



# Variance reduction for Metropolis–Hastings samplers

Angelos Alexopoulos<sup>1</sup> · Petros Dellaportas<sup>2,3,4</sup> · Michalis K. Titsias<sup>5</sup>

Received: 23 March 2022 / Accepted: 12 November 2022  
© The Author(s) 2022

## Abstract

We introduce a general framework that constructs estimators with reduced variance for random walk Metropolis and Metropolis-adjusted Langevin algorithms. The resulting estimators require negligible computational cost and are derived in a post-process manner utilising all proposal values of the Metropolis algorithms. Variance reduction is achieved by producing control variates through the approximate solution of the Poisson equation associated with the target density of the Markov chain. The proposed method is based on approximating the target density with a Gaussian and then utilising accurate solutions of the Poisson equation for the Gaussian case. This leads to an estimator that uses two key elements: (1) a control variate from the Poisson equation that contains an intractable expectation under the proposal distribution, (2) a second control variate to reduce the variance of a Monte Carlo estimate of this latter intractable expectation. Simulated data examples are used to illustrate the impressive variance reduction achieved in the Gaussian target case and the corresponding effect when target Gaussianity assumption is violated. Real data examples on Bayesian logistic regression and stochastic volatility models verify that considerable variance reduction is achieved with negligible extra computational cost.

**Keywords** Bayesian inference · Control variates · Markov chain Monte Carlo · Logistic regression · Poisson equation · Stochastic volatility

## 1 Introduction

Statistical methods for reducing the bias and the variance of estimators have played a prominent role in Monte Carlo based numerical algorithms. Variance reduction via control variates has a long and well studied history introduced as early as the work of Kahn and Marshall (1953), whereas an early non-

parametric estimate of bias, subsequently renamed jackknife and broadly used for bias reduction, was first presented by Quenouille (1956). However, the corresponding theoretical developments in the more complicated, but extremely popular and practically important, estimators based on MCMC algorithms has been rather limited. The major impediment is the fact that the MCMC estimators are based on ergodic averages of dependent samples produced by simulating a Markov chain.

We provide a general methodology to construct control variates for any discrete time random walk Metropolis (RWM) and Metropolis-adjusted Langevin algorithm (MALA) Markov chains that can achieve, in a post-processing manner and with a negligible additional computational cost, impressive variance reduction when compared to the standard MCMC ergodic averages. Our proposed estimators are based on an approximate, but accurate, solution of the Poisson equation for a multivariate Gaussian target density of any dimension.

Suppose that we have a sample of size  $n$  from an ergodic Markov chain  $\{X_n\}_{n \geq 0}$  with continuous state space  $\mathbf{X} \subseteq \mathbb{R}^d$ , transition kernel  $P$  and invariant measure  $\pi$ . A standard estimator of the mean  $E_\pi[F] := \pi(F) = \int F d\pi$  of a real-

✉ Angelos Alexopoulos  
angelos@mrc-bsu.cam.ac.uk

Petros Dellaportas  
p.dellaportas@ucl.ac.uk

Michalis K. Titsias  
mtitsias@google.com

<sup>1</sup> MRC Biostatistics Unit, University of Cambridge Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK

<sup>2</sup> Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK

<sup>3</sup> Department of Statistics, Athens University of Economics and Business, Athens, Greece

<sup>4</sup> The Alan Turing Institute, London, UK

<sup>5</sup> DeepMind, London, UK

valued function  $F$  defined on  $\mathbf{X}$  under  $\pi$  is the ergodic mean

$$\mu_n(F) := \frac{1}{n} \sum_{i=0}^{n-1} F(X_i).$$

which satisfies, for any initial distribution of  $X_0$ , a central limit theorem of the form

$$\begin{aligned} \sqrt{n} [\mu_n(F) - \pi(F)] \\ = n^{-1/2} \sum_{i=0}^{n-1} [F(X_i) - \pi(F)] \xrightarrow{D} N(0, \sigma_F^2), \end{aligned}$$

with the asymptotic variance given by

$$\sigma_F^2 := \lim_{n \rightarrow \infty} n E_\pi [(\mu_n(F) - \pi(F))^2].$$

Interesting attempts on variance reduction methods for Markov chain samplers include the use of antithetic variables (Barone and Frigessi 1990; Green and Han 1992; Craiu et al. 2005), Rao-Blackwellization (Gelfand and Smith 1990), Riemann sums (Philippe and Robert 2001) or autocorrelation reduction (Mira and Geyer 2000; Van Dyk and Meng 2001; Yu and Meng 2011).

Control variates have played an outstanding role in the MCMC variance reduction quiver. A strand of research is based on Assaraf and Caffarel (1999) who noticed that a Hamiltonian operator together with a trial function are sufficient to construct an estimator with zero asymptotic variance. They considered a Hamiltonian operator of Schrödinger-type that led to a series of zero-variance estimators studied by Valle and Leisen (2010), Mira et al. (2013) and Papamarkou et al. (2014). The estimation of the optimal parameters of the trial function is conducted by ignoring the Markov chain sample dependency, an issue that was dealt with by Belomestny et al. (2020) by utilizing spectral methods. The main barrier for the wide applicability of zero-variance estimators is that their computational complexity increases with  $d$ , see South et al. (2018). Another approach to construct control variates is a non-parametric version of the methods presented by Mira et al. (2013) and Papamarkou et al. (2014) which lead to the construction of control functionals (Oates et al. 2017; Barp et al. 2018; South et al. 2020). Although their computational cost with respect to  $d$  is low, their general applicability is prohibited due to the cubic computational cost with respect to  $n$  (South et al. 2018; Oates et al. 2019) and the possibility to suffer from the curse of dimensionality that is often met in non-parametric methods (Wasserman 2006). Finally, Hammer and Tjelmeland (2008) proposed constructing control variates by expanding the state space of the Metropolis-Hastings algorithm.

An approach which is closely related to our proposed methodology attempts to minimise the asymptotic variance

$\sigma_F^2$ . This seems a hard problem since a closed form expression of  $\sigma_F^2$  is not available and therefore a loss function to be minimised is not readily available; see, for example, Flégel et al. (2010). However, there has been a recent research activity based on the following observation by Andradóttir et al. (1993). If a solution  $\hat{F}$  to the Poisson equation for  $F$  was available, that is if for every  $x \in \mathbf{X}$

$$F(x) + P\hat{F}(x) - \hat{F}(x) = \pi(F) \quad (1)$$

where

$$PF(x) := E_x[F(X_1)] := E_x[F(X_1)|X_0 = x],$$

then one could construct a function equal to  $F(x) + P\hat{F}(x) - \hat{F}(x)$  which is constant and equal to  $\pi(F)$ . It is then immediate that a zero-variance and zero-bias estimator for  $F$  is given by

$$\mu_{n,\hat{F}}(F) := \frac{1}{n} \sum_{i=0}^{n-1} \{F(X_i) + P\hat{F}(X_i) - \hat{F}(X_i)\}$$

which can be viewed as an enrichment of the estimator  $\mu_n(F)$  with the (optimal) control variate  $P\hat{F} - \hat{F}$ . Of course, solving (1) is extremely hard for continuous state space Markov chains, even if we assume that  $E_\pi[F]$  is known, because it involves solving a non-standard integral equation. Interestingly, a solution of this equation (also called the fundamental equation) produces zero-variance estimators suggested by Assaraf and Caffarel (1999) for a specific choice of Hamiltonian operator. One of the rare examples that (1) has been solved exactly for discrete time Markov chains is the random scan Gibbs sampler where the target density is a multivariate Gaussian density, see Dellaportas and Kontoyiannis (2012), Dellaportas and Kontoyiannis (2009). They advocated that this solution provides a good approximation to (1) for posterior densities often met in Bayesian statistics that are close to multivariate Gaussian densities. Indeed, since direct solution of (1) is not available, approximating  $\hat{F}$  has been also suggested by Andradóttir et al. (1993), Atchadé and Perron (2005), Henderson (1997), Meyn (2008).

Tsourti (2012) attempted to extend the work by Dellaportas and Kontoyiannis (2012) to RWM samplers. The resulting algorithms produced estimators with lower variance but the computational cost required for the post-processing construction of these estimators counterbalance the variance reduction gains. We build on the work by Tsourti (2012) here but we differ in that (i) we build new, appropriately chosen to facilitate analytic computations, non-linear  $d$ -dimensional approximations to  $\hat{F}(x)$  rather than linear combinations of 1-dimensional functions and (ii) we produce efficient Monte Carlo approximations of the  $d$ -dimensional integral  $P\hat{F}(x)$  so that no extra computation is required for its evaluation.

Finally, Mijatović et al. (2018) approximate numerically the solution of (1) for 1-dimensional RWM samplers and Mijatović and Vogrinc (2019) construct control variates for large  $d$  by employing the solution of (1) that is associated with the Langevin diffusion in which the Markov chain converges as the dimension of its state space tends to infinity (Roberts et al. 1997); this requires very expensive Monte Carlo estimation methods so it is prohibited for realistic statistical applications.

We follow this route and add to this literature by extending the work of Dellaportas and Kontoyiannis (2012) and Tsourtis (2012) to RWM and MALA algorithms by producing estimators for the posterior means of each co-ordinate of a  $d$ -dimensional target density with reduced asymptotic variance and negligible extra computational cost. Our Monte Carlo estimator to compute the expectation  $\pi(F)$  makes use of three components:

- (a) An approximation  $G(x)$  to the solution of the Poisson equation associated with the target  $\pi(x)$ , transition kernel  $P$  and function  $F(x)$ .
- (b) An approximation of the target  $\pi(x)$  with a Gaussian density  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$  and then specifying  $G(x)$  in (a) by an accurate approximation to the solution of the Poisson equation for the approximate target  $\tilde{\pi}(x)$ .
- (c) An additional control variate, referred to as *static* control variate, that is based on the same Gaussian approximation  $\tilde{\pi}(x)$  and allows to reduce the variance of a Monte Carlo estimator for the intractable expectation  $PG(x)$ .

In Section 2 we provide full details of the above steps. We start by discussing, in Section 2.1, how all the above ingredients are put together to eventually arrive at the general form of our proposed estimator in equation (7). In Section 3 we present extensive simulation studies that verify that our methodology performs very well with multi-dimensional Gaussian targets and it stops reducing the asymptotic variance when we deal with a multimodal 50-dimensional target density with distinct, remote modes. Moreover, we apply our methodology to real data examples consisting of a series of logistic regression examples with parameter vectors up to 25 dimensions and two stochastic volatility examples with 53 and 103 parameters. In all cases we have produced estimators with considerable variance reduction with negligible extra computational cost.

### 1.1 Some notation

In the remainder of the paper we use a simplified notation where both  $d$ -dimensional random variables and their values are denoted by lower case letters, such as  $x = (x^{(1)}, \dots, x^{(d)})$  and where  $x^{(j)}$  is the  $j$ th dimension or coor-

dinate,  $j = 1, \dots, d$ ; the subscript  $i$  refers to the  $i$ th sample drawn by using an MCMC algorithm, that is  $x_i^{(j)}$  is the  $i$ th sample for the  $j$ th coordinate of  $x$ ; the density of the  $d$ -variate Gaussian distribution with mean  $m$  and covariance matrix  $S$  is denoted by  $\mathcal{N}(\cdot|m, S)$ ; for a function  $f(x)$  we set  $\nabla =: (\partial f/\partial x^{(1)}, \dots, \partial f/\partial x^{(d)})$ ;  $I_d$  is the  $d \times d$  identity matrix and the superscript  $\top$  in a vector or matrix denotes its transpose;  $\|\cdot\|$  denotes the Euclidean norm; all the vectors are understood as column vectors.

## 2 Metropolis–Hastings estimators with control variates from the Poisson equation

### 2.1 The general form of estimators for arbitrary targets

Consider an arbitrary intractable target  $\pi$  from which we have obtained a set of correlated samples by simulating a Markov chain with transition kernel  $P$  obtained by a Metropolis–Hastings kernel invariant to  $\pi$ . To start with, assume a function  $G(x)$ . By following the observation of Henderson (1997) the function  $PG(x) - G(x)$  has zero expectation with respect to  $\pi$  because the kernel  $P$  is invariant to  $\pi$ . Therefore, given  $n$  correlated samples from the target, i.e.  $x_i \sim \pi$  with  $i = 0, \dots, n - 1$ , the following estimator is unbiased

$$\mu_{n,G}(F) := \frac{1}{n} \sum_{i=0}^{n-1} \{F(x_i) + \underbrace{PG(x_i) - G(x_i)}_{\text{Poisson control variate}}\}. \tag{2}$$

For general Metropolis–Hastings algorithms the kernel  $P$  is such that the expectation  $PG(x)$  takes the form

$$\begin{aligned} PG(x) &= \int P(x, dy)G(y) \\ &= \int \alpha(x, y)q(y|x)G(y)dy \\ &\quad + \left(1 - \int \alpha(x, y)q(y|x)dy\right)G(x) \\ &= G(x) + \int \alpha(x, y)(G(y) - G(x))q(y|x)dy, \tag{3} \end{aligned}$$

where

$$\alpha(x, y) = \min\{1, r(x, y)\}, \quad r(x, y) = \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)} \tag{4}$$

and  $q(y|x)$  is the proposal distribution. By substituting (3) back into estimator (2) we obtain

$$\mu_{n,G}(F) := \frac{1}{n} \sum_{i=0}^{n-1} \left\{ F(x_i) + \underbrace{\int \alpha(x_i, y)(G(y) - G(x_i))q(y|x_i)dy}_{\text{Poisson control variate}} \right\}. \quad (5)$$

To use this estimator we need to overcome two obstacles: (i) we need to specify the function  $G(x)$  and (ii) we need to deal with the intractable integral associated with the control variate.

Regarding (i) there is a theoretical *best choice* which is to set  $G(x)$  to the function  $\hat{F}(x)$  that solves the Poisson equation,

$$\int \alpha(x, y)(\hat{F}(y) - \hat{F}(x))q(y|x)dy = -F(x) + \pi(F), \quad (6)$$

for every  $x \sim \pi$ , where we have substituted in the general form of the Poisson equation from (1) the Metropolis-Hastings kernel. For such optimal choice for  $G$  the estimator in (5) has zero variance, i.e. it equals to the exact expectation  $\pi(F)$ . Nevertheless, getting  $\hat{F}$  for general high-dimensional intractable targets is not feasible, and hence we need to compromise with an inferior choice for  $G$  that can only approximate  $\hat{F}$ . To get such  $G$ , we make use of a Gaussian approximation to the intractable target, as indicated by the assumption below.

**Assumption 1** The target  $\pi(x)$  is approximated by a multivariate Gaussian  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$  and the covariance matrix of the proposal  $q(y|x)$  is proportional to  $\Sigma$ .

The main purpose of the above assumption is to establish the ability to construct an efficient RWM or MALA sampler. Indeed, it is well-known that efficient implementation of these Metropolis-Hastings samplers when  $d > 1$  requires that the covariance matrix of  $q(y|x)$  should resemble as much as possible the shape of  $\Sigma$ . In adaptive MCMC (Roberts and Rosenthal 2009), such a shape matching is achieved during the adaptive phase where  $\Sigma$  is estimated. If  $\pi(x)$  is a smooth differentiable function,  $\Sigma$  could be alternatively estimated by a gradient-based optimisation procedure and it is then customary to choose a proposal covariance matrix of the form  $c^2\Sigma$  for a tuned scalar  $c$ .

We then aim to solve the Poisson equation for the Gaussian approximation by finding the function  $\hat{F}_{\tilde{\pi}}(x)$  that satisfies,

$$\int \tilde{\alpha}(x, y)(\hat{F}_{\tilde{\pi}}(y) - \hat{F}_{\tilde{\pi}}(x))q(y|x)dy = -F(x) + \tilde{\pi}(F),$$

for every  $x \sim \tilde{\pi}$ . It is useful to emphasize the difference between this new Poisson equation and the original Poisson equation in (6). This new equation involves the approximate Gaussian target  $\tilde{\pi}$  and the corresponding ‘‘approximate’’

Metropolis-Hastings transition kernel  $\tilde{P}$ , which now has been modified so that the ratio  $\tilde{\alpha}(x, y)$  is obtained by replacing the exact target  $\pi$  with the approximate target  $\tilde{\pi}$  while the proposal  $q(y|x)$  is also modified if needed.<sup>1</sup> Clearly, this modification makes  $\tilde{P}$  invariant to  $\tilde{\pi}$ . When  $\tilde{\pi}$  is a good approximation to  $\pi$ , we expect also  $\hat{F}_{\tilde{\pi}}$  to closely approximate the ideal function  $\hat{F}$ . Therefore, in our method we propose to set  $G$  to  $\hat{F}_{\tilde{\pi}}$  (actually to an analytic approximation of  $\hat{F}_{\tilde{\pi}}$ ) and then use it in the estimator (5).

Having chosen  $G(x)$ , we now discuss the second challenge (ii), i.e. dealing with the intractable expectation  $\int \alpha(x_i, y)(G(y) - G(x_i))q(y|x_i)dy$ . Given that for any drawn sample  $x_i$  of the Markov chain there is also a corresponding proposed sample  $y_i$  that is generated from the proposal, we can unbiasedly approximate the integral with a single-sample Monte Carlo estimate,

$$\int \alpha(x_i, y)(G(y) - G(x_i))q(y|x_i)dy \approx \alpha(x_i, y_i)(G(y_i) - G(x_i)),$$

where  $y_i \sim q(y|x_i)$ . Although  $\alpha(x_i, y_i)(G(y_i) - G(x_i))$  is a unbiased stochastic estimate of the Poisson-type control variate, it can have high variance that needs to be reduced. We introduce a second control variate based on some function  $h(x_i, y_i)$ , that correlates well with  $\alpha(x_i, y_i)(G(y_i) - G(x_i))$ , and it has analytic expectation  $E_{q(y|x_i)}[h(x_i, y)]$ . We refer to this control variate as *static* since it involves a standard Monte Carlo problem with exact samples from the tractable proposal density  $q(y|x)$ . To construct  $h(x_i, y)$  we rely again on the Gaussian approximation  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$  as we describe in Sect. 2.3.

With  $G(x)$  and  $h(x, y)$  specified, we can finally write down the general form of the proposed estimator that can be efficiently computed only from the MCMC output samples  $\{x_i\}_{i=0}^{n-1}$  and the corresponding proposed samples  $\{y_i\}_{i=0}^{n-1}$ :

$$\mu_{n,G}(F) := \frac{1}{n} \sum_{i=0}^{n-1} \left\{ F(x_i) + \underbrace{\alpha(x_i, y_i)(G(y_i) - G(x_i))}_{\text{Stochastic Poisson control variate}} + \underbrace{h(x_i, y_i) - E_{q(y|x_i)}[h(x_i, y)]}_{\text{Static control variate}} \right\}. \quad (7)$$

In practice we use a slightly modified version of this estimator by adding a set of adaptive regression coefficients  $\theta_n$  to further reduce the variance following Dellaportas and Kontoyiannis (2012); see Sect. 2.4.

<sup>1</sup> For the standard RWM algorithm  $q(y|x)$  remains exactly the same, while for MALA it needs to be modified by replacing the gradient  $\nabla \log \pi(x)$  with  $\nabla \log \tilde{\pi}(x)$ .

## 2.2 Approximation of the Poisson equation for Gaussian targets

### 2.2.1 Standard Gaussian case

In this section we construct an analytical approximation to the exact solution of the Poisson equation for the standard Gaussian  $d$ -variate target  $\tilde{\pi}_0(x) = \mathcal{N}(x|0, I_d)$  and for the function  $F(x) = x^{(j)}$  where  $1 \leq j \leq d$ . We use the function  $F(x) = x^{(j)}$  in the remainder of the paper which corresponds to approximating the mean value  $E_{\pi}[x^{(j)}]$ , while other choices of  $F$  are left for future work. We denote the exact unknown solution by  $\hat{F}_{\tilde{\pi}_0}$  and the analytical approximation by  $G_0$ . Given this target and some choice for  $G_0$  we express the expectation in (3) as

$$PG_0(x) = G_0(x)(1 - a(x)) + a_g(x),$$

where

$$a(x) = \int \min \left\{ 1, e^{-\frac{1}{2}(y^\top y - x^\top x)} \frac{q(x|y)}{q(y|x)} \right\} q(y|x) dy, \tag{8}$$

$$a_g(x) = \int \min \left\{ 1, e^{-\frac{1}{2}(y^\top y - x^\top x)} \frac{q(x|y)}{q(y|x)} \right\} G_0(y) q(y|x) dy. \tag{9}$$

The calculation of  $PG_0(x)$  reduces thus to the calculation of the integrals  $a(x)$  and  $a_g(x)$ . In both integrals  $x^\top x$  is just a constant since the integration is with respect to  $y$ . Moreover, the MCMC algorithm we consider is either RWM or MALA with proposal

$$q(y|x) = \mathcal{N}(y|rx, c^2I), \tag{10}$$

where  $r = 1$  corresponds to RWM and  $r = 1 - c^2/2$  to MALA while  $c > 0$  is the step-size. Both  $a(x)$  and  $a_g(x)$  are expectations under the proposal distribution  $q(y|x)$ .

One key observation is that for any dimension  $d$ ,  $y^\top y$  is just an univariate random variable with law induced by  $q(y|x)$ . Then,  $y^\top y$  together with  $\log \frac{q(x|y)}{q(y|x)}$  can induce an overall tractable univariate random variable so that the computation of  $a(x)$  in (8) can be performed analytically. The computation of  $a_g(x)$  is more involved since it depends on the form of  $G_0$ . Therefore, we propose an approximate  $G_0$  by first introducing a parametrised family that leads to tractable and efficient closed form computation of  $a_g(x)$ . In particular, we consider the following weighted sum of exponential functions

$$\sum_{k=1}^K w_k \exp\{\beta_k^\top x - \gamma_k(x - \delta_k)^\top (x - \delta_k)\}, \tag{11}$$

where  $w_k$  and  $\gamma_k$  are scalars whereas  $\beta_k$  and  $\delta_k$  are  $d$ -dimensional vectors. It turns out that using the form in (11) for  $G_0$  we can analytically compute the expectation  $PG_0$  as stated in Proposition 1. The proof of this proposition and the proofs of all remaining propositions and remarks presented throughout Sect. 2 are given in the ‘‘Appendix’’.

**Proposition 1** *Let  $a(x)$  and  $a_g(x)$  given by (8) and (9) respectively and  $G_0$  in  $a_g(x)$  to have the form in (11). Then,*

$$a(x) = E_f \left[ \min \left( 1, \exp \left\{ -\frac{c^2 \tau^2 (f - x^\top x / c^2)}{2} \right\} \right) \right],$$

where  $\tau^2 = 1$  in the case of RWM and  $\tau^2 = c^2/4$  in the case of MALA and  $f$  follows the non-central chi-squared distribution with  $d$  degrees of freedom and non-central parameter  $x^\top x / c^2$ , and

$$a_g(x) = \sum_{k=1}^K A_k(x) E_{f_{k,g}} \left[ \min \left\{ 1, \exp \left\{ -\frac{\tau^2 s_k^2}{2} (f_{k,g} - x^\top x / s_k^2) \right\} \right\} \right],$$

where  $f_{k,g}$  follows the non-central chi-squared distribution with  $d$  degrees of freedom and non-central parameter  $m_k(x)^\top m_k(x) / c^2$  and  $A_k(x) = (1 + 2c^2 \gamma_k)^{-d/2} \exp \left\{ -\frac{r^2 x^\top x - \gamma_k \delta_k^\top \delta_k + \frac{m_k(x)^\top m_k(x)}{2c^2(1+2\gamma_k c^2)}}{1 + 2c^2 \gamma_k} \right\}$ ,  $m_k(x) = \frac{rx + c^2(\beta_k + \gamma_k \delta_k)}{1 + 2c^2 \gamma_k}$  and  $s_k^2 = c^2 / (1 + 2c^2 \gamma_k)$ .

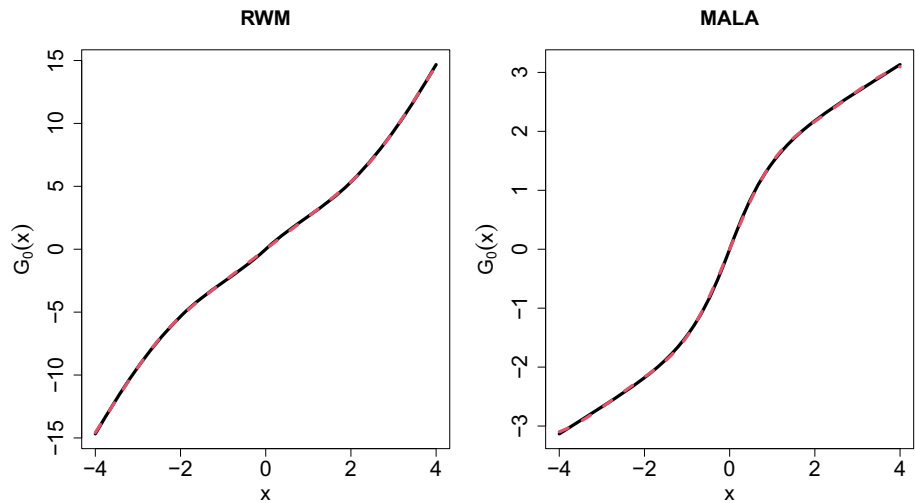
Proposition 1 states that the calculation of  $a_g(x)$  and  $a(x)$  is based on the cdf of the non-central chi-squared distribution and allows, for  $d$ -variate standard normal targets, the exact computation of the modified estimator  $\mu_{n,G}$  given by (2).

Having a family of functions for which we can calculate analytically the expectation  $PG_0$  we turn to the problem of specifying a particular member of this family to serve as an accurate approximation to the solution of the Poisson equation for the standard Gaussian distribution. We first provide the following proposition which states that  $\hat{F}_{\tilde{\pi}_0}$  satisfies certain symmetry properties.

**Proposition 2** *Given  $F(x) = x^{(j)}$ , the exact solution  $\hat{F}_{\tilde{\pi}_0}(x)$  is: (i) (holds for  $d \geq 1$ ) Odd function in the dimension  $x^{(j)}$ . (ii) (holds for  $d \geq 2$ ) Even function over any remaining dimension  $x^{(j')}$ ,  $j' \neq j$ . (iii) (holds for  $d \geq 3$ ) Permutation invariant over the remaining dimensions.*

To construct an approximation model family that incorporates the symmetry properties of Proposition 2 we make the following assumptions for the parameters in (11). We set  $K = 4$  and we assume that  $w_k \in \mathbb{R}$  and  $\gamma_k > 0$  for each  $k = 1, 2, 3, 4$  whereas we set  $w_1 = -w_2 = b_0$ ,  $w_3 = -w_4 = c_0$ ,  $\gamma_1 = \gamma_2 = b_2$  and  $\gamma_3 = \gamma_4 = c_1$ . Moreover, for the  $d$ -dimensional vectors  $\beta_k$  and  $\delta_k$  we assume that  $\beta_1 = -\beta_2$ ,  $\beta_3 = \beta_4 = \delta_1 = \delta_2 = 0$  and  $\delta_3 = -\delta_4$ ; we

**Fig. 1** Numerical solution of the Poisson equation (black solid lines) and its approximation (red dashed lines) in the case of univariate standard Gaussian target simulated by using the random walk Metropolis (RWM) algorithm and the Metropolis-adjusted Langevin algorithm (MALA). (Color figure online)



set the vectors  $\beta_1$  and  $\delta_3$  to be filled everywhere with zeros except from their  $j$ th element which is equal to  $b_1$  and  $c_2$  respectively. We specify thus the function  $G_0 : \mathbb{R}^d \rightarrow \mathbb{R}$  as

$$G_0(x) = b_0(e^{b_1x^{(j)}} - e^{-b_1x^{(j)}}) \times e^{-b_2\|x\|^2} + c_0(e^{-c_1(x^{(j)}-c_2)^2} - e^{-c_1(x^{(j)}+c_2)^2}) \times e^{-c_1 \sum_{j' \neq j} (x^{(j')})^2}. \tag{12}$$

We note that the above choices for the parameters of  $G_0$  are not the only ones that result in a function that obeys the symmetries properties of Proposition 2. By imposing, however, the described restrictions on the parameters of  $G_0$ , we keep the number of free parameters low allowing, thus, the efficient identification of optimal parameter values.

To identify optimal parameters for the function  $G_0$  in (12) such that  $G_0 \approx \hat{F}_{\tilde{\pi}_0}$  we first simulate a Markov chain with large sample size  $n$  from the  $d$ -variate standard Gaussian distribution by employing the RWM algorithm and the MALA. Then, for each algorithm we minimize the loss function

$$\mathcal{L} = (1/n) \sum_{i=1}^n (G_0(x_i) - PG_0(x_i) - x_i^{(1)})^2, \tag{13}$$

with respect to the parameters  $b_0, b_1, b_2, c_0, c_1$  and  $c_2$  by employing the Broyden-Fletcher-Goldfarb-Shanno method (Broyden 1970) as implemented by the routine `optim` in the statistical software R (R Core Team 2021). Figure 1 provides an illustration of the achieved approximation to  $\hat{F}_{\tilde{\pi}_0}$  in the univariate case where  $d = 1$  and the model in (12) simplifies as

$$G_0(x) = b_0(e^{b_1x-b_2x^2} - e^{-b_1x-b_2x^2}) + c_0(e^{-c_1(x-c_2)^2} - e^{-c_1(x+c_2)^2}).$$

For such case, we can visualize our optimised  $G_0$  and compare it against the numerical solution from Mijatović et al. (2018). Figure 1 shows this comparison which provides clear evidence that for  $d = 1$  our approximation is very accurate.

### 2.2.2 General Gaussian case

Given the general  $d$ -variate Gaussian target  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$  we denote by  $\hat{F}_{\tilde{\pi}}$  the exact solution of the Poisson equation and by  $G$  the approximation that we wish to construct. To approximate  $\hat{F}_{\tilde{\pi}}$  we apply a change of variables transformation from the standard normal, as motivated by the following proposition and remark.

**Proposition 3** *Suppose the standard normal target  $\tilde{\pi}_0(x) = \mathcal{N}(x|0, I_d)$ , the function  $F(x) = x^{(1)}$  and  $\hat{F}_{\tilde{\pi}_0}$  the associated solution of the Poisson equation for either RWM with proposal  $q(y|x) = \mathcal{N}(y|x, c^2I)$  or MALA with proposal  $q(y|x) = \mathcal{N}(y|(1 - c^2/2)x, c^2I)$ . Then, the solution  $\hat{F}_{\tilde{\pi}}$  for the general Gaussian target  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$  and Metropolis-Hastings proposal*

$$q(y|x) = \begin{cases} \mathcal{N}(y|x, c^2\Sigma) & \text{if RWM} \\ \mathcal{N}(y|x + (c^2/2)\Sigma \nabla \log \tilde{\pi}(x), c^2\Sigma) & \text{if MALA,} \end{cases} \tag{14}$$

is  $\hat{F}_{\tilde{\pi}}(x) = L_{11}\hat{F}_{\tilde{\pi}_0}(L^{-1}(x - \mu))$ , where  $L$  is a lower triangular Cholesky matrix such that  $\Sigma = LL^T$  and  $L_{11}$  is its first diagonal element.

**Remark 1** To apply Proposition 3 for  $F(x) = x^{(j)}$ ,  $j \neq 1$ , the vector  $x$  needs to be permuted such that  $x^{(j)}$  becomes its first element; the corresponding permutation has also to be applied to the mean  $\mu$  and covariance matrix  $\Sigma$ .

Proposition 3 implies that we can obtain the exact solution of the Poisson equation for any  $d$ -variate Gaussian target by applying a change of variables transformation to the solution

of the standard normal  $d$ -variate target. Therefore, based on this theoretical result we propose to obtain an approximation  $G$  of the Poisson equation in the general Gaussian case by simply transforming the approximation  $G_0$  in (12) from the standard normal case so as

$$G(x) = G_0(L^{-1}(x - \mu)). \tag{15}$$

The constant  $L_{11}$  is omitted since it can be absorbed by the regression coefficient  $\theta$ ; see Sect. 2.4. Note that Remark 1 provides guidelines for the solution of the Poisson equation associated with  $\tilde{\pi}$  for  $F(x) = x^{(j)}$  for any  $j = 2, \dots, d$ . However, if we need to perform variance reduction in the estimation of the means of all or a large subset of the marginal distributions of a high-dimensional target, a computationally more efficient method to conduct the desired variance reductions is available: see ‘‘Appendix D’’ for details.

### 2.3 Construction of the static control variate $h(x, y)$

Suppose we have constructed a Gaussian approximation  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$ , where  $\Sigma = LL^T$ , to the intractable target  $\pi(x)$  and also have obtained the function  $G$  from (15) needed for the proposed, general, estimator in (7). What remains is to specify the function  $h(x, y)$ , labelled as *static control variate* in (7), which should correlate well with  $\alpha(x, y)(G(y) - G(x))$ . The intractable term in this function is the Metropolis-Hastings probability  $\alpha(x, y)$  in (4) where the Metropolis-Hastings ratio  $r(x, y)$  contains the intractable target  $\pi$ . This suggests to choose  $h(x, y)$  as

$$h(x, y) = \min\{1, \tilde{r}(x, y)\}[G(y) - G(x)], \tag{16}$$

where  $\tilde{r}(x, y)$  is the acceptance ratio in a M-H algorithm that targets the Gaussian approximation  $\tilde{\pi}(x)$ , that is

$$\tilde{r}(x, y) = \min \left\{ 1, \frac{\tilde{\pi}(y)\tilde{q}(x|y)}{\tilde{\pi}(x)\tilde{q}(y|x)} \right\}, \tag{17}$$

and  $\tilde{q}(\cdot|\cdot)$  is the proposal distribution that we would use for the Gaussian target  $\tilde{\pi}(x)$  as defined by equation (14). Importantly, by assuming that  $\tilde{\pi}$  serves as an accurate approximation to  $\pi$ , the ratio  $\tilde{r}(x, y)$  approximates accurately the exact M-H ratio  $r(x, y)$  and  $E_q[h(x, y)]$  can be calculated analytically. In particular, using (15) we have that

$$\begin{aligned} E_q[h(x, y)] &= \int h(x, y)q(y|x)dy \\ &= \int \min\{1, \tilde{r}(x, y)\}[G_0(L^{-1}(y - \mu)) \\ &\quad - G_0(L^{-1}(x - \mu))]q(y|x)dy. \end{aligned}$$

This integral can be computed efficiently as follows. We reparametrize the integral according to the new variable  $\tilde{y} = L^{-1}(y - \mu)$  and also use the shortcut  $\tilde{x} = L^{-1}(x - \mu)$  where  $x$  is an MCMC sample. After this reparametrization, the above expectation becomes under the distribution

$$q(\tilde{y}|\tilde{x}) = \begin{cases} \mathcal{N}(\tilde{y}|\tilde{x}, c^2I) & \text{if RWM} \\ \mathcal{N}(\tilde{y}|\tilde{x} + \frac{c^2}{2}L^T \nabla \log \pi(x), c^2I) & \text{if MALA,} \end{cases} \tag{18}$$

where we condition on  $\tilde{x}$  with a slightly abuse of notation since the term  $\nabla \log \pi(x)$  is the exact pre-computed gradient for the sample  $x$  of the intractable target. Thus, the calculation of  $E_q[h(x, y)]$  reduces to the evaluation of the following integral

$$\int \min \left\{ 1, \exp\left\{-\frac{1}{2}(\tilde{y}^T \tilde{y} - \tilde{x}^T \tilde{x})\right\} \frac{\tilde{q}(\tilde{x}|\tilde{y})}{\tilde{q}(\tilde{y}|\tilde{x})} \right\} [G_0(\tilde{y}) - G_0(\tilde{x})]q(\tilde{y}|\tilde{x})d\tilde{y}. \tag{19}$$

Note also that inside the Metropolis-Hastings ratio  $\tilde{q}(\tilde{y}|\tilde{x}) = \mathcal{N}(\tilde{y}|r\tilde{x}, c^2I)$  with  $r$  as in (10). In the case of RWM and by noting that the density  $q(\tilde{y}|\tilde{x})$  in (18) coincides with the density  $\tilde{q}(\tilde{y}|\tilde{x})$  in (10) we have that the calculation of the integral in (19) reduces to the calculation of the integrals in (8) and (9) and, thus, can be conducted by utilizing Proposition 1. The calculation of the integral in (19) for the MALA is slightly different as highlighted by the following remark.

**Remark 2** In the case of MALA the mean of the density  $q(\tilde{y}|\tilde{x})$  in (18) is different from the mean of  $\tilde{q}(\tilde{y}|\tilde{x})$  due to the presence of the term  $\frac{c^2}{2}L^T \nabla \log \pi(x)$  and the formulas in Proposition (1) are modified accordingly.

Finally, we note that except from the tractability in the calculations which is offered by the particular choice of  $h(x, y)$ , there is also the following intuition for its effectiveness. If the Gaussian approximation is exact, then the overall control variate, defined in equation (7) as the sum of a stochastic and a static control variate, becomes the exact ‘‘Poisson control variate’’ that we would compute if the initial target was actually Gaussian. Thus, we expect that the function  $h(x, y)$ , as a static control variate in a non-Gaussian target, enables effective variance reduction under the assumption that the target is well-approximated by a Gaussian distribution.

### 2.4 The modified estimator with regression coefficients

As pointed out by Dellaportas and Kontoyiannis (2012) the fact that the proposed estimator  $\mu_{n,G}(F)$  is based on an approximation  $G$  of the true solution  $\hat{F}_\pi$  of the Poisson equa-

tion implies that we need to modify  $\mu_{n,G}(F)$  as

$$\mu_{n,G}(F_{\hat{\theta}_n}) := \frac{1}{n} \sum_{i=0}^{n-1} \left\{ F(x_i) + \hat{\theta}_n \left\{ \underbrace{\alpha(x_i, y_i)(G(y_i) - G(x_i))}_{\text{Stochastic Poisson control variate}} + \underbrace{h(x_i, y_i) - E_{q(y|x_i)}[h(x_i, y)]}_{\text{Static control variate}} \right\} \right\} \tag{20}$$

where  $\hat{\theta}_n$  estimates the optimal coefficient  $\theta$  that further minimizes the variance of the overall estimator. Dellaportas and Kontoyiannis (2012) show that for reversible MCMC samplers, the optimal estimator  $\hat{\theta}_n$  of the true coefficient  $\theta$  can be constructed solely from the MCMC output. By re-writing the estimator in (20) as

$$\mu_{n,G}(F_{\hat{\theta}_n}) := \frac{1}{n} \sum_{i=0}^{n-1} \{F(x_i) - \hat{\theta}_n \{G(x_i) - \widehat{PG}(x_i)\}\},$$

where the term

$$\widehat{PG}(x_i) = G(x_i) + a(x_i, y_i)(G(y_i) - G(x_i)) + h(x_i, y_i) - E_{q(y|x_i)}[h(x_i, y)], \tag{21}$$

approximates  $PG(x_i)$ , we can estimate  $\hat{\theta}_n$  as

$$\hat{\theta}_n = \frac{\mu_n(F(G + \widehat{PG})) - \mu_n(F)\mu_n(G + \widehat{PG})}{\frac{1}{n} \sum_{i=1}^{n-1} (G(x_i) - \widehat{PG}(x_{i-1}))^2}. \tag{22}$$

Notice that, as in the case of control variates for Monte Carlo integration (Glasserman 2004), the denominator in (22) is the standard empirical estimator of the variance of the control variate  $G(x_i) - \widehat{PG}(x_i)$ . However, in contrast to the standard Monte Carlo case, the numerator is not the usual estimator of the covariance between the function  $F$  and the control variate since this latter covariance, in the case of a Markov Chain, is non-tractable. Therefore, the numerator in (22) has been constructed by Dellaportas and Kontoyiannis (2012) based on an alternative, tractable, form of the stationary covariance between the function  $F$  and the control variate  $G(x_i) - \widehat{PG}(x_i)$ .

The resulting estimator  $\mu_{n,G}(F_{\hat{\theta}_n})$  in (20) is evaluated by using solely the output of the MCMC algorithm and under some regularity conditions converges to  $\pi(F)$  a.s. as  $n \rightarrow \infty$ , see Tsourti (2012).

### 2.5 Algorithmic summary

In summary, the proposed variance reduction approach can be applied a posteriori to the MCMC output samples  $\{x_i\}_{i=0}^{n-1}$  obtained from either RWM or MALA with proposal density given by (14). The extra computations needed involve

the evaluation of  $\widehat{PG}(x_i)$  given by (21). This is efficient since it relies on quantities that are readily available such as the values  $G(x_i)$  and  $G(y_i)$ , where  $y_i$  is the value generated from the proposal  $q(y|x_i)$  during the main MCMC algorithm, as well as on the acceptance probability  $a(x_i, y_i)$  which has been also computed and stored at each MCMC iteration. The evaluation of  $\widehat{PG}(x_i)$  requires also the construction of the static control variate  $h(x_i, y_i)$  defined by (16). This depends on the ratio  $\tilde{r}(x, y)$  given by (17) and on the expectation  $E_{q(y|x_i)}[h(x_i, y)]$ . The calculation of the latter expectation is tractable since  $\tilde{r}(x, y)$  is the acceptance ratio of Metropolis-Hastings algorithm that targets the Gaussian target  $\tilde{\pi}(x) = N(x|\mu, \Sigma)$ , where  $\mu$  and  $\Sigma$  are estimators of the mean and covariance matrix respectively of the target  $\pi(x)$ ; see Assumption 1. It is important to note that the calculation of the covariance matrix  $\Sigma$  as well as its Cholesky factor  $L$  do not increase the computational cost of the proposed variance reduction technique since they are calculated during the main MCMC algorithm. Finally, we compute  $\hat{\theta}_n$  using (22) and evaluate the proposed estimator  $\mu_{n,G}(F_{\hat{\theta}_n})$  from (20). Algorithm 1 summarizes the steps of the variance reduction procedure.

---

#### Algorithm 1 Variance reduction for Metropolis-Hasting samplers

---

**Inputs:** The samples  $x_i, i = 0, \dots, n - 1$ , simulated by using RWM or MALA with proposal distribution given by (14); the proposed samples  $y_i$  generated from the proposal during the MCMC; the M-H probabilities  $\alpha(x_i, y_i)$  calculated during the MCMC; estimators  $\mu$  and  $\Sigma$  of the mean and covariance matrix respectively of the target.

**Returns:** An estimate for the mean of the  $j$ th coordinate of the target.

- 1: Set  $F(x) = x^{(j)}$ .
  - 2: Calculate  $h(x_i, y_i)$  given by (16).
  - 3: Calculate  $E_{q(y|x_i)}[h(x_i, y)]$  by utilising Propositions 1 and 3.
  - 4: Calculate  $\widehat{PG}(x_i)$  given by (21) for each  $i = 1, \dots, n$ .
  - 5: Calculate  $\hat{\theta}_n$  given by (22).
  - 6: Return  $\mu_{n,G}(F_{\hat{\theta}_n})$  given by (20).
- 

### 3 Application on real and simulated data

We present results from the application of the proposed methodology on real and simulated data examples. First we consider multivariate Gaussian targets for which we have shown that the function  $G$  in (12) allows the explicit calculation of the expectation  $PG$  defined by (3). Section 3.1 presents variance reduction factors in the case of  $d$ -variate standard Gaussian densities, simulated by employing the RWM and MALA, up to  $d = 100$  dimensions. In Sects. 3.2, 3.3 and 3.4 we examine the efficiency of our proposed methodology in targets that depart from the Gaussian distribution and the expectation  $PG$  is not analytically available. Assumption 1 and Algorithm 1 suggest that our proposed



methodology depends on estimators  $\mu$  and  $\Sigma$  of the mean and covariance matrix of the target distribution respectively. Since the technique that we developed is a post-processing procedure which takes as input samples drawn by either the MALA or the RWM algorithm, we utilise these samples to estimate  $\mu$ , the variance of which we aim to reduce without spending more computational resources than those used to run the MCMC algorithm. The choice of  $\Sigma$  is implied by the construction of the proposed methodology and it is required to be the covariance matrix of the proposal distribution of the MCMC algorithm.

To conduct all the experiments we set the parameters  $b_0, b_1, b_2, c_0, c_1$  and  $c_2$  of the function  $G_0$  in (12) in the values given by Table 1 which were estimated by minimizing the loss function in (13) for  $d = 2$ . In practice we observe that such values lead to good performance across all real data experiments, including those with  $d > 2$ .

To estimate the variance of  $\mu_n(F)$  in each experiment we obtained  $T = 100$  different estimates  $\mu_n^{(i)}(F), i = 1, \dots, T$ , for  $\mu_n(F)$  based on  $T$  independent MCMC runs. Then, the variance of  $\mu_n(F)$  has been estimated by

$$\frac{1}{T-1} \sum_{i=1}^T \{\mu_n^{(i)}(F) - \bar{\mu}_n(F)\}^2,$$

where  $\bar{\mu}_n(F)$  is the average of  $\mu_n^{(i)}(F)$ . We estimated similarly the variance of the proposed estimator  $\mu_{n,G}(F)$ .

### 3.1 Simulated data: Gaussian targets

The target distribution is a  $d$ -variate standard Gaussian distribution and we are interested in estimating the expected value of the first coordinate of the target by setting  $F(x) = x^{(1)}$ . Samples of size  $n$  were drawn from target densities by utilising the proposal distribution in (10) with  $c^2 = 2.38^2/d$  for the RWM case and by tuning  $c^2$  during the burn-in period to achieve acceptance rate between 55% and 60% in the MALA

case; we initiated all the MCMC algorithms by drawing an initial parameter vector from the stationary distribution. Table 2 presents factors by which the variance of  $\mu_n(F)$  is greater than the variance of  $\mu_{n,G}(F)$  in the case of the RWM and MALA. Variance reduction is considerable even for  $d = 100$ . Figure 2 shows typical realizations of the sequences of estimates obtained by the standard estimators  $\mu_n(F)$  and the proposed  $\mu_{n,G}(F_\theta)$  for different dimensions of the standard Gaussian target and Fig. 3 provides a visualization of the distribution of the estimators  $\mu_n(F)$  and  $\mu_{n,G}(F_\theta)$ . Note that in these experiments the covariance matrix of the target is assumed known; In the ‘‘Appendix’’ we repeat these experiments by relaxing this assumption.

### 3.2 Simulated data: mixtures of Gaussian distributions

It is important to investigate how our proposed methodology performs when the target density departs from normality. We used as  $\pi(x)$  a mixture of  $d$ -variate Gaussian distributions with density

$$\pi(x) = \frac{1}{2} \mathcal{N}(x|m, \Sigma) + \frac{1}{2} \mathcal{N}(x|-m, \Sigma), \tag{23}$$

where, following Mijatović and Vogrinc (2019), we set  $m$  to be the  $d$ -dimensional vector  $(h/2, 0, \dots, 0)$  and  $\Sigma$  is  $d \times d$  covariance matrix randomly drawn from an inverse Wishart distribution by requiring its largest eigenvalue to be equal to 25. More precisely, we simulated one matrix  $\Sigma$  for each different value of  $d$  and we used the same matrix across the different choices for  $h$ .

We drew samples from the target distribution by using the Metropolis-Hastings algorithm with proposal distribution  $q(y|x) = \mathcal{N}(y|x, c^2\Sigma)$  where by setting  $c^2 = 2.38^2/d$  we achieve an acceptance ratio between 23% and 33%. We also note that the covariance matrix  $\Sigma$  of the target was fixed across the  $T$  independent MCMC runs used to estimate the

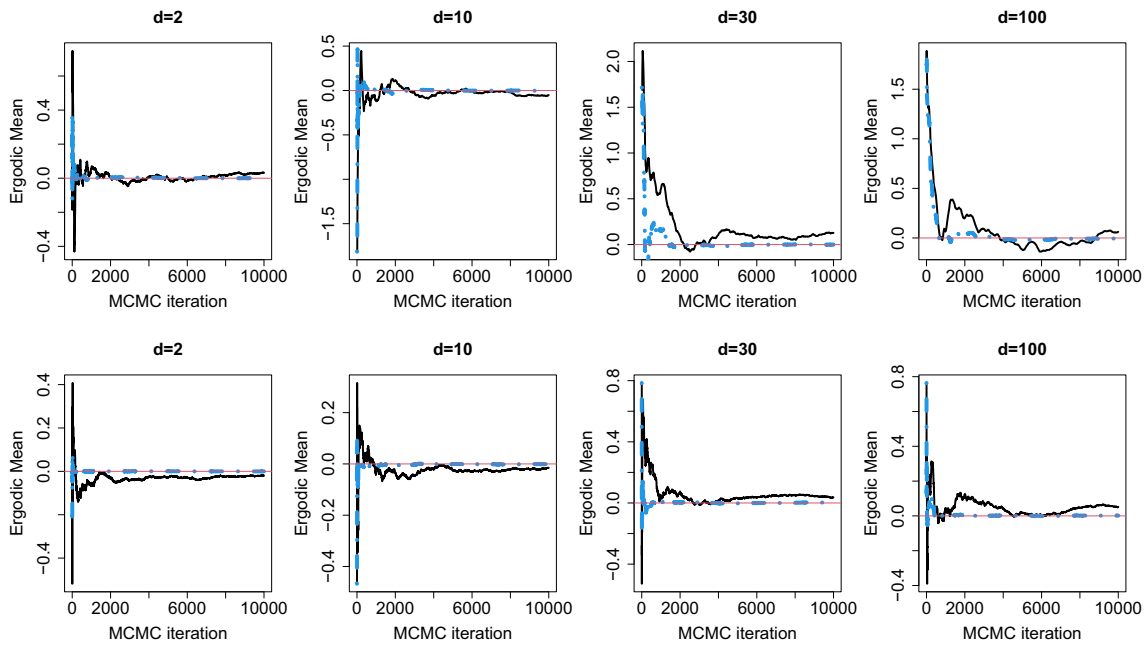
**Table 1** Optimal values for the parameters of the function  $G_0$  in (12)

	$b_0$	$b_1$	$b_2$	$c_0$	$c_1$	$c_2$
RWM	8.7078	0.2916	0.0001	-3.5619	0.1131	3.9162
MALA	7.6639	0.0613	0.0096	-14.8086	0.3431	-0.0647

**Table 2** Estimated factors by which the variance of  $\mu_n(F)$  is larger than the variance of  $\mu_{n,G}(F)$  for standard Gaussian  $d$ -variate target

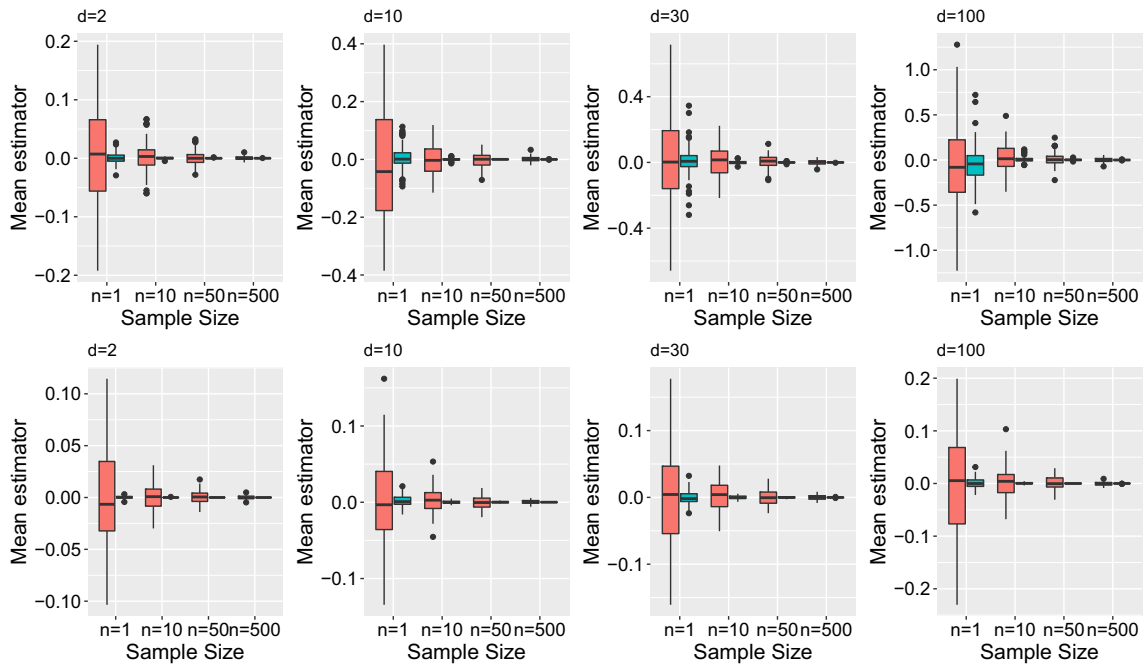
	RWM				MALA			
	d = 2	d = 10	d = 30	d = 100	d = 2	d = 10	d = 30	d = 100
n = 1000	93	26	10	5	1345	64	57	97
n = 10,000	278	173	112	27	3572	81	88	316
n = 50,000	541	445	177	94	4628	92	103	274
n = 500,000	531	820	370	263	4997	83	157	286

We collect  $n$  samples after the first 10,000 iterations of the RWM and the MALA



**Fig. 2** Sequence of the standard ergodic averages (black solid lines) and the proposed estimates (blue dashed lines). The red lines indicate the mean of the  $d$ -variate standard Gaussian target. The values are based on

samples drawn by employing either the RWM (top row) or the MALA (bottom row) with 10,000 iterations discarded as burn-in period. (Color figure online)



**Fig. 3** Each pair of boxplots is consisted of 100 values for the estimators  $\mu_n(F)$  (left boxplot) and  $\mu_{n,G}(F_\theta)$  (right boxplot) for the  $d$ -variate standard Gaussian target. The estimators have been calculated by using

$n \times 10^3$  samples drawn by employing either the RWM (top row) or the MALA (bottom row) and discarded the first 10,000 samples as burn-in period

**Table 3** Estimated factors by which the variance of  $\mu_n(F)$  is larger than the variance of  $\mu_{n,G}(F)$  for a mixture of  $d$ -variate Gaussian distributions with density given by (23) for different values of the mean  $m$

	h = 2	h = 4	h = 6
d = 10	20.73	2.39	1.26
d = 50	7.88	1.35	1.01

We collect  $n = 200,000$  samples after the first 10,000 iterations of the RWM algorithm

variance of  $\mu_n(F)$  and  $\mu_{n,G}(F_\theta)$ . When  $h > 6$  the MCMC algorithm struggles to converge. Table 3 presents the factors by which the variance of  $\mu_n(F)$  is greater than the variance of the modified estimator  $\mu_{n,G}(F)$  for dimensions  $d = 10$  and  $d = 50$  and for different values of  $h$ . It is very reassuring that even in the very non-Gaussian scenario ( $h = 6$ ) our modified estimator achieved a slight variance reduction.

### 3.3 Real data: Bayesian logistic regressions

We tested the variance reduction of our modified estimators on five datasets that have been commonly used in MCMC applications, see e.g. Girolami and Calderhead (2011), Titsias and Dellaportas (2019). They are consisted of one  $N$ -dimensional binary response variable and an  $N \times d$  matrix with covariates including a column of ones; see Table 4 for the names of the datasets and details on the specific samples sizes and dimensions. We consider a Bayesian logistic regression model by setting an improper prior for the regression coefficients  $\gamma \in \mathbb{R}^d$  of the form  $p(\gamma) \propto 1$ .

#### 3.3.1 Variance reduction for RWM

We draw samples from the posterior distribution of  $\gamma$  by employing the Metropolis-Hastings algorithm with proposal distribution

$$q(\gamma'|\gamma) = \mathcal{N}(\gamma'|\gamma, c^2\hat{\Sigma}),$$

where  $c^2 = 2.38^2/d$  and  $\hat{\Sigma}$  is the maximum likelihood estimator of the covariance of  $\gamma$ . Table 5 presents the range of factors by which the variance of  $\mu_n(F)$  is greater than the variance of  $\mu_{n,G}(F)$  for all parameters  $\gamma$ . It is clear that our modified estimators achieve impressive variance reductions when compared with the standard RWM ergodic estimators.

**Table 4** Summary of datasets for logistic regression

Dataset	d	N
Ripley	3	250
Pima Indian	8	532
Heart	14	270
Australian	15	690
German	25	1000

#### 3.3.2 Variance reduction for MALA

We draw samples from the posterior distribution of  $\gamma$  by employing the Metropolis-Hastings algorithm with proposal distribution

$$q(\gamma'|\gamma) = \mathcal{N}(\gamma'|\gamma + \frac{1}{2}c^2\hat{\Sigma}\nabla \log \pi(\gamma), c^2\hat{\Sigma}),$$

where  $c^2$  is tuned during the burn-in period in order to achieve an acceptance ratio between 55% and 60%,  $\hat{\Sigma}$  is maximum likelihood estimator of the covariance of  $\gamma$  and  $\pi(\gamma)$  denotes the density of the posterior distribution of  $\gamma$ . Table 6 presents the range of factors by which the variance of  $\mu_n(F)$  is greater than the variance of  $\mu_{n,G}(F)$  for all parameters  $\gamma$ . Again, there is considerable variance reduction for all modified estimators.

#### 3.4 Simulated data: a stochastic volatility model

We use simulated data from a standard stochastic volatility model often employed in econometric applications to model the evolution of asset prices over time (Kim et al. 1998; Kastner and Frühwirth-Schnatter 2014). By denoting with  $r_t, t = 1, \dots, N$ , the  $t$ th observation (usually log-return of an asset) the model assumes that  $r_t = \exp\{h_t/2\}\epsilon_t$ , where  $\epsilon_t \sim N(0, 1)$  and  $h_t$  is an autoregressive AR(1) log-volatility, process:  $h_t = m + \phi(h_{t-1} - m) + s\eta_t, \eta_t \sim N(0, 1)$  and  $h_0 \sim N(m, s^2/(1 - \phi^2))$ . To conduct Bayesian inference for the parameters  $m \in \mathbb{R}, \phi \in (-1, 1)$  and  $s^2 \in (0, \infty)$  we specify commonly used prior distributions (Kastner and Frühwirth-Schnatter 2014; Alexopoulos et al. 2021):  $m \sim N(0, 10), (\phi + 1)/2 \sim Beta(20, 1/5)$  and  $s^2 \sim Gam(1/2, 1/2)$ . The posterior of interest is

$$\begin{aligned} \pi(m, \phi, s^2, h) &= p(m, \phi, s^2, h|r) \\ &\propto p(m)p(s^2)p(\phi)\mathcal{N}(h_0|m, s^2/(1 - \phi^2)) \\ &\quad \times \prod_{t=1}^N \mathcal{N}(r_t|0, e^{h_t})\mathcal{N}(h_t|m + \phi(h_{t-1} - m), s^2), \end{aligned} \tag{24}$$

where  $h = (h_0, \dots, h_N)$  and  $r = (r_1, \dots, r_N)$ .

To assess the proposed variance reduction methods we simulated daily log-returns of a stock for  $d$  days by using values for the parameters of the model that have been previously estimated in real data applications (Kim et al. 1998; Alexopoulos et al. 2021)  $\phi = 0.98, \mu = -0.85$  and  $s = 0.15$ . To draw samples from the  $d$ -dimensional,  $d = N + 3$ , target posterior in (24) we first transform the parameters  $\phi$  and  $s^2$  to real-valued parameters  $\tilde{\phi}$  and  $\tilde{s}^2$  by taking the logit and logarithm transformations and we assign Gaussian prior distributions by matching the first two moments of the Gaussian distributions with the corresponding moments of the beta and gamma distributions used as priors for the parameters of the original formulation. Then, we set  $x = (m, \tilde{\phi}, \tilde{s}^2, h)$  and we

**Table 5** Range of estimated factors by which the variance of  $\mu_n(F)$  is larger than the variance of  $\mu_{n,G}(F_\theta)$  for the posterior distribution of logistic regression models applied on the datasets indicated by the first column

Dataset	$n = 1000$	$n = 10,000$	$n = 50,000$	$n = 200,000$
Ripley	27.07–34.06	26.89–91.96	34.42–105.35	34.48–137.68
Pima Indian	14.62–25.91	84.16–137.35	99.38–218.06	99.16–241.38
Heart	8.26–13.70	16.63–40.81	23.53–64.07	18.68–76.62
Australian	6.14–15.27	25.91–80.65	33.43–89.56	23.43–92.76
German	4.72–10.20	19.61–54.63	33.60–119.73	25.61–148.54

We collect  $n$  samples after the first 10,000 iterations of the RWM algorithm

**Table 6** Estimated factors by which the variance of  $\mu_n(F)$  is larger than the variance of  $\mu_{n,G}(F_\theta)$  for the posterior distribution of logistic regression models applied on the datasets indicated by the first column

Dataset	$n = 1000$	$n = 10,000$	$n = 50,000$	$n = 200,000$
Ripley	10.89–15.99	14.83–24.76	12.13–26.06	10.38–20.14
Pima Indian	23.50–51.64	34.95–52.42	34.78–73.51	36.64–72.75
Heart	10.04–17.31	7.74–18.36	10.11–18.07	11.83–20.51
Australian	9.32–22.78	8.56–22.92	6.86–18.93	7.45–21.74
German	11.79–33.29	11.39–42.46	9.80–52.24	8.21–40.72

We collect  $n$  samples after the first 10,000 iterations of the MALA

**Table 7** Estimated factors by which the variance of  $\mu_n(F)$  is larger than the variance of  $\mu_{n,G}(F)$  for the parameters of  $d$ -dimensional stochastic volatility model

	$n = 10,000$				$n = 50,000$				$n = 200,000$			
	$h$	$m$	$\phi$	$s^2$	$h$	$m$	$\phi$	$s^2$	$h$	$m$	$\phi$	$s^2$
$d = 50$	7.18–15.24	14.07	17.44	4.36	7.49–15.92	13.02	19.27	2.86	7.46–16.36	16.97	14.16	2.43
$d = 100$	1.06–7.66	7.99	3.46	1.26	1.09–7.02	6.50	7.80	1.69	1.52–9.54	5.40	4.44	1.07

We collect  $n$  samples after the first 10,000 of the MALA

draw the desired samples using a Metropolis-Hastings algorithm with proposal distribution

$$q(y|x) = \mathcal{N}(y|x + \frac{c^2}{2} \hat{\Sigma} \nabla \log \pi(x), c^2 \hat{\Sigma}),$$

where  $y = (m', \tilde{\phi}', \tilde{s}^2', h')$  are the proposed values,  $c^2$  is tuned during the burn-in period in order to achieve an acceptance ratio between 55% and 60% and  $\hat{\Sigma}$  is the maximum a posteriori estimate of the covariance matrix of  $(m, \phi, s^2, h)$ . Table 7 presents the factors by which the variance of  $\mu_n(F)$  is greater than the variance of the proposed estimator  $\mu_{n,G}(F_\theta)$ . We report variance reduction for all static parameters of the volatility process and the range of reductions achieved for the  $N$ -dimensional latent path  $h$ . All estimators have achieved considerable variance reduction.

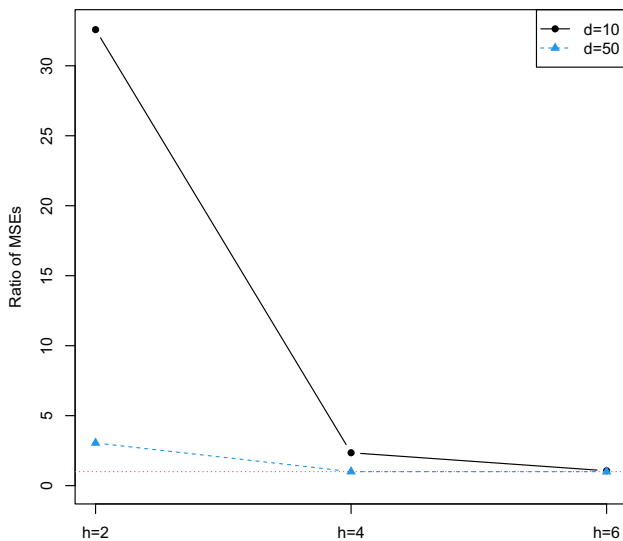
### 3.5 Comparison with alternative methods

We compare the proposed variance reduction methodology with the zero variance (ZV) estimators considered, among others, by Mira et al. (2013) and South et al. (2018). We consider the first order ZV control variates as a competitive variance reduction method since their computational cost is less than all the other ZV estimators and, thus, comparable

with the negligible computational cost of our methodology. The comparison that we perform is twofold. First, we compare the computational complexity of our proposed techniques with the one of the first order ZV estimators and then we present the mean squared error (MSE) in the estimation of the mean  $\pi(F)$  obtained by the two different approaches.

We compare the computational efficiency of the two methods by assuming that it is determined by their computational complexity based on target (and its derivatives) evaluations. First note that our proposed technique relies on the following three ingredients: (i) a Monte Carlo integration with sample size one, (ii) the computation of the cdf of the non-central chi-squared distribution and (iii) the calculation of the coefficient  $\hat{\theta}_n$  in Eq. (22). All these steps do not require any extra target and/or gradient evaluations rather than the pre-computed evaluations required by the RWM algorithm. The computation of the transformation  $L^{-1}(x - \mu)$  can be also achieved by an efficient post-processing manner without extra target evaluations, see ‘‘Appendix F’’ for details.

On the other hand, although the first order ZV control variates do not depend on extra evaluations of the target, they require the extra (in the case of the RWM algorithm) evaluation of its gradients. Furthermore, the first order ZV methods are based on polynomials in which the number of



**Fig. 4** MSEs of the estimator  $\mu_{n,G}(F)$  over MSEs of the first order ZV estimator for a mixture of  $d$ -variate Gaussian distributions with density given by Eq. (23) for different values of the mean  $m$  indicated by the choice of the parameter  $h$  in the  $x$ -axis. The red dotted line indicates the value 1 in the  $y$ -axis. (Color figure online)

terms increases with the dimension of the target and thus the inversion of a  $d \times d$  matrix is required for each sample drawn from the target distribution.

We additionally compare the two methods in terms of mean squared error (MSE) when estimating  $\pi(F)$ . In particular, by using the MCMC output we calculate, for each one of the examples in Sects. 3.2–3.4, the MSEs of the proposed estimator  $\mu_{n,G}(F)$  and of the first order ZV estimator; we employed the R-package developed by South (2021) to conduct the calculation of the first order ZV estimators.

Figure 4 displays the ratio of the MSE of the proposed estimator  $\mu_{n,G}(F)$  over the MSE of the first order ZV estimator

for the mixture of  $d$ -variate Gaussian distributions with density given by Eq. (23). It indicates that the proposed variance methodology is more robust with respect to non-Gaussian targets as well as to the dimension of the target. In particular, the ratio of the MSEs is getting closer to one as the dimension  $d$  and/or the parameter  $h$  of the target increase. Notice that given the difference of the computational complexity, a ratio of MSEs less or equal than one implies better overall efficiency for the proposed estimator  $\mu_{n,G}(F)$ . We also note that the closer the target to the Gaussian distribution (small  $h$ ), the lower MSE for the ZV estimator compared to our proposed estimator. This is a consequence of the fact that the ZV estimators are exact in the case of Gaussian targets. Tables 8, 9, 10 present ratios of the MSEs of the estimators that we compare in the case of the logistic regression and stochastic volatility models described in Sects. 3.3 and 3.4 respectively.

The combination of the ratios displayed by the Tables together with the computational complexity analysis of the competitive variance reductions methods provides evidence of the overall advantage of our variance reduction techniques over the first order ZV estimators for the estimation of the mean  $\pi(F)$ .

### 4 Discussion

Typical variance reduction strategies for MCMC algorithms study ways to produce new estimators which have smaller variance than the standard ergodic averages by performing a post-processing manipulation of the drawn samples. Here we studied a methodology that constructs such estimators but our development was based on the essential requirement of a negligible post-processing cost. In turn, this feature allows the effortless variance reduction for MCMC estimators that

**Table 8** Range of MSEs of the estimator  $\mu_{n,G}(F)$  over MSEs of the first order ZV estimator in logistic regression models applied to the datasets indicated by the first column

Dataset	$n = 1000$	$n = 10,000$	$n = 50,000$	$n = 200,000$
Ripley	0.86–2.21	0.59–0.93	0.70–0.78	0.84–1.02
Pima Indian	2.41–27.09	0.81–4.59	1.02–2.96	0.98–1.27
Heart	2.43–9.79	1.20–2.84	1.00–2.04	1.04–1.91
Australian	2.03–17.78	0.99–6.50	0.82–3.81	0.98–3.18
German	2.29–75.40	0.93–20.60	0.94–9.53	0.93–5.33

The MSEs are based on  $n$  samples collected after the first 10,000 iterations of the RMW algorithm

**Table 9** Range of MSEs of the estimator  $\mu_{n,G}(F)$  over MSEs of the first order ZV estimator in logistic, where the MSEs are based on  $n$  samples collected after the first 10,000 iterations of MALA

Dataset	$n = 1000$	$n = 10,000$	$n = 50,000$	$n = 200,000$
Ripley	1.99–4.46	1.58–2.70	0.84–7.83	0.48–12.22
Pima Indian	2.11–8.85	1.20–6.40	0.86–3.30	0.82–1.58
Heart	2.33–5.99	2.03–8.09	0.53–5.76	0.89–4.42
Australian	2.83–16.56	1.71–12.21	0.73–12.18	0.57–10.21
German	1.80–29.48	1.31–28.20	1.09–13.54	0.96–7.29

**Table 10** Range of MSEs of the estimator  $\mu_{n,G}(F)$  over MSEs of the first order ZV estimator for the parameters of  $d$ -dimensional stochastic volatility model

	$n = 10,000$				$n = 50,000$				$n = 200,000$			
	$h$	$m$	$\phi$	$s^2$	$h$	$m$	$\phi$	$s^2$	$h$	$m$	$\phi$	$s^2$
$d=50$	3.15–14.66	54.95	<0.01	1.44	1.08–11.17	12.92	<0.01	1.42	0.45–17.09	3.37	<0.01	1.37
$d=100$	0.02–0.79	0.04	0.01	<0.01	0.01–0.23	0.01	<0.01	<0.01	<0.01–0.06	0.01	<0.01	<0.01

We collect  $n$  samples after the first 10,000 iterations of MALA

are used in a wide spectrum of Bayesian inference applications. We investigated both the applicability of our strategy in high dimensions and the robustness to departures of normality in the target densities by using simulated and real data examples.

There are many directions for future work. We limited ourselves to the simplest cases of linear functions such as  $F(x) = x^{(j)}$ , but higher moments and indicator functions seem interesting avenues to be investigated next. The extension of the proposed method for other functions  $F$  requires the construction of an approximation of the solution  $\hat{F}_{\pi_0}$  of the Poisson equation associated with a standard Gaussian target, i.e, a function which will play the role of the function  $G$  defined by Eq. (15). Importantly, this function should be chosen such that the integral in Eq. (9) can be calculated analytically. We think that the form of  $G$  used in the present paper can serve as a starting point for this direction of research.

The developed variance reduction technique can be applied on any output from the MALA and the RWM algorithm. Other Metropolis samplers such as the independent Metropolis or the Metropolis-within-Gibbs are also obvious candidates for future work. Finally, an issue that was discussed in some detail in Dellaportas and Kontoyiannis (2009) but has not yet studied with the care it deserves is the important problem of reducing the estimation bias of the MCMC samplers which depends on the initial point of the chain  $X_0 = x$  and vanishes asymptotically. As also noted by Dellaportas and Kontoyiannis (2009), control variables have probably an important role to play in this setting.

*Supplementary information* The R code for reproducing the experiments is available at <https://gitlab.com/aggelisalexopoulos/variance-reduction>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copy-

right holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Proof of Proposition 1

**Proof** We need to calculate the integrals  $a(x)$  and  $a_g(x)$  in Eq. (8) and (9) for  $G_0(x)$  given by (12). We have that for  $q(y|x)$  given by (10)

$$\exp\left\{-\frac{1}{2}(y^\top y - x^\top x)\right\} \frac{q(x|y)}{q(y|x)} = \exp\left\{-\frac{\tau^2}{2}(y^\top y - x^\top x)\right\},$$

where  $\tau^2 = 1$  in the case of RWM and  $\tau^2 = c^2/4$  in the case of MALA.

To compute  $a(x)$  we set  $z = (y - rx)/c$ , where  $r$  as in (10). Then, we have that

$$\begin{aligned} & \exp\left\{-\frac{\tau^2}{2}(y^\top y - x^\top x)\right\} \\ &= \exp\left\{-\frac{\tau^2 c^2}{2}[(z + \kappa)^\top(z + \kappa) - x^\top x/c^2]\right\}, \end{aligned} \quad (25)$$

where  $\kappa = rx/c$ . By setting  $f = (z + \kappa)^\top(z + \kappa)$  we have that  $f$  follows the non-central chi-squared distribution with  $d$  degrees of freedom and non-central parameter  $r^2 x^\top x/c^2$ . Eq. (25) implies that  $\alpha(x)$  in (8) becomes

$$\begin{aligned} a(x) &= \int \min\{1, \exp\{-\frac{\tau^2 c^2}{2}[f - x^\top x/c^2]\}\} p(f) df \\ &= E_f\left[\min\{1, e^{-\frac{c^2 \tau^2}{2}(f - x^\top x/c^2)}\}\right] \\ &= \int_{-\infty}^{\frac{x^\top x}{c^2}} e^{-\frac{c^2 \tau^2}{2}(f - x^\top x/c^2)} p(f) df + \int_{\frac{x^\top x}{c^2}}^{\infty} p(f) df \end{aligned} \quad (26)$$

where  $p(f)$  is the density of the random variable  $f$  and writes

$$p(f) = \sum_{j=0}^{\infty} \text{Pois}(j|\kappa/2) \text{Gam}(f; d/2 + j, 2).$$

Notice that the second term in (26) can be calculated by using the cdf of the non-central chi squared distribution. For

the first term after some algebra we have that

$$e^{-\frac{c^2\tau^2}{2}(f-x^\top x/c^2)} p(f) = \sum_{j=0}^\infty \frac{\text{Pois}(j|\kappa/2)}{(c^2\tau^2 + 1)^{d/2+j}} \text{Gam}(f; d/2 + j, \frac{2}{c^2\tau^2+1}). \tag{27}$$

To compute  $a_g(x)$  we first note that

$$a_g(x) = \sum_{k=1}^K w_k a_{g_k}(x), \tag{28}$$

where

$$a_{g_k}(x) = \int \min \left\{ 1, \exp \left\{ -\frac{1}{2}(y^\top y - x^\top x) \right\} \frac{q(x|y)}{q(y|x)} \right\} g_k(y) q(y|x) dy,$$

and  $g_k(x) = \exp\{\beta_k^\top x - \gamma_k(x - \delta_k)^\top(x - \delta_k)\}$ .

Then, we calculate the  $a_{g_k}(x)$  by noting that

$$g_k(y) q(y|x) = A_k(x) \mathcal{N}(y|m_k(x), s_k^2),$$

where

$$A_k(x) = (1 + 2c^2\gamma_k)^{-d/2} \exp \left\{ -\frac{r^2 x^\top x}{2c^2} - \gamma_k \delta_k^\top \delta_k + \frac{m_k(x)^\top m_k(x)}{2c^2(1 + 2\gamma_k c^2)} \right\},$$

$$m_k(x) = \frac{rx + c^2(\beta_k + \gamma_k \delta_k)}{1 + 2c^2\gamma_k} \text{ and } s_k^2 = c^2/(1 + 2c^2\gamma_k).$$

By setting  $z_{k,g} = (x - m_k(x))/s$  and  $f_{k,g} = (z_{k,g} + \zeta_{k,g})^\top(z_{k,g} + \zeta_{k,g})$ , where  $\zeta_{k,g} = m_k(x)/s_k$ , we work as in (25) and have that

$$a_{g_k}(x) = A_k(x) \int \min \left\{ 1, \exp \left\{ -\frac{\tau^2 s_k^2}{2} [f_{k,g} - x^\top x/s_k^2] \right\} p(f_{k,g}) df_{k,g} \right\} = A_k(x) E_{f_{k,g}} \left[ \min \left\{ 1, \exp \left\{ -\frac{\tau^2 s_k^2}{2} (f_{k,g} - x^\top x/s_k^2) \right\} \right\} \right],$$

where the random variable  $f_{k,g}$  follows the chi-squared distribution with  $d$  degrees of freedom and non-central parameters  $m_k(x)^\top m_k(x)/s_k^2$  and the expectation is calculated by utilizing the cdf of  $f_{k,g}$  as in Eq. (26)-(27). Finally, from Eq. (28) we have that

$$a_g(x) = \sum_{k=1}^K A_k(x) E_{f_{k,g}} \left[ \min \left\{ 1, \exp \left\{ -\frac{\tau^2 s_k^2}{2} (f_{k,g} - x^\top x/s_k^2) \right\} \right\} \right]. \tag{29}$$

□

## B Proof of Proposition 2

**Proof** Let  $q(y|x)$  be the proposal distribution defined by (10). We have that

$$\frac{\tilde{\pi}_0(y)q(x|y)}{\tilde{\pi}_0(x)q(y|x)} = \exp\left\{-\frac{\tau^2}{2}(y^\top y - x^\top x)\right\},$$

where  $\tau^2 = 1$  in the case of RWM and  $\tau^2 = c^2/4$  in the case of MALA. We assume that  $F(x) = x^{(j)}$  and we show that i)  $\hat{F}_\pi^0(-x^{(j)}, x^{(j)}) = -\hat{F}_\pi^0(x)$  and that ii)  $\hat{F}_\pi^0(x^{(j)}, x^{(j)}) = \hat{F}_\pi^0(x^{(j)}, \Pi x^{(j)})$ , where  $x^{(j)}$  denotes the vector  $x \in \mathbb{R}^d$  without its  $j$ th coordinate and  $\Pi$  is a permutation matrix.

Since  $\hat{F}_\pi^0$  satisfies the Poisson equation we have that

$$\int \min \left\{ 1, \exp \left\{ -\frac{\tau^2}{2}(y^\top y - x^\top x) \right\} \right\} [\hat{F}_\pi^0(x) - \hat{F}_\pi^0(y)] q(y|x) dy = x^{(j)}, \tag{30}$$

which implies that

$$-\hat{F}_\pi^0(x) \int \alpha(x, y) q(y|x) dy - \int \alpha(x, y) [-\hat{F}_\pi^0(y)] q(y|x) dy = -x^{(j)}, \tag{31}$$

where  $\alpha(x, y) = \min \left\{ 1, \exp \left\{ -\frac{\tau^2}{2}(y^\top y - x^\top x) \right\} \right\}$ . Let also  $z$  and  $\tilde{z}$  be  $d$ -dimensional vectors such that  $z^{(j)} = -x^{(j)}$  and  $z^{(-j)} = x^{(-j)}$  and  $\tilde{z}^{(j)} = -y^{(j)}$ ,  $\tilde{z}^{(j')} = y^{(j')}$ . Then, by noting that the Jacobian of the transformations is equal to one, (31) becomes

$$-\hat{F}_\pi^0(-z^{(j)}, z^{(j)}) \int \alpha(z, \tilde{z}) q(\tilde{z}|z) d\tilde{z} - \int \alpha(z, \tilde{z}) [-\hat{F}_\pi^0(-z^{(j)}, z^{(j)})] q(\tilde{z}|z) d\tilde{z} = z^{(j)}, \tag{32}$$

where  $\alpha(z, \tilde{z}) = \min \left( 1, e^{-\frac{c^2+r^2-1}{2c^2}(\tilde{z}^\top \tilde{z} - z^\top z)} \right)$  and  $q(\tilde{z}|z) = \mathcal{N}(\tilde{z}|rz, c^2I)$ . Equation (32) implies that  $-\hat{F}_\pi^0(-z^{(j)}, z^{(j)})$  is solution of the Poisson equation and from the uniqueness of the solution we have i).

To prove ii) we denote by  $z$  the  $d$ -dimensional vector such that  $z^{(j)} = x^{(j)}$  and  $z^{(j')} = \Pi x^{(j')}$  and we apply the following transformation on (30); we set  $\tilde{z}$  to be  $d$ -dimensional vector such that  $\tilde{z}^{(j)} = y^{(j)}$  and  $\tilde{z}^{(j')} = \Pi y^{(j')}$ . Then, we have that

$$\hat{F}_\pi^0(z^{(j)}, \Pi^{-1}z^{(j')}) \int \alpha(z, \tilde{z}) q(\tilde{z}|z) d\tilde{z} - \int \alpha(z, \tilde{z}) [\hat{F}_\pi^0(z^{(j)}, \Pi^{-1}z^{(j')})] q(\tilde{z}|z) d\tilde{z} = z^{(j)}, \tag{33}$$

where  $\alpha(z, \tilde{z})$  and  $q(\tilde{z}|z)$  as in (32) since they are invariant to arbitrary permutations of  $z$  and/or  $\tilde{z}$  and  $\Pi^{-1}$  is permutation matrix such that  $\Pi^{-1}\Pi z^{(j')} = z^{(j')}$ . From (33) we have that  $\hat{F}_{\tilde{\pi}}^0(z^{(j)}, \Pi^{-1}z^{(j)})$  is solution of the Poisson equation and then ii) holds again due to the uniqueness of the solution of the Poisson equation.  $\square$

### C Proof of Proposition 3

**Proof** Let  $\tilde{\pi}(z) = \mathcal{N}(z|0, I)$  be the target of a Metropolis-Hastings algorithm with proposal  $q(\tilde{z}|z) = \mathcal{N}(\tilde{z}|rz, c^2I)$ , it easy to see that  $r = 1$  corresponds to the RWM algorithm and  $r = 1 - c^2/2$  to MALA. Let also  $\hat{F}_{\tilde{\pi}_0}$  the solution of the associated Poisson equation and  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$  be a bivariate Gaussian density with mean  $\mu$  and covariance matrix  $\Sigma$ . We assume that  $F(z) = z^{(1)}$  in (1) which becomes

$$\left( \int \alpha_z(z, \tilde{z})q(\tilde{z}|z)d\tilde{z} \right) \hat{F}_{\tilde{\pi}_0}(z) - \int \alpha_z(z, \tilde{z})q(\tilde{z}|z)\hat{F}_{\tilde{\pi}_0}(\tilde{z})d\tilde{z} = z^{(1)}, \tag{34}$$

where

$$\alpha_z(z, \tilde{z}) = \min \left\{ 1, e^{-\frac{c^2+r^2-1}{2c^2}(\tilde{z}^\top \tilde{z} - z^\top z)} \right\} \tag{35}$$

Let  $x = \mu + Lz$  and  $y = \mu + L\tilde{z}$ , where  $L$  such that  $\Sigma = LL^\top$ . From the properties of the Gaussian distribution we have that

$$q(y|x) = \mathcal{N}(y|r(x - \mu) + \mu, c^2\Sigma)$$

Moreover, equation (35) becomes

$$\begin{aligned} &\alpha_z(L^{-1}(x - \mu), L^{-1}(y - \mu)) \\ &= \min \left\{ 1, e^{-\frac{c^2+r^2-1}{2c^2}[(y-\mu)^\top \Sigma^{-1}(y-\mu) - (x-\mu)^\top \Sigma^{-1}(x-\mu)]} \right\} \\ &= \tilde{\alpha}(x, y) \end{aligned}$$

Then, Eq. (34) becomes

$$\begin{aligned} &\left( \int \tilde{\alpha}(x, y)q(y|x)dy \right) \hat{F}_{\tilde{\pi}_0}(L^{-1}(x - \mu)) \\ &- \int \tilde{\alpha}(x, y)q(y|x)\hat{F}_{\tilde{\pi}_0}(L^{-1}(y - \mu))dy = L_{11}^{-1}(x^{(1)} - \mu^{(1)}), \end{aligned} \tag{36}$$

where  $L_{11}$  is the first diagonal element of  $L$ . Since  $L_{11} = 1/L_{11}^{-1}$ . Equation (36) implies that the function

$$\hat{F}_{\tilde{\pi}}(x) = L_{11}\hat{F}_{\tilde{\pi}_0}(L^{-1}(x - \mu))$$

is the solution of the Poisson equation associated to the Metropolis-Hastings algorithm with target  $\tilde{\pi}(x)$  and proposal  $q(y|x)$ .  $\square$

### D Variance reduction across all the marginal distributions

Here we present a computationally efficient method in order to perform variance reduction in the estimation of the mean  $E_\pi[x^{(j)}]$ , for each  $j = 1, \dots, d$ . To achieve this we work under the assumption that  $F(x) = x$  in the Poisson equation (1). Following the same steps as in the main paper we first construct an approximation to its solution in the case of a standard Gaussian target and then we move to the general Gaussian case by extending Proposition 3 in the case where  $F(x) = x$ . Finally, we construct a static control variate for this choice of  $F$  and thus we extend the proposed estimator in (20) to be an estimator for  $E_\pi[x]$ .

We start by noting that by approximating the solution  $\hat{F}_{\tilde{\pi}_0}$  of the Poisson equation associated with the standard Gaussian target with the function  $G_0$  suggests that a natural approximation of the solution  $\hat{\mathcal{F}}_{\tilde{\pi}_0}$  of the Poisson equation in the case where  $F(x) = x$  is the function  $\mathcal{G}_0 : \mathbb{R}^d \rightarrow \mathbb{R}^d$  where  $\mathcal{G}_{0,j}(x) = G_0(x)$ . Having approximated  $\hat{\mathcal{F}}_{\tilde{\pi}_0}$  we prove Remark 3 which shows that the solution  $\hat{\mathcal{F}}_\pi$  of the Poisson equation associated with the general Gaussian target can be obtained from  $\hat{\mathcal{F}}_{\tilde{\pi}_0}$  by applying a change of variables like the one presented by Proposition 3 for the  $F(x) = x^{(j)}$  case.

**Remark 3** Under the same assumptions with Proposition 3 and for  $F(x) = x$  we have that the function  $\hat{\mathcal{F}}_\pi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  with  $\hat{\mathcal{F}}_\pi(x) = L\hat{\mathcal{F}}_{\tilde{\pi}_0}(L^{-1}(x - \mu))$ , where  $L$  is a lower triangular Cholesky matrix such that  $\Sigma = LL^\top$  and  $\tilde{\pi}_0(x) = \mathcal{N}(x|0, I)$  is the standard Gaussian target, is the solution of the Poisson equation associated with the general Gaussian target  $\tilde{\pi}(x) = \mathcal{N}(x|\mu, \Sigma)$ .

To prove Remark 3 we work as in the proof of Proposition 3. More precisely, Eq. (34) becomes

$$\begin{aligned} &\left( \int \alpha_z(z, \tilde{z})q(\tilde{z}|z)d\tilde{z} \right) \hat{\mathcal{F}}_{\tilde{\pi}_0}(z) \\ &- \int \alpha_z(z, \tilde{z})q(\tilde{z}|z)\hat{\mathcal{F}}_{\tilde{\pi}_0}(\tilde{z})d\tilde{z} = z, \end{aligned} \tag{37}$$

which is now a system of  $d$  independent equations. Then, by conducting the same change of variables as in the proof of Proposition 3, Eq. (37) becomes

$$\begin{aligned} &\left( \int \tilde{\alpha}(x, y)q(y|x)dy \right) \hat{\mathcal{F}}_{\tilde{\pi}_0}(L^{-1}(x - \mu)) \\ &- \int \tilde{\alpha}(x, y)q(y|x)\hat{\mathcal{F}}_{\tilde{\pi}_0}(L^{-1}(y - \mu))dy = L^{-1}(x - \mu). \end{aligned} \tag{38}$$

Equation (38) implies that the function

$$\hat{\mathcal{F}}_\pi(x) = L\hat{\mathcal{F}}_{\tilde{\pi}_0}(L^{-1}(x - \mu))$$



is the solution of the Poisson equation associated to the Metropolis-Hastings algorithm with target  $\tilde{\pi}(x)$  and proposal  $q(y|x)$ .

It is also important that Remark 3 implies that the function  $\mathcal{F}_{\tilde{\pi}_0}(L^{-1}(x - \mu))$  is the solution of the Poisson equation associated with the target  $\tilde{\pi}$  in the case where  $F(x) = L^{-1}x$ . Based on this result we construct a static control variate for this choice of  $F$ . Similarly to the case where  $F(x) = x^{(j)}$  the static control variate should be highly correlated with the function  $\alpha(x, y)(\mathcal{G}(y) - \mathcal{G}(x))$ , where  $\mathcal{G} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and based on Remark 3 we have assumed that  $\mathcal{G}(x) = \mathcal{G}_0(L^{-1}(x - \mu))$ . Therefore, relying on the same arguments presented in Sect. 2.3, we choose as a static control variate the function  $H : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that

$$H(x, y) = \min\{1, \tilde{r}(x, y)\}[\mathcal{G}(y) - \mathcal{G}(x)],$$

Importantly, it is easy to check that  $E_q[H(x, y)]$  is analytically available since its computation reduces to the calculation of integrals of the form in Eq. (19). Finally, analogously to the estimator  $\mu_{n,G}(F_{\hat{\theta}_n})$  in (20) for  $F(x) = x^{(j)}$  in the case where  $F(x) = L^{-1}x$  we propose to estimate  $E_{\pi}[x]$  with

$$\begin{aligned} \mu_{n,\mathcal{G}}(F_{\hat{\theta}_n}) &:= \frac{1}{n} \sum_{i=0}^{n-1} \left\{ L^{-1}x_i + \hat{\Theta}_n \circ \underbrace{\left\{ \alpha(x_i, y_i)(\mathcal{G}(y_i) - \mathcal{G}(x_i)) \right\}}_{\text{Stochastic Poisson control variate}} \right. \\ &\quad \left. + \underbrace{H(x_i, y_i) - E_{q(y|x_i)}[H(x_i, y)]}_{\text{Static control variate}} \right\}, \end{aligned} \tag{39}$$

where  $\hat{\Theta}_n$  is a  $d$ -dimensional vector in which the  $j$ th element is the coefficient  $\hat{\theta}_n$  in Eq. (22) and  $\circ$  denotes the Hadamard product. It follows that for  $F(x_i) = x_i$  the proposed modified estimator of  $E_{\pi}[F]$  is the vector  $L\mu_{n,\mathcal{G}}$  which can be computed with negligible extra computations than those already conducted during the main MCMC algorithm.

### E Calculations for Remark 2

As noted in Sect. 2.3 the calculation of  $E_q[h(x, y)]$  requires the to compute the following integral

$$\int \min \left\{ 1, \exp\left\{-\frac{1}{2}(\tilde{y}^\top \tilde{y} - \tilde{x}^\top \tilde{x})\right\} \frac{\tilde{q}(\tilde{x}|\tilde{y})}{\tilde{q}(\tilde{y}|\tilde{x})} \right\} [G_0(\tilde{y}) - G_0(\tilde{x})]q(\tilde{y}|\tilde{x})d\tilde{y}, \tag{40}$$

where  $\tilde{x}$  and  $\tilde{y}$  as defined in Sect. 2.3. In the case of the RWM algorithm the calculation of the integral above is conducted

by using the results in Proposition 1 since the the densities  $\tilde{q}(\tilde{y}|\tilde{x})$  and  $q(\tilde{y}|\tilde{x})$  coincide and, thus, (40) is consisted of the integrals in (8) and (9).

In the case of the MALA  $\tilde{q}(\tilde{y}|\tilde{x})$ , which is given by (10), has mean  $r\tilde{x}$  whereas the mean of the distribution with density  $q(\tilde{y}|\tilde{x})$  in (18) is  $k(\tilde{x}) = \tilde{x} + (c^2/2)L^\top \nabla \log \pi(x)$ . However, the calculation of

$$a^h(x) = \int \min \left\{ 1, \exp\left\{-\frac{1}{2}(\tilde{y}^\top \tilde{y} - \tilde{x}^\top \tilde{x})\right\} \frac{\tilde{q}(\tilde{x}|\tilde{y})}{\tilde{q}(\tilde{y}|\tilde{x})} \right\} q(\tilde{y}|\tilde{x})$$

is conducted similarly to the calculation of  $a(x)$  in the Proof of Proposition 2 and, more precisely, we have that  $a^h(x)$  is calculated from equation (26) where  $f$  follows the chi-squared distribution with  $d$  degrees of freedom and non-central parameter  $k(\tilde{x})^\top k(\tilde{x})/c^2$ . To compute the integral

$$a_g^h(x) = \int \min \left\{ 1, \exp\left\{-\frac{1}{2}(\tilde{y}^\top \tilde{y} - \tilde{x}^\top \tilde{x})\right\} \frac{\tilde{q}(\tilde{x}|\tilde{y})}{\tilde{q}(\tilde{y}|\tilde{x})} G_0(\tilde{y}) \right\} dy$$

we work again as in the proof of Proposition 2 for the calculation of  $a_g(x)$  and we find that  $a_g^h(x)$  is given by equation (29) for

$$A_k(x) = \frac{\exp \left\{ -\frac{k(\tilde{x})^\top k(\tilde{x})}{2c^2} - \gamma_k \delta_k^\top \delta_k + \frac{m_k(x)^\top m_k(x)}{2c^2(1+2\gamma_k c^2)} \right\}}{(1 + 2c^2\gamma_k)^{d/2}}$$

and

$$m_k(x) = \frac{k(x) + c^2(\beta_k + \gamma_k \delta_k)}{1 + 2c^2\gamma_k}.$$

### F Efficient post-processing

The proposed method requires the transformation  $z = L^{-1}(x - \mu)$  to be applied in all the samples of the MCMC output,  $x_0, \dots, x_{n-1}$ . This calculation can be conducted efficiently as follows. We assume that the samples are drawn from the target  $\pi$  by employing the RWM algorithm. The first sample  $x_0$  is standardized according to  $z_0 = L^{-1}(x_0 - \mu)$ . Then, any other  $z_j, j = 1, \dots, n - 1$ , is obtained recursively by

$$z_j = z_{j-1} + \alpha_j c \epsilon_j.$$

where  $\epsilon_j$  is the  $j$ -th standard normal noise vector generated with the proposal  $y_j = x_j + cL\epsilon_j$ , and  $\alpha_j \in \{0, 1\}$  is the binary acceptance decision. Therefore, the calculation of the standardized vectors can be conducted for all the MCMC samples with negligible cost (only  $O(n^2)$  to standardize the initial  $z_0$ ) by relying only on quantities computed during the MCMC. A similar computational efficiency also holds for the case of the MALA.

**Table 11** Estimated factors by which the variance of  $\mu_n(F)$  is larger than the variance of  $\mu_{n,G}(F)$  for standard Gaussian  $d$ -variate target

	RWM				MALA			
	d = 2	d = 10	d = 30	d = 100	d = 2	d = 10	d = 30	d = 100
n = 1000	92	23	7	–	1,258	60	48	–
n = 10,000	272	165	114	28	3556	79	82	263
n = 50,000	540	439	182	98	4604	92	104	156
n = 500,000	440	828	377	278	5031	263	273	284

We collect  $n$  samples after the first 10,000 iterations of the RWM and the MALA and we apply the proposed variance reduction methods by assuming that the covariance of the target is unknown and we use the sample covariance of the MCMC output as its estimator. In the case  $d = 100$  and  $n = 1,000$  the application of the method was impossible due to numerical issues in the estimation of the target covariance

**Table 12** Factors by which the effective sample (ESS) size that corresponds to the proposed estimator  $\mu_{n,G}(F)$  is larger than the ESS for the standard estimator  $\mu_n(F)$  where the latter is multiplied by the extra computational time needed for the proposed post-processing

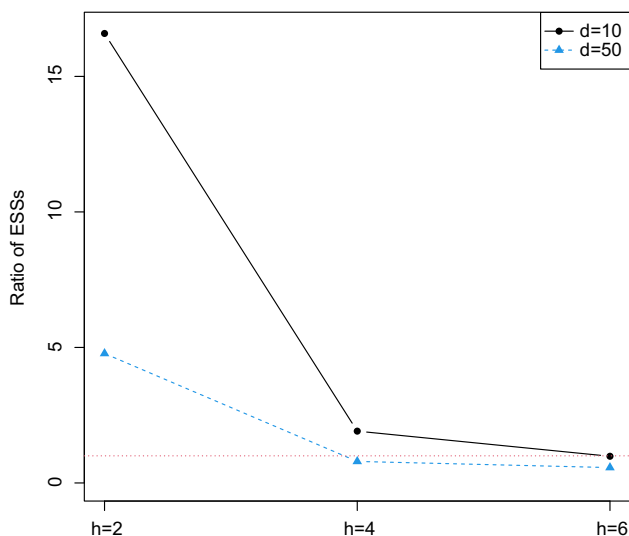
	RWM				MALA			
	d = 2	d = 10	d = 30	d = 100	d = 2	d = 10	d = 30	d = 100
n = 1000	18.60	5.19	1.44	0.29	224.32	10.62	9.44	16.23
n = 10,000	27.84	15.76	5.60	0.30	446.55	8.97	9.76	31.57
n = 50,000	49.20	29.65	6.32	0.94	578.52	9.17	8.59	22.79
n = 500,000	36.74	54.64	10.89	2.03	416.42	6.92	13.05	47.65

The target was a standard  $d$ -variate Gaussian distribution and we collected  $n$  samples after the first 10,000 iterations of the RWM and the MALA and we were interested in estimating the mean of the first coordinate of the target, i.e.,  $F(x) = x^{(1)}$

## G Additional experiments

### G.1 Gaussian target with unknown covariance

We repeat the experiments presented in Sect. 3.1 where the target distribution is a  $d$ -variate standard Gaussian distribu-



**Fig. 5** Effective sample sizes (ESSs) that correspond to the estimator  $\mu_{n,G}(F)$  over ESSs for  $\mu_n(F)$  for a mixture of  $d$ -variate Gaussian distributions with density given by Eq. (23) for different values of the mean  $m$  indicated by the choice of the parameter  $h$  in the x-axis. The ESS for  $\mu_n(F)$  is multiplied by the extra computational time needed for the proposed post-processing. The red dotted line indicates the value 1 in the y-axis. (Color figure online)

tion and we are interested in estimating the expected value of the first coordinate of the target, i.e., we set  $F(x) = x^{(1)}$ . However, we now relax the assumption that the target covariance matrix is known and we use the sample covariance of the MCMC output as an estimator of the true covariance of the target. Table 11 displays the factors by which the variance  $\mu_n(F)$  is greater than the variance of  $\mu_{n,G}(F)$  in the case of the MALA and RWM algorithm. A comparison of Tables 11 and 2 indicates that the error in the estimation of the covariance of the target has negligible effect on the achieved variance reduction.

### G.2 Effective sample size

We examine the effective sample sizes (ESSs) that correspond to the two competitive estimators  $\mu_n(F)$  and  $\mu_{n,G}(F)$ . The ESS in the case of an MCMC estimator can be calculated as the ratio of the variance of the estimator assuming independent samples over its variance estimated from the dependent samples of the MCMC output; see for example Vehtari et al. (2021) for more details. Therefore, to compute the ratio of the ESSs that correspond to  $\mu_n(F)$  and  $\mu_{n,G}(F)$  we only need the estimates of their variances calculated as explained in Sect. 3.

Table 12 presents the ratios of the ESSs that correspond to  $\mu_n(F)$  and  $\mu_{n,G}(F)$  for the experiment presented by Sect. 3.1 where the target was a multivariate standard Gaussian distribution. To take into account any extra computational

**Table 13** Factors by which the effective sample (ESS) size that corresponds to the proposed estimator  $\mu_{n,G}(F)$  is larger than the ESS for the standard estimator  $\mu_n(F)$  where the latter is multiplied by the extra computational time needed for the proposed post-processing

Dataset	$n = 1000$	$n = 10,000$	$n = 50,000$	$n = 200,000$
Ripley	20.98–26.40	17.02–58.20	19.34–59.19	18.15–72.46
Pima Indian	12.08–21.41	57.64–94.08	56.79–124.61	53.31–129.77
Heart	6.31–10.46	8.44–20.72	9.34–25.42	6.89–28.27
Australian	5.12–12.72	15.51–48.29	16.23–43.48	10.65–42.16
German	4.00–8.64	11.47–31.95	15.63–55.69	11.13–64.58

The target was the posterior distribution of the logistic regression models applied on the datasets indicated by the first column. We collect  $n$  samples after the first 10,000 iterations of the RWM algorithm

**Table 14** Factors by which the effective sample (ESS) size that corresponds to the proposed estimator  $\mu_{n,G}(F)$  is larger than the ESS for the standard estimator  $\mu_n(F)$  where the latter is multiplied by the extra computational time needed for the proposed post-processing

Dataset	$n = 1000$	$n = 10,000$	$n = 50,000$	$n = 200,000$
Ripley	10.18–14.94	12.90–21.53	9.94–21.36	8.37–16.24
Pima Indian	21.96–48.26	29.62–44.42	27.17–57.43	27.97–55.53
Heart	9.30–16.03	6.14–14.57	7.17–12.82	6.10–10.57
Australian	8.47–20.71	6.48–17.36	4.57–12.62	3.76–10.98
German	11.02–31.11	8.97–33.43	6.81–36.28	5.47–27.15

The target was the posterior distribution of the logistic regression models applied on the datasets indicated by the first column. We collect  $n$  samples after the first 10,000 iterations of the MALA

**Table 15** Factors by which the effective sample (ESS) size that corresponds to the proposed estimator  $\mu_{n,G}(F)$  is larger than the ESS for the standard estimator  $\mu_n(F)$  where the latter is multiplied by the extra computational time needed for the proposed post-processing

	$n = 10,000$				$n = 50,000$				$n = 200,000$			
	$h$	$m$	$\phi$	$s^2$	$h$	$m$	$\phi$	$s^2$	$h$	$m$	$\phi$	$s^2$
d=50	7.18–15.24	10.42	12.92	3.23	4.86–10.34	8.45	12.51	1.86	4.52–9.92	10.28	8.58	1.47
d=100	0.82–5.89	6.39	2.77	1.01	0.73–4.71	4.36	5.23	1.13	0.87–5.45	3.09	2.54	0.61

The target was the posterior distribution of the  $d$ -dimensional stochastic volatility model. We collect  $n$  samples after the first 10,000 iterations of the MALA

cost for the proposed post-processing of the MCMC output we have multiplied the ESSs for  $\mu_n(F)$  by the extra computational time needed for the proposed post-processing of the MCMC output. Similarly, Figure 5 and Tables 13, 14, 15 present the ratio of the ESSs normalised with respect to the computational time that correspond to the examples presented by Sects. 3.2–3.4. It is clear from the visual inspection of Figure 5 and the Tables that the application of the proposed variance reduction techniques results in an increase of the ESS per computational time in the vast majority of the examples.

**References**

Alexopoulos, A., Dellaportas, P., Papaspiliopoulos, O.: Bayesian prediction of jumps in large panels of time series data. *Bayesian Anal.* **1**(1), 1–33 (2021)

Andradóttir, S., Heyman, D.P., Ott, T.J.: Variance reduction through smoothing and control variates for Markov chain simulations. *ACM Trans. Model. Comput. Simul. (TOMACS)* **3**(3), 167–189 (1993)

Assaraf, R., Caffarel, M.: Zero-variance principle for Monte Carlo algorithms. *Phys. Rev. Lett.* **83**(23), 4682 (1999)

Atchadé Y.F., Perron, F.: Improving on the independent Metropolis–Hastings algorithm. *Stat. Sin.*, 3–18 (2005)

Barone, P., Frigessi, A.: Improving stochastic relaxation for Gaussian random fields. *Probab. Eng. Inf. Sci.* **4**(3), 369–389 (1990)

Barp, A., Oates, C., Porcu, E., Girolami, M., et al.: A Riemannian–Stein kernel method. **1**(5), 6–9 (2018). arXiv preprint [arXiv:1810.04946](https://arxiv.org/abs/1810.04946)

Belomestny, D., Iosipoi, L., Moulines, E., Naumov, A., Samsonov, S.: Variance reduction for Markov chains with application to MCMC. *Stat. Comput.* **30**(4), 973–997 (2020)

Broyden, C.G.: The convergence of a class of double-rank minimization algorithms 1. General considerations. *IMA J. Appl. Math.* **6**(1), 76–90 (1970)

Craiu, R.V., Meng, X.-L., et al.: Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo. *Ann. Stat.* **33**(2), 661–697 (2005)

Dellaportas, P., Kontoyiannis, I.: Notes on using control variates for estimation with reversible mcmc samplers (2009). arXiv preprint [arXiv:0907.4160](https://arxiv.org/abs/0907.4160)

Dellaportas, P., Kontoyiannis, I.: Control variates for estimation based on reversible Markov chain Monte Carlo samplers. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **74**(1), 133–161 (2012)

Flegal, J.M., Jones, G.L., et al.: Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Stat.* **38**(2), 1034–1070 (2010)

Gelfand, A.E., Smith, A.F.: Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**(410), 398–409 (1990)

- Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **73**(2), 123–214 (2011)
- Glasserman, P.: *Monte Carlo Methods in Financial Engineering*, vol. 53. Springer, Berlin (2004)
- Green, P.J., Han, X.-l.: Metropolis methods, Gaussian proposals and antithetic variables. In: *Stochastic Models, Statistical methods, and Algorithms in Image Analysis*, pp. 142–164. Springer (1992)
- Hammer, H., Tjelmeland, H.: Control variates for the Metropolis–Hastings algorithm. *Scand. J. Stat.* **35**(3), 400–414 (2008)
- Henderson, S.G.: *Variance reduction via an approximating Markov process*. Ph.D. thesis, Stanford University (1997)
- Kahn, H., Marshall, A.W.: Methods of reducing sample size in Monte Carlo computations. *J. Oper. Res. Soc. Am.* **1**(5), 263–278 (1953)
- Kastner, G., Frühwirth-Schnatter, S.: Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Comput. Stat. Data Anal.* **76**, 408–423 (2014)
- Kim, S., Shephard, N., Chib, S.: Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65**(3), 361–393 (1998)
- Meyn, S.: *Control Techniques for Complex Networks*. Cambridge University Press, Cambridge (2008)
- Mijatović, A., Vogrinc, J.: Asymptotic variance for Random Walk Metropolis chains in high dimensions: logarithmic growth via the Poisson equation. *Adv. Appl. Probab.* **51**(4), 994–1026 (2019)
- Mijatović, A., Vogrinc, J., et al.: On the Poisson equation for Metropolis–Hastings chains. *Bernoulli* **24**(3), 2401–2428 (2018)
- Mira, A., Geyer, C.J.: On Non-reversible Markov Chains. *Monte Carlo Methods*, pp. 95–110. Fields Institute/AMS (2000)
- Mira, A., Solgi, R., Imparato, D.: Zero variance Markov chain Monte Carlo for Bayesian estimators. *Stat. Comput.* **23**(5), 653–662 (2013)
- Oates, C.J., Girolami, M., Chopin, N.: Control functionals for Monte Carlo integration. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **79**(3), 695–718 (2017)
- Oates, C.J., Cockayne, J., Briol, F.-X., Girolami, M., et al.: Convergence rates for a class of estimators based on Stein’s method. *Bernoulli* **25**(2), 1141–1159 (2019)
- Papamarkou, T., Mira, A., Girolami, M., et al.: Zero variance differential geometric Markov chain Monte Carlo algorithms. *Bayesian Anal.* **9**(1), 97–128 (2014)
- Philippe, A., Robert, C.P.: Riemann sums for MCMC estimation and convergence monitoring. *Stat. Comput.* **11**(2), 103–115 (2001)
- Quenouille, M.H.: Notes on bias in estimation. *Biometrika* **43**(3/4), 353–360 (1956)
- R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna (2021)
- Roberts, G., Rosenthal, J., et al.: Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2**, 13–25 (1997)
- Roberts, G.O., Rosenthal, J.S.: Examples of adaptive MCMC. *J. Comput. Graph. Stat.* **18**(2), 349–367 (2009)
- South, L.F.: ZVCV: Zero-Variance Control Variates. R Package Version **2**(1), 1 (2021)
- South, L.F., Oates, C.J., Mira, A., Drovandi, C.: Regularised zero-variance control variates for high-dimensional variance reduction (2018). arXiv preprint [arXiv:1811.05073](https://arxiv.org/abs/1811.05073)
- South, L.F., Karvonen, T., Nemeth, C., Girolami, M., Oates, C., et al.: Semi-exact control functionals from sard’s method (2020). arXiv preprint [arXiv:2002.00033](https://arxiv.org/abs/2002.00033)
- Titsias, M., Dellaportas, P.: Gradient-based adaptive Markov chain Monte Carlo. In: *Advances in Neural Information Processing Systems*, pp. 15704–15713 (2019)
- Tsourti, Z.: *On variance reduction for Markov chain Monte Carlo*. Ph.D. thesis, Athens University of Economics and Business (2012)
- Valle, L.D., Leisen, F.: A new multinomial model and a zero variance estimation. *Commun. Stat. Simul. Comput.* **39**(4), 846–859 (2010)
- Van Dyk, D.A., Meng, X.-L.: The art of data augmentation. *J. Comput. Graph. Stat.* **10**(1), 1–50 (2001)
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C.: Rank-normalization, folding, and localization: an improved R for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16**(2), 667–718 (2021)
- Wasserman, L.: *All of Nonparametric Statistics*. Springer (2006)
- Yu, Y., Meng, X.-L.: To center or not to center: that is not the question—an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *J. Comput. Graph. Stat.* **20**(3), 531–570 (2011)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.