h(X_j)||^2) \\ &\leq \kappa^2(\mathrm{E}(||\sqrt{V}Z_j - X_j||^2))^\gamma \\ &= \kappa^2||M_j||^{2\gamma}. \end{aligned}$$

The second equation follows from the relations $\mathrm{Cov}(\sqrt{V}Z_j) = V$ and $\mathrm{Cov}(X_j) \leq V$, and the last equation follows from (2.5). Thus, for any random sequence $i_0, \ldots, i_{n-1}$,

$$\mathrm{E}(\beta_j^2 | i_0, \ldots, i_j) \leq \kappa^2 ||M_j||^{2\gamma}.$$

The first inequality in the lemma then follows by taking expectations and using the tower law. The second inequality follows from the first one and the Cauchy-Schwartz inequality. □

Combining Lemmas 4.2 and 4.3 yields the following.

Lemma 4.4. *Let b, n, and δ be integers, with $0 \leq \delta \leq b < n$. If $i_0, \ldots, i_{n-1}$ are independent random variables uniformly distributed over $\{1, \ldots, d\}$, then*

$$\mathrm{E}((\frac{\sum_{j=b}^{n-1} h(X_j)}{n-b} - m)^2) \leq \frac{1}{n-b}(8\delta\Sigma^2 + 18\kappa^2 \sum_{j=\delta}^{n-1} \mathrm{E}(||M_j||^{2\gamma}).$$

Proof. Since $M_{j,l} \sim M_{l-j} \sim M_{l-j}^T$, for any fixed $j \geq 0$,

$$\sum_{l=j+\delta}^{n-1} \mathrm{E}(||M_{j,l}||^{2\gamma}) \leq \sum_{j=\delta}^{n-1} \mathrm{E}(||M_j||^{2\gamma}),$$

and

$$\sum_{l=j+\delta}^{n-1} \mathrm{E}(||M_{j,l}^T||^{2\gamma}) \leq \sum_{j=\delta}^{n-1} \mathrm{E}(||M_j||^{2\gamma}).$$

Hence, by Lemma 4.2,

$$\mathrm{E}((\sum_{j=b}^{n-1} \hat{h}(Z_j))^2) \leq 4(n-b)\delta\Sigma^2 + 8(n-b)\kappa^2 \sum_{j=\delta}^{n-1} \mathrm{E}(||M_j||^{2\gamma}).$$

As

$$(\sum_{j=b}^{n-1} \hat{h}(Y_j))^2 \leq 2\beta^2 + 2(\sum_{j=b}^{n-1} \hat{h}(Z_j))^2,$$

it follows by Lemma 4.3 that

$$\mathrm{E}((\sum_{j=b}^{n-1} \hat{h}(Y_j))^2) \leq 8(n-b)\delta\Sigma^2 + 18(n-b)\kappa^2 \sum_{j=\delta}^{n-1} \mathrm{E}(||M_j||^{2\gamma}).$$

Since $\hat{h}(Y_j) = h(X_j) - m$, this concludes the proof. □

By applying Lemma 4.4 with $b = \delta = 0$ and using (3.1), we get the following upper bound on MSE(n).

Theorem 4.1. *Let h be a (κ, V)-Lipschitz function on $\mathbb{R}^d$, with $m = \mathrm{E}(h(X))$, where $X \sim N(0, V)$. If $i_0, \ldots, i_{n-1}$ are independent random variables uniformly distributed over $\{1, \ldots, d\}$, then*

$$\mathrm{E}((\frac{\sum_{j=0}^{n-1} h(X_j)}{n} - m)^2) \leq 18\kappa^2 \frac{d^2}{n}. \tag{4.3}$$

4.1 Tightness of MSE bound

We now give an example where the bound on the mean square error in Theorem 4.1 is optimal, up to a multiplicative constant. Let $V = I$ and let $h(x) = ||x||$ for $x \in \mathbb{R}^d$. Thus h is a 1-Lipschitz function on $\mathbb{R}^d$. By (Forbes, Evans, Hastings and Peacock 2011, Sec. 11.3),

$$m = \frac{\sqrt{2}\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})},$$

which implies by induction that $m \geq \sqrt{d/2}$. Furthermore, it follows from (1.2) and by induction on n that X_n has at most n non-zero components, and that the non-zero components of X_n are independent standard Gaussian random variables. Thus $\mathrm{E}(||X_n||^2) \leq n$, and so $\mathrm{E}(||X_n||) \leq \sqrt{n}$ for $n \geq 0$. Thus,

$$\mathrm{E}(\sum_{j=0}^{n-1} ||X_j||) \leq n^{3/2}.$$

Hence, for $n = d/4$,

$$E(m - \frac{\sum_{j=0}^{n-1} h(X_j)}{n}) \geq \frac{\sqrt{2}-1}{2}\sqrt{d},$$

and so

$$\mathrm{E}((\frac{\sum_{j=0}^{n-1} h(X_j)}{n} - m)^2) \geq \frac{d}{25}.$$

Thus, the LHS of (4.3) is within an absolute constant from its RHS.

5 The positive definite case

Let h be a (κ, γ, V)-Lipschitz function. Define m, Σ and $\hat{h}$ as in Section 4. This section assumes that V is positive definite and that, i_n, $n \geq 0$, are independent random variables uniformly distributed over $\{1, \ldots, d\}$. Denote by λ the smallest eigenvalue of V, and set

$$\kappa' = \frac{2\kappa d^{1+\gamma}}{\lambda\gamma}.$$

The following lemma, combined with Lemma 2.2, implies a geometric bound on the Wasserstein distance between the distribution of X_n and $N(0, V)$ if V is positive semi-definite.

Lemma 5.1. *For $j \geq 0$, $\mathrm{E}(||M_j||^2) \leq d^2(1 - \lambda d^{-1})^j$.*

Proof. We use the same notation as in the proof of Theorem 3.1. Since the largest eigenvalue of V is at most $\mathrm{tr}(V) = d$, it follows from (3.4) that

$$\mathrm{E}(v^T M_j^T V M_j v) \leq dE(||v_j||^2).$$

On the other hand, (3.3) implies that

$$\mathrm{E}(||v_j||^2) - \mathrm{E}(||v_{j+1}||^2) \geq \lambda d^{-1}\mathrm{E}(||v_j||^2),$$

and so $\mathrm{E}(||v_j||^2) \leq (1 - \lambda d^{-1})^j$. Hence, for any unit-vector v,

$$v^T \mathrm{E}(M_j^T V M_j) v \leq d(1 - \lambda d^{-1})^j.$$

Thus each diagonal element of $\mathrm{E}(M_j^T V M_j)$ is at most $d(1 - \lambda d^{-1})^j$, and so $\mathrm{tr}(\mathrm{E}(M_j^T V M_j)) \leq d^2(1 - \lambda d^{-1})^j$. This completes the proof. □

As noted before, $n^{-1}\sum_{j=0}^{n-1} h(\sqrt{V}Z_j)$ is an unbiased estimator of m. Lemma 5.2 below implies that, if $c > 0$, the variance of this estimator is $\Theta(n^{-1})$ as n goes to infinity. The proof of Lemma 5.2 shows that the series in (5.1) is absolutely convergent.

Lemma 5.2. *As n goes to infinity, $n^{-1}\mathrm{E}((\sum_{j=0}^{n-1} \hat{h}(Z_j))^2)$ converges to a real number c, and*

$$c = \mathrm{Var}(h(\sqrt{V}Z_0)) + 2\sum_{j=1}^{\infty} \mathrm{Cov}(h(\sqrt{V}Z_0), h(\sqrt{V}Z_j)). \tag{5.1}$$

Theorem 5.1 below implies that if $c > 0$, $\mathrm{MSE}(n) \sim cn^{-1}$ as n goes to infinity.

Theorem 5.1. *As n goes to infinity, $n\mathrm{E}((\frac{\sum_{j=0}^{n-1} h(X_j)}{n} - m)^2)$ converges to c .*

Proof. Define β_j and β via 4.2, with $b = 0$. Let $\theta = (1 - \lambda d^{-1})^\gamma$. By Lemma 4.3,

$$\begin{aligned} \mathrm{E}(\beta_j^2) &\leq \kappa^2 \mathrm{E}(||M_j||^{2\gamma}) \\ &\leq \kappa^2 \mathrm{E}(||M_j||^2)^\gamma \\ &\leq \kappa^2 d^{2\gamma}\theta^j. \end{aligned}$$

The second inequality follows from Jensen's inequality and the last one from Lemma 5.1. Hence, by the Cauchy-Schwartz inequality,

$$\begin{aligned} \mathrm{E}(\beta^2) &\leq (\sum_{j=0}^{n-1} \theta^{j/2})\mathrm{E}(\sum_{j=0}^{n-1} \theta^{-j/2}\beta_j^2) \\ &\leq \frac{\kappa^2 d^{2\gamma}}{(1 - \theta^{1/2})^2} \\ &\leq \kappa'^2. \end{aligned}$$

The last inequality follows from the relation $\theta^{1/2} \leq 1 - \lambda\gamma d^{-1}/2$ (which is a consequence of Taylor's formula with Lagrange remainder). Moreover,

$$\begin{aligned} \mathrm{E}((\sum_{j=0}^{n-1} \hat{h}(Y_j))^2) &= \mathrm{E}((-\beta + \sum_{j=0}^{n-1} \hat{h}(Z_j))^2) \\ &= \mathrm{E}(\beta^2) - 2\mathrm{E}(\beta \sum_{j=0}^{n-1} \hat{h}(Z_j)) + \mathrm{E}((\sum_{j=0}^{n-1} \hat{h}(Z_j))^2). \end{aligned}$$

But, by the Cauchy-Schwartz inequality and Lemma 5.2, for sufficiently large n,

$$\begin{aligned} |\mathrm{E}(\beta(\sum_{j=0}^{n-1} \hat{h}(Z_j)))| &\leq \kappa' \sqrt{\mathrm{E}\big((\sum_{j=0}^{n-1} \hat{h}(Z_j))^2\big)} \\ &\leq \kappa' \sqrt{(c+1)n}. \end{aligned}$$

Using Lemma 5.2 once again, it follows that $n^{-1}\mathrm{E}((\sum_{j=0}^{n-1} \hat{h}(Y_j))^2)$ converges to c as n goes to infinity. □

We now show that the estimator $(2/n)\sum_{j=n/2}^{n-1} h(X_j)$ of m has an exponentially decreasing bias. We also give a bound on the mean square error of this estimator which, in certain cases, is smaller than the RHS of (4.3).

Theorem 5.2. *Set* $\delta = \lceil 4(\lambda\gamma)^{-1} d \ln(\kappa d/\Sigma) \rceil$. *For* $d \geq 3$ *and even* $n > 0$,

$$|\mathrm{E}(\frac{\sum_{j=n/2}^{n-1} h(X_j)}{n/2} - m)| \leq 2\kappa' \frac{e^{-\lambda\gamma n/(4d)}}{n}. \tag{5.2}$$

Furthermore, if $n > 2\delta$,

$$\mathrm{E}((\frac{\sum_{j=n/2}^{n-1} h(X_j)}{n/2} - m)^2) \leq 34 \frac{\delta\Sigma^2}{n}. \tag{5.3}$$

Proof. Set $b = n/2$ and define β via (4.2). Using calculations similar to the proof of Theorem 5.1, it follows that

$$\mathrm{E}(\beta^2) \leq \kappa'^2 (1 - \lambda d^{-1})^{\gamma n/2}.$$

Since $1 + x \leq e^x$ for $x \in \mathbb{R}$, it follows that

$$|\mathrm{E}(\beta)| \leq \kappa' e^{-\lambda\gamma n/(4d)}.$$

This implies (5.2) since the $\hat{h}(Z_j)$'s are centered.

We now prove (5.3). We first note that by applying (4.1) with $X \sim N(0, V)$, $X' \sim N(0, V)$, X and X' independent, if follows after some calculations that

$$\Sigma^2 \leq \kappa^2 d, \tag{5.4}$$

and so

$$\delta \geq 2(\lambda\gamma)^{-1} d. \tag{5.5}$$

Moreover, by Lemma (5.1) and Jensen's inequality,

$$\begin{aligned} \mathrm{E}(||M_j||^{2\gamma}) &\leq d^{2\gamma}(1 - \lambda d^{-1})^{\gamma j} \\ &\leq d^{2\gamma}(1 - \lambda\gamma d^{-1})^{j}. \end{aligned}$$

The second equation follows from the inequality $(1-\lambda d^{-1})^{\gamma} \leq 1-\lambda\gamma d^{-1}$. Thus,

$$\begin{aligned}\sum_{j=\delta}^{n-1} \mathrm{E}(||M_j||^{2\gamma}) &\leq \frac{d^{2\gamma+1}}{\lambda\gamma}(1-\lambda\gamma d^{-1})^{\delta} \\ &\leq \frac{d^3}{\lambda\gamma}\exp(-\lambda\gamma\delta d^{-1}) \\ &\leq \frac{d\Sigma^2}{\lambda\gamma\kappa^2}.\end{aligned}$$

Thus, by applying Lemma 4.4, it follows that

$$\mathrm{E}((\frac{\sum_{j=n/2}^{n-1} h(X_j)}{n/2} - m)^2) \leq \frac{2}{n}(8\delta\Sigma^2 + 18\frac{d\Sigma^2}{\lambda\gamma}).$$

By (5.5), this implies (5.3). □

6 Examples

Let h be a real-valued Borel function of d variables that can be calculated at any point in $O(d)$ time. Assume that V is positive definite, and that both $m = \mathrm{E}(h(X))$ and $\Sigma^2 = \mathrm{Var}(h(X))$ exist and are finite, where $X \sim N(0, V)$. Denote by MCMC the algorithm that generates $X_0, \ldots, X_{n-1}$ via (1.2), where the i_j's are independent and identically distributed over $\{1, \ldots, d\}$, and estimates m via

$$h_{n,b} = \frac{\sum_{j=b}^{n-1} h(X_j)}{n-b},$$

where b is a burn-in period. The standard Monte Carlo algorithm, referred to thereafterr, as MC, first calculates a lower-triangular matrix A satisfying (1.1) in $\Theta(d^3)$ time via the procedure described in (Glasserman 2004, Subsection 2.3.3). The MC algorithm then generates n' independent d-dimensional vectors of independent standard Gaussian random variables $Z_1, \ldots, Z_{n'}$, and estimates m by taking the average of $h(AZ_j)$, $1 \leq j \leq n'$. The variance of this estimator is $V_{\mathrm{MC}}(n') = \Sigma^2/n'$.

6.1 Comparison of the MC and MCMC methods

The mean square error of the $h_{n,b}$ estimator of m is defined as

$$\mathrm{MSE}(n; b) = \mathrm{E}((h_{n,b} - m)^2).$$

Given $\epsilon \in (0, \Sigma)$, $n' = \Sigma^2/\epsilon^2$ samples of the MC algorithm are needed to ensure that $V_{\mathrm{MC}}(n') = \epsilon^2$ (ignoring rounding issues). Calculating the Cholesky decomposition and $h(AZ_j)$, $1 \leq j \leq n'$, takes

$$\tau_{\mathrm{MC}}(\epsilon) = \Theta(d^3 + \frac{\Sigma^2}{\epsilon^2}d^2) \tag{6.1}$$

time. On the other hand, for $\epsilon > 0$, if h is (κ, γ, V)-Lipschitz and $\xi \in \{0, 1/2\}$, denote by $\tau_{\mathrm{MCMC}}(\epsilon, \xi)$ the running time of the MCMC algorithm needed to ensure that $\mathrm{MSE}(n; b) \leq \epsilon^2$ using burn-in period $b = \xi n$. If $\gamma = 1$, by Theorem 4.1, after $n = \lceil 18\kappa^2 d^2/\epsilon^2 \rceil$ steps of the MCMC algorithm, $\mathrm{MSE}(n) \leq \epsilon^2$. By (5.4), it follows that, for $\epsilon < \Sigma$,

$$\tau_{\mathrm{MCMC}}(\epsilon, 0) = O(\kappa^2 \frac{d^3}{\epsilon^2}).$$

Thus, if there is a constant $\phi \geq 1$ independent of d such that $\kappa^2 d \leq \phi \Sigma^2$ (this is equivalent to saying that (5.4) is tight, up to a constant), then, for fixed $\epsilon/\Sigma < 1$,

$$\tau_{\text{MCMC}}(\epsilon, 0) = O(d^2). \tag{6.2}$$

Similarly, under the assumptions of Theorem 5.2, for $\epsilon < \Sigma$,

$$\tau_{\text{MCMC}}(\epsilon, 1/2) = O(\frac{\ln(\kappa d/\Sigma) d^2}{\lambda \gamma} \frac{\Sigma^2}{\epsilon^2}).$$

Hence, if there are positive constants ϕ and ϕ' independent of d such that $\kappa \leq d^{\phi}\Sigma$ and $\lambda\gamma \geq \phi'$, then, for fixed $\epsilon/\Sigma < 1$,

$$\tau_{\text{MCMC}}(\epsilon, 1/2) = O(d^2 \ln(d)). \tag{6.3}$$

Examples where (6.2) or (6.3) holds are given below. Other examples requiring the simulation of large Gaussian vectors can be found in (Diggle and Ribeiro 2007, Chap. 6).

6.2 A Basket option

Consider a set of d stocks $S_1, \ldots, S_d$. For $t \geq 0$, denote by $S_i(t)$ the price of S_i at time t. Assume that $S_1(0) = S_2(0) = \cdots = S_d(0) = 1$. A Basket call option with maturity T and strike K is a financial derivative that pays the amount $((S_1(T) + \cdots + S_d(T))/d - K)^+$ at time T. Under a standard pricing model (Glasserman 2004, Subsection 3.2.3), the price of a basket option is $\mathrm{E}(h(U))$, where U is a centered Gaussian vector with covariance matrix V given by $V_{ij} = \text{Correl}(\ln(S_i(T)), \ln(S_j(T)))$ for $1 \leq i \leq j \leq d$, and for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$h(x) = (d^{-1} \sum_{i=1}^{d} \exp(-\frac{\sigma_i^2}{2} T + \sigma_i \sqrt{T} x_i) - K e^{-rT})^+,$$

where r is the risk-free rate, and σ_i is the volatility of S_i. Assume that the σ_i's are bounded by a constant independent of d. It follows from Lemma 6.1 below that h is (κ, V)-Lipschitz, where $\kappa = O(d^{-1/2})$ as d goes to infinity.

Lemma 6.1. *Let $g(x_1, \ldots, x_d) = \max(\sum_{i=1}^d w_i e^{\sigma_i x_i} - K, 0)$, where $w_i \geq 0$ for $1 \leq i \leq d$. Then g is (κ, V)-Lipschitz, where $\kappa = \sqrt{\sum_{i=1}^d w_i^2 e^{2\sigma_i^2}(4\sigma_i^2 + 1)}$.*

Thus, by Theorem 4.1, $n = O(d/\epsilon^2)$ steps of the MCMC algorithm are sufficient to ensure that $\text{MSE}(n) \leq \epsilon^2$, and so $\tau_{\text{MCMC}}(\epsilon, 0) = O(d^2/\epsilon^2)$. Furthermore, if $\Sigma = \Theta(1)$ (which is the case (Hull 2012, Sec. 25.14) if the volatilities and correlations are lower-bounded by a constant and $K = 0$, for instance), then (6.2) holds and $\tau_{\text{MC}}(\epsilon) = \Theta(d^3 + d^2/\epsilon^2)$. In practice, though, d is quite small.

6.3 The multivariate normal function

Let $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$, and $\hat{a} = \min_{1 \leq i \leq d} |a_i|$. Set $h(x) = 1_{x \leq a}$ for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, where $x \leq a$ if and only if $x_i \leq a_i$ for $1 \leq i \leq d$. The following lemma and the analysis in Subsection 6.1 show that for $0 < \epsilon < \Sigma$,

$$\tau_{\text{MCMC}}(\epsilon, 1/2) = O(\frac{\ln(d^4/(\hat{a}\Sigma^3)) d^2}{\lambda} \frac{\Sigma^2}{\epsilon^2}).$$

Thus, if there are positive constants ϕ and ϕ' independent of d such that $\hat{a} \geq d^{-\phi}$, $\Sigma \geq d^{-\phi}$, and $\lambda \geq \phi'$, then (6.3) holds for fixed $\epsilon/\Sigma < 1$.

Lemma 6.2. *The function h is $(3(d/\hat{a})^{1/3}, 1/3, W)$-Lipschitz for any $d \times d$ positive semi-definite matrix W.*

Proof. Let $\nu > 0$. It is easy to see that if $||X - Y|| < \nu$ and $|X_i| \notin [|a_i|, |a_i| + \nu]$ and $|Y_i| \notin [|a_i|, |a_i| + \nu]$ for $1 \leq i \leq d$, then $h(X) = h(Y)$. Hence

$$|h(X) - h(Y)| \leq 1_{\nu \leq ||X-Y||} + \sum_{i=1}^{d} (1_{|X_i| \in [|a_i|, |a_i|+\nu]} + 1_{|Y_i| \in [|a_i|, |a_i|+\nu]}).$$

By Chebyshev's inequality, $\Pr(\nu \leq ||X - Y||) \leq \nu^{-2} E(||X - Y||^2)$. A simple calculation shows that for $z > 0$, the density of any centered Gaussian random variable at z is at most $1/z$. Hence,

$$\Pr(|X_i| \in [|a_i|, |a_i| + \nu]) \leq 2\nu/\hat{a},$$

and a similar relation holds for Y_i. Thus,

$$E(|h(X) - h(Y)|) \leq \nu^{-2} E(||X - Y||^2) + 4d\nu/\hat{a}.$$

Minimizing over ν implies that

$$E(|h(X) - h(Y)|) \leq 9(d/\hat{a})^{2/3} (E(||X - Y||^2))^{1/3}.$$

□

6.4 The maximum function

For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, let $h(x) = \max_{1 \leq i \leq d} x_i$. Then h is 1-Lipschitz. Let $X \sim N(0, V)$, where V is a correlation matrix. Standard calculations show that $\Pr(h(X) > z) \leq de^{-z^2/2}$ for $z > 0$. Thus, it follows after some calculations that $\Pr(h(X)^2 > z \ln(d)) \leq e^{-z/4}$ for $z \geq 2$ and $d \geq 3$, and so $\mathrm{E}(h(X)^2) \leq 6 \ln d$. Furthermore, since

$$\Pr(X_1 \in [4\sqrt{\ln d}, 5\sqrt{\ln d}]) \geq \frac{d^{-15}}{\sqrt{2\pi}},$$

where X_1 is the first coordinate of X,

$$\Pr(h(X) \geq 4\sqrt{\ln d}) \geq \frac{d^{-15}}{\sqrt{2\pi}}.$$

Since $E(h(X)) \leq 3\sqrt{\ln d}$, we conclude that $\Sigma^2 \geq d^{-15}/\sqrt{2\pi}$. Hence (6.3) holds for fixed $\epsilon/\Sigma < 1$ if there is a positive constant ϕ' independent of d such that $\lambda \geq \phi'$.

6.5 A numerical example

In (Gel, Raftery and Gneiting 2004), the temperatures $Y(s_1), \cdots, Y(s_d)$ at a set of d locations $s_1, \cdots, s_d$ in $\mathbb{R}^2$ and a future time are modelled as a Gaussian vector where $\mathrm{E}(Y(s_i))$ is a known function of s_i, with $\mathrm{Var}(Y(s_i)) = \varrho$ and, for two different locations s_i and s_j,

$$\mathrm{Cov}(Y(s_i), Y(s_j)) = \sigma^2 \exp(-\frac{||s_i - s_j||}{r}),$$

where ϱ, σ and r are positive constants with $\sigma^2 \leq \varrho$. By simulating the vector $(Y(s_1), \cdots, Y(s_d))$, we can estimate the expected maximum temperature at these d locations.

For simplicity we assume thereafter that $Y(s_i)$ is centered for $1 \leq i \leq d$, and so $X_i = \varrho^{-1/2} Y(s_i)$ is a standard Gaussian random variable. The correlation matrix V of the Gaussian vector $X = (X_1, \ldots, X_d)^T$ is given by

$$V_{ij} = \frac{\sigma^2}{\varrho} \exp(-\frac{||s_i - s_j||}{r}),$$

for $i \neq j$. Since the matrix $(\exp(-||s_i - s_j||/r))_{1 \leq i,j \leq d}$ is positive semi-definite (Cressie 2015, Section 2.5), $\lambda \geq 1 - (\sigma^2/\varrho)$. We use the MC and MCMC algorithms to estimate $\mathrm{E}(\max_{1 \leq i \leq d} Y(s_i))$. Note that $\max_{1 \leq i \leq d} Y(s_i) = h(X)$, where $X = (X_1, \ldots, X_d)$, and $h(x) = \sqrt{\varrho} \max(x_1, \ldots, x_d)$ for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The analysis in Subsection 6.4 shows that if $\sigma^2 < \varrho$, then (6.3) holds for fixed $\epsilon/\Sigma < 1$.

Our numerical simulations assume that for $1 \leq i \leq d$, the first (resp. second) coordinate of s_i equals $\lfloor i/d' \rfloor / d'$ (resp. $(i \mod d')/d'$), where $d' = \lceil \sqrt{d} \rceil$. Our experiments were performed on a desktop PC with an Intel Pentium 2.90 GHz processor and 4 GB of RAM, running Windows 7 Professional. The codes were written in the C++ programming language, and the compiler used was Microsoft Visual C++ 2013. Computing times are given in seconds. The Cholesky factorization was implemented using the method described in (Glasserman 2004, Fig. 2.16). Table 1 compares the MC and MCMC methods for d up to 10^4 by assuming that $r = 10$, $\varrho = 8$, $\sigma^2 = 7.44$. After scaling, these parameters are close to those estimated in (Gel, Raftery and Gneiting 2004). Running the Cholesky factorization for $d = 10^5$ without external storage causes memory overflow. Extrapolating the results in Table 1 shows that the Cholesky factorization would take a few weeks for $d = 10^5$ if enough internal memory were available. Following the discussion surrounding (6.1), $\tau_{\mathrm{MC}}(\epsilon)$ was calculated in Table 1 as

$$\tau_{\mathrm{MC}}(\epsilon) = \mathrm{Chol} + \frac{\Sigma^2}{\epsilon^2} \mathrm{Simul},$$

where Chol is the time in seconds to perform the Cholesky decomposition, and Simul is the time in seconds to simulate Z_1 and to calculate $h(AZ_1)$. When $n' = \Sigma^2/\epsilon^2$, rounded to the nearest integer, the empirical standard deviation of the MC method, based on 100 replications, is within 11% from ϵ. For the tested parameters, the MCMC method is more efficient than the MC method, and its efficiency increases with d. For $d = 10^5$, $n = 100d$ and $b = n/2$, the MCMC average is 4.01, and is calculated in 8969 seconds.

Table 1: The MCMC and MC methods for estimating $\mathrm{E}(\max_{1 \leq i \leq d} Y(s_i))$, with $r = 10$, $\varrho = 8$, $\sigma^2 = 7.44$, $n = 100d$ and burnin $= n/2$. The second column gives the time to perform the Cholesky decomposition. The third column gives the time to simulate $Z_1, \ldots, Z_{n'}$ and calculate $h(AZ_i)$, $1 \leq i \leq n'$, where $n' = 10^4$, and does not incorporate the Cholesky decomposition running time. The MCMC RMSE is an estimate of $\sqrt{\mathrm{MSE}(n; n/2)}$, which is calculated as explained in Section F. The standard deviation Σ is estimated using the MC method with 10^4 samples. The last column gives $\tau_{\mathrm{MC}}(\epsilon)/\tau_{\mathrm{MCMC}}(\epsilon, 1/2)$, where $\epsilon = \sqrt{\mathrm{MSE}(n; n/2)}$, and $\tau_{\mathrm{MC}}(\epsilon)$ and $\tau_{\mathrm{MCMC}}(\epsilon, 1/2)$ are calculated in seconds.

d	Cholesky decomposition	MC simulations	MCMC Average	MCMC RMSE	MCMC comp. time	Σ	$\tau_{\mathrm{MC}}/\tau_{\mathrm{MCMC}}$
10^2	0.001	0.19	2.38	0.119	0.002	2.7	4
10^3	2.3	17	3.02	0.069	0.19	2.7	26
10^4	2306	1621	3.57	0.049	16	2.7	171

Table 2 compares the MC and MCMC methods by setting $\sigma^2 = \varrho$ and using the same methodology as in Table 1. For large values of d, the mean squared errors of the MCMC method are higher in Table 2 than in Table 1, but they have the same order of magnitude.

Table 3 reports the maximum difference, in absolute value, between the entries of V and the corresponding entries of the empirical covariance matrices of X_n and of AZ_1, respectively, with $n = 100d$. The maximum difference is of order 10^{-2} for both the MC and MCMC methods. This suggests that each entry of the matrix $V - \mathrm{Cov}(X_n)$ is of order 10^{-2} or less, in absolute value, which is lower than the bound implied by the discussion preceding Theorem 3.1.

Table 2: The MCMC and MC methods for estimating $\mathrm{E}(\max_{1\le i\le d} Y(s_i))$, with $r = 10$, $\varrho = \sigma^2 = 8$, $n = 100d$, and burnin $= n/2$.

d	Cholesky decomposition	MC simulations	MCMC Average	MCMC RMSE	MCMC comp. time	Σ	$\tau_{\mathrm{MC}}/\tau_{\mathrm{MCMC}}$
10^2	0.001	0.19	1.34	0.107	0.003	2.8	5
10^3	2.7	17	1.50	0.092	0.20	2.8	22
10^4	2130	1648	1.62	0.092	17	2.8	136

Table 3: Maximum difference between the empirical covariance matrices and V for the MCMC and MC methods. The covariance matrices were calculated using 10000 independent samples.

	$r = 10$, $\varrho = 8$, $\sigma^2 = 7.44$		$r = 10$, $\varrho = \sigma^2 = 8$	
d	MCMC	MC	MCMC	MC
10	0.019	0.017	0.019	0.012
10^2	0.034	0.026	0.025	0.016
10^3	0.032	0.033	0.024	0.019

Lemma 6.2 shows that, if $\sigma^2 < \varrho$, the MCMC method can also be used to calculate the probability that the maximum temperature over a set of d points exceeds a certain level.

7 Conclusion

We have shown how to simulate a Markov chain X_n, $n \ge 0$, such that the Wasserstein distance between the distribution of X_n and $N(0, V)$ is at most $d/\sqrt{n}$. It takes $O(d)$ time to generate each step of the chain. Whereas the standard Monte Carlo simulation method has $\Theta(d^2)$ storage cost, the additional storage cost of our method is $\Theta(d)$. Furthermore, by running the chain n steps, our method can estimate $\mathrm{E}(h(X))$, where X is a centered Gaussian vector with covariance matrix V and h is a real-valued function of d variables. Under certain conditions, we give an explicit upper bound on the mean square error of our estimate, and show that it is inversely proportional to the running time. We also prove that, in certain cases, the total time needed by our method to obtain a given standardized mean square error is $O^*(d^2)$ time, whereas the standard Monte Carlo method takes $\Theta(d^3)$ time.

A Proof of Lemma 2.1

Recall first that if Z is a centered d-dimensional random vector such that $\mathrm{E}(||Z||^2)$ is finite, then $\mathrm{E}(||Z||^2) = \mathrm{tr}(\mathrm{Cov}(Z))$.

It can be shown by induction that Y_n is a linear combination of $g_0, \ldots, g_{n-1}$, and so Y_n is a centered Gaussian vector. Hence X_n is also centered and Gaussian. Furthermore, Z_n is a centered Gaussian vector since it is a linear combination of $g_0, \ldots, g_{n-1}$ and of Z_0. Thus, Z_n and g_n are independent. For $0 \le l \le n$,

$$\begin{aligned} Z_{l+1} &= Y_{l+1} + M_{l+1} Z_0 \\ &= P_l Y_l + g_l f_l + P_l M_l Z_0 \\ &= P_l Z_l + g_l f_l. \end{aligned} \tag{A.1}$$

Thus, since Z_l and g_l are independent,

$$\begin{aligned}\mathrm{Cov}(Z_{l+1}) &= \mathrm{Cov}(P_l Z_l) + \mathrm{Cov}(g_l f_l) \\ &= P_l \mathrm{Cov}(Z_l) P_l^T + \mathrm{E}(g_l^2) f_l f_l^T \\ &= P_l \mathrm{Cov}(Z_l) P_l + f_l f_l^T.\end{aligned}$$

It follows by induction that $\mathrm{Cov}(Z_l) = I$, and so $Z_l \sim N(0, I)$. Thus, (2.3) holds when $j = n$. Furthermore, since g_l and Z_j are independent for $0 \le j \le l \le n-1$, it follows from (A.1) that $\mathrm{E}(Z_{l+1} Z_j^T) = P_l \mathrm{E}(Z_l Z_j^T)$. It follows by induction on l that $\mathrm{E}(Z_l Z_j^T) = M_{j,l}$ for $0 \le j \le l \le n$, Hence (2.3).

Since Y_n is a linear combination of $g_0, \ldots, g_{n-1}$, the vectors Z_0 and Y_n are independent. Thus, as $\mathrm{Cov}(Z_n) = I$ and $\mathrm{Cov}(M_n Z_0) = M_n M_n^T$, it follows from (2.2) that

$$\mathrm{Cov}(Y_n) = I - M_n M_n^T.$$

Hence (2.4), which implies that $\mathrm{Cov}(X_n) \le V$.

By (2.2), $\sqrt{V} Z_n - X_n = \sqrt{V} M_n Z_0$ and so,

$$\sqrt{V} Z_n - X_n \sim N(0, \sqrt{V} M_n M_n^T \sqrt{V}).$$

Hence,

$$\begin{aligned}\mathrm{E}(||X_n - \sqrt{V} Z_n||^2) &= \mathrm{tr}(\sqrt{V} M_n M_n^T \sqrt{V}) \\ &= ||M_n||^2.\end{aligned}$$

The second equation follows from the relation $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Using (2.4), this implies (2.5). □

B Proof of Lemma 4.1

Let (X, X') be a Gaussian vector in $\mathbb{R}^2$, with $X \sim N(0, \nu)$, $X' \sim N(0, \nu')$ and $\nu' \le \nu$. We show the following, which immediately implies Lemma 4.1:

$$\mathrm{E}((e^X - e^{X'})^2) \le (\nu + \nu' + 1/2)(e^{2\nu} + e^{2\nu'}) \mathrm{E}((X - X')^2). \tag{B.1}$$

Let

$$\rho = \mathrm{E}((X - X')^2) = \nu + \nu' - 2\mathrm{Cov}(X, X').$$

Since $\mathrm{E}(e^Z) = e^{\frac{1}{2}\mathrm{Var}(Z)}$ for any centered Gaussian random variable Z,

$$\begin{aligned}\mathrm{E}((e^X - e^{X'})^2) &= \mathrm{E}(e^{2X} + e^{2X'} - 2e^{X+X'}) \\ &= e^{2\nu} + e^{2\nu'} - 2e^{\frac{1}{2}\mathrm{Var}(X+X')} \\ &= e^{2\nu} + e^{2\nu'} - 2e^{\nu/2 + \nu'/2 + \mathrm{Cov}(X,X')} \\ &= 2e^{\nu+\nu'}(\cosh(\nu - \nu') - e^{-\rho/2}) \\ &\le e^{\nu+\nu'}((\nu - \nu')^2 \cosh(\nu - \nu') + \rho).\end{aligned}$$

The last equation follows from the inequalities $1 - x \le e^{-x}$ and $\cosh(x) \le 1 + x^2 \cosh(x)/2$ (which is a consequence of Taylor's formula with Lagrange remainder) for any real number x. Furthermore, $(\sqrt{\nu} - \sqrt{\nu'})^2 \le \rho$ since $\mathrm{Cov}(X, X') \le \sqrt{\nu\nu'}$, and so

$$\begin{aligned}(\nu - \nu')^2 &= (\sqrt{\nu} + \sqrt{\nu'})^2 (\sqrt{\nu} - \sqrt{\nu'})^2 \\ &\le \rho(\sqrt{\nu} + \sqrt{\nu'})^2 \\ &\le 2\rho(\nu + \nu').\end{aligned} \tag{B.2}$$

Hence, as $1 \leq \cosh(x)$ for any real number x,

$$\begin{aligned} \mathrm{E}((e^X - e^{X'})^2) &\leq \rho e^{\nu+\nu'}(2(\nu+\nu')\cosh(\nu-\nu')+1) \\ &\leq \rho e^{\nu+\nu'}\cosh(\nu-\nu')(2\nu+2\nu'+1). \end{aligned}$$

This implies (B.1).

C Proof of Lemma 4.2

We first prove the following lemma which implies that, under certain conditions, the covariances between the components of Y and Y' can be used to bound the covariance between $\hat{h}_1(Y)$ and $\hat{h}_2(Y')$.

Lemma C.1. *Let* $\begin{pmatrix} Y \\ Y' \end{pmatrix}$ *be a Gaussian column vector in* $\mathbb{R}^{2d}$*, with* $Y \sim Y' \sim N(0, I)$ *and* $\mathrm{E}(YY'^T) = AB^T$*, where* A *and* B *are* $d \times d$ *matrices, with* $AA^T \leq I$ *and* $BB^T \leq I$*. Let* h_1 *and* h_2 *be two* (κ, γ, V)*-Lipschitz functions on* $\mathbb{R}^d$*. Then*

$$|\mathrm{E}(\hat{h}_1(Y)\hat{h}_2(Y'))| \leq \kappa^2(||A||^{2\gamma} + ||B||^{2\gamma}).$$

Proof. We first note that, if $Z \sim Z' \sim N(0, I)$ and Z, Z' are independent, then

$$\mathrm{E}(\hat{h}_1(Z)\hat{h}_2(Z')) = 0. \tag{C.1}$$

Furthermore, if (Y, Y') is a centered Gaussian vector in $\mathbb{R}^{2d}$ with $\mathrm{Cov}(Y) \leq I$ and $\mathrm{Cov}(Y') \leq I$, and h is a (κ, γ, V)-Lipschitz function on $\mathbb{R}^d$, then

$$\begin{aligned} \mathrm{E}((\hat{h}(Y) - \hat{h}(Y'))^2) &\leq \kappa^2(\mathrm{E}(||\sqrt{V}Y - \sqrt{V}Y'||^2))^\gamma \\ &= \kappa^2(\mathrm{tr}(\sqrt{V}\mathrm{Cov}(Y-Y')\sqrt{V}))^\gamma \\ &= \kappa^2(\mathrm{tr}(V\mathrm{Cov}(Y-Y')))^\gamma. \end{aligned} \tag{C.2}$$

The second equation follows from the fact that $\sqrt{V}(Y - Y')$ is a centered Gaussian vector with covariance matrix $\sqrt{V}\mathrm{Cov}(Y - Y')\sqrt{V}$.

We now prove the lemma. Let G, G', G'', G_1 and G_2 be independent d-dimensional Gaussian vectors such that $G \sim G' \sim G'' \sim N(0, I)$, $G_1 \sim N(0, I - AA^T)$, and $G_2 \sim N(0, I - BB^T)$. Note that G_1 and G_2 exist since $I - AA^T$ and $I - BB^T$ are positive semi-definite. Since $AG \sim N(0, AA^T)$ and is independent of G_1, $AG + G_1 \sim N(0, I)$. Similarly, $BG + G_2 \sim N(0, I)$. Also, since G, G_1 and G_2 are independent and centered,

$$\begin{aligned} \mathrm{E}((AG + G_1)(BG + G_2)^T) &= \mathrm{E}((AG)(BG)^T) \\ &= AB^T. \end{aligned}$$

Thus, the covariance matrix of the Gaussian vector $\begin{pmatrix} AG + G_1 \\ BG + G_2 \end{pmatrix}$ is $\begin{pmatrix} I & AB^T \\ BA^T & I \end{pmatrix}$. Hence the centered Gaussian vectors $\begin{pmatrix} AG + G_1 \\ BG + G_2 \end{pmatrix}$ and $\begin{pmatrix} Y \\ Y' \end{pmatrix}$ have the same covariance matrix, and so they have the same distribution. Thus,

$$\begin{aligned} \mathrm{E}(\hat{h}_1(Y)\hat{h}_2(Y')) &= \mathrm{E}(\hat{h}_1(AG + G_1)\hat{h}_2(BG + G_2)) \\ &= \mathrm{E}((\hat{h}_1(AG + G_1) - \hat{h}_1(AG' + G_1))(\hat{h}_2(BG + G_2) - \hat{h}_2(BG'' + G_2)) \\ &\leq \frac{1}{2}(\mathrm{E}((\hat{h}_1(AG + G_1) - \hat{h}_1(AG' + G_1))^2) + \\ &\qquad \mathrm{E}((\hat{h}_2(BG + G_2) - \hat{h}_2(BG'' + G_2))^2)) \\ &\leq \kappa^2((\mathrm{tr}(VAA^T))^\gamma + (\mathrm{tr}(VBB^T))^\gamma). \\ &= \kappa^2((\mathrm{tr}(A^TVA))^\gamma + (\mathrm{tr}(B^TVB))^\gamma). \end{aligned}$$

The second equation follows by applying (C.1) to each of the pairs $(AG + G_1, BG'' + G_2)$, $(AG' + G_1, BG'' + G_2)$, and $(AG' + G_1, BG + G_2)$. The fourth equation follows from (C.2) and the relations $\mathrm{Cov}(AG - AG') = 2AA^T$ and $\mathrm{Cov}(BG - BG'') = 2BB^T$.

Hence

$$\mathrm{E}(\hat{h}_1(Y)\hat{h}_2(Y')) \le \kappa^2(||A||^{2\gamma} + ||B||^{2\gamma}).$$

Replacing h_2 with $-h_2$ completes the proof. □

We now prove Lemma 4.2. Assume first that the sequence $i_0, \ldots, i_{n-1}$, is deterministic. By Lemma 2.1, $Z_j \sim Z_l \sim N(0, I)$ for $0 \le j \le l \le n$, and

$$\mathrm{E}(Z_l Z_j^T) = M_{j,l}.$$

Let $A = M_{j',l}$ and $B = M_{j,j'}^T$, with $j' = \lfloor (j+l)/2 \rfloor$. Since A^T is the product of $l - j'$ projection matrices, it follows by induction on l that $||A^T x|| \le ||x||$ for $x \in \mathbb{R}^d$, and so $AA^T \le I$. Similarly, $BB^T \le I$. Since $M_{j,l} = AB^T$, it follows from Lemma C.1 that

$$|\mathrm{E}(\hat{h}(Z_j)\hat{h}(Z_l))| \le \kappa^2(||M_{j',l}||^{2\gamma} + ||M_{j,j'}^T||^{2\gamma}). \tag{C.3}$$

Thus,

$$\begin{aligned}
\mathrm{E}((\sum_{j=b}^{n-1} \hat{h}(Z_j))^2) &\le 2\sum_{j=b}^{n-1}\sum_{l=j}^{n-1} \mathrm{E}(\hat{h}(Z_j)\hat{h}(Z_l)) \\
&= 2\sum_{b\le j\le l\le n-1,\, l-j<2\delta} \mathrm{E}(\hat{h}(Z_j)\hat{h}(Z_l)) + 2\sum_{b\le j\le l\le n-1,\, l-j\ge 2\delta} \mathrm{E}(\hat{h}(Z_j)\hat{h}(Z_l)) \\
&\le 4(n-b)\delta\Sigma^2 + 2\kappa^2 \sum_{b\le j\le l\le n-1,\, l-j\ge 2\delta}^{n-1} ||M_{j',l}||^{2\gamma} + ||M_{j,j'}^T||^{2\gamma} \\
&\le 4(n-b)\delta\Sigma^2 + 4\kappa^2 \sum_{b\le j\le l\le n-1,\, l-j\ge \delta}^{n-1} ||M_{j,l}||^{2\gamma} + ||M_{j,l}^T||^{2\gamma}. \qquad \text{(C.4)}
\end{aligned}$$

The third equation follows from observing that $\hat{h}(Z_j)$ and $\hat{h}(Z_l)$ are centered and have standard deviation Σ. The last equation follows from the fact that each term $M_{j',l}$ (resp. $M_{j,j'}$) occurs at most twice in the last line.

Assume now that the sequence $i_0, \ldots, i_{n-1}$, is random. By (C.4),

$$\mathrm{E}((\sum_{j=b}^{n-1} \hat{h}(Z_j))^2 | i_0, \ldots, i_{n-1}) \le 4(n-b)\delta\Sigma^2 + 4\kappa^2 \sum_{b\le j\le l\le n-1,\, l-j\ge\delta} ||M_{j,l}||^{2\gamma} + ||M_{j,l}^T||^{2\gamma}.$$

We conclude the proof by taking expectations and using the tower law. □

D Proof of Lemma 5.2

Let $a_0 = \mathrm{E}((\hat{h}(Z_0))^2)$ and, for $j > 0$, let $a_j = 2\mathrm{E}(\hat{h}(Z_0)\hat{h}(Z_j))$. Since (Z_j), $j \ge 0$, is a time-homogeneous Markov Chain and $Z_j \sim N(0, I)$, $2\mathrm{E}(\hat{h}(Z_k)\hat{h}(Z_{k+j})) = a_j$ for $j > 0$. Hence

$$\begin{aligned}
\mathrm{E}((\sum_{j=0}^{n-1} \hat{h}(Z_j))^2) &= \sum_{j=0}^{n-1} \mathrm{E}((\hat{h}(Z_j)^2) + 2\sum_{0\le k<k+j<n} \mathrm{E}(\hat{h}(Z_k)\hat{h}(Z_{k+j})) \\
&= na_0 + \sum_{0\le k<k+j<n} a_j,
\end{aligned}$$

and so

$$n^{-1}\mathrm{E}(\sum_{j=0}^{n-1}\hat{h}(Z_j))^2 = \sum_{j=0}^{n-1}\frac{n-j}{n}a_j. \tag{D.1}$$

Moreover, by applying (C.3) with $j=0$ and $l=j$, it follows that

$$\begin{aligned}|\mathrm{E}(\hat{h}(Z_0)\hat{h}(Z_j))| &\leq \kappa^2\mathrm{E}(||M_{j',j}||^{2\gamma}+||M_{j'}^T||^{2\gamma})\\ &\leq \kappa^2((\mathrm{E}(||M_{j',j}||^2))^\gamma+(\mathrm{E}(||M_{j'}^T||^2))^\gamma),\end{aligned}$$

with $j'=\lfloor j/2\rfloor$. The second equation follows from Jensen's inequality. But, since $M_{j',j}\sim M_{j-j'}$,

$$\mathrm{E}(||M_{j',j}||^2)=\mathrm{E}(||M_{j-j'}||^2).$$

Furthermore, as $j'\leq j-j'$, it follows from Theorem 3.1 that

$$\mathrm{E}(||M_{j-j'}||^2)\leq\mathrm{E}(||M_{j'}||^2).$$

Since $M_{j'}\sim M_{j'}^T$, we conclude that $|a_j|\leq 2\kappa^2(\mathrm{E}(||M_{j'}||^2))^\gamma$. Hence, by Lemma 5.1, the series $\sum_{j=0}^\infty a_j$ is absolutely convergent. Thus, the RHS of (D.1) converges to c as n goes to infinity, where $c=\sum_{j=0}^\infty a_j$. But $\mathrm{E}(\hat{h}(Z_0)\hat{h}(Z_j))=\mathrm{Cov}(h(\sqrt{V}Z_0),h(\sqrt{V}Z_j))$ since $\mathrm{E}(h(\sqrt{V}Z_0))=\mathrm{E}(h(\sqrt{V}Z_j))=m$, which implies (5.1). □

E Proof of Lemma 6.1

Since the function $\max(z-K,0)$ is 1-Lipschitz with respect to z, we can assume without loss of generality that $K=0$. Let $\begin{pmatrix}U\\U'\end{pmatrix}$ be a centered Gaussian vector with $\mathrm{Cov}(U)\leq V$ and $\mathrm{Cov}(U')\leq V$, where U and U' have dimension d. Let U_i (resp. U_i') be the i-th component of U (resp. U') and $\lambda_i=(4\sigma_i^2+1)e^{2\sigma_i^2}$. It follows from Lemma 4.1 that

$$\mathrm{E}((e^{\sigma_iU_i}-e^{\sigma_iU_i'})^2)\leq\lambda_i\mathrm{E}((U_i-U_i')^2).$$

Moreover, by the Cauchy-Schwartz inequality,

$$\begin{aligned}(g(U)-g(U'))^2 &= \left(\sum_{i=1}^d(\lambda_i^{1/2}w_i)(\frac{e^{\sigma_iU_i}-e^{\sigma_iU_i'}}{\lambda_i^{1/2}})\right)^2\\ &\leq (\sum_{i=1}^d\lambda_iw_i^2)\sum_{i=1}^d\frac{(e^{\sigma_iU_i}-e^{\sigma_iU_i'})^2}{\lambda_i}.\end{aligned}$$

Taking expectations, it follows that

$$\begin{aligned}\mathrm{E}((g(U)-g(U'))^2) &\leq (\sum_{i=1}^d\lambda_iw_i^2)(\sum_{i=1}^d\mathrm{E}((U_i-U_i')^2))\\ &= \kappa^2\mathrm{E}(||U-U'||^2),\end{aligned}$$

as desired. □

F Estimating the mean square error

This section describes a numerical method to estimate the mean square error of the MCMC method for moderate values of d. The method first calculates a matrix A satisfying (1.1) using Cholesky factorization. An unbiased estimator of m is then computed as follows. Define the Markov chain (X'_n), $n \geq 0$, by setting $X'_0 = AZ_0$ and, for $n \geq 0$,

$$X'_{n+1} = X'_n + (g_n - e_n^T X'_n)(Ve_n). \tag{F.1}$$

Thus X'_n satisfies the same recursion as X_n. Using calculations similar to those in the beginning of Section 2, it can be shown by induction that $X'_n \sim N(0, V)$. Thus,

$$h'_{n,b} = \frac{\sum_{j=b}^{n-1} h(X'_j)}{n-b}$$

is an unbiased estimator of m. Since $\mathrm{MSE}(n;b)$ equals the variance of $h_{n,b}$ plus its square bias,

$$\mathrm{MSE}(n;b) = \mathrm{Var}(h_{n,b}) + (\mathrm{E}(h_{n,b} - h'_{n,b}))^2. \tag{F.2}$$

In Subsection 6.5, $\mathrm{MSE}(n;b)$ is estimated for $d \leq 10^4$ using (F.2), each term in the RHS of (F.2) being calculated via 100 independent simulations of $h_{n,b}$ and $h'_{n,b}$. For any $j \in \{0, \ldots, n-1\}$, the same random variables g_j and i_j are used to calculate $h_{n,b}$ and $h'_{n,b}$ via (1.2) and (F.1).

Acknowledgments

This research has been presented at the Paris Bachelier Seminar, November 2016, and at the 2nd IMA Conference on the Mathematical Challenges of Big Data, London, December 2016. The author thanks Nicolas Chopin, Petros Dellaportas, Peter Diggle, Peter Glynn, Emmanuel Gobet, Benjamin Jourdain, Didier Marteau, seminar and conference participants for helpful conversations. He is grateful to two anonymous referees and an associate editor for insightful comments and suggestions. This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) under the reference ANR-10-LABX-0095. It benefitted from a French government support managed by the National Research Agency (ANR).

References

Aune, E., Eidsvik, J. and Pokern, Y. (2013). Iterative numerical methods for sampling from high dimensional Gaussian distributions, *Statistics and Computing* **23**(4): 501–521.

Bélisle, C. J., Romeijn, H. E. and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions, *Mathematics of Operations Research* **18**(2): 255–266.

Cousins, B. and Vempala, S. (2016). A practical volume algorithm, *Mathematical Programming Computation* **8**(2): 133–160.

Cressie, N. (2015). *Statistics for spatial data*, John Wiley & Sons, New York.

Dellaportas, P. and Roberts, G. O. (2003). An introduction to MCMC, *Spatial statistics and computational methods*, Springer New York, pp. 1–41.

Diaconis, P. (2009). The Markov chain Monte Carlo revolution, *Bulletin of the American Mathematical Society* **46**(2): 179–205.

Dietrich, C. and Newsam, G. N. (1997). Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM Journal on Scientific Computing* **18**(4): 1088–1107.

Diggle, P. J. and Ribeiro Jr, P. J. (2007). *Model-based Geostatistics*, Springer Series in Statistics, Springer, New York.

Diggle, P. J., Ribeiro Jr, P. J. and Christensen, O. F. (2003). An introduction to model-based geostatistics, *Spatial statistics and computational methods*, Springer, New York, pp. 43–86.

Dowson, D. and Landau, B. (1982). The Fréchet distance between multivariate normal distributions, *Journal of multivariate analysis* **12**(3): 450–455.

Durmus, A. and Moulines, E. (2016). Sampling from strongly log-concave distributions with the unadjusted Langevin algorithm, *arXiv preprint 1605.01559* .

Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2011). *Statistical distributions*, fourth edn, John Wiley & Sons.

Gel, Y., Raftery, A. E. and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method, *Journal of the American Statistical Association* **99**(467): 575–583.

Gillman, D. (1998). A Chernoff bound for random walks on expander graphs, *SIAM Journal on Computing* **27**(4): 1203–1220.

Glasserman, P. (2004). *Monte Carlo methods in financial engineering*, Vol. 53, Springer, New York.

Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*, fourth edn, JHU Press, Baltimore.

Hull, J. (2012). *Options, Futures and Other Derivatives*, eigth edn, Pearson Education Limited, England.

Kahale, N. (1997a). Large deviation bounds for Markov chains, *Combinatorics, Probability and Computing* **6**(04): 465–474.

Kahale, N. (1997b). A semidefinite bound for mixing rates of Markov chains, *Random Structures & Algorithms* **11**(4): 299–313.

Lovász, L. (1999). Hit-and-run mixes fast, *Mathematical Programming* **86**(3): 443–461.

Lovász, L. and Vempala, S. (2006). Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization, *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, IEEE, pp. 57–68.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2): 325–338.

Russo, D. and Van Roy, B. (2014). Learning to optimize via posterior sampling, *Mathematics of Operations Research* **39**(4): 1221–1243.

Russo, D. and Van Roy, B. (2016). An information-theoretic analysis of Thompson sampling, *Journal of Machine Learning Research* **17**(68): 1–30.

Sinclair, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow, *Combinatorics, Probability and Computing* **1**(04): 351–370.

Smith, R. L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions, *Operations Research* **32**(6): 1296–1308.

Stein, M. L. (2002). Fast and exact simulation of fractional Brownian surfaces, *Journal of Computational and Graphical Statistics* **11**(3): 587–599.

Villani, C. (2009). *Optimal transport: old and new*, Vol. 338, Springer-Verlag Berlin Heidelberg.

Wood, A. T. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$, *Journal of Computational and Graphical Statistics* **3**(4): 409–432.