Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator

Arnaud Doucet
Department of Statistics,
University of Oxford,
doucet@stats.ox.ac.uk

Michael Pitt
Department of Economics,
University of Warwick
m.pitt@warwick.ac.uk

George Deligiannidis
Department of Statistics,
University of Oxford
deligian@stats.ox.ac.uk

Robert Kohn
Department of Economics,
University of New South Wales
r.kohn@unsw.edu.au

September 13, 2014

Abstract

When an unbiased estimator of the likelihood is used within a Metropolis–Hastings chain, it is necessary to trade off the number of Monte Carlo samples used to construct this estimator against the asymptotic variances of averages computed under this chain. Many Monte Carlo samples will typically result in Metropolis–Hastings averages with lower asymptotic variances than the corresponding Metropolis–Hastings averages using fewer samples. However, the computing time required to construct the likelihood estimator increases with the number of Monte Carlo samples. Under the assumption that the distribution of the additive noise introduced by the log-likelihood estimator is Gaussian with variance inversely proportional to the number of Monte Carlo samples and independent of the parameter value at which it is evaluated, we provide guidelines on the number of samples to select. We demonstrate our results by considering a stochastic volatility model applied to stock index returns.

Keywords: Intractable likelihood, Metropolis-Hastings algorithm, Particle filter, Sequential Monte Carlo, State-space model.

1 Introduction

The use of unbiased estimators within the Metropolis–Hastings algorithm was initiated by Lin et al. (2000), with a surge of interest in these ideas since their introduction in Bayesian statistics by Beaumont (2003). In a Bayesian context, an unbiased likelihood estimator is commonly constructed using importance sampling as in Beaumont (2003) or particle filters as in Andrieu et al. (2010). Andrieu & Roberts (2009) call this method the pseudo-marginal algorithm, and establish some of its theoretical properties.

Apart from the choice of proposals inherent to any Metropolis–Hastings algorithm, the main practical issue with the pseudo-marginal algorithm is the choice of the number, N, of Monte Carlo samples or particles used to estimate the likelihood. For any fixed N, the transition kernel of the pseudo-marginal algorithm leaves the posterior distribution of interest invariant. Using many Monte Carlo samples usually results in pseudo-marginal averages with asymptotic variances lower than the corresponding averages using fewer samples, as established by Andrieu & Vihola (2014) for likelihood estimators based on importance sampling. Empirical evidence suggests this result also holds when the likelihood is estimated by particle filters. However, the computing cost of constructing the likelihood estimator increases with N. We aim to select N so as to minimize the computational resources necessary to achieve a specified asymptotic variance for a particular pseudo-marginal average. This quantity, which is referred to as the computing time, is typically proportional to N times the asymptotic variance of this average, which is itself a function of N. Assuming that the distribution of the additive noise introduced by the log-likelihood estimator is Gaussian, with a variance inversely proportional to N and independent of the parameter value at which it is evaluated, this minimization was carried out in Pitt et al. (2012) and in Sherlock et al. (2013). However, Pitt et al. (2012) assume that the Metropolis–Hastings proposal is the posterior density, whereas Sherlock et al. (2013) relax the Gaussian noise assumption, but restrict themselves to an isotropic normal random walk proposal and assume that the posterior density factorizes into d independent and identically distributed components and $d \to \infty$.

Our article addresses a similar problem but considers general proposal and target densities and relaxes the Gaussian noise assumption. In this more general setting, we cannot minimize the computing time, and instead minimize explicit upper bounds on it. Quantitative results are presented under a Gaussian assumption. In this scenario, our guidelines are that N should be chosen such that the standard deviation of the log-likelihood estimator should be around 1.0 when the Metropolis–Hastings algorithm using the exact likelihood is efficient and around 1.7 when it is inefficient. In most practical scenarios, the efficiency of the Metropolis–Hastings algorithm using the exact likelihood is unknown as it cannot be implemented. In these cases, our results suggest selecting a standard deviation around 1.2.

2 Metropolis–Hastings method using an estimated likelihood

We briefly review how an unbiased likelihood estimator may be used within a Metropolis–Hastings scheme in a Bayesian context. Let $y \in \mathsf{Y}$ be the observations and $\theta \in \Theta \subseteq \mathbb{R}^d$ the parameters of interest. The likelihood of the observations is denoted by $p(y \mid \theta)$ and the prior for θ admits a density $p(\theta)$ with respect to Lebesgue measure so the posterior density of interest is $\pi(\theta) \propto p(y \mid \theta)p(\theta)$. We slightly abuse notation by using the same symbols for distributions and densities.

The Metropolis–Hastings scheme to sample from π simulates a Markov chain according to the transition kernel

$$Q_{\text{EX}}(\theta, \mathrm{d}\vartheta) = q(\theta, \vartheta)\, \alpha_{\text{EX}}(\theta, \vartheta)\mathrm{d}\vartheta + \{1 - \varrho_{\text{EX}}(\theta)\}\, \delta_\theta(\mathrm{d}\vartheta), \tag{1}$$

where

$$\alpha_{\text{EX}}(\theta, \vartheta) = \min\{1, r_{\text{EX}}(\theta, \vartheta)\}, \quad \varrho_{\text{EX}}(\theta) = \int q(\theta, \vartheta)\, \alpha_{\text{EX}}(\theta, \vartheta)\mathrm{d}\vartheta, \tag{2}$$

with $r_{\text{EX}}(\theta, \vartheta) = \pi(\vartheta)q(\vartheta, \theta) / \{\pi(\theta)q(\theta, \vartheta)\}$. This Markov chain cannot be simulated if $p(y \mid \theta)$ is intractable.

Assume $p(y \mid \theta)$ is intractable, but we have access to a non-negative unbiased estimator $\widehat{p}(y \mid \theta, U)$ of $p(y \mid \theta)$, where $U \sim m(\cdot)$ represents all the auxiliary random variables used to obtain this estimator. In this case, we introduce the joint density $\overline{\pi}(\theta, u)$ on $\Theta \times \mathcal{U}$, where

$$\overline{\pi}(\theta, u) = \pi(\theta)m(u)\widehat{p}(y \mid \theta, u)/p(y \mid \theta). \tag{3}$$

This joint density admits the correct marginal density $\pi(\theta)$, because $\widehat{p}(y \mid \theta, U)$ is unbiased. The pseudo-marginal algorithm is a Metropolis–Hastings scheme targeting (3) with proposal density $q(\theta, \cdot)\, m(\cdot)$, yielding the acceptance probability

$$\min\left\{1, \frac{\widehat{p}(y \mid \vartheta, v)p(\vartheta)\, q(\vartheta, \theta)}{\widehat{p}(y \mid \theta, u)p(\theta)\, q(\theta, \vartheta)}\right\} = \min\left\{1, \frac{\widehat{p}(y \mid \vartheta, v)/p(y \mid \vartheta)}{\widehat{p}(y \mid \theta, u)/p(y \mid \theta)} r_{\text{EX}}(\theta, \vartheta)\right\}, \tag{4}$$

for a proposal (ϑ, v). In practice, we only record $\{\theta, \log \widehat{p}(y \mid \theta, u)\}$ instead of $\{\theta, u\}$. We follow Andrieu & Roberts (2009) and Pitt et al. (2012) and analyze this scheme using additive noise,

$Z = \log \widehat{p}(y \mid \theta, U) - \log p(y \mid \theta) = \psi(\theta, U)$, in the log-likelihood estimator, rather than U. In this parameterization, the target density on $\Theta \times \mathbb{R}$ becomes

$$\overline{\pi}(\theta, z) = \pi(\theta) \exp(z)\, g(z \mid \theta), \tag{5}$$

where $g(z \mid \theta)$ is the density of Z when $U \sim m(\cdot)$ and the transformation $Z = \psi(\theta, U)$ is applied.

To sample from $\overline{\pi}(\theta, z)$, we could use the scheme previously described to sample from $\overline{\pi}(\theta, u)$ and then set $z = \psi(\theta, u)$. We can equivalently use the transition kernel

$$\begin{aligned} Q\{(\theta, z), (\mathrm{d}\vartheta, \mathrm{d}w)\} = q(\theta, \vartheta)\, g(w \mid \vartheta) \alpha_Q \{(\theta, z), (\vartheta, w)\} \mathrm{d}\vartheta \mathrm{d}w \\ + \{1 - \varrho_Q(\theta, z)\} \delta_{(\theta, z)}(\mathrm{d}\vartheta, \mathrm{d}w), \end{aligned} \tag{6}$$

where

$$\alpha_Q\{(\theta, z), (\vartheta, w)\} = \min\{1, \exp(w - z)\; r_{\mathrm{EX}}(\theta, \vartheta)\} \tag{7}$$

is (4) expressed in the new parameterization. Henceforth, we make the following assumption.

Assumption 1. *The noise density is independent of θ and is denoted by $g(z)$.*

Under this assumption, the target density (5) factorizes as $\pi(\theta)\pi_{\mathrm{Z}}(z)$, where

$$\pi_{\mathrm{Z}}(z) = \exp(z)\, g(z). \tag{8}$$

Assumption 1 allows us to analyze in detail the performance of the pseudo-marginal algorithm. This simplifying assumption is not satisfied in practical scenarios. However, in the stationary regime, we are concerned with the noise density at values of the parameter which arise from the target density $\pi(\theta)$ and the marginal density of the proposals at stationarity $\int \pi(\mathrm{d}\vartheta)\, q(\vartheta, \theta)$. If the noise density does not vary significantly in regions of high probability mass of these densities, then we expect this assumption to be a reasonable approximation. In Section 4, we examine experimentally how the noise density varies against draws from $\pi(\theta)$ and $\int \pi(\mathrm{d}\vartheta)\, q(\vartheta, \theta)$.

3 Main results

3.1 Outline

This section presents the main contributions of the paper. All the proofs are in Appendix 1 and in the Supplementary Material. We minimize upper bounds on the computing time of the pseudo-marginal algorithm, as discussed in Section 1. This requires establishing upper bounds on the asymptotic variance of an ergodic average under the kernel Q given in (6). To obtain these bounds, we introduce a new Markov kernel Q^*, where

$$\begin{aligned} Q^*\{(\theta, z), (\mathrm{d}\vartheta, \mathrm{d}w)\} = q(\theta, \vartheta)\, g(w) \alpha_{Q^*} \{(\theta, z), (\vartheta, w)\} \mathrm{d}\vartheta \mathrm{d}w \\ + \{1 - \varrho_{\mathrm{EX}}(\theta)\, \varrho_{\mathrm{Z}}(z)\} \delta_{(\theta, z)}(\mathrm{d}\vartheta, \mathrm{d}w), \end{aligned} \tag{9}$$

and

$$\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} = \alpha_{\mathrm{EX}}(\theta, \vartheta) \alpha_{\mathrm{Z}}(z, w), \quad \alpha_{\mathrm{Z}}(z, w) = \min\{1, \exp(w - z)\}, \tag{10}$$

$$\varrho_{\mathrm{Z}}(z) = \int g(w)\, \alpha_{\mathrm{Z}}(z, w)\, \mathrm{d}w. \tag{11}$$

As Q and Q^* are reversible with respect to $\overline{\pi}$ and the acceptance probability (10) is always smaller than (7), an application of the theorem in Peskun (1973) ensures that the variance of an ergodic average under Q^* is greater than or equal to the variance under Q. We obtain an exact expression for the variance under the bounding kernel Q^* and simpler upper bounds by exploiting a non-standard representation of this variance, the factor form of the acceptance probability (10) and the spectral properties of an auxiliary Markov kernel.

3.2 Inefficiency of Metropolis–Hastings type chains

This section recalls and establishes various results on the integrated autocorrelation time of Markov chains, henceforth referred to as the inefficiency. In particular, we present a novel representation of the inefficiency of Metropolis–Hastings type chains, which is the basic component of the proof of our main result.

Consider a Markov kernel Π on the measurable space $(\mathsf{X},\mathcal{X}) = \{\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\}$, where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-algebra on $\mathbb{R}^n$. For any measurable real-valued function f, measurable set A and probability measure μ, we use the standard notation: $\mu(f) = \int_{\mathsf{X}} \mu(\mathrm{d}x) f(x)$, $\mu(A) = \mu\{\mathbb{I}_A(\cdot)\}$, $\Pi f(x) = \int_{\mathsf{X}} \Pi(x, \mathrm{d}y) f(y)$ and for $n \geq 2$, $\Pi^n(x, \mathrm{d}y) = \int_{\mathsf{X}} \Pi^{n-1}(x, \mathrm{d}z) \Pi(z, \mathrm{d}y)$, with $\Pi^1 = \Pi$. We introduce the Hilbert spaces

$$L^2(\mathsf{X},\mu) = \left\{f : \mathsf{X} \to \mathbb{R} : \mu\left(f^2\right) < \infty\right\}, \ L_0^2(\mathsf{X},\mu) = \left\{f : \mathsf{X} \to \mathbb{R} : \mu(f) = 0, \ \mu\left(f^2\right) < \infty\right\}$$

equipped with the inner product $\langle f, g\rangle_\mu = \int f(x) g(x) \mu(\mathrm{d}x)$. A μ-invariant and ψ-irreducible Markov chain is said to be ergodic; see Tierney (1994) for the definition of ψ-irreducibility. The next result follows directly from Kipnis & Varadhan (1986) and Theorem 4 and Corollary 6 in Häggström & Rosenthal (2007).

Proposition 1. *Suppose Π is a μ-reversible and ergodic Markov kernel. Let $(X_i)_{i\geqslant 1}$ be a stationary Markov chain evolving according to Π and let $h \in L^2(\mathsf{X},\mu)$ be such that $\mu\left(\bar{h}^2\right) > 0$ where $\bar{h} = h - \mu(h)$. Write $\phi_n(h,\Pi) = \left\langle \bar{h}, \Pi^n \bar{h}\right\rangle_\mu / \mu\left(\bar{h}^2\right)$ for the autocorrelation at lag $n \geq 0$ of $\{h(X_i)\}_{i\geq 1}$ and $\mathrm{IF}(h,\Pi) = 1 + 2\sum_{n=1}^{\infty} \phi_n(h,\Pi)$ for the associated inefficiency. Then,*

(i) there exists a probability measure $e(h,\Pi)$ on $[-1,1)$ such that the autocorrelation and inefficiency satisfy the spectral representations

$$\phi_n(h,\Pi) = \int_{-1}^{1} \lambda^n e(h,\Pi)(\mathrm{d}\lambda), \quad \mathrm{IF}(h,\Pi) = \int_{-1}^{1} (1+\lambda)(1-\lambda)^{-1} e(h,\Pi)(\mathrm{d}\lambda); \tag{12}$$

(ii) if $\mathrm{IF}(h,\Pi) < \infty$, then as $n \to \infty$

$$n^{-1/2} \sum_{i=1}^{n} \{h(X_i) - \mu(h)\} \longrightarrow \mathcal{N}\left\{0; \mu\left(\bar{h}^2\right) \mathrm{IF}(h,\Pi)\right\}, \tag{13}$$

in distribution, where $\mathcal{N}\left(a; b^2\right)$ denotes the normal distribution with mean a and variance b^2.

When estimating $\mu(h)$, equation (13) implies that we need approximately $n\ \mathrm{IF}(h,\Pi)$ samples from the Markov chain $(X_i)_{i\geq 1}$ to obtain an estimator of the same precision as an average of n independent draws from μ.

We consider henceforth a μ-reversible kernel given by

$$P(x, \mathrm{d}y) = q(x, \mathrm{d}y) \alpha(x,y) + \{1 - \varrho(x)\} \delta_x(\mathrm{d}y), \quad \varrho(x) = \int q(x, \mathrm{d}y) \alpha(x,y),$$

where the proposal kernel is selected such that $q(x,\{x\}) = 0$, $\alpha(x,y)$ is the acceptance probability and we assume there does not exist an x such that $\mu(\{x\}) = 1$. We refer to P as a Metropolis–Hastings type kernel since it is structurally similar to the Metropolis–Hastings kernel, but we do not require $\alpha(x,y)$ to be the Metropolis–Hastings acceptance probability. This generalization is required when studying the kernel Q^* as the acceptance probability $\alpha_{Q^*}\{(\theta,z),(\vartheta,w)\}$ in (10) is not the Metropolis–Hastings acceptance probability.

Let $(X_i)_{i\geq 1}$ be a Markov chain evolving according to P. We now establish a non-standard expression for $\mathrm{IF}(h,P)$ derived from the associated jump chain representation $(\widetilde{X}_i, \tau_i)_{i\geq 1}$ of $(X_i)_{i\geq 1}$. In this representation, $(\widetilde{X}_i)_{i\geq 1}$ corresponds to the sequence of accepted proposals and $(\tau_i)_{i\geq 1}$ the associated sojourn times, that is $\widetilde{X}_1 = X_1 = \cdots = X_{\tau_1}$, $\widetilde{X}_2 = X_{\tau_1+1} = \cdots = X_{\tau_1+\tau_2}$ etc., with $\widetilde{X}_{i+1} \neq \widetilde{X}_i$. Some properties of this jump chain are now stated; see Lemma 1 in Douc & Robert (2011).

Lemma 1. *Let P be ψ-irreducible. Then $\varrho(x) > 0$ for any $x \in \mathsf{X}$ and $(\widetilde{X}_i, \tau_i)_{i\geq 1}$ is a Markov chain with a $\overline{\mu}$-reversible transition kernel $\overline{P}$, where*

$$\overline{P}\{(x,\tau), (\mathrm{d}y, \zeta)\} = \widetilde{P}(x, \mathrm{d}y)\, G\{\zeta; \varrho(y)\}, \quad \overline{\mu}(\mathrm{d}x, \tau) = \widetilde{\mu}(\mathrm{d}x)\, G\{\tau; \varrho(x)\}, \tag{14}$$

with

$$\widetilde{P}(x, \mathrm{d}y) = \frac{q(x,\mathrm{d}y)\alpha(x,y)}{\varrho(x)}, \quad \widetilde{\mu}(\mathrm{d}x) = \frac{\mu(\mathrm{d}x)\,\varrho(x)}{\mu(\varrho)}, \tag{15}$$

and $G(\cdot; \upsilon)$ denotes the geometric distribution with parameter υ.

The next proposition gives the relationship between $\mathrm{IF}(h, P)$ and $\mathrm{IF}(h/\varrho, \widetilde{P})$.

Proposition 2. *Assume that P and $\widetilde{P}$ are ergodic, that $h \in L_0^2(\mathsf{X}, \mu)$ and that $\mathrm{IF}(h, P) < \infty$. Then $h/\varrho \in L_0^2(\mathsf{X}, \tilde{\mu})$,*

$$\mu\left(h^2\right)\{1 + \mathrm{IF}(h, P)\} = \mu(\varrho)\, \widetilde{\mu}\left(h^2/\varrho^2\right)\left\{1 + \mathrm{IF}(h/\varrho, \widetilde{P})\right\}, \tag{16}$$

and $\mathrm{IF}(h/\varrho, \widetilde{P}) \leq \mathrm{IF}(h, P)$.

Lemma 1 and Proposition 2 are used in Section 3.3 to establish a representation of the inefficiency for the kernel $P = Q^*$.

We conclude this section by establishing some results on the positivity of the Metropolis–Hastings kernel and its associated jump kernel. Recall that a μ-invariant Markov kernel Π is positive if $\langle \Pi h, h\rangle_\mu \geq 0$ for any $h \in L^2(\mathsf{X}, \mu)$. If Π is reversible, then positivity is equivalent to $e(h, \Pi)([0,1)) = 1$ for all $h \in L^2(\mathsf{X}, \mu)$, where $e(h, \Pi)$ is the spectral measure, and it implies that $\mathrm{IF}(h, \Pi) \geq 1$; see, for example, Geyer (1992). The positivity of the jump kernel $\widetilde{P}$ associated with a Metropolis-Hastings kernel P is useful here as several bounds on the inefficiency established subsequently require the spectral measure of $\widetilde{P}$ to be supported on $[0,1)$. We now give sufficient conditions ensuring this property by extending Lemma 3.1 of Baxendale (2005). This complements results of Rudolf & Ullrich (2013).

Proposition 3. *Assume $\alpha(x,y)$ is the Metropolis–Hastings acceptance probability and $\mu(\mathrm{d}x) = \mu(x)\,\mathrm{d}x$. If P is ψ-irreducible, then $\widetilde{P}$ and P are both positive if one of the following two conditions is satisfied:*

(i) $q(x, \mathrm{d}y) = q(x,y)\mathrm{d}y$ is a ν-reversible kernel with $\nu(\mathrm{d}x) = \nu(x)\mathrm{d}x$, μ is absolutely continuous with respect to ν, and there exists $r : \mathsf{X} \times \mathsf{Z} \to \mathbb{R}^+$ such that $\nu(x)\, q(x,y) = \int r(x,z) r(y,z) \chi(\mathrm{d}z)$, where χ is a measure on Z;

(ii) $q(x, \mathrm{d}y) = q(x,y)\mathrm{d}y$ and there exists $s : \mathsf{X} \times \mathsf{Z} \to \mathbb{R}^+$ such that $q(x,y) = \int s(x,z)s(y,z)\chi(\mathrm{d}z)$, where χ is a measure on Z.

Remark 1. *Condition (i) is satisfied for an independent proposal $q(x,y) = \nu(y)$ by taking $\mathsf{Z} = \{1\}$, $\chi(\mathrm{d}z) = \delta_1(\mathrm{d}z)$ and $r(x, 1) = \nu(x)$. It is also satisfied for autoregressive positively correlated proposals with normal or Student-t innovations. Condition (ii) holds if $q(x,y)$ is a symmetric random walk proposal whose increments are multivariate normal or Student-t.*

3.3 Inefficiency of the bounding chain

This section applies the results of Section 3.2 to establish an exact expression for $\mathrm{IF}(h, Q^*)$. The next lemma shows that $\mathrm{IF}(h, Q^*)$ is an upper bound on $\mathrm{IF}(h, Q)$.

Lemma 2. *The kernel Q^* is $\overline{\pi}$-reversible and $\mathrm{IF}(h, Q) \leq \mathrm{IF}(h, Q^*)$ for any $h \in L^2(\Theta \times \mathbb{R}, \overline{\pi})$.*

In practice, we are only interested in functions $h \in L^2(\Theta, \pi)$. To simplify notation, we write $\mathrm{IF}(h, Q)$ in this case, instead of introducing the function $\widetilde{h} \in L^2(\Theta \times \mathbb{R}, \overline{\pi})$ satisfying $\widetilde{h}(\theta, z) = h(\theta)$ for all $z \in \mathbb{R}$ and writing $\mathrm{IF}(\widetilde{h}, Q)$. Proposition 2 shows that it is possible to express $\mathrm{IF}(h, Q^*)$ as a function of the inefficiency of its jump kernel $\widetilde{Q}^*$, which is particularly useful as $\widetilde{Q}^*$ admits a simple structure.

Lemma 3. *Assume Q^* is $\overline{\pi}$-irreducible. The jump kernel $\widetilde{Q}^*$ associated with Q^* is*

$$\widetilde{Q}^* \{(\theta, z), (\mathrm{d}\vartheta, \mathrm{d}w)\} = \widetilde{Q}_{\mathrm{EX}}(\theta, \mathrm{d}\vartheta) \widetilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w), \tag{17}$$

where

$$\widetilde{Q}_{\mathrm{EX}}(\theta, \mathrm{d}\vartheta) = \frac{q(\theta, \vartheta)\, \alpha_{\mathrm{EX}}(\theta, \vartheta) \mathrm{d}\vartheta}{\varrho_{\mathrm{EX}}(\theta)}, \quad \widetilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w) = \frac{g(w)\, \alpha_{\mathrm{Z}}(z, w)\, \mathrm{d}w}{\varrho_{\mathrm{Z}}(z)}. \tag{18}$$

The kernel $\widetilde{Q}_{\mathrm{EX}}(\theta, \mathrm{d}\vartheta)$ is reversible with respect to $\widetilde{\pi}(\mathrm{d}\theta)$ and the kernel $\widetilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w)$ is positive and reversible with respect to $\widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}z)$, where

$$\widetilde{\pi}(\mathrm{d}\theta) = \frac{\pi(\mathrm{d}\theta)\, \varrho_{\mathrm{EX}}(\theta)}{\pi(\varrho_{\mathrm{EX}})}, \quad \widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}z) = \frac{\pi_{\mathrm{Z}}(\mathrm{d}z)\, \varrho_{\mathrm{Z}}(z)}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})}.$$

If Q^* is ergodic, $h \in L_0^2(\Theta, \pi)$, $\mathrm{IF}(h, Q^*) < \infty$ and $\widetilde{Q}^*$ is ergodic, then $h/\varrho_{\mathrm{EX}} \in L_0^2(\Theta, \widetilde{\pi})$, $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) < \infty$, $\mathrm{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\} < \infty$ and

$$\pi\left(h^2\right)\{1 + \mathrm{IF}(h, Q^*)\} = \pi(\varrho_{\mathrm{EX}})\, \pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}})\, \widetilde{\pi}\left(h^2/\varrho_{\mathrm{EX}}^2\right)\left[1 + \mathrm{IF}\left\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\right\}\right]. \tag{19}$$

Additionally, $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) < \infty$ ensures that $\widetilde{Q}_{\mathrm{Z}}$ is geometrically ergodic and $\mathrm{IF}(1/\varrho_{\mathrm{Z}}, \widetilde{Q}_{\mathrm{Z}}) < \infty$.

The following theorem provides an expression for $\mathrm{IF}(h, Q^*)$ which decouples the contributions of the parameter and the noise components. The proof exploits the relationships between $\mathrm{IF}(h, Q_{\mathrm{EX}})$ and $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$, $\mathrm{IF}(h, Q^*)$ and $\mathrm{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\}$ and the spectral representation (12) of $\mathrm{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\}$. This spectral representation admits a simple structure due to the product form (17) of $\widetilde{Q}^*$.

Theorem 1. *Let $h \in L^2(\Theta, \pi)$. Assume that Q_{EX}, $Q^*, \widetilde{Q}_{\mathrm{EX}}$, $\widetilde{Q}^*$ are ergodic with $\mathrm{IF}(h, Q^*) < \infty$. Then, $\mathrm{IF}(h, Q) \leq \mathrm{IF}(h, Q^*)$ and*

$$\begin{aligned}\mathrm{IF}(h, Q^*) = {} & \frac{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})} - 1 \\ & + \frac{2\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - \frac{1}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})}\right\} \sum_{n=0}^{\infty} \phi_n(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \phi_n(1/\varrho_{\mathrm{Z}}, \widetilde{Q}_{\mathrm{Z}}).\end{aligned} \tag{20}$$

Remark 2. *If $q(\theta, \vartheta) = \pi(\vartheta)$, then $\mathrm{IF}(h, Q_{\mathrm{EX}}) = \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) = 1$ and $\phi_n(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) = 0$ for $n \geq 1$. It follows from Theorem 1 that $\mathrm{IF}(h, Q^*) = 2\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1$. This result was established in Lemma 4 of Pitt et al. (2012).*

Theorem 1 requires $Q_{\mathrm{EX}}, Q^*, \widetilde{Q}_{\mathrm{EX}}$ and $\widetilde{Q}^*$ to be ergodic. The following proposition, generalizing Theorem 2.2 of Roberts & Tweedie (1996), provides sufficient conditions ensuring this.

Proposition 4. *Suppose $\pi(\theta)$ is bounded away from 0 and ∞ on compact sets, and there exist $\delta > 0$ and $\varepsilon > 0$ such that, for every θ,*

$$|\theta - \vartheta| \leq \delta \Rightarrow q(\theta, \vartheta) \geq \varepsilon. \tag{21}$$

Then $Q_{\mathrm{EX}}, Q^*, \widetilde{Q}_{\mathrm{EX}}$ and $\widetilde{Q}^*$ are ergodic.

3.4 Bounds on the relative inefficiency of the pseudo-marginal chain

For any kernel Π, we define the relative inefficiency $\mathrm{RIF}(h,\Pi) = \mathrm{IF}(h,\Pi)/\mathrm{IF}(h,Q_{\mathrm{EX}})$, which measures the inefficiency of Π compared to that of Q_{EX}. This section provides tractable upper bounds for $\mathrm{RIF}(h,Q)$. From Lemma 2, $\mathrm{RIF}(h,Q) \le \mathrm{RIF}(h,Q^*)$, but the expression of $\mathrm{RIF}(h,Q^*)$ that follows from Theorem 1 is intricate and depends on the autocorrelation sequence $\{\phi_n(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})\}_{n\ge 1}$, as well as other terms. The next corollary provides upper bounds on $\mathrm{RIF}(h,Q)$ that depend only on $\mathrm{IF}(h,Q_{\mathrm{EX}})$. To simplify the notation, we write $\phi_{\mathrm{Z}} = \phi_1(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}})$.

Corollary 1. *Under the assumptions of Theorem 1,*

1. $\mathrm{RIF}(h,Q) \le \mathrm{uRIF}_1(h)$, *where*
$$\begin{aligned}\mathrm{uRIF}_1(h) = {} & \{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}[\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) + (1-\phi_{\mathrm{Z}})\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})\}] \\ & - 1/\mathrm{IF}(h,Q_{\mathrm{EX}});\end{aligned} \tag{22}$$

2. *if, in addition,* $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}}) \ge 1$, *then* $\mathrm{RIF}(h,Q) \le \mathrm{uRIF}_2(h) \le \mathrm{uRIF}_1(h)$, *where*
$$\mathrm{uRIF}_2(h) = \{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}\,\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1/\mathrm{IF}(h,Q_{\mathrm{EX}}). \tag{23}$$

Proposition 3 gives sufficient conditions for the condition $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}}) \ge 1$ of Part 2 of Corollary 1 to hold.

Remark 3. *The bounds above are tight in two cases. First, if* $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) \to 1$, *then* $\mathrm{RIF}(h,Q)$, $\mathrm{uRIF}_1(h)$, $\mathrm{uRIF}_2(h) \to 1$. *Second, if* $q(\theta,\vartheta) = \pi(\vartheta)$, *then* $\mathrm{RIF}(h,Q) = \mathrm{uRIF}_2(h)$.

We now provide upper bounds on $\mathrm{RIF}(h,Q)$ and lower bounds on $\mathrm{RIF}(h,Q^*)$ in terms of $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})$.

Corollary 2. *Under the assumptions of Theorem 1,*

1. $\mathrm{RIF}(h,Q) \le \mathrm{uRIF}_3(h)$, *where*
$$\begin{aligned}\mathrm{uRIF}_3(h) = {} & \left\{1+\frac{1}{\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\right\}\left[\frac{1}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})} + \phi_{\mathrm{Z}}\left\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - \frac{1}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})}\right\}\right] \\ & + 2\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})\}(1-\phi_{\mathrm{Z}})/\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}}) - 1/\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}});\end{aligned} \tag{24}$$

2. $\mathrm{RIF}(h,Q) \le \mathrm{uRIF}_4(h)$, *where*
$$\begin{aligned}\mathrm{uRIF}_4(h) = {} & \frac{\left\{1+1/\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})\right\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})\}\{1+\mathrm{IF}(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}})\} \\ & + 1/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}}) + \frac{1}{\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\frac{1}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})} - 1\right\};\end{aligned} \tag{25}$$

3. *if* $\widetilde{Q}_{\mathrm{EX}}$ *is positive, then* $\mathrm{RIF}(h,Q^*) \ge \mathrm{lRIF}_1(h)$, *where*
$$\mathrm{lRIF}_1(h) = \frac{1}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})} + \frac{2}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})\}; \tag{26}$$

4. $\mathrm{RIF}(h,Q^*) \ge \mathrm{lRIF}_2$, *where*
$$\mathrm{lRIF}_2 = 1/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}}), \tag{27}$$
and $\mathrm{RIF}(h,Q^*), \mathrm{uRIF}_4(h) \to \mathrm{lRIF}_2$ *as* $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}}) \to \infty$.

Proposition 3 gives sufficient conditions for $\widetilde{Q}_{\mathrm{EX}}$ to be positive. Section 3.5 discusses these bounds in more detail.

3.5 Optimizing the computing time under a Gaussian assumption

This section provides quantitative guidelines on how to select the standard deviation σ of the noise density, under the following assumption.

Assumption 2. *The noise density is* $g^{\sigma}(z) = \varphi\left(z; -\sigma^2/2, \sigma^2\right)$*, where* $\varphi(z; a, b^2)$ *is a univariate normal density with mean* a *and variance* b^2*.*

Assumption 2 ensures that $\int \exp(z)\, g^{\sigma}(z)\, dz = 1$ as required by the unbiasedness of the likelihood estimator. Consider a time series $y_{1:T} = (y_1, \ldots, y_T)$, where the likelihood estimator $\widehat{p}(y_{1:T} \mid \theta)$ of $p(y_{1:T} \mid \theta)$ is computed through a particle filter with N particles. Theorem 1 of an unpublished technical report (arXiv:1307.0181) by Bérard et al. shows that, under regularity assumptions, the log-likelihood error is distributed according to a normal density with mean $-\delta\gamma^2/2$ and variance $\delta\gamma^2$ as $T \to \infty$, for $N = \delta^{-1}T$. Hence, in this important scenario, the noise distribution satisfies approximately the form specified in Assumption 2 for large T and the variance is asymptotically inversely proportional to the number of samples. This assumption is also made in Pitt et al. (2012), where it is justified experimentally. Section 4 below provides additional experimental results.

The next result is Lemma 4 in Pitt et al. (2012) and follows from Assumption 2, equation (8) and Remark 2. We now make the dependence on σ explicit in our notation.

Corollary 3. *Under Assumption 2,* $\pi_{\mathrm{Z}}^{\sigma}(z) = \varphi\left(z; \sigma^2/2, \sigma^2\right)$,

$$\varrho_{\mathrm{Z}}^{\sigma}(z) = 1 - \Phi(z/\sigma + \sigma/2) + \exp(-z)\Phi(z/\sigma - \sigma/2), \quad \pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma}) = \int \frac{\varphi(w; 0, 1)}{1 - \overline{\varrho}_{\mathrm{Z}}^{\sigma}(w)} \mathrm{d}w,$$

where $\overline{\varrho}_{\mathrm{Z}}^{\sigma}(w) = \Phi(w + \sigma) - \exp(-w\sigma - \sigma^2/2)\Phi(w)$ *and* $\Phi(\cdot)$ *is the standard Gaussian cumulative distribution function. Additionally,* $\pi_{\mathrm{Z}}^{\sigma}(\varrho_{\mathrm{Z}}^{\sigma}) = 2\Phi(-\sigma/\sqrt{2})$.

The terms $\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma})$, $\phi_{\mathrm{Z}}^{\sigma}$ and $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \widetilde{Q}_{\mathrm{Z}})$, appearing in the bounds of Corollaries 1 and 2, do not admit analytic expressions, but can be computed numerically. We note that $\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma})$ is finite, and thus by Lemma 3 $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \widetilde{Q}_{\mathrm{Z}})$ is also finite. Consequently, for specific values of σ, $\mathrm{IF}(h, Q_{\mathrm{EX}})$ and $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$, these bounds can be calculated.

We now use these bounds to guide the choice of σ. The quantity we aim to minimize is the relative computing time for Q defined as $\mathrm{RCT}(h, Q; \sigma) = \mathrm{RIF}(h, Q; \sigma)/\sigma^2$ because $1/\sigma^2$ is usually approximately proportional to the number of samples N used to estimate the likelihood and the computational cost at each iteration is typically proportional to N, at least in the particle filter scenario described previously. We define $\mathrm{RCT}(h, Q^*; \sigma)$ similarly. As $\mathrm{RIF}(h, Q; \sigma)$ is intractable, we instead minimize the upper bounds $\mathrm{uRCT}_i(h; \sigma) = \mathrm{uRIF}_i(h; \sigma)/\sigma^2$, for $i = 1, \ldots, 4$. We similarly define the quantities $\mathrm{lRCT}_1(h; \sigma) = \mathrm{lRIF}_1(h; \sigma)/\sigma^2$ and $\mathrm{lRCT}_2(\sigma) = \mathrm{lRIF}_2(\sigma)/\sigma^2$, which bound $\mathrm{RCT}(h, Q^*; \sigma)$ from below. Figure 1 plots these bounds against σ for different values of $\mathrm{IF}(h, Q_{\mathrm{EX}})$ and $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$.

Prior to discussing how these results guide the selection of σ, we outline some properties of the bounds. First, as the corresponding inefficiency increases, the upper bounds $\mathrm{uRCT}_i(h; \sigma)$ displayed in Fig. 1 become flatter as functions of σ, and the corresponding minimizing argument σ_{opt} increases. This flattening effect suggests less sensitivity to the choice of σ for the pseudo-marginal algorithm. Second, for given σ, all the upper bounds are decreasing functions of the corresponding inefficiency, which suggests that the penalty from using the pseudo-marginal algorithm drops as the exact algorithm becomes more inefficient. Third, in the case discussed in Remark 2, where $q(\theta, \vartheta) = \pi(\vartheta)$, so that $\mathrm{IF}(h, Q_{\mathrm{EX}}) = \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) = 1$, we obtain $\mathrm{uRCT}_2(h; \sigma) = \mathrm{uRCT}_3(h; \sigma) = \mathrm{RCT}(h, Q^*; \sigma) = \mathrm{RCT}(h, Q; \sigma)$. Fourth, $\mathrm{uRCT}_4(h; \sigma)$ agrees with the lower bound $\mathrm{lRCT}_2(\sigma)$ as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \to \infty$ as indicated by Part 2 of Corollary 2. In this case, these two bounds, as well as $\mathrm{uRCT}_1(h; \sigma)$, are sharp for $\mathrm{RCT}(h, Q^*; \sigma)$. Fifth, $\mathrm{uRCT}_2(h; \sigma)$ is sharper than $\mathrm{uRCT}_1(h; \sigma)$ for $\mathrm{RCT}(h, Q^*; \sigma)$, but requires a mild additional assumption.

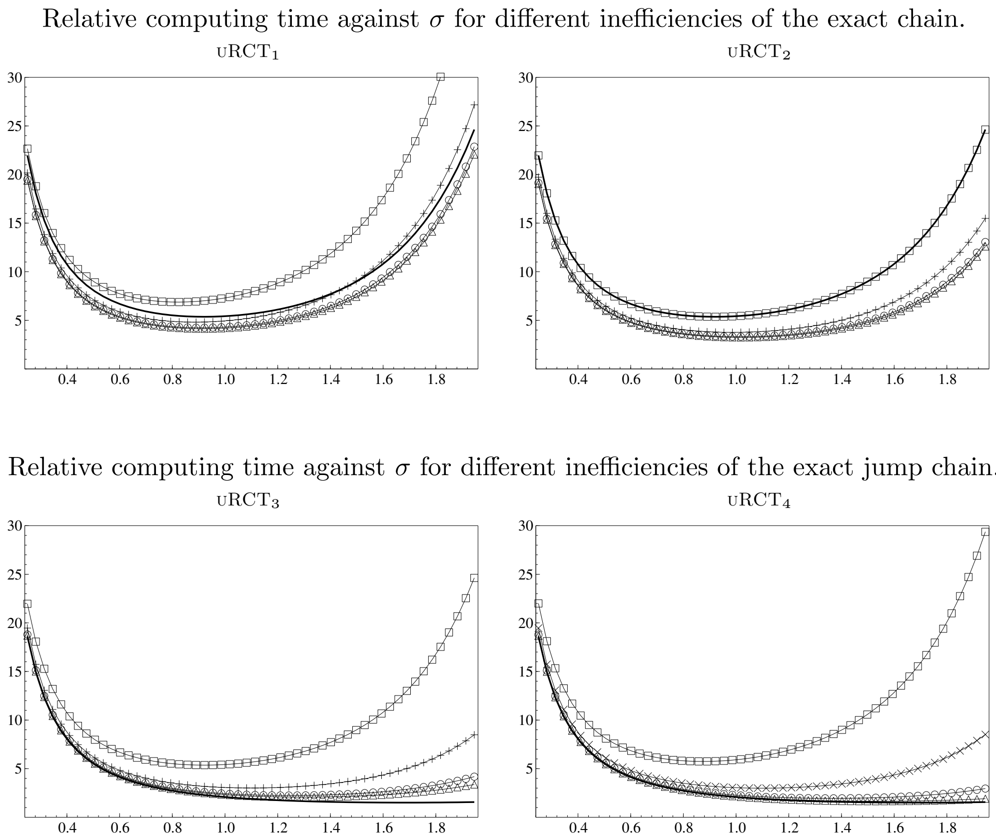

Figure 1: Theoretical results for relative computing time against σ. Top: The bounds $\text{uRCT}_1(h;\sigma)$ (left) and $\text{uRCT}_2(h;\sigma)$ (right) are displayed. Different values of $\text{IF}\,(h, Q_{\text{EX}})$ are taken as 1 (squares), 4 (crosses), 20 (circles) and 80 (triangles). The solid line corresponds to the perfect proposal, as discussed in Remark 2. Bottom: The lower bound $\text{lRCT}_2(\sigma)$ (solid line) is shown together with $\text{uRCT}_3(h;\sigma)$ (left) and $\text{uRCT}_4(h;\sigma)$ (right). Different values of $\text{IF}(h/\varrho_{\text{EX}}, \widetilde{Q}_{\text{EX}})$ are taken as 1 (squares), 4 (crosses), 20 (circles) and 80 (triangles).

As the likelihood is intractable, it is necessary to make a judgment on how to choose σ, because $\text{IF}\,(h, Q_{\text{EX}})$ and $\text{IF}(h/\varrho_{\text{EX}}, \widetilde{Q}_{\text{EX}})$ are unknown and cannot be easily estimated. Consider two extreme scenarios. The first is the perfect proposal $q\,(\theta, \vartheta) = \pi\,(\vartheta)$, so that by Corollary 3 and Remark 2, $\text{RCT}\,(h, Q; \sigma) = \{2\pi_{\text{Z}}^{\sigma}(1/\varrho_{\text{Z}}^{\sigma}) - 1\}/\sigma^2$, which we denote by $\text{RCT}\,(h, Q_\pi; \sigma)$, is minimized at $\sigma_{\text{opt}} = 0.92$. The second scenario considers a very inefficient proposal corresponding to Part 4 of Corollary 2 so that $\text{RCT}\,(h, Q^*; \sigma) = \text{lRCT}_2(\sigma)$, which is minimized at $\sigma_{\text{opt}} = 1.68$. If we choose $\sigma_{\text{opt}} = 1.68$ over $\sigma_{\text{opt}} = 0.92$ in scenario 1, then $\text{RCT}\,(h, Q_\pi; \sigma)$ rises from 5.36 to 12.73. Conversely, if we choose $\sigma_{\text{opt}} = 0.92$ over $\sigma_{\text{opt}} = 1.68$ in scenario 2, the relative computing time $\text{RCT}\,(h, Q^*; \sigma)$ rises from 1.51 to 2.29. This suggests that the penalty in choosing the wrong value is much more severe if we incorrectly assume we are in scenario 2 than if we incorrectly assume we are in scenario 1. This is because as $\text{IF}(h/\varrho_{\text{EX}}, \widetilde{Q}_{\text{EX}})$ increases, $\text{lRCT}_2(\sigma)$ is very flat relative to $\text{RCT}\,(h, Q_\pi; \sigma)$, as a function of σ. In practice, choosing σ_{opt} slightly greater than 1.0 appears sensible. For example, a value of $\sigma = 1.2$ leads to an increase in $\text{RCT}(h, Q_\pi; \sigma)$ from the minimum value of 5.36 to 6.10 and an increase in $\text{lRCT}_2(\sigma)$ from the minimum value of 1.51 to 1.75. In Appendix 2, we compute lower and upper bounds for the minimizing argument of $\text{RCT}(h, Q^*; \sigma)$ for various values of $\text{IF}(h/\varrho_{\text{EX}}, \widetilde{Q}_{\text{EX}})$.

Some caution should be exercised in interpreting these results as the lower bounds apply to $\text{RCT}(h, Q^*; \sigma)$, but not in general to $\text{RCT}(h, Q; \sigma)$. Similarly, whilst $\text{uRCT}_4(h;\sigma)$ and the lower bounds become exact for $\text{RCT}(h, Q^*; \sigma)$ as $\text{IF}(h/\varrho_{\text{EX}}, \widetilde{Q}_{\text{EX}}) \to \infty$, they only provide upper bounds for $\text{RCT}(h, Q; \sigma)$.

However, in an important class of problems $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$ is large, for instance when $q(\theta, \vartheta)$ is a random walk proposal with small step size. In this case, we expect that as the step size gets smaller the acceptance probability α_{EX} of Q_{EX} will tend towards unity and hence asymptotically $\alpha_{Q^*} = \alpha_Q$. This suggests that, for small enough step size, $\mathrm{RCT}(h, Q^*; \sigma) \approx \mathrm{RCT}(h, Q; \sigma)$.

The numerical results in this section are based on Assumption 2. However, the bounds on the relative inefficiences of Q and Q^* presented in Corollaries 1 and 2 can be calculated for any other noise distribution $g(z)$, subject to $\int \exp(z)\, g(z)\, dz = 1$. These bounds can in turn be used to construct corresponding bounds on the relative computing times of Q and Q^*, provided that an appropriate penalization term is employed to account for the computational effort of obtaining the likelihood estimator.

3.6 Discussion

We now compare informally the bound $\mathrm{LRIF}_2(\sigma) = 1/\{2\Phi(-\sigma/\sqrt{2})\}$ of Part 4 of Corollary 2 to the results in Sherlock et al. (2013). These authors make Assumption 1, assume that the target factorises into d independent and identically distributed components and that the proposal is an isotropic Gaussian random walk of jump size $d^{-1/2}l$. In the Gaussian noise case, for $h(\theta) = \theta_1$ where $\theta = (\theta_1, ..., \theta_d)$, their results and a standard calculation with their diffusion limit, suggest that as $d \to \infty$ the relative inefficiency satisfies

$$\frac{\mathrm{IF}\,(h, Q; \sigma, l)}{\mathrm{IF}\,(h, Q_{\mathrm{EX}}; l)} = \mathrm{RIF}(h, Q; \sigma, l) \to \mathrm{ARIF}(\sigma, l) = \frac{J_{\sigma^2=0}(l)}{J_{\sigma^2}(l)} = \frac{\Phi(-l/2)}{\Phi\left\{-\left(2\sigma^2 + l^2\right)^{1/2}/2\right\}}, \tag{28}$$

where the expression for $J_{\sigma^2}(l)$ is given by equations (3.3) and (3.4) of Sherlock et al. (2013). We observe that $\mathrm{ARIF}(\sigma, l)$ converges to $\mathrm{LRIF}_2(\sigma)$ as $l \to 0$. This is unsurprising. As $d \to \infty$, we conjecture that in this scenario the conditions of Part 4 of Corollary 2 apply, in particular that $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \to \infty$ for any $l > 0$. Therefore, in this case, $\mathrm{RIF}\,(h, Q^*; \sigma, l) \to \mathrm{LRIF}_2(\sigma)$. As $l \to 0$, we have informally that $\varrho_{\mathrm{EX}}(\theta) \to 1$, so that it is reasonable to conjecture that $\mathrm{RIF}\,(h, Q; \sigma, l)\,/\mathrm{RIF}\,(h, Q^*; \sigma, l) \to 1$. If one of these limits holds uniformly, then $\mathrm{ARIF}(\sigma, l) \to \mathrm{LRIF}_2(\sigma)$.

4 Application

4.1 Stochastic volatility model and pseudo-marginal algorithm

This section examines a multivariate partially observed diffusion model, which was introduced by Chernov et al. (2003), and discussed in Huang & Tauchen (2005). The regularly observed log price $P(t)$ evolves according to,

$$\begin{aligned}
&\mathrm{d}\log P(t) = \mu_y \mathrm{d}t + \text{s-}\exp\left[\{v_1(t) + \beta_2 v_2(t)\}/2\right] \mathrm{d}B(t),\\
&\mathrm{d}v_1(t) = -k_1\{v_1(t) - \mu_1\}\,\mathrm{d}t + \sigma_1 \mathrm{d}W_1(t),\ \mathrm{d}v_2(t) = -k_2 v_2(t)\mathrm{d}t + \{1 + \beta_{12} v_2(t)\}\,\mathrm{d}W_2(t),
\end{aligned}$$

and the leverage parameters corresponding to the correlations between the driving Brownian motions are $\phi_1 = \mathrm{corr}\{B(t), W_1(t)\}$ and $\phi_2 = \mathrm{corr}\{B(t), W_2(t)\}$. The function s-exp $(\cdot)$ is a spliced exponential function to ensure non-explosive growth, see Huang & Tauchen (2005). The two components for volatility allow for quite sudden changes in log price whilst retaining long memory in volatility. We note that the Brownian motion of the price process may be expressed as $\mathrm{d}B(t) = a_1 \mathrm{d}W_1(t) + a_2 \mathrm{d}W_2(t) + \sqrt{b}\mathrm{d}\overline{B}(t)$, where $a_1 = \phi_1(1 - \phi_2^2)/(1 - \phi_1^2\phi_2^2)$, $a_2 = \phi_2(1 - \phi_1^2)/(1 - \phi_1^2\phi_2^2)$ and $b = (1 - \phi_1^2)(1 - \phi_2^2)/(1 - \phi_1^2\phi_2^2)$. Here $\overline{B}(t)$ is an independent Brownian motion. Suppose the log prices are observed at equally spaced times $\tau_1 < \tau_2 < \tau_2 < \ldots < \tau_T < \tau_{T+1}$ and $\Delta = \tau_{s+1} - \tau_s$ for any s which gives returns $Y_s = \log P(\tau_{s+1}) - \log P(\tau_s)$, for $s = 1, \ldots, T$. The distribution of these returns conditional upon the volatility paths and the driving processes $W_1(t)$ and $W_2(t)$ is available in closed form as $Y_s \sim \mathcal{N}\left(\mu_y\Delta + a_1 Z_{1,s} + a_2 Z_{2,s}; b\sigma_s^{2*}\right)$, where

$$Z_{1,s} = \int_{\tau_s}^{\tau_{s+1}} \sigma(u)\mathrm{d}W_1(u),\ Z_{2,s} = \int_{\tau_s}^{\tau_{s+1}} \sigma(u)\mathrm{d}W_2(u),\ \sigma_s^{2*} = \int_{\tau_s}^{\tau_{s+1}} \sigma^2(u)du, \tag{29}$$

and $\sigma(t) =$s-exp$\left[\{v_1(t) + \beta_2 v_2(t)\}/2\right]$. An Euler scheme is used to approximate the evolution of the volatilities $v_1(t)$ and $v_2(t)$ by placing a number, $M-1$, of latent points between τ_s and τ_{s+1}. The volatility components are denoted by $v_{1,1}^s, ..., v_{1,M-1}^s$ and $v_{2,1}^s, ..., v_{2,M-1}^s$. For notational convenience, the start and end points are set to $v_{1,0}^s = v_1(\tau_s)$ and $v_{1,M}^s = v_1(\tau_{s+1})$, and similarly for $v_2(t)$. These latent points are evenly spaced in time by $\delta = \Delta/M$. The equation for the Euler evolution, starting at $v_{1,0}^s = v_{1,M}^{s-1}$ and $v_{2,0}^s = v_{2,M}^{s-1}$, is

$$\begin{aligned} v_{1,m+1}^s &= v_{1,m}^s - k_1(v_{1,m}^s - \mu_1)\delta + \sigma_1\sqrt{\delta}u_{1,m}, \\ v_{2,m+1}^s &= v_{2,m}^s - k_2 v_{2,m}^s\delta + \left(1 + \beta_{12}v_{2,m}^s\right)\sqrt{\delta}u_{2,m},\ m = 0, \dots, M-1, \end{aligned}$$

where $u_{1,m} \sim \mathcal{N}(0,1)$ and $u_{2,m} \sim \mathcal{N}(0,1)$. Conditional upon these trajectories and the innovations, the distribution of the returns has a closed form so that $Y_s \sim \mathcal{N}\left(\mu_y\Delta + a_1\widehat{Z}_{1,s} + a_2\widehat{Z}_{2,s}; b\widehat{\sigma}_s^{2*}\right)$, where $\widehat{Z}_{1,s}$, $\widehat{Z}_{2,s}$ and $\widehat{\sigma}_s^{2*}$ are the Euler approximations to the corresponding expression in (29).

We consider T daily returns, $y = (y_1, ..., y_T)$, from the S&P 500 index. Bayesian inference is performed on the 9-dimensional parameter vector $\theta = (k_1, \mu_1, \sigma_1, k_2, \beta_{12}, \beta_2, \mu_y, \phi_1, \phi_2)$ to which we assign a vague prior. We simulate from the posterior density $\pi(\theta)$ using the pseudo-marginal algorithm where the likelihood is estimated using the bootstrap particle filter with N particles. A multivariate Student-t random walk proposal on the parameter components transformed to the real line is used.

4.2 Empirical results for the error of the log-likelihood estimator

This section investigates empirically Assumptions 1 and 2 by examining the behaviour of $Z = \log\widehat{p}_N(y \mid \theta) - \log p(y \mid \theta)$ for $T = 40$, 300 and 2700. Corresponding values of N are selected in each case to ensure that the variance of Z evaluated at the posterior mean $\overline{\theta}$ is approximately unity. We use $\delta = 0.5$ in the Euler scheme.

The three plots on the left of Fig. 2 display the histograms corresponding to the density of Z for $\theta = \overline{\theta}$ denoted $g_N(z \mid \overline{\theta})$, which is obtained by running $S = 6000$ particle filters at this value. As $p(y \mid \overline{\theta})$ is unknown, it is estimated by averaging these estimates. The Metropolis–Hastings algorithm is then used to obtain the histograms corresponding to $\pi_N(z \mid \overline{\theta}) = \exp(z)\, g_N(z \mid \overline{\theta})$. We overlay on each histogram a kernel density estimate together with the corresponding assumed density, $g_Z^\sigma(z)$ or $\pi_Z^\sigma(z)$, where σ^2 is the sample variance of Z over the S particle filters. For $T = 40$, there is a discrepancy between the assumed Gaussian densities and the true histograms representing $g_N(z \mid \overline{\theta})$ and $\pi_N(z \mid \overline{\theta})$. In particular, whilst $g_N(z \mid \overline{\theta})$ is well approximated over most of its support, it is slightly lighter tailed than the assumed Gaussian in the right tail and much heavier tailed in the left tail. This translates into a smaller discrepancy between $g_N(z \mid \theta)$ and $\pi_N(z \mid \theta)$ and a higher acceptance rate for the pseudo-marginal algorithm than the Gaussian assumption suggests. For $T = 300$ and $T = 2700$, the assumed Gaussian densities are very accurate.

We also examine Z when θ is distributed according to $\pi(\theta)$. We record 200 samples from $\pi(\theta)$, for $T = 40$, 300 and 2700. For each of these samples, we run the particle filter 300 times in order to estimate the true likelihood at these values. The resulting histograms, corresponding to the densities $\int\pi(\mathrm{d}\theta)\, g_N(z \mid \theta)$ and $\int\pi(\mathrm{d}\theta)\, \pi_N(z \mid \theta)$, are displayed in the middle column of Fig. 2. We similarly examine the density of Z when θ is distributed according to the marginal proposal density in the stationary regime $\int\pi(\mathrm{d}\vartheta)\, q(\vartheta, \theta)$. Here $q(\vartheta, \theta)$ is a multivariate Student-t random walk proposal, with step size proportional to $T^{-1/2}$. The right hand column of Fig. 2 shows the resulting histograms. In both scenarios, Assumptions 1 and 2 are problematic for $T = 40$ as $g_N(z \mid \overline{\theta})$ is not close to being Gaussian as T is too small for the central limit theorem to provide a good approximation. Moreover, since T is small, $\pi(\theta)$ and $\int \pi\,(\mathrm{d}\vartheta)\, q\,(\vartheta,\theta)$ are relatively diffuse. Consequently, $g_N(z \mid \overline{\theta})$ is not close to $g_N(z \mid \theta)$ marginalized over $\pi(\theta)$ or $\int \pi\,(\mathrm{d}\vartheta)\, q\,(\vartheta,\theta)$. For $T = 300$ and $T = 2700$, the assumed densities $g_{\mathrm{Z}}^{\sigma}(z)$ and $\pi_{\mathrm{Z}}^{\sigma}(z)$ are close to the corresponding histograms and Assumptions 1 and 2 appear to capture reasonably well the salient features of the densities associated with Z. In particular, the approximation suggested by the central limit theorem becomes very good. Additionally, $\pi(\theta)$ and $\int \pi\,(\mathrm{d}\vartheta)\, q\,(\vartheta,\theta)$ are sufficiently concentrated to ensure that the variance of Z as a function of θ exhibits little variability.

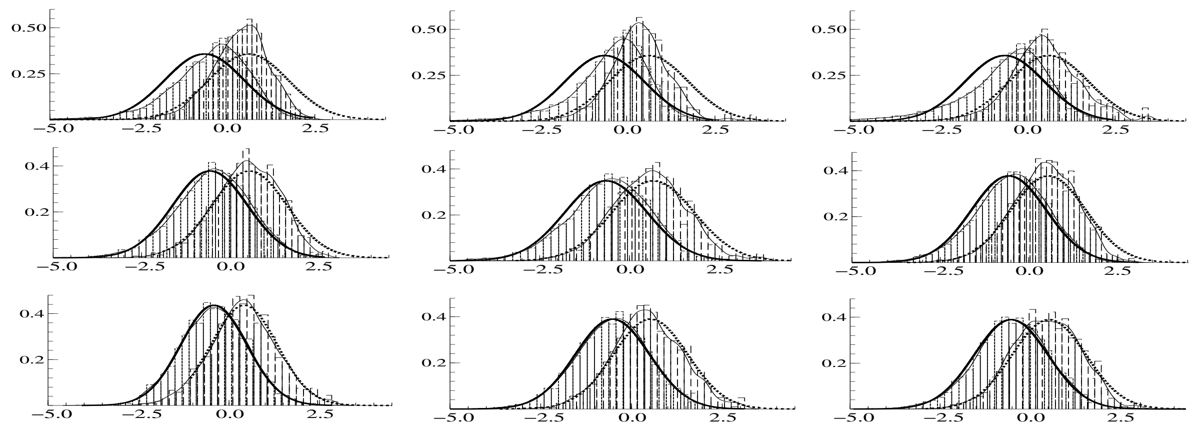

Figure 2: Huang and Tauchen two factor model for S&P 500 data. Top to bottom: $T = 40$, $N = 4$ (top), $T = 300$, $N = 80$ (middle), $T = 2700$, $N = 700$ (bottom). Left to right: histograms and theoretical densities associated with $g_N(z \mid \theta)$ and $\pi_N(z \mid \theta)$ evaluated at the posterior mean $\overline{\theta}$ (left), over values from the posterior $\pi(\theta)$ (middle) and over values from $\int \pi(\mathrm{d}\vartheta)q(\vartheta,\theta)$ (right). The densities $g_{\mathrm{Z}}^{\sigma}(z)$ and $\pi_{\mathrm{Z}}^{\sigma}(z)$ are overlaid (solid lines).

4.3 Empirical results for the pseudo-marginal algorithm

We apply the pseudo-marginal algorithm with $\delta = 0.05$, $T = 300$ and various values of N. The standard deviation $\sigma\left(\overline{\theta}; N\right)$ of $\log \widehat{p}_N(y \mid \overline{\theta})$ is evaluated by Monte Carlo simulations, where $\overline{\theta}$ is the posterior mean. For each value of N, we compute the inefficiencies, denoted by IF, and the corresponding approximate relative computing times, denoted by RCT, of all parameter components. The quantity RCT is computed as $\mathrm{IF}/\sigma^2\left(\overline{\theta}; N\right)$ divided by the inefficiency of Q when $N = 2000$, the latter being an approximation of the inefficiency of Q_{EX}. The results are very similar for all parameter components and so, for ease of presentation, Fig. 3 shows the average quantities over the 9 components. For most parameters, the optimal value for $\sigma\left(\overline{\theta}; N\right)$ is between 1.2 and 1.5, corresponding to $N = 40$ and 60. The results agree with the bound $\mathrm{uRCT}_4(h;\sigma)$ in Section 3.5. This can be partly explained because the inefficiencies associated with $\widetilde{Q}$ for $N = 2000$ are large, suggesting that the inefficiencies associated with $\widetilde{Q}_{\mathrm{EX}}$ are large.

As all the bounds in the paper are based on Q^*, it is useful to assess the discrepancy between Q and Q^*. One approach to explore this discrepancy is to examine the marginal acceptance probability $\overline{\pi}(\varrho_{\mathrm{Q}})$ under Q against $\sigma = \sigma(\overline{\theta}, N)$ as N varies. Using the acceptance criterion (10) of Q^*, we obtain under Assumptions 1 and 2 that $\overline{\pi}(\varrho_{\mathrm{Q}}) \geq 2\Phi(-\sigma/\sqrt{2})\pi(\varrho_{\mathrm{EX}})$. If Q and Q^* are close in the sense of having similar marginal acceptance probabilities, then we expect $\overline{\pi}(\varrho_{\mathrm{Q}})$ to have a similar shape as its lower bound where $\pi(\varrho_{\mathrm{EX}})$ is approximated using $\overline{\pi}(\varrho_{\mathrm{Q}})$ with $N = 2000$. For this model, the two functions on either side of the inequality, displayed in Fig. 3, are similar.

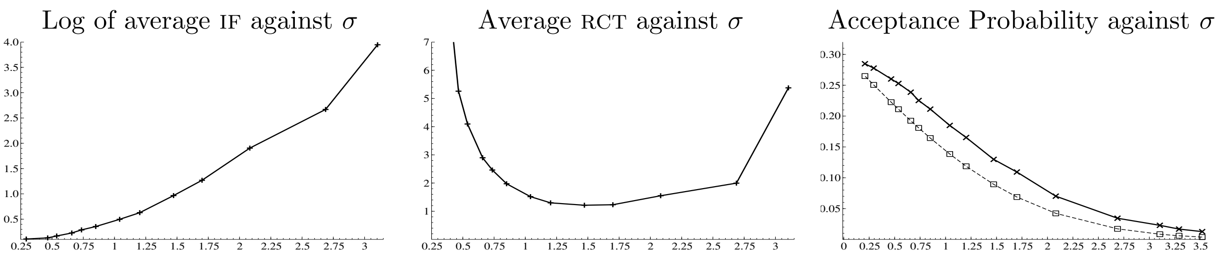

Figure 3: Huang and Tauchen two factor model for S&P 500 data, $T = 300$. Inefficiencies (IF) and Relative Computing Times (RCT) against σ, where IF is computed by averaging over the 9 parameter components. Right panel: The marginal acceptance probability $\overline{\pi}(\varrho_{\mathrm{Q}})$ (crosses) against σ together with the lower bound (squares) $2\Phi(-\sigma/\sqrt{2})\pi(\varrho_{\mathrm{EX}})$.

Acknowledgements

The authors would like to thank the editor, the associate editor and the reviewers for their comments which helped to improve the paper significantly. Arnaud Doucet was partially supported by EPSRC and Robert Kohn was partially supported by an ARC Discovery grant.

Appendix 1

Proof of Lemma 2. It is straightforward to establish that Q^* is $\overline{\pi}$-reversible. Moreover, for any $a, b \geq 0$, $\min(1, a)\min(1, b) \leq \min(1, ab)$ so $\alpha_{Q^*}\left\{(\theta, z), (\vartheta, w)\right\} \leq \alpha_{Q}\left\{(\theta, z), (\vartheta, w)\right\}$ for any θ, z, ϑ, w. Hence, Theorem 4 in Tierney (1998), which is a general state-space extension of Peskun (1973), applies and yields the result. $\square$

Proof of Theorem 1. Without loss of generality, let $h \in L_0^2(\Theta, \pi)$. By Theorem 6 of Andrieu & Vihola (2012), $\mathrm{IF}\left(h, Q_{\mathrm{EX}}\right) \leq \mathrm{IF}\left(h, Q\right)$ and, by Lemma 2, $\mathrm{IF}\left(h, Q\right) \leq \mathrm{IF}\left(h, Q^*\right)$, where $\mathrm{IF}\left(h, Q^*\right) < \infty$ by assumption. Hence, $\mathrm{IF}\left(h, Q_{\mathrm{EX}}\right) < \infty$ and Proposition 2 applied to Q_{EX} yields that $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$, $\widetilde{\pi}\left(h^2/\varrho_{\mathrm{EX}}^2\right) < \infty$ and

$$\pi\left(h^2\right)\left\{1 + \mathrm{IF}\left(h, Q_{\mathrm{EX}}\right)\right\} = \pi\left(\varrho_{\mathrm{EX}}\right)\widetilde{\pi}\left(h^2/\varrho_{\mathrm{EX}}^2\right)\left\{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})\right\}. \tag{30}$$

Since the assumptions of Lemma 3 are satisfied, we can substitute (30) into (19) to obtain

$$1 + \mathrm{IF}\left(h, Q^*\right) = \pi_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}\right)\frac{\left\{1 + \mathrm{IF}\left(h, Q_{\mathrm{EX}}\right)\right\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})}\left[1 + \mathrm{IF}\left\{h/\left(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}\right), \widetilde{Q}^*\right\}\right]. \tag{31}$$

We now provide a spectral representation for $\mathrm{IF}\{h/\left(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}\right), \widetilde{Q}^*\}$. With $\widetilde{\pi}\otimes\widetilde{\pi}_{\mathrm{Z}}\left(\mathrm{d}\theta, \mathrm{d}z\right) = \widetilde{\pi}\left(\mathrm{d}\theta\right)\widetilde{\pi}_{\mathrm{Z}}\left(\mathrm{d}z\right)$,

$$\begin{aligned}
\mathrm{IF}\left\{h/\left(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}\right), \widetilde{Q}^*\right\} &= 1 + 2\sum_{n=1}^{\infty}\frac{\left\langle \varrho_{\mathrm{EX}}^{-1}\varrho_{\mathrm{Z}}^{-1}h, \left(\widetilde{Q}^*\right)^n \varrho_{\mathrm{EX}}^{-1}\varrho_{\mathrm{Z}}^{-1}h\right\rangle_{\widetilde{\pi}\otimes\widetilde{\pi}_{\mathrm{Z}}}}{\widetilde{\pi}\otimes\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-2}\varrho_{\mathrm{EX}}^{-2}h^2\right)} \\
&= 1 + 2\sum_{n=1}^{\infty}\frac{\left\langle \varrho_{\mathrm{Z}}^{-1}, \left(\widetilde{Q}_{\mathrm{Z}}\right)^n \varrho_{\mathrm{Z}}^{-1}\right\rangle_{\widetilde{\pi}_{\mathrm{Z}}}\left\langle \varrho_{\mathrm{EX}}^{-1}h, \left(\widetilde{Q}_{\mathrm{EX}}\right)^n \varrho_{\mathrm{EX}}^{-1}h\right\rangle_{\widetilde{\pi}}}{\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-2}\right)\widetilde{\pi}\left(\varrho_{\mathrm{EX}}^{-2}h^2\right)}
\end{aligned} \tag{32}$$

and, as $\widetilde{Q}_{\mathrm{Z}}$ and $\widetilde{Q}_{\mathrm{EX}}$ are reversible, the following spectral representations, as in (12), hold

$$\phi_n(1/\varrho_{\mathrm{Z}}, \widetilde{Q}_{\mathrm{Z}}) = \frac{\left\langle \varrho_{\mathrm{Z}}^{-1}, \left(\widetilde{Q}_{\mathrm{Z}}\right)^n \varrho_{\mathrm{Z}}^{-1} \right\rangle_{\widetilde{\pi}_{\mathrm{Z}}} - \left\{\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\right\}^2}{\mathbb{V}_{\widetilde{\pi}_{\mathrm{Z}}}\left(\varrho_{\mathrm{Z}}^{-1}\right)} = \int_{-1}^{1} \lambda^n \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda),$$
$$\phi_n(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) = \frac{\left\langle \varrho_{\mathrm{EX}}^{-1} h, \left(\widetilde{Q}_{\mathrm{EX}}\right)^n \varrho_{\mathrm{EX}}^{-1} h \right\rangle_{\widetilde{\pi}}}{\widetilde{\pi}\left(\varrho_{\mathrm{EX}}^{-2} h^2\right)} = \int_{-1}^{1} \omega^n \widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega), \tag{33}$$

where we define $\mathbb{V}_{\widetilde{\pi}_{\mathrm{Z}}}\left(\varrho_{\mathrm{Z}}^{-1}\right) = \widetilde{\pi}_{\mathrm{Z}}\left[\left\{\varrho_{\mathrm{Z}}^{-1} - \widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\right\}^2\right]$, $\widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda) = e(\varrho_{\mathrm{Z}}^{-1}, \widetilde{Q}_{\mathrm{Z}})(\mathrm{d}\lambda)$ and $\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) = e(\varrho_{\mathrm{EX}}^{-1} h, \widetilde{Q}_{\mathrm{EX}})(\mathrm{d}\omega)$ to simplify notation. Using $\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right) = 1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)$, we can rewrite (32) as

$$\begin{aligned}
\mathrm{IF}\left\{h/\left(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}\right), \widetilde{Q}^*\right\} &= 1 + 2\sum_{n=1}^{\infty} \frac{1}{\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-2}\right)} \left\{\mathbb{V}_{\widetilde{\pi}_{\mathrm{Z}}}\left(\varrho_{\mathrm{Z}}^{-1}\right) \int \lambda^n \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda) + \frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)^2}\right\} \int \omega^n \widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) \\
&= 1 + 2\left(1-\gamma\right) \int \frac{\omega}{1-\omega} \widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) + 2\gamma \iint \frac{\lambda\omega}{1-\lambda\omega} \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) \\
&= -1 + 2\left(1-\gamma\right) \int \left(1 + \frac{\omega}{1-\omega}\right) \widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) + 2\gamma \iint \left(1 + \frac{\omega\lambda}{1-\lambda\omega}\right) \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega),
\end{aligned} \tag{34}$$

where the second expression is finite since $\int \left(1+\omega\right)\left(1-\omega\right)^{-1} \widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) = \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) < \infty$ and

$$\gamma = \mathbb{V}_{\widetilde{\pi}_{\mathrm{Z}}}\left(\varrho_{\mathrm{Z}}^{-1}\right)/\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-2}\right) = \left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right) - 1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right). \tag{35}$$

Rearranging (34), we obtain

$$\begin{aligned}
1 + \mathrm{IF}\left\{h/\left(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}\right), \widetilde{Q}^*\right\} &= \left\{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})\right\}\left(1-\gamma\right) + \gamma\beta \\
&= \frac{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)} + \left\{\frac{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right) - 1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)}\right\}\beta,
\end{aligned} \tag{36}$$

with

$$\frac{\beta}{2} = \iint \frac{\widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega)}{1-\omega\lambda} = \sum_{n=0}^{\infty} \phi_n(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})\phi_n(1/\varrho_{\mathrm{Z}}, \widetilde{Q}_{\mathrm{Z}}). \tag{37}$$

By substituting (36) into (31), we obtain the result since

$$\begin{aligned}
\mathrm{IF}\left(h, Q^*\right) &= \pi_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}\right) \frac{\left\{1 + \mathrm{IF}\left(h, Q^{\mathrm{EX}}\right)\right\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})} \left[1 + \mathrm{IF}\left\{h/\left(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}\right), \widetilde{Q}^*\right\}\right] - 1 \\
&= \frac{\left\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\right\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})} \left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right) - \frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}\right\}\beta + \frac{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)} - 1.
\end{aligned} \tag{38}$$

□

Proof of Corollary 1. Dividing (38) by $\mathrm{IF}(h, Q_{\mathrm{EX}})$, we obtain

$$\mathrm{RIF}\left(h, Q^*\right) = \frac{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\left\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\right\}}{\mathrm{IF}(h, Q_{\mathrm{EX}})\left\{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})\right\}} A - \frac{1}{\mathrm{IF}(h, Q_{\mathrm{EX}})}, \tag{39}$$

where A is the quantity in (36) and can be expressed in terms of γ, defined in (35), as

$$A = 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) - 2\gamma \iint \left\{\frac{1}{(1-\omega)} - \frac{1}{(1-\lambda\omega)}\right\} \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega)$$

$$= 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) - 2\gamma \iint \frac{\omega\,(1-\lambda)}{(1-\omega)\,(1-\omega\lambda)} \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega).$$

Lemma 3 ensures that the kernel $\widetilde{Q}_{\mathrm{Z}}$ is positive, implying that $\widetilde{e}_{\mathrm{Z}}\{[0,1)\} = 1$. Hence,

$$\iint \left\{ \frac{\omega\,(1-\lambda)}{(1-\omega)\,(1-\omega\lambda)} - \frac{\omega\,(1-\lambda)}{(1-\omega)} \right\} \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) = \iint \frac{\omega^2(1-\lambda)\lambda}{(1-\omega)\,(1-\omega\lambda)} \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) \geq 0.$$

We can now bound A from above by

$$\begin{aligned}
A &\leq 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) - 2\gamma \iint \frac{\omega\,(1-\lambda)}{(1-\omega)} \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) \\
&= 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) - \gamma \left\{ 1 - \int \lambda \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda) \right\} \int \frac{2\omega}{(1-\omega)} \widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) \\
&= 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) - \gamma(1-\phi_{\mathrm{Z}}) \left\{ \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) - 1 \right\} \\
&= \left\{ 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \right\} \{\phi_{\mathrm{Z}} + (1-\phi_{\mathrm{Z}})(1-\gamma)\} + 2(1-\phi_{\mathrm{Z}})\gamma \qquad (40) \\
&\leq \left\{ 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \right\} \{\phi_{\mathrm{Z}} + (1-\phi_{\mathrm{Z}})(1-\gamma) + 2(1-\phi_{\mathrm{Z}})\gamma\} \\
&= \left\{ 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \right\} \left\{ 2(1-\phi_{\mathrm{Z}}/2) - \frac{(1-\phi_{\mathrm{Z}})}{\pi_{\mathrm{Z}}\,(1/\varrho_{\mathrm{Z}})\,\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})} \right\},
\end{aligned}$$

where we have used the identity $\phi_{\mathrm{Z}} = \int \lambda \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)$. The last inequality is established by noting that $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$ and γ are non-negative. Substituting the expression into (39) establishes Part 1. To establish the inequality of Part 2, we note that if $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \geq 1$, then (40) is bounded from above by

$$\begin{aligned}
&\left\{ 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \right\} \left[\phi_{\mathrm{Z}} + \frac{(1-\phi_{\mathrm{Z}})}{\pi_{\mathrm{Z}}\,(1/\varrho_{\mathrm{Z}})\,\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})} + (1-\phi_{\mathrm{Z}}) \left\{ 1 - \frac{1}{\pi_{\mathrm{Z}}\,(1/\varrho_{\mathrm{Z}})\,\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})} \right\} \right] \\
&= 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}).
\end{aligned}$$

□

Proof of Corollary 2. We establish the upper bound $\mathrm{uRIF}_3\,(h)$ of Part 1 by first noting that (40) implies

$$A \leq \left\{ 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \right\} \{\phi_{\mathrm{Z}} + (1-\phi_{\mathrm{Z}})(1-\gamma)\} + 2(1-\phi_{\mathrm{Z}})\gamma,$$

with A is the quantity in (36), γ given by (35) and $\phi_{\mathrm{Z}} = \int \lambda \widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)$. Upon substituting into (39), we obtain

$$\begin{aligned}
\mathrm{RIF}\,(h, Q^*) + \frac{1}{\mathrm{IF}(h, Q_{\mathrm{EX}})} \leq{}& \frac{\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\}}{\mathrm{IF}(h, Q_{\mathrm{EX}})} \left\{ \phi_{\mathrm{Z}} \pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right) + \frac{(1-\phi_{\mathrm{Z}})}{\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})} \right\} \\
&+ \frac{2(1-\phi_{\mathrm{Z}})\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\}}{\mathrm{IF}(h, Q_{\mathrm{EX}}) \left\{ 1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \right\}} \left\{ \pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right) - \frac{1}{\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})} \right\},
\end{aligned}$$

and, after further manipulations,

$$\begin{aligned}
\mathrm{RIF}\,(h, Q^*) \leq{}& \phi_{\mathrm{Z}} \{\pi_{\mathrm{Z}}\,(1/\varrho_{\mathrm{Z}}) - 1/\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})\} + 1/\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}}) \\
&+ \frac{1}{\mathrm{IF}(h, Q_{\mathrm{EX}})} \left[\phi_{\mathrm{Z}} \{\pi_{\mathrm{Z}}\,(1/\varrho_{\mathrm{Z}}) - 1/\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})\} + \frac{1}{\pi_{\mathrm{Z}}\,(\varrho_{\mathrm{Z}})} - 1 \right]
\end{aligned}$$

$$
\begin{aligned}
&+2\frac{\{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}(1-\phi_{\mathrm{Z}})\\
&\le \phi_{\mathrm{Z}}\left\{\pi_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}+1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\\
&\quad+\frac{1}{\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left[\phi_{\mathrm{Z}}\left\{\pi_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}+\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}-1\right]\\
&\quad+\frac{2}{\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}(1-\phi_{\mathrm{Z}}),
\end{aligned}
$$

as $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})\le \mathrm{IF}(h,Q_{\mathrm{EX}})$ from Proposition 2.

To establish the upper bound $\mathrm{uRIF}_4\left(h\right)$ of Part 2, we use that, in the right hand side of the equality of (39), the term β defined in (37) and appearing in A satisfies the inequality

$$
\beta=\iint\frac{2}{(1-\lambda\omega)}\widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)\widetilde{e}_{\mathrm{EX}}(\mathrm{d}\omega)\le\int\frac{2}{(1-\lambda)}\widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)=1+\mathrm{IF}(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}}), \tag{41}
$$

where $\mathrm{IF}(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}})=\int(1+\lambda)/(1-\lambda)\widetilde{e}_{\mathrm{Z}}(\mathrm{d}\lambda)<\infty$, by assumption. Therefore, upon substituting into (39), we obtain

$$
\begin{aligned}
\mathrm{RIF}\left(h,Q^{*}\right)&\le\frac{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\{1+\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{\mathrm{IF}(h,Q_{\mathrm{EX}})\{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})\}}\left\{1-\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}\right\}\{1+\mathrm{IF}(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}})\}\\
&\quad+\frac{\{1+\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{\mathrm{IF}(h,Q_{\mathrm{EX}})}\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}-\frac{1}{\mathrm{IF}(h,Q_{\mathrm{EX}})}\\
&=\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}+\frac{\{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}\{1+\mathrm{IF}(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}})\}\\
&\quad+\frac{1}{\mathrm{IF}(h,Q_{\mathrm{EX}})}\left\{\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}-1\right\}\\
&\le\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}+\frac{\left\{1+1/\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})\right\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}\{1+\mathrm{IF}(1/\varrho_{\mathrm{Z}},\widetilde{Q}_{\mathrm{Z}})\}\\
&\quad+\frac{1}{\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}-1\right\},
\end{aligned}
$$

as $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})\le \mathrm{IF}(h,Q_{\mathrm{EX}})$.

To establish the inequality of Part 3, we combine (36) and (39) to obtain

$$
\begin{aligned}
\mathrm{RIF}\left(h,Q^{*}\right)&=\frac{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\gamma\beta+(1-\gamma)\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}-\frac{1}{\mathrm{IF}(h,Q_{\mathrm{EX}})}\\
&=\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}+\frac{\{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}\beta+\frac{\{1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)-1\}}{\mathrm{IF}(h,Q_{\mathrm{EX}})}\\
&\ge\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}+\frac{2\{1+1/\mathrm{IF}(h,Q_{\mathrm{EX}})\}}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}+\frac{\{1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)-1\}}{\mathrm{IF}(h,Q_{\mathrm{EX}})}\\
&\ge\frac{1}{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)}+\frac{2}{1+\mathrm{IF}(h/\varrho_{\mathrm{EX}},\widetilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)-1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)\right\}.
\end{aligned} \tag{42}
$$

The first inequality follows because the identity for β given in (41) shows that $\beta\ge 2$ when $\widetilde{Q}_{\mathrm{EX}}$ is positive. The second inequality follows from $\mathrm{IF}(h,Q_{\mathrm{EX}})\ge 0$.

From (39), we have $\mathrm{RIF}\left(h,Q^{*}\right)\ge 1/\pi_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}\right)$ as the second and third terms on the left hand side of the inequality (42) are both positive. This establishes the inequality of Part 4. We examine the limit of $\mathrm{RIF}\,(h, Q^*)$ as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \to \infty$, again noting that $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \le \mathrm{IF}(h, Q_{\mathrm{EX}})$. Using the inequality for β given by (41) and the fact that $\mathrm{IF}(1/\varrho_{\mathrm{Z}}, \widetilde{Q}_{\mathrm{Z}}) < \infty$ by Lemma 3, we obtain the limiting form, as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}}) \to \infty$, given by (27) for $\mathrm{RIF}\,(h, Q^*)$. □

Appendix 2

We exploit the two upper bounds $\mathrm{uRCT}_3(h;\sigma)$ and $\mathrm{uRCT}_4(h;\sigma)$, together with the lower bound $\mathrm{lRCT}_1(h;\sigma)$, in order to find an interval where the optimal value σ_{opt} for $\mathrm{RCT}(h, Q^*;\sigma)$ lies. We consider how this interval varies as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$ increases. To do this, we compute the interval where $\mathrm{lRCT}_1(h;\sigma)$ lies below the minimum of $\inf_\sigma \mathrm{uRCT}_3(h;\sigma)$, and $\inf_\sigma \mathrm{uRCT}_4(h;\sigma)$. Table 1 displays this interval together with the minimum of the two upper bounds and the minimum of the lower bound. It is straightforward to see that σ_{opt} is contained in this interval and $\mathrm{RCT}(h, Q^*;\sigma_{\mathrm{opt}})$ is contained in the corresponding interval in Table 1. It is apparent that the intervals tighten as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$ increases. Similarly the endpoints of the interval containing $\mathrm{RCT}(h, Q^*;\sigma_{\mathrm{opt}})$ both decrease whilst the lower endpoint of the interval containing σ_{opt} increases.

Table 1. *Sandwiching results based upon different values of* $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$. *These are based upon the upper bounds for* $\mathrm{RCT}(h, Q^*;\sigma)$ *given by* $\mathrm{uRCT}_3(h;\sigma)$ *and* $\mathrm{uRCT}_4(h;\sigma)$ *and upon the lower bound* $\mathrm{lRCT}_1(h;\sigma)$.

$\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \widetilde{Q}_{\mathrm{EX}})$	1	10	25	100	1000
$\mathrm{RCT}(h, Q^*;\sigma_{\mathrm{opt}})$	(3.201, 5.327)	(2.020, 2.256)	(1.773, 1.876)	(1.595, 1.625)	(1.518, 1.522)
σ_{opt}	(0.548, 1.572)	(1.018, 1.598)	(1.205, 1.658)	(1.421, 1.730)	(1.607, 1.730)

Supplementary Material

A Contents

This supplement provides some technical proofs and an additional example for the paper "Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator". Section B presents the proof of Proposition 2. Section C presents the proofs of Propositions 3 and 4 and Lemmas 1 and 3. Section D presents some auxiliary technical results. Section E illustrates the upper bound on the inefficiency of Part 4 of Corollary 2 and compares it to the results in Sherlock et al. (2013). Section F applies the pseudo-marginal algorithm to a linear Gaussian state-space model and presents additional simulation results for the stochastic volatility model discussed in the main paper. Section G explains how the bounds on the inefficiency introduced in Section 3.5 of the main paper are computed.

All code was implemented in the `Ox` language with pre-compiled `C` code for computationally intensive routines.

B Proof of Proposition 2

The proof of Proposition 2 relies on Lemmas 5 to 8, which are given below. Lemmas 5 to 7 establish that $h/\varrho \in L^2(\mathsf{X}, \tilde{\mu})$ and $\mathrm{IF}(h/\varrho, \tilde{P}) < \infty$ whenever $\mathrm{IF}(h, P) < \infty$. To prove this result, we define the map that sends the functional h to h/ϱ as a linear operator between two Hilbert spaces, $\mathcal{H}$ and $\tilde{\mathcal{H}}$ defined below. The space $\mathcal{H}$, respectively $\tilde{\mathcal{H}}$, corresponds to the set of functions having finite inefficiencies under P, respectively under $\tilde{P}$. We then exploit the structure of the Metropolis–Hastings type kernel P to prove that this linear operator is bounded on a dense subspace $\mathcal{H}_P \subset \mathcal{H}$, which allows us to extend the operator to $\mathcal{H}$. The proof is then completed by checking that the unique extension constructed this way is the one required. Lemma 8 is a general result on the central limit theorem for reversible and ergodic Markov chains which are not started in their stationary regime. The proof of Proposition 2 uses these preliminary results to establish the identity of interest.

Using the notation of Proposition 2, we write $\|\cdot\|_\mu$, $\langle\cdot,\cdot\rangle_\mu$ for the norm and inner product of $L^2(\mathsf{X},\mu)$, with a similar notation for $L^2(\mathsf{X},\tilde{\mu})$. By reversibility of P and $\tilde{P}$ with respect to μ and $\tilde{\mu}$ respectively, it is easy to check that $(I-P)$ and $(I-\tilde{P})$ are positive, self-adjoint operators on $L^2(\mathsf{X},\mu)$ and $L^2(\mathsf{X},\tilde{\mu})$ respectively. By Theorem 13.11 in Rudin (1991), the inverses $(I-P)^{-1}$ and $(I-\tilde{P})^{-1}$ are densely defined and self-adjoint. They are also positive, since for any $f \in \mathrm{Domain}\{(I-P)^{-1}\}$, there exists a function g such that $f = (I-P)g$, and thus

$$\langle (I-P)^{-1}f, f\rangle_\mu = \langle (I-P)^{-1}(I-P)g, (I-P)g\rangle_\mu = \langle g, (I-P)g\rangle_\mu \geq 0,$$

since $I-P$ is positive. Therefore, by Theorem 13.31 in Rudin (1991), there exists a unique, self-adjoint, positive operator $(I-P)^{-1/2}$ such that $(I-P)^{-1} = (I-P)^{-1/2}(I-P)^{-1/2}$. Finally, since $(I-P)^{-1}$ is densely defined, so is $(I-P)^{-1/2}$. Similar considerations show the existence and uniqueness of the positive, self-adjoint operator $(I-\tilde{P})^{-1/2}$, which is densely defined on $L^2(\mathsf{X},\tilde{\mu})$.

We now introduce the inner product spaces $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ and $(\tilde{\mathcal{H}}, \langle\cdot,\cdot\rangle_{\tilde{\mathcal{H}}})$, where

$$\begin{aligned}
\mathcal{H} &= \{f \in L_0^2(\mathsf{X},\mu) : \|f\|_\mu^2 + \|(I-P)^{-1/2}f\|_\mu^2 < \infty\},\\
\langle f, g\rangle_{\mathcal{H}} &= \langle f, g\rangle_\mu + \langle (I-P)^{-1/2}f, (I-P)^{-1/2}g\rangle_\mu,\\
\tilde{\mathcal{H}} &= \{f \in L_0^2(\mathsf{X},\tilde{\mu}) : \|f\|_{\tilde{\mu}}^2 + \|(I-\widetilde{P})^{-1/2}f\|_{\tilde{\mu}}^2 < \infty\},\\
\langle f, g\rangle_{\tilde{\mathcal{H}}} &= \langle f, g\rangle_{\tilde{\mu}} + \langle (I-\widetilde{P})^{-1/2}f, (I-\widetilde{P})^{-1/2}g\rangle_{\tilde{\mu}}.
\end{aligned}$$

Clearly the space $\mathcal{H}$, respectively $\tilde{\mathcal{H}}$, corresponds to the set of functions having finite inefficiencies under P, respectively under $\tilde{P}$.

Lemma 4. *Let P and $\widetilde{P}$ be ergodic. Then $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ and $(\tilde{\mathcal{H}}, \langle\cdot,\cdot\rangle_{\tilde{\mathcal{H}}})$ are Hilbert spaces.*

Proof. Since P and $\widetilde{P}$ are ergodic, the only solutions in $L^2(\mathsf{X},\mu)$ and $L^2(\mathsf{X},\tilde{\mu})$, of $h = Ph$, respectively $g = \widetilde{P}g$, are almost surely constant with respect to μ and $\tilde{\mu}$. If $f = Pf$ $\mu-$almost surely, then

$$0 = \|f - Pf\|_\mu^2 = \int_{-1}^{1} (1-\lambda)^2 e(f, P)(\mathrm{d}\lambda),$$

where $e(f,P)$ is the spectral measure of P with respect to the function f, and therefore $e(f,P)$ must be an atom at 1, which is impossible as P is ergodic; see the proof of Lemma 17 in Häggström & Rosenthal (2007) and Proposition 17.4.1 in Meyn & Tweedie (2009). Since $I-P$ and $I-\tilde{P}$ are injective in $L_0^2(\mu)$ and $L_0^2(\tilde{\mu})$ respectively, $(I-P)^{1/2}$ and $(I-\widetilde{P})^{1/2}$ must also be injective on the corresponding spaces, because $(I-P)^{1/2}h = 0$ implies $(I-P)h = 0$. In addition, as mentioned above, these operators are self-adjoint and thus their inverses, $(I-P)^{-1/2}$ and $(I-\widetilde{P})^{-1/2}$, are densely defined and self-adjoint by Theorem 13.11 in Rudin (1991).

By Theorem 13.9 in Rudin (1991), $(I-P)^{-1/2}$ and $(I-\widetilde{P})^{-1/2}$ are *closed operators* on $L_0^2(\mathsf{X},\mu)$ and $L_0^2(\mathsf{X},\tilde{\mu})$ respectively because they are self-adjoint. By Section 13.1 in Rudin (1991), a possibly unbounded operator T on a Hilbert space $\mathcal{F}$ is said to be closed if and only if its graph

$$\mathfrak{G}(T) = \{(x, Tx) : x \in \mathcal{F}\},$$

is a closed subset of $\mathcal{F} \times \mathcal{F}$. Equivalently T is closed if $x_n \to x$ and $Tx_n \to y$ implies $Tx = y$. In particular, x is in the domain of T. It follows that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ and $(\tilde{\mathcal{H}}, \langle\cdot,\cdot\rangle_{\tilde{\mathcal{H}}})$ are Hilbert spaces by Proposition 1.4 in Schmüdgen (2012). □

Lemma 5. *The linear space*

$$\mathcal{H}_P = \text{Range}\big\{(I-P)\big\} = \{h \in L_0^2(\mathsf{X},\mu) : h = (I-P)g,\ g \in L^2(\mathsf{X},\mu)\}$$

is dense in $\mathcal{H}$ in the norm induced by $\langle\cdot,\cdot\rangle_{\mathcal{H}}$.

Proof. For $h \in \mathcal{H}$, we have

$$\|(I-P)^{-1/2}h\|_\mu = \int_{-1}^{1} \frac{e(h,P)(\mathrm{d}\lambda)}{1-\lambda} < \infty,$$

where $e(h,P)$ is the spectral measure associated with h and P. For $\epsilon > 0$, define

$$h_\epsilon = (I-P)\big\{(1+\epsilon)I - P\big\}^{-1} h \in \mathcal{H}_P.$$

Then,

$$\begin{aligned}
\|(I-P)^{-1/2}(h_\epsilon - h)\|_\mu^2 &= \Big\|(I-P)^{-1/2}\Big[(I-P)\{(1+\epsilon)I-P\}^{-1} - I\Big]h\Big\|_\mu^2 \\
&= \int_{-1}^{1} \frac{1}{1-\lambda}\Big(\frac{1-\lambda}{1+\epsilon-\lambda} - 1\Big)^2 e(h,P)(\mathrm{d}\lambda) \\
&= \int_{-1}^{1} (1-\lambda)\left(\frac{1}{1+\epsilon-\lambda} - \frac{1}{1-\lambda}\right)^2 e(h,P)(\mathrm{d}\lambda) \\
&= \int_{-1}^{1} \frac{\epsilon^2 e(h,P)(\mathrm{d}\lambda)}{(1+\epsilon-\lambda)^2(1-\lambda)}.
\end{aligned}$$

The integrand is bounded above by $1/(1-\lambda)$, since $|\lambda| \le 1$ implies that $\epsilon^2/(1+\epsilon-\lambda)^2 \le 1$, and thus, by dominated convergence, the integral vanishes as $\epsilon \to 0$. Since $I-P$ is bounded, $\|h_\epsilon - h\|_\mu$ also vanishes. Therefore, $h_\epsilon \to h$ in $\mathcal{H}$. In particular, $\mathcal{H}_P$ is dense in $\mathcal{H}$. □

Lemma 6. *If* $\mathrm{IF}(h,P) < \infty$, *then* $h/\varrho \in L^2(\mathsf{X},\tilde{\mu})$ *and* $\mathrm{IF}(h/\varrho,\tilde{P}) < \infty$.

Proof. For $h \in \mathcal{H}_P$, there exists $g \in L^2(\mathsf{X},\mu)$ such that

$$h(x) = (I-P)g(x) = \varrho(x)(I-\widetilde{P})g(x).$$

Therefore, $h(x)/\varrho(x) = (I-\widetilde{P})g(x) \in \tilde{\mathcal{H}}$, since $\|g\|_{\tilde{\mu}}^2 \le \|g\|_\mu^2/\mu(\varrho)$. Thus, we can define the multiplication operator $T : \mathcal{H}_P \to \tilde{\mathcal{H}}$ by $T : h \to h/\varrho$.

Let $h\,(x) = (I-P)g(x)$. Then,

$$\|h\|_{\mathcal{H}}^2 = \|h\|_\mu^2 + \langle h, (I-P)^{-1}(I-P)g\rangle_\mu \ge \langle h, g\rangle_\mu,$$

because $I-P$ is self-adjoint. Similarly,

$$\begin{aligned}
\|Th\|_{\tilde{\mathcal{H}}}^2 &= \|h/\varrho\|_{\tilde{\mu}}^2 + \|(I-\widetilde{P})^{-1/2}(h/\varrho)\|_{\tilde{\mu}}^2 \\
&= \|(I-\widetilde{P})g\|_{\tilde{\mu}}^2 + \|(I-\widetilde{P})^{1/2}g\|_{\tilde{\mu}}^2 \le K\|(I-\widetilde{P})^{1/2}g\|_{\tilde{\mu}}^2,
\end{aligned}$$

where $K = 1 + \Big\|(I-\widetilde{P})^{1/2}\Big\|$ with $\|(I-\widetilde{P})^{1/2}\|$ the finite norm of the operator $(I-\widetilde{P})^{1/2}$. Recalling that $h(x) = (I-P)g(x) = \varrho(x)(I-\widetilde{P})g(x)$, we obtain

$$\|(I-\widetilde{P})^{1/2}g\|_{\tilde{\mu}}^2 = \int g(x)(I-\widetilde{P})g(x)\frac{\varrho(x)\mu(\mathrm{d}x)}{\mu(\varrho)} = \frac{\langle g, h\rangle_\mu}{\mu(\varrho)}.$$

It follows that $T:\mathcal{H}_P \to \tilde{\mathcal{H}}$ is bounded as

$$\sup_{h\in\mathcal{H}_P}\frac{\|Th\|^2_{\tilde{\mathcal{H}}}}{\|h\|^2_{\mathcal{H}}} \le \frac{K\|(I-\widetilde{P})^{1/2}g\|^2_{\tilde{\mu}}}{\|h\|^2_\mu + \langle h,g\rangle_\mu} = \frac{K}{\mu(\varrho)}\frac{\langle g,h\rangle_\mu}{\|h\|^2_\mu + \langle g,h\rangle_\mu} \le \frac{K}{\mu(\varrho)}.$$

Since $\mathcal{H}_P$ is dense, given $h\in\mathcal{H}$, there is a sequence $h_n\in\mathcal{H}_P$ such that $\|h_n-h\|_{\mathcal{H}}\to 0$, as $n\to\infty$. This, in particular, implies that h_n is a Cauchy sequence in $\mathcal{H}$, that is

$$\|h_n - h_m\|_{\mathcal{H}} \to 0, \quad \text{as } n\ge m\to\infty.$$

Since h_n and h_n-h_m are in $\mathcal{H}_P$, $Th_n, T(h_n-h_m)\in\tilde{\mathcal{H}}$ and, from the above calculation,

$$\|Th_n - Th_m\|_{\tilde{\mathcal{H}}} \le \frac{K}{\mu(\varrho)}\|h_n-h_m\|_{\mathcal{H}} \to 0,$$

as $m,n\to\infty$. Therefore, Th_n forms a Cauchy sequence in $\tilde{\mathcal{H}}$; in particular h_n and $(I-\widetilde{P})^{-1/2}h_n$ are Cauchy in $L^2(\mathsf{X},\tilde{\mu})$. Since $L^2(\mathsf{X},\tilde{\mu})$ is complete, we have $h_n\to g\in L^2(\mathsf{X},\tilde{\mu})$ and $(I-\widetilde{P})^{-1/2}h_n\to f\in L^2(\mathsf{X},\tilde{\mu})$. Since $Q=(I-\widetilde{P})^{-1/2}$ is a closed operator, we can conclude that

$$g\in \text{Domain}\left\{Q\right\}, \quad Qg=f,$$

and, in particular, $g\in\tilde{\mathcal{H}}$.

To complete the proof, we need to show that $g=h/\varrho$. Recall that $h_n\to h$ in $\mathcal{H}$ implies that $\|h_n-h\|_\mu\to 0$. We can then choose a subsequence $n(k)$ such that $h_{n(k)}\to h$ μ-almost surely. Since $\tilde{\mu}$ is absolutely continuous with respect to μ, we also have $h_{n(k)}/\varrho\to h/\varrho$ $\tilde{\mu}$-almost surely.

In addition, we know that $Th_n = h_n/\varrho\to g$ in $\tilde{\mathcal{H}}$ and thus in $L^2(\mathsf{X},\tilde{\mu})$. Therefore, $h_{n(k)}/\varrho\to g$ in $L^2(\mathsf{X},\tilde{\mu})$. We can now choose a further subsequence $n'(k)$ such that $h_{n'(k)}/\varrho\to g$ $\tilde{\mu}$-almost surely. Since $h_{n(k)}/\varrho$ also converges to h/ϱ $\tilde{\mu}$-almost surely, and $n'(k)$ is a subsequence of $n(k)$, we conclude that $g=h/\varrho$ $\tilde{\mu}$-almost surely. □

Lemma 7. *Assume Π is μ-reversible and ergodic, $h\in L^2_0(\mathsf{X},\mu)$ and $\mathrm{IF}(h,\Pi)<\infty$. Let $(X_i)_{i\ge1}$ be a Markov chain evolving according to Π. If $X_1\sim\nu$, where ν is absolutely continuous with respect to μ then, as $n\to\infty$,*

$$n^{-1/2}\sum_{i=1}^n h(X_i)\longrightarrow \mathcal{N}\left\{0;\mu\left(h^2\right)\mathrm{IF}(h,\Pi)\right\}.$$

Proof of Lemma 7. Let $e(h,\Pi)(\mathrm{d}\lambda)$ be the associated spectral measure and define $S_n=\sum_{i=1}^n h\left(X_i\right)$. Then,

$$\begin{aligned}\frac{1}{n}\mathrm{E}_\mu\left\{\mathrm{E}\left(S_n|\,X_1\right)^2\right\} &= \frac{1}{n}\int\mu\left(dx\right)\left\{\sum_{i=0}^{n-1}\Pi^i h\left(x\right)\right\}^2 = \frac{1}{n}\int_{-1}^1\left(\sum_{i=0}^{n-1}\lambda^i\right)^2 e(h,\Pi)(\mathrm{d}\lambda)\\ &= \int_{-1}^1\left(\frac{1}{n}\sum_{i=0}^{n-1}\lambda^i\right)\frac{1-\lambda^n}{1-\lambda}e(h,\Pi)(\mathrm{d}\lambda)\to 0\end{aligned}$$

as $n\to\infty$ by dominated convergence, since $\int\left(1-\lambda\right)^{-1}e(h,\Pi)(\mathrm{d}\lambda)<\infty$ by assumption. Hence, equation (4) in Wu & Woodroofe (2004) holds with $\sigma_n^2=\mathrm{E}_\mu\left(S_n^2\right)\sim\sigma^2 n$, where $\sigma^2=\mu\left(h^2\right)\mathrm{IF}(h,\Pi)$. It is straightforward to check, with calculations similar to the above, that the solution to the approximate Poisson equation given in the proof of Theorem 1.3 in Kipnis & Varadhan (1986),

$$h_n\left(x\right)=\left\{\left(1+\frac{1}{n}\right)I-\Pi\right\}^{-1}h\left(x\right),$$

satisfies equation (5) in Wu & Woodroofe (2004), while equation (1.10) in Kipnis & Varadhan (1986) shows that $H_n(x_0, x_1) := h_n(x_1) - \Pi h_n(x_0)$ converges in $L^2(\mathsf{X} \times \mathsf{X}, \mu \otimes \Pi)$. Therefore, the conditions of Corollary 2 in Wu & Woodroofe (2004) are satisfied so the statement of the lemma follows from their equation (10); see their comments after this equation. □

Proof of Proposition 2. Let $(X_i)_{i\geq 1}$ be a Markov chain evolving according to P and $(\widetilde{X}_i, \tau_i)_{i\geq 1}$ the associated jump chain representation evolving according to $\overline{P}$, as defined in Lemma 1. We denote by $\mathrm{P}_{\nu,\Pi}$ the law of a Markov chain with initial distribution ν and transition kernel Π. By Theorem 1.3 in Kipnis & Varadhan (1986), we have under $\mathrm{P}_{\mu,P}$

$$S_n = \sum_{i=1}^{n} h(X_i) = M_n + \xi_n, \tag{43}$$

where M_n is a square integrable martingale with respect to the natural filtration of $(X_i)_{i\geq 1}$, while we have the following convergence in probability

$$n^{-1/2} \sup_{1\leq i\leq n} |\xi_i| \xrightarrow{\mathrm{P}_{\mu,P}} 0. \tag{44}$$

Define $T_n = \tau_1 + \cdots + \tau_n$. The kernel $\overline{P}$ is ergodic because $\widetilde{P}$ is ergodic. Hence, $\mathrm{P}_{\widetilde{\mu},\widetilde{P}}$−almost surely,

$$\frac{T_n}{n} \to \widetilde{\mu}(1/\varrho) = \frac{1}{\mu(\varrho)}. \tag{45}$$

The above limit also holds $\mathrm{P}_{\mu,\widetilde{P}}$-almost surely, since $\mathrm{P}_{\mu,\widetilde{P}}$ is absolutely continuous with respect to $\mathrm{P}_{\widetilde{\mu},\widetilde{P}}$. We first show that

$$\{n/\mu(\varrho)\}^{-1/2} \left(M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right) \xrightarrow{\mathrm{P}_{\mu,P}} 0. \tag{46}$$

Let $\epsilon > 0$ be arbitrary and define the event

$$A_n = \left\{(1-\epsilon)\frac{n}{\mu(\varrho)} \leq T_n < (1+\epsilon)\frac{n}{\mu(\varrho)}\right\}.$$

By (45), we have $\mathrm{P}_{\mu,P}(A_n) \to 1$. The following inequality holds on the event A_n,

$$\begin{aligned}|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}| &\leq |M_{T_n} - M_{\lfloor (1-\epsilon)n/\mu(\varrho)\rfloor}| + |M_{\lfloor n/\mu(\varrho)\rfloor} - M_{\lfloor (1-\epsilon)n/\mu(\varrho)\rfloor}| \\ &\leq 2 \sup_{1\leq i\leq 2\lfloor \epsilon n/\mu(\varrho)\rfloor+1} |\tilde{M}_i|,\end{aligned}$$

where $\tilde{M}_i := M_{\lfloor (1-\epsilon)n/\mu(\varrho)\rfloor + i} - M_{\lfloor (1-\epsilon)n/\mu(\varrho)\rfloor}$ is a square integrable martingale with stationary increments. Thus, for any $\delta > 0$,

$$\begin{aligned}&\mathrm{P}_{\mu,P}\left(\left|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right| > \delta\{n/\mu(\varrho)\}^{1/2}\right) \\ &\qquad\leq \mathrm{P}_{\mu,P}\left(\left\{\left|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right| > \delta\{n/\mu(\varrho)\}^{1/2}\right\} \cap A_n\right) + \mathrm{P}_{\mu,P}(A_n^c) \\ &\qquad\leq \mathrm{P}_{\mu,P}\left(2 \sup_{1\leq i\leq 2\lfloor \epsilon n/\mu(\varrho)\rfloor+1} |\tilde{M}_i| > \delta\{n/\mu(\varrho)\}^{1/2}\right) + o(1) \\ &\qquad\leq C_1 \frac{\mathrm{E}_{\mu,P}\left(\tilde{M}^2_{2\lfloor \epsilon n/\mu(\varrho)\rfloor+1}\right)}{\delta^2 n/\mu(\varrho)} + o(1)\end{aligned}$$

$$\leq C_2 \frac{\epsilon n/\mu(\varrho)}{\delta^2 n/\mu(\varrho)} + o(1) \leq C_3\epsilon/\delta^2 + o(1),$$

where $C_1, C_2, C_3 < \infty$. The third inequality follows from Doob's maximal inequality. The last inequality follows because, for any square integrable martingale $(N_i)_{i\geq 1}$ with stationary increments, $\mathrm{E}_{\mu,P}\left(N_n^2\right) = \mathrm{E}_{\mu,P}\left(N_1^2\right) n$ holds. This bound is uniform in n, and therefore

$$\limsup_{n\to\infty} \mathrm{P}_{\mu,P}\left(\left|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right| > \delta\left\{n/\mu(\varrho)\right\}^{1/2}\right) \leq \epsilon/\delta^2.$$

As $\epsilon > 0$ is arbitrary,

$$\lim_{n\to\infty} \mathrm{P}_{\mu,P}\left(\left|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right| > \delta\left\{n/\mu(\varrho)\right\}^{1/2}\right) = 0,$$

for any $\delta > 0$, and therefore (46) holds. Now, by Proposition 1, $n^{-1/2}S_n \longrightarrow \mathcal{N}\left\{0; \mu\left(h^2\right)\mathrm{IF}(h,P)\right\}$. By the asymptotic negligibility (44) of ξ_n and (46), we have by Slutsky's theorem that

$$\left\{n/\mu(\varrho)\right\}^{-1/2} M_{T_n} \longrightarrow \mathcal{N}\left\{0; \mu\left(h^2\right)\mathrm{IF}(h,P)\right\},$$

equivalently $n^{-1/2}M_{T_n} \longrightarrow \mathcal{N}\left\{0; \mu\left(h^2\right)\mathrm{IF}(h,P)/\mu(\varrho)\right\}$. Finally, note that for any $\delta > 0$,

$$\begin{aligned}
\mathrm{P}_{\mu,P}(|\xi_{T_n}| > \delta n^{1/2}) &\leq \mathrm{P}_{\mu,P}(\{|\xi_{T_n}| > \delta n^{1/2}\} \cap A_n) + \mathrm{P}_{\mu,P}(A_n^c) \\
&\leq \mathrm{P}_{\mu,P}\Big(\sup_{1\leq i\leq \lfloor(1+\epsilon)n/\mu(\varrho)\rfloor} |\xi_i| > \delta n^{1/2}\Big) + o(1) \to 0 \text{ by (44).}
\end{aligned}$$

Therefore, using (43) and Slutsky's theorem, $n^{-1/2}S_{T_n} \to \mathcal{N}\left\{0; \mu\left(h^2\right)\mathrm{IF}(h,P)/\mu(\varrho)\right\}$ when $\widetilde{X}_1 = X_1 \sim \mu$. However, this result also holds when $\widetilde{X}_1 \sim \widetilde{\mu}$, as established in Lemma 7. In particular, the asymptotic variance is the same. Moreover, $(\tilde{X}_i, \tau_i)_{i\geq 1}$ is reversible and ergodic, while Lemma 6 guarantees that $h/\varrho \in L_0^2(\mathsf{X}, \widetilde{\mu})$ and $\mathrm{IF}(h/\varrho, \widetilde{P}) < \infty$. Hence, Proposition 1 applied to $(\tilde{X}_i, \tau_i)_{i\geq 1}$ ensures that the asymptotic variance is also given by the integrated autocovariance time. Equating the two expressions, we obtain

$$\begin{aligned}
\mu\left(h^2\right)\mathrm{IF}(h,P)/\mu(\varrho) &= \overline{\mu}\left(\tau^2 h^2\right) + 2\sum_{n\geq 1}\left\langle \tau h, \overline{P}^n \tau h\right\rangle_{\overline{\mu}} \\
&= \widetilde{\mu}\left(\frac{2-\varrho}{\varrho^2}h^2\right) + 2\sum_{n\geq 1}\left\langle \frac{h}{\varrho}, \widetilde{P}^n \frac{h}{\varrho}\right\rangle_{\widetilde{\mu}} \\
&= \widetilde{\mu}\left(h^2/\varrho^2\right) + \widetilde{\mu}\left(h^2/\varrho^2\right)\mathrm{IF}(h/\varrho, \widetilde{P}) - \mu\left(h^2\right)/\mu(\varrho),
\end{aligned}$$

where the equality in the second line follows from the expression of $\overline{\mu}$ and $\overline{P}$, given in Lemma 1, and the properties of the geometric distribution. This yields the equality of Proposition 2, which can also be written as

$$\frac{1 + \mathrm{IF}(h/\varrho, \widetilde{P})}{1 + \mathrm{IF}(h,P)} = \frac{\mu\left(h^2\right)}{\mu(\varrho)\,\widetilde{\mu}\left(h^2/\varrho^2\right)} = \frac{\mu\left(h^2\right)}{\mu\left(h^2/\varrho\right)} \leq 1;$$

as $0 < \varrho \leq 1$, implying that $\mathrm{IF}(h/\varrho, \widetilde{P}) \leq \mathrm{IF}(h,P)$. □

C Proofs of other technical results in the main paper

Proof of Lemma 1. As P is ψ-irreducible, it is also μ-irreducible as it is μ-invariant; see, for example, Tierney (1994), p. 1759. Hence, for any $x \in \mathsf{X}$ and $A \in \mathcal{X}$ with $\mu(A) > 0$, there exists an $n \geq 1$ such that $P^n(x, A) > 0$. As μ is not concentrated on a single point by assumption, this implies that $\varrho(x) > 0$ for any $x \in \mathsf{X}$. The rest of the proposition follows directly from Lemma 1 in Douc & Robert (2011). □

Proof of Lemma 3. Equations (17) and (18) and the expressions of their associated invariant distributions follow from a direct application of Lemma 1. The positivity of $\widetilde{Q}_{\mathrm{Z}}$ follows directly from Proposition 3, see Remark 1. We write $\widetilde{\pi} \otimes \widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}\theta, \mathrm{d}z) = \widetilde{\pi}(\mathrm{d}\theta)\, \widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}z)$. By applying Proposition 2 to Q^*, we obtain for any $h \in L_0^2(\Theta, \pi)$ that $h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}) \in L_0^2(\Theta \times \mathbb{R}, \widetilde{\pi} \otimes \widetilde{\pi}_{\mathrm{Z}})$, IF $\left\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\right\} < \infty$ and

$$
\begin{aligned}
\pi\left(h^2\right)\{1 + \mathrm{IF}(h, Q^*)\} &= \overline{\pi}(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}})\;\; \widetilde{\pi} \otimes \widetilde{\pi}_{\mathrm{Z}}\left\{h^2/\left(\varrho_{\mathrm{EX}}^2\varrho_{\mathrm{Z}}^2\right)\right\}\left[1 + \mathrm{IF}\left\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\right\}\right] \\
&= \pi(\varrho_{\mathrm{EX}})\, \pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})\, \widetilde{\pi}\left(h^2/\varrho_{\mathrm{EX}}^2\right) \widetilde{\pi}_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}^2\right)\left[1 + \mathrm{IF}\left\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \widetilde{Q}^*\right\}\right].
\end{aligned}
$$

The identity follows easily as $\widetilde{\pi}_{\mathrm{Z}}\left(1/\varrho_{\mathrm{Z}}^2\right) = \pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}})/\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})$ and $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) < \infty$.

To prove the geometric ergodicity of $\widetilde{Q}_{\mathrm{Z}}$, we follow Meyn & Tweedie (2009, Chapter 15). Notice first that

$$
\widetilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w) = \frac{g(\mathrm{d}w)\, \alpha(z, w)}{\varrho_{\mathrm{Z}}(z)} \geq g(\mathrm{d}w)\left\{e^{w-z}\mathbb{I}(w < z) + \mathbb{I}(w \geq z)\right\},
$$

and consider the set $C = (-\infty, z_0]$, where $z_0 > 0$ and $\int_0^{z_0} g(w)\mathrm{d}w > 0$. For any $z \in C$ and $w \geq 0$,

$$
\begin{aligned}
\widetilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w) &\geq g(\mathrm{d}w)\left\{e^{w-z_0}\mathbb{I}(w < z) + \mathbb{I}(w \geq z)\right\} \\
&\geq e^{-z_0} g(\mathrm{d}w) = \varepsilon\, \nu(\mathrm{d}w),
\end{aligned}
$$

where

$$
\varepsilon = e^{-z_0}\int_0^{z_0} g(\mathrm{d}w) \leq 1,
$$

and ν is the probability measure concentrated on $[0, z_0] \subset C$, given by

$$
\nu(\mathrm{d}w) = \frac{g(\mathrm{d}w)\,\mathbb{I}(0 \leq w \leq z_0)}{\int_0^{z_0} g(\mathrm{d}w)}.
$$

Hence, C is a small set.

To complete the proof of geometric ergodicity of $\widetilde{Q}_{\mathrm{Z}}$, we check that $V(z) = 1/\varrho_{\mathrm{Z}}(z)$ satisfies a geometric drift condition. Note that $V(z) \geq 1$ for any z, and

$$
\begin{aligned}
\frac{\int \widetilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w)\, V(w)}{V(z)} &= e^{-z}\int_{-\infty}^{z} \frac{e^w g(\mathrm{d}w)}{\varrho_{\mathrm{Z}}(w)} + \int_z^{\infty} \frac{g(\mathrm{d}w)}{\varrho_{\mathrm{Z}}(w)} \\
&= e^{-z}\int_{-\infty}^{z} \frac{\pi_{\mathrm{Z}}(\mathrm{d}w)}{\varrho_{\mathrm{Z}}(w)} + \int_z^{\infty} \frac{g(\mathrm{d}w)}{\varrho_{\mathrm{Z}}(w)}.
\end{aligned} \tag{47}
$$

We have $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) < \infty$, as established earlier, because $\mathrm{IF}(h, Q^*) < \infty$ by assumption. It follows that the first integral on the right hand side of (47) is bounded. To prove that the second integral is bounded, we use the fact that $\varrho_{\mathrm{Z}}(z)$ is a non-increasing function. We have

$$
\varrho_{\mathrm{Z}}(z) = 1 - G(z) + e^{-z}\Pi(z),
$$

where G, respectively Π, is the cumulative distribution function of g, respectively π_{z}, so its derivative with respect to z is equal to

$$\varrho_{\mathrm{z}}'(z) = -g(z) + e^{-z}\pi(z) - e^{-z}\Pi(z) = -e^{-z}\Pi(z) \leq 0.$$

It follows that the second term on the right hand side of (47) is bounded by

$$\begin{aligned}\int_z^\infty \frac{g(\mathrm{d}w)}{\varrho_{\mathrm{z}}(w)} &\leq \int_{-\infty}^\infty \frac{g(\mathrm{d}w)}{\varrho_{\mathrm{z}}(w)} = \int_{-\infty}^0 \frac{g(\mathrm{d}w)}{\varrho_{\mathrm{z}}(w)} + \int_0^\infty \frac{g(\mathrm{d}w)}{\varrho_{\mathrm{z}}(w)} \\ &\leq \frac{1}{\varrho_{\mathrm{z}}(0)} \int_{-\infty}^0 g(\mathrm{d}w) + \int_0^\infty \frac{e^{-w}\pi_{\mathrm{z}}(\mathrm{d}w)}{\varrho_{\mathrm{z}}(w)} < \infty.\end{aligned}$$

Therefore, for any $0 < \lambda < 1$, there exists $z_0' > 0$ such that

$$\frac{\int \widetilde{Q}_{\mathrm{z}}(z, \mathrm{d}w)\, V(w)}{V(z)} \leq \lambda,$$

for all $z \geq z_0'$. We now establish that

$$\sup_{z \leq z_0'} \int \widetilde{Q}_{\mathrm{z}}(z, \mathrm{d}w)\, V(w) < \infty.$$

As $\varrho_{\mathrm{z}}(z)$ is a non-increasing function, it follows that for $z \leq z_0'$

$$\begin{aligned}\int \widetilde{Q}_{\mathrm{z}}(z, \mathrm{d}w)\, V(w) &= \int \frac{g(\mathrm{d}w)\,\alpha(z,w)}{\varrho_{\mathrm{z}}(z)\,\varrho_{\mathrm{z}}(w)} \\ &\leq \frac{1}{\varrho_{\mathrm{z}}(z_0')} \int \frac{g(\mathrm{d}w)\,\alpha(z,w)}{\varrho_{\mathrm{z}}(w)} \\ &= \frac{1}{\varrho_{\mathrm{z}}(z_0')} \frac{\int \widetilde{Q}_{\mathrm{z}}(z, \mathrm{d}w)\, V(w)}{V(z)}.\end{aligned}$$

We now show that

$$\sup_z \frac{\int \widetilde{Q}_{\mathrm{z}}(z, \mathrm{d}w)\, V(w)}{V(z)} < \infty.$$

The first term on the right hand side of (47) is bounded by

$$e^{-z} \int_{-\infty}^z \frac{\pi_{\mathrm{z}}(\mathrm{d}w)}{\varrho_{\mathrm{z}}(w)} \leq \frac{e^{-z}}{\varrho_{\mathrm{z}}(z)} \int_{-\infty}^z \pi_{\mathrm{z}}(\mathrm{d}w) \leq \frac{e^{-z}}{e^{-z}\Pi(z)}\Pi(z) = 1,$$

while we have already shown that the second term on right hand side of (47) is bounded.

Hence, we can conclude that, for any $0 < \lambda < 1$, there exists $z_0 > 0$ and $b < \infty$ such that

$$\int \widetilde{Q}_{\mathrm{z}}(z, \mathrm{d}w)\, V(w) \leq \lambda V(z) + b\mathbb{I}_C(z),$$

where $C = (-\infty, z_0]$.

The inequality $\mathrm{IF}(1/\varrho_{\mathrm{z}}, \widetilde{Q}_{\mathrm{z}}) < \infty$ now follows because $\widetilde{Q}_{\mathrm{z}}$ is geometrically ergodic with drift function $V(z) = 1/\varrho_{\mathrm{z}}(z)$ and $\widetilde{\pi}_{\mathrm{z}}(1/\varrho_{\mathrm{z}}^2) < \infty$. □

Proof of Proposition 3. If $\langle f, \widetilde{P}f\rangle_{\widetilde{\mu}} \geq 0$ for any $f \in L^2(\mathsf{X}, \widetilde{\mu})$, then $\widetilde{P}$ is positive by definition, implying the positivity of P as $L^2(\mathsf{X}, \mu) \subseteq L^2(\mathsf{X}, \widetilde{\mu})$ and

$$\langle f, Pf\rangle_\mu = \mu(\varrho)\,\langle f, \widetilde{P}f\rangle_{\widetilde{\mu}} + \mu\left\{(1-\varrho)\, f^2\right\}.$$

For a proposal of the form $q(x,y) = \int s(x,z)\, s(y,z)\, \chi(\mathrm{d}z)$, Lemma 3.1 in Baxendale (2005) establishes that $\langle f, \widetilde{P} f\rangle_{\widetilde{\mu}} \geq 0$ for any $f \in L^2(\mathsf{X}, \widetilde{\mu})$. For a ν-reversible proposal such that $\nu(x)\, q(x,y) = \int r(x,z)\, r(y,z)\, \chi(\mathrm{d}z)$, we have for any $f \in L^2(\mathsf{X}, \widetilde{\mu})$,

$$
\begin{aligned}
&\mu(\varrho) \langle f, \widetilde{P} f\rangle_{\widetilde{\mu}} \\
&= \iint f(x) f(y) \nu(x) q(x,y) \min\left\{\frac{\mu(x)}{\nu(x)}, \frac{\mu(y)}{\nu(y)}\right\} \mathrm{d}x\mathrm{d}y \\
&= \iint f(x) f(y) \nu(x) q(x,y) \left[\int_0^\infty \mathbb{I}_{\{0,\mu(x)/\nu(x)\}}(t)\, \mathbb{I}_{\{0,\mu(y)/\nu(y)\}}(t)\, \mathrm{d}t\right] \mathrm{d}x\mathrm{d}y \\
&= \int_0^\infty \left[\iiint f(x) r(x,z) f(y) r(y,z)\, \mathbb{I}_{\{0,\mu(x)/\nu(x)\}}(t)\, \mathbb{I}_{\{0,\mu(y)/\nu(y)\}}(t)\, \mathrm{d}x\mathrm{d}y\ \chi(\mathrm{d}z)\right] \mathrm{d}t \\
&= \int_0^\infty \left(\int \left[\int f(x) r(x,z)\, \mathbb{I}_{\{0,\mu(x)/\nu(x)\}} \mathrm{d}x\right]^2 \chi(\mathrm{d}z)\right) \mathrm{d}t \geq 0,
\end{aligned}
$$

by a repeated application of Fubini's theorem. □

Proof of Proposition 4. Theorem 2.2 in Roberts & Tweedie (1996) establishes the ergodicity of Q_{EX}. We extend their argument to prove the ergodicity of Q^*. For the ball $B(\theta, L)$ centred at θ of radius L, we define

$$
\eta(\theta, L) = \left\{\sup_{\vartheta \in B(\theta,L)} \pi(\vartheta)\right\}^{-1} \inf_{\vartheta \in B(\theta,L)} \pi(\vartheta),
$$

which, by assumption, is such that $0 < \eta(\theta, L) < \infty$. Then, we have for any $(\theta, z) \in \Theta \times \mathbb{R}$, $\vartheta \in B(\theta, \delta)$ and $w \in \mathbb{R}$,

$$
\begin{aligned}
Q^*\{(\theta, z), (d\vartheta, \mathrm{d}w)\} &\geq q(\theta, \vartheta)\, \alpha_{\mathrm{EX}}(\theta, \vartheta) g(w) \alpha_{\mathrm{Z}}(z, w) \mathrm{d}\vartheta\mathrm{d}w \\
&\geq \varepsilon \eta(\theta, \delta) \min\{g(w), e^{-z}\pi_{\mathrm{Z}}(w)\} \mathrm{d}\vartheta\mathrm{d}w, \qquad (48)
\end{aligned}
$$

which is strictly positive on $S := \{w : g(w) > 0\} = \{w : \pi_{\mathrm{Z}}(w) > 0\}$. Hence, the n-step density part of $(Q^*)^n$ is strictly positive for all $(\vartheta, z) \in B(\theta, n\delta) \times S$. This establishes the $\mathrm{d}\vartheta \times \pi_{\mathrm{Z}}(\mathrm{d}z)$ irreducibility of Q^*, and hence its ergodicity as it is $\overline{\pi}$-invariant. For $\widetilde{Q}^*$, we have for any $(\theta, z) \in \Theta \times \mathbb{R}$, $\vartheta \in B(\theta, \delta)$ and $w \in \mathbb{R}$,

$$
\begin{aligned}
\widetilde{Q}^*\{(\theta, z), (d\vartheta, \mathrm{d}w)\} &= \frac{q(\theta, \vartheta)\, \alpha_{\mathrm{EX}}(\theta, \vartheta) g(w) \alpha_{\mathrm{Z}}(z, w)}{\varrho_{\mathrm{EX}}(\theta)\, \varrho_{\mathrm{Z}}(z)} \mathrm{d}\vartheta\mathrm{d}w \\
&\geq \varepsilon \eta(\theta, \delta) \min\{g(w), e^{-z}\pi_{\mathrm{Z}}(w)\} \mathrm{d}\vartheta\mathrm{d}w,
\end{aligned}
$$

using calculations as in (48) and the fact that $0 < \varrho_{\mathrm{EX}}(\theta)\, \varrho_{\mathrm{Z}}(z) \leq 1$ for any $(\theta, z) \in \Theta \times \mathbb{R}$, as Q^* is irreducible. Finally, the ergodicity of $\widetilde{Q}_{\mathrm{EX}}$ follows, using similar arguments, from the ergodicity of Q_{EX}. □

D Statements and proofs of auxiliary technical results

Proposition 5. *Define the relative computing time* $\mathrm{URCT}_2(h; \sigma)$

$$
\mathrm{URCT}_2(h; \sigma) = \frac{\mathrm{URIF}_2(h; \sigma)}{\sigma^2},
$$

where $\mathrm{URIF}_2(h; \sigma)$ *is the relative inefficiency. Using the same assumptions as in Theorem 1,*

(i) If $\mathrm{IF}(h, Q_{\mathrm{EX}}) = 1$*, then* $\mathrm{URCT}_2(h; \sigma)$ *is minimized at* $\sigma_{opt} = 0.92$ *and* $\mathrm{RIF}(h, Q; \sigma_{opt}) = \mathrm{IF}(h, Q_\pi; \sigma_{opt}) = 4.54$, $\pi_{\mathrm{Z}}^{\sigma_{opt}}\left(\varrho_{\mathrm{Z}}^{\sigma_{opt}}\right) = 0.51$.

(ii) *If* $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \geq 1$, σ_{opt} *increases to* $\sigma_{opt} = 1.02$ *as* $\mathrm{IF}(h, Q_{\mathrm{EX}}) \longrightarrow \infty$.

(iii) $\mathrm{uRIF}_2(h;\sigma)$ *and* $\mathrm{uRCT}_2(h;\sigma)$ *are decreasing functions of* $\mathrm{IF}(h, Q_{\mathrm{EX}})$.

Proof of Proposition 5. We consider minimizing $\mathrm{uRCT}_2(h;\sigma)$ with respect to σ. Then,

$$\mathrm{uRCT}_2(h;\sigma) = \frac{\{1 + \mathrm{IF}\,(h, Q_{\mathrm{EX}})\}}{2\mathrm{IF}\,(h, Q_{\mathrm{EX}})} \frac{\mathrm{IF}(h, Q_\pi;\sigma)}{\sigma^2} + \frac{\mathrm{IF}\,(h, Q_{\mathrm{EX}}) - 1}{2\sigma^2 \mathrm{IF}\,(h, Q_{\mathrm{EX}})}.$$

To obtain Part (i), we note that $\mathrm{uRCT}_2(h;\sigma) = \mathrm{IF}(h, Q_\pi;\sigma)/\sigma^2$ when $\mathrm{IF}\,(h, Q_{\mathrm{EX}}) = 1$. We define $H(\sigma) = \mathrm{IF}(h, Q_\pi;\sigma)/\sigma^2$. Using Lemma 5 in Pitt et al. (2012), one can verify that $H(\sigma)$ is minimized at $\sigma_{\mathrm{opt}} = 0.92$ and that $\partial^2 \{H(\sigma)\}/(\partial\sigma)^2 > 0$. The numerical values of Part (i) at $\sigma_{\mathrm{opt}} = 0.92$ can be found in Pitt et al. (2012). To obtain Part (ii), we note that

$$\begin{aligned} \partial \mathrm{uRCT}_2(h;\sigma)/\partial\sigma &= \{1 + \mathrm{IF}\,(h, Q_{\mathrm{EX}})\} / \{2\mathrm{IF}\,(h, Q_{\mathrm{EX}})\}\, \partial H(\sigma)/\partial\sigma \\ &\quad - \{\mathrm{IF}\,(h, Q_{\mathrm{EX}}) - 1\} / \left\{\sigma^3 \mathrm{IF}\,(h, Q_{\mathrm{EX}})\right\}, \\ \partial^2 \mathrm{uRCT}_2(h;\sigma)/(\partial\sigma)^2 &= (1 + \mathrm{IF}\,(h, Q_{\mathrm{EX}}))/(2\mathrm{IF}\left(h, Q^{\mathrm{EX}}\right))\partial^2 H(\sigma)/(\partial\sigma)^2 \\ &\quad + 3\{\mathrm{IF}\,(h, Q_{\mathrm{EX}}) - 1\} / \left\{\sigma^4 \mathrm{IF}\,(h, Q_{\mathrm{EX}})\right\}, \end{aligned} \tag{49}$$

so that $\partial^2 \mathrm{uRCT}_2(h;\sigma)/(\partial\sigma)^2 > 0$ if $\mathrm{IF}\,(h, Q_{\mathrm{EX}}) \geq 1$. For the limiting case of Part (ii),

$$\lim_{\mathrm{IF}_{\mathrm{EX}} \uparrow \infty} \partial \mathrm{uRCT}_2(h;\sigma)/\partial\sigma = \{\partial H(\sigma)/\partial\sigma\}/2 - 1/\sigma^3,$$

which we can verify numerically is equal to 0 at $\sigma_{\mathrm{opt}} = 1.02$. For general values of $\mathrm{IF}_{\mathrm{EX}}$,

$$\begin{aligned} &\partial \left\{ \partial \mathrm{uRCT}_2(h;\sigma)/\partial\sigma|_{\sigma=\sigma_{\mathrm{opt}}} \right\} / \partial \mathrm{IF}\,(h, Q_{\mathrm{EX}}) \\ &\qquad = -1/\left\{\mathrm{IF}\,(h, Q_{\mathrm{EX}})\right\}^2 \left(\partial H(\sigma)/\partial\sigma|_{\sigma=\sigma_{\mathrm{opt}}}/2 + 1/\sigma^3_{opt} \right) < 0, \end{aligned}$$

where $\partial H(\sigma)/\partial\sigma > 0$ for $\sigma > 0.92$. Hence, σ_{opt} increases with $\mathrm{IF}\,(h, Q_{\mathrm{EX}})$, which verifies Part (ii). Finally, to obtain Part (iii), it is straightforward to see that

$$\mathrm{uRIF}_2(h;\sigma) = \frac{\mathrm{IF}(h, Q_\pi;\sigma) + 1}{2} + \frac{\mathrm{IF}(h, Q_\pi;\sigma) - 1}{2\mathrm{IF}\,(h, Q_{\mathrm{EX}})},$$

so that $\mathrm{uRIF}_2(h;\sigma)$ and $\mathrm{uRCT}_2(h;\sigma) = \mathrm{uRIF}_2(h;\sigma)/\sigma^2$ are decreasing functions of $\mathrm{IF}\,(h, Q_{\mathrm{EX}})$, holding σ constant. □

E Asymptotic upper bound

This section illustrates, in the Gaussian noise case, the lower bound on the inefficiency $\mathrm{lRIF}_2(\sigma) = 1/\{2\Phi(-\sigma/\sqrt{2})\}$ and the exact relative inefficiency $\mathrm{aRIF}(\sigma, l) = \Phi(-l/2)/\Phi\left\{-\left(2\sigma^2 + l^2\right)^{1/2}/2\right\}$ obtained in Sherlock et al. (2013) and discussed in Section 3.6 of the main paper. Recall that $\mathrm{aRIF}(\sigma, l) \to \mathrm{lRIF}_2(\sigma)$ as $l \to 0$ and note that $\mathrm{aRIF}(\sigma, l) \to \Psi(\sigma) = \exp\left(\sigma^2/4\right)/\sigma^2$ as $l \to \infty$. Figure 4 displays the corresponding relative computing times $\mathrm{lRCT}_2(\sigma) = \mathrm{lRIF}_2(\sigma)/\sigma^2$ and $\mathrm{aRCT}(\sigma; l) = \mathrm{aRIF}(\sigma, l)/\sigma^2$. They are very similar in shape as a function of σ, regardless of l, and are also minimized at similar values: $\mathrm{lRCT}_2(\sigma)$ is minimized at $\sigma_1 = 1.68$ and $\Psi(\sigma)$ is minimized at $\sigma_2 = 2.00$, and $\mathrm{lRCT}_2(\sigma_1) = 1.51$, $\mathrm{lRCT}_2(\sigma_2) = 1.59$, $\Psi(\sigma_1) = 0.72$, $\Psi(\sigma_2) = 0.68$.

F Simulation results

This section applies the pseudo-marginal algorithm to a linear Gaussian state-space model and presents additional simulation results for the stochastic volatility model discussed in the main paper. The linear Gaussian state-space model we consider is a first order autoregression AR(1) observed with noise. In this case, $Y_t = X_t + \sigma_\varepsilon \varepsilon_t$, and the state evolution is $X_{t+1} = \mu_x(1-\phi) + \phi X_t + \sigma_\eta \eta_t$, where ε_t and η_t are standard normal and independent. We take $\phi = 0.8$, $\mu_x = 0.5$, $\sigma_\eta^2 = 1 - \phi^2$, so that the marginal variance σ_x^2 of the state X_t is 1. We consider a series of length T, where $\sigma_\varepsilon^2 = 0.5$ is assumed known. The parameters of interest are therefore $\theta = (\phi, \mu_x, \sigma_x)$. The analysis is very similar to that of Section 4 of the main paper. However, for this state-space model, the likelihood can be calculated by using the Kalman filter. This facilitates the analysis of sections F.1 and F.2 in two ways. First, in the calculation of the log-likelihood error $Z = \log \widehat{p}_N(y \mid \theta) - \log p(y \mid \theta)$, the true likelihood term is known rather than estimated. Second, because the likelihood is known, we can directly examine the exact chain Q_{EX} and estimate the inefficiency $\text{IF}(h, Q_{\text{EX}})$.

F.1 Empirical results for the error of the log-likelihood estimator

The analysis in this section mirrors that of Section 4.2 of the main paper. We investigate empirically Assumptions 1 and 2 by examining the behaviour of $Z = \log \widehat{p}_N(y \mid \theta) - \log p(y \mid \theta)$ for $T = 40$, 300 and 2700. Corresponding values of N are selected in each case to ensure that the variance of Z evaluated at the posterior mean $\overline{\theta}$ is approximately unity. The three plots on the left of Fig. 5 display the histograms corresponding to the density $g_N(z \mid \overline{\theta})$ of Z for $\theta = \overline{\theta}$, which is obtained by running $S = 6000$ particle filters at this value. We overlay on each histogram a kernel density estimate together with the corresponding assumed density, $g_{\text{Z}}^{\sigma}(z)$ of Assumption 2, where σ^2 is the sample variance of Z over the S particle filters. For $T = 40$, there is a slight discrepancy between the assumed Gaussian densities and the true histograms representing $g_N(z \mid \overline{\theta})$. In particular, whilst $g_N(z \mid \overline{\theta})$ is well approximated over most of its support, it is heavier tailed in the left tail. For $T = 300$ and $T = 2700$, the assumed Gaussian densities are very accurate.

We also examine Z when θ is distributed according to $\pi(\theta)$. We record 100 samples from $\pi(\theta)$, for $T = 40$, 300 and 2700. For each of these samples, we run the particle filter 300 times in order to estimate the true likelihood at these values. The resulting histograms, corresponding to the density $\int \pi(\mathrm{d}\theta)\, g_N(z \mid \theta)$ are displayed on the right panel of Fig. 5. For $T = 300$ and $T = 2700$, the assumed densities $g_{\text{Z}}^{\sigma}(z)$ are close to the corresponding histograms and Assumptions 1 and 2 again appear to capture reasonably well the salient features of the densities associated with Z.

It is important, in examining departures from Assumption 1, to consider the heterogeneity of the conditional density $g_N(z \mid \theta)$ as θ varies over $\pi(\theta)$. In Fig. 6, the conditional moments associated with the density $g_N(z \mid \theta)$ are estimated, based on running the particle filter independently $S = 300$ times for each of 100 values of θ from $\pi(\theta)$. We record the estimates of the mean, the variance and the third and fourth central moments at each value of θ, for $T = 300$ and $T = 2700$. There is a small degree of variability for $T = 300$ around the values that we would expect which are -0.5, $1, 0$ and 3 corresponding to $g_{\text{Z}}^{\sigma}(z)$ where $\sigma = 1$. This variability reduces as T rises to 2700. A small degree variability is expected as these are moments estimated from $S = 300$ samples. This lack of heterogeneity explains why the values of Z, marginalized over $\pi(\theta)$, on the right hand side of Fig. 5, are close to $g_{\text{Z}}^{\sigma}(z)$ for time series of moderate and large length. Figure 7 records a similar experiment for the stochastic volatility model and data considered in Section 4 of the main paper. There is rather more variability as the true value of the likelihood in this case is unknown and has to be estimated. However, the results are similar and the variability again reduces as T rises to 2700.

F.2 Empirical results for the pseudo-marginal algorithm

The pseudo-marginal algorithm is applied to $T = 300$ data. The true likelihood of the data is computed by the Kalman filter as the model is a linear Gaussian state space model. This allows the exact Metropolis–Hastings scheme Q_{EX} to be implemented so that the corresponding inefficiency IF (h, Q_{EX}) can be easily estimated. We consider varying N so that the standard deviation $\sigma(\overline{\theta}; N)$ of the log-likelihood estimator varies. The grid of values that we consider for N is {11, 16, 22, 31, 43, 60, 83, 116, 161, 224, 312}, see Table 2. The value $N = 60$ results in $\sigma(\overline{\theta}; N) = 0.92$.

Table 2. *AR(1) plus noise example with proposal parameter* $\rho = 0, T = 300, \phi = 0.8, \mu = 0.5, \sigma_x^2 = 1$ *and* $\sigma_\varepsilon^2 = 0.5$. *Inefficiencies* (IF) *and computing times* (CT=N × IF) *shown for* (ϕ, μ, σ_x) *and marginal probabilities of acceptance. See Fig. 8 and Fig. 9.*

Q_{EX}		IF(ϕ)	IF(μ)	IF(σ_x)				pr(Acc)
		2.5845	2.5040	2.4163				0.7678
Q								
N	$\sigma(\overline{\theta}; N)$	IF(ϕ)	IF(μ)	IF(σ_x)	CT(ϕ)	CT(μ)	CT(σ_x)	pr(Acc)
11	2.2886	136.32	132.41	128.66	1499.5	1456.5	1415.3	0.11424
16	1.8692	61.403	63.756	66.609	982.45	1020.1	1065.7	0.19036
22	1.6063	37.256	40.486	37.367	819.63	890.68	822.07	0.25549
31	1.3412	15.880	18.099	19.135	492.29	561.08	593.20	0.32622
43	1.1096	11.320	9.7400	10.710	486.75	418.82	460.54	0.39347
60	0.9197	7.5040	8.0428	7.6168	450.24	482.57	457.01	0.45933
83	0.8058	5.7253	5.5841	5.9348	475.20	463.48	492.59	0.50885
116	0.6828	4.3756	4.7106	4.1693	507.57	546.43	483.63	0.56621
161	0.5828	3.8112	4.2379	3.6388	613.61	682.30	585.84	0.60160
224	0.4838	3.2711	3.1605	3.3134	732.73	707.94	742.19	0.63562
312	0.4096	3.0774	3.4768	2.8355	960.14	1084.8	884.67	0.65793

Table 3. *AR(1) plus noise example with proposal parameter* $\rho = 0.9$. *Other settings identical to Table 2.*

Q_{EX}		IF(ϕ)	IF(μ)	IF(σ_x)				pr(Acc)
		25.59	22.21	24.44				0.87717
Q								
N	$\sigma(\overline{\theta}; N)$	IF(ϕ)	IF(μ)	IF(σ_x)	CT(ϕ)	CT(μ)	CT(σ_x)	pr(Acc)
11	2.2886	594.64	488.30	639.04	6541.1	5371.3	7029.5	0.12579
16	1.8692	157.49	183.78	182.07	2519.9	2940.4	2913.1	0.20410
22	1.6063	126.87	115.84	125.37	2791.2	2548.6	2758.2	0.27279
31	1.3412	69.541	67.421	71.982	2155.9	2089.9	2231.5	0.35385
43	1.1096	53.053	62.344	58.002	2281.1	2680.9	2494.0	0.42577
60	0.9197	49.351	47.476	45.194	2961.1	2848.6	2711.6	0.49610
83	0.8058	37.709	29.550	38.266	3129.8	2452.7	3176.1	0.55764
116	0.6828	29.360	36.943	34.892	3405.8	4285.4	4047.4	0.61174
161	0.5828	28.277	27.883	29.864	4552.6	4489.2	4808.1	0.65704
224	0.4838	27.770	29.471	30.533	6220.5	6601.5	6839.4	0.69674
312	0.4096	29.231	25.549	29.967	9120.2	7971.4	9349.8	0.73057

We transform each of the parameters to the real line so that $\Psi = k(\theta)$, where both θ and Ψ are three dimensional vectors, and place a Gaussian prior on Ψ centred at zero with a large variance.

We use the autoregressive Metropolis proposal $q(\Psi, \Psi^*)$

$$\Psi^* = (1-\rho)\widehat{\Psi} + \rho\Psi + (1-\rho^2)^{1/2}\{(\nu-2)/\nu\}^{1/2}\Sigma^{1/2}t_\nu,$$

for both the pseudo-marginal algorithm and exact likelihood schemes, where $\widehat{\Psi}$ is the mode of the log-likelihood obtained from the Kalman filter and the covariance Σ is the negative inverse of the second derivative of the log-likelihood at the mode. Here t_ν denotes a standard multivariate t-distributed random variable with ν degrees of freedom. We set $\nu = 5$. We use this autoregressive proposal with the scalar autoregressive parameter ρ chosen as one of $\{0,\ 0.4,\ 0.6,\ 0.9\}$. We first apply this proposal, for the four values of ρ, using the known likelihood in the Metropolis scheme and estimate the inefficiency for each of the parameters $\theta = (\mu_x, \phi, \sigma_\eta)$.

Figure 8 shows the acceptance probability for the pseudo-marginal algorithm against $\sigma(\overline{\theta}; N)$ for the four values of the proposal parameter ρ. The lower bound for the acceptance probabilities, as discussed at the end of Section 4.3 of the main paper, is also displayed and there is close correspondence in all cases. The histograms for the accepted and rejected values of Z, for $N = 60$ when $\sigma(\overline{\theta}; N) = 0.92$, are also displayed. The approximating asymptotic Gaussian densities, with $\sigma = 0.92$, are superimposed. This figure shows that the approximating densities correspond very closely to the two histograms. It should be noted that these are the marginal values for Z over the draws from the posterior $\pi(\theta)$ obtained by running the pseudo-marginal scheme, rather than being based upon a fixed value of the parameters.

Tables 2 and 3 show the pseudo-marginal algorithm results for $\rho = 0$ and $\rho = 0.9$. For the independent Metropolis–Hastings proposal, it is clear from Table 2, that the computing time in minimised around $N = 43$ or 60, depending on which parameter is examined, with the corresponding values of $\sigma(\overline{\theta}; N)$ being 1.11 and 0.92, supporting the findings that when an efficient proposal is used the optimal value of σ is close to unity. This is again supported by Fig. 9, for which the relative computing time ($\rho = 0$ is the top right plot) is shown against $\sigma(\overline{\theta}; N)$. We note that the relative inefficiencies and computing times are straightforward to calculate as the exact chain inefficiencies for the three parameters have been calculated and are given in the top row of Table 2. Table 3 shows the results for the more persistent proposal where $\rho = 0.9$. In this case, for all three parameters the optimal value of N is around 31, at which $\sigma(\overline{\theta}; N) = 1.34$, the corresponding graph of the relative computing time being given by the bottom right of Fig. 9. It is clear that again the findings are consistent with the discussion of 3.5 in the main paper. In particular, as ρ increases, then IF (h, Q^{EX}) should increase, and, from Fig. 9, it is clear that the optimal value of $\sigma(\overline{\theta}; N)$ increases, the relative computing time decreases for any given $\sigma(\overline{\theta}; N)$. In addition, the relative computing time becomes more flat as a function of $\sigma(\overline{\theta}; N)$ as ρ increases.

G Numerical procedures

Under the Gaussian assumption, Corollary 3 specifies the function $\varrho_{\mathrm{Z}}^{\sigma}$ and the term $\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma})$ can be accurately evaluated using numerical quadrature. This section explains how we numerically evaluate the terms $\phi_{\mathrm{Z}}^{\sigma}$ and $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \widetilde{Q}^{\mathrm{Z}})$ which appear in the bounds of Corollaries 1 and 2. The inefficiency $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \widetilde{Q}^{\mathrm{Z}})$ is finite by Lemma 3, because $\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma})$ is finite. The autocorrelations quickly descend to zero as a function of n, for all σ. Hence, it is straightforward to estimate $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \widetilde{Q}^{\mathrm{Z}})$ by the appropriate summation of the autocorrelations, and to tabulate it against σ for use in the bounds of Corollaries 1 and 2. The autocorrelation $\phi_{\mathrm{Z}}^{\sigma}$, for $n = 1$, is similarly tabulated.

From Lemma 3,

$$\widetilde{Q}^{\mathrm{Z}}(z, \mathrm{d}w) = \frac{g(w)\min\{1, \exp(w-z)\}\mathrm{d}w}{\varrho_{\mathrm{Z}}(z)}, \quad \widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}z) = \frac{\pi_{\mathrm{Z}}(\mathrm{d}z)\,\varrho_{\mathrm{Z}}(z)}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})},$$

so the autocorrelation at lag n is

$$\phi_n\left(\varrho_{\mathrm{Z}}^{-1}, \widetilde{Q}^{\mathrm{Z}}\right) = \frac{\left\langle \varrho_{\mathrm{Z}}^{-1}, \left(\widetilde{Q}^{\mathrm{Z}}\right)^n \varrho_{\mathrm{Z}}^{-1}\right\rangle_{\widetilde{\pi}_{\mathrm{Z}}} - \left\{\widetilde{\pi}_{\mathrm{Z}}\left(\varrho_{\mathrm{Z}}^{-1}\right)\right\}^2}{\mathbb{V}_{\widetilde{\pi}_Z}\left(\varrho_{\mathrm{Z}}^{-1}\right)}$$

with

$$\begin{aligned}\left\langle \varrho_{\mathrm{Z}}^{-1}, \left(\widetilde{Q}^{\mathrm{Z}}\right)^n \varrho_{\mathrm{Z}}^{-1}\right\rangle_{\widetilde{\pi}_{\mathrm{Z}}} &= \int \varrho_{\mathrm{Z}}^{-1}(z_0)\varrho_{\mathrm{Z}}^{-1}(z_n)\widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}z_0)\left(\widetilde{Q}^{\mathrm{Z}}\right)^n(z_0, \mathrm{d}z_n) \\ &= \pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})^{-1}\int \varrho_{\mathrm{Z}}^{-1}(z_n)\widetilde{\pi}_{\mathrm{Z}}(\mathrm{d}z_0)\left(\widetilde{Q}^{\mathrm{Z}}\right)^n(z_0, \mathrm{d}z_n)\,. \end{aligned} \tag{50}$$

The term $\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})$ can be computed by quadrature. The term (50) can be also accurately calculated by Monte Carlo integration, by simulating a large number M of i.i.d. samples $Z_0^i \sim \pi_{\mathrm{Z}}$ and then propagating each sample through the transition kernel $\widetilde{Q}^{\mathrm{Z}}$ n times to obtain $Z_n^i \sim \pi_{\mathrm{Z}}(\widetilde{Q}^{\mathrm{Z}})^n$, yielding the estimate $\frac{1}{M}\sum_{i=1}^{M} \varrho_{\mathrm{Z}}^{-1}(Z_n^i)$.

References

Andrieu, C., Doucet, A. & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **72**, 1–33.

Andrieu, C. & Vihola, M. (2014). Establishing some order amongst exact approximations of MCMCs, arXiv:1404.6909.

Andrieu, C. & Vihola, M. (2012). Convergence properties of pseudo-marginal Markov Chain Monte Carlo algorithms. *Annals of Applied Probability, to appear.*

Andrieu, C. & Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics* **37**, 697–725.

Baxendale, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability* **15**, 700–738.

Beaumont, M. (2003a). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139.

Chernov, M., Gallant, A. R., Ghysels, E. & Tauchen, G. (2003). Alternative models of stock price dynamics. *Journal of Econometrics* **116**, 225–257.

Douc, R. & Robert, C. P. (2011). A vanilla Rao–Blackwellization of Metropolis–Hastings algorithms. *The Annals of Statistics* **39**, 261–277.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* **7**, 473–483.

Häggström, O. & Rosenthal, J. S. (2007). On variance conditions for Markov chain central limit theorems. *Electronic Communications in Probability* **12**, 454–64.

Huang, X. & Tauchen, G. (2005). The relative contribution of jumps to total price variation. *Journal of Financial Econometrics* **3**, 456–499.

Lin, L., Liu, K. F. & Sloan, J. (2000). A noisy Monte Carlo algorithm. *Physical Review D* **61**, 074505.

Peskun, P. H. (1973). Optimum Monte–Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.

Pitt, M. K., Silva, R., Giordani, P. & Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics* **171**, 134–151.

Roberts, G. & Tweedie, R. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.

Rudolf, D. & Ullrich, M. (2013). Positivity of hit-and-run and related algorithms. *Electronic Communications in Probability* **18**, 1–8.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* **21**, 1701–62.

Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *The Annals of Applied Probability* **8**, 1–9.

Kipnis, C. & Varadhan, S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics* **104**, 1–19.

Meyn, S. & Tweedie, R.L. (1991). *Markov Chains and Stochastic Stability.* Second edition. Cambridge University Press, Cambridge.

Rudin, W. (1991). *Functional Analysis.* International series in pure and applied mathematics. McGraw-Hill, Inc., New York.

Sherlock, C., Thiery, A., Roberts, G. & Rosenthal, J. (2013). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *ArXiv e-prints* URL: http://arxiv.org/abs/1309.7209.

Schmüdgen, K. (2012). *Unbounded Self-adjoint Operators on Hilbert Space.* Graduate Text in Mathematics, Springer.

Wu, W. & Woodroofe, M. (2004). Martingale approximations for sums of stationary processes. *The Annals of Probability* **32**, 1674–1690.

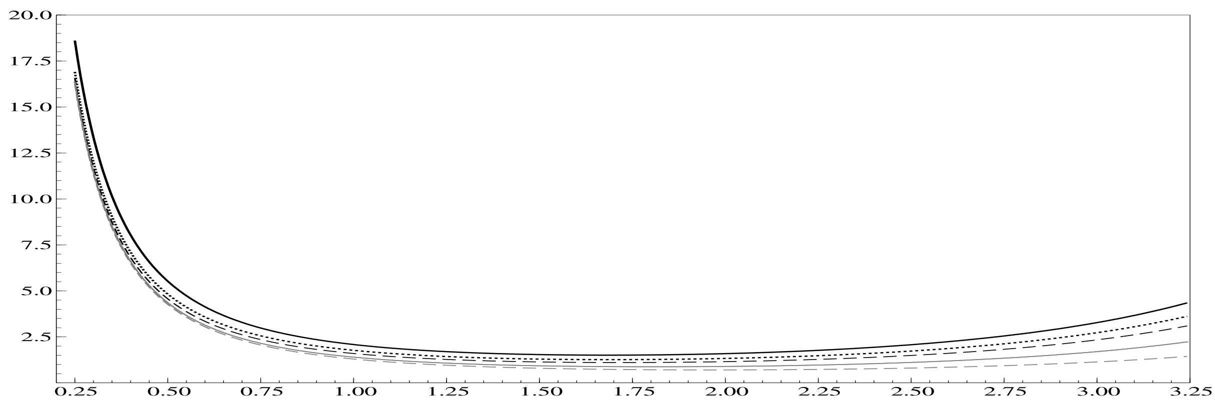

Figure 4: Theoretical results for relative computing time. $\text{LRCT}_2(\sigma)$ (solid black) is displayed together with $\text{ARCT}(\sigma, l)$ against σ. $\text{ARCT}(\sigma, l)$, the relative computing time for the limiting case of a random walk proposal, is evaluated for $l = 0.5$ (dotted black), 1 (dashed black), 2.5 (solid grey) and 10 (dashed grey), where l is the scaling factor in the proposal.

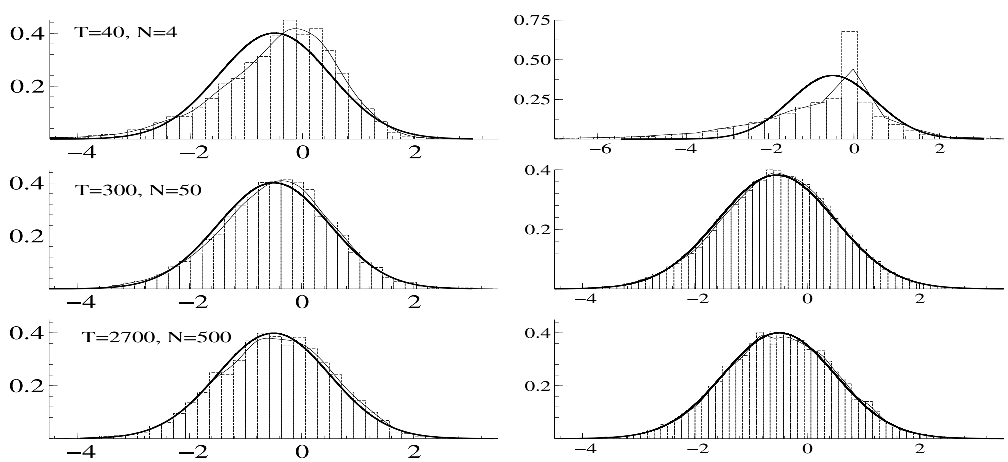

Figure 5: AR(1) plus noise model experiment. Top to bottom: $T = 40$, $N = 4$ (top), $T = 300$, $N = 50$ (middle), $T = 2700$, $N = 500$ (bottom). Left to right: histograms and theoretical densities associated with $g_N(z \mid \theta)$ evaluated at the posterior mean $\overline{\theta}$ (left), over values from the posterior $\pi(\theta)$ (right). The densities $g_Z^\sigma(z)$ are overlaid (solid lines).

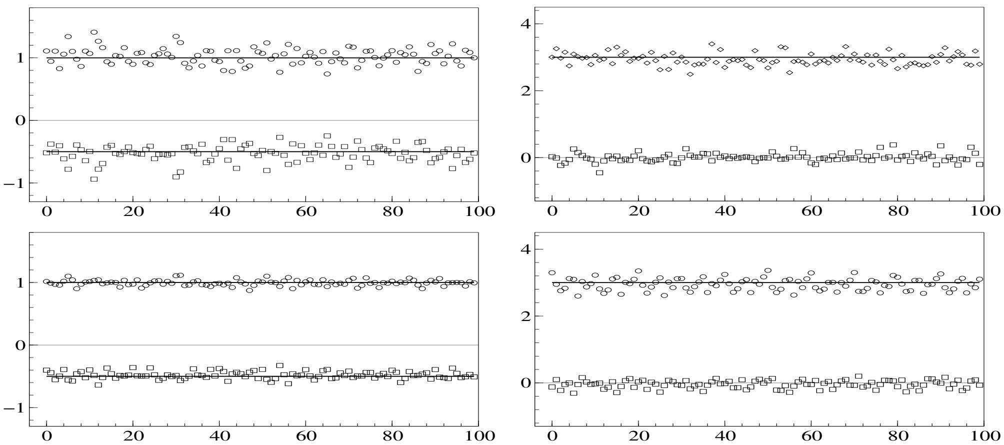

Figure 6: AR(1) plus noise model experiment. Top to bottom: $T = 300$, $N = 50$ (top), $T = 2700$, $N = 500$ (bottom). Left to right: mean (squares) and variance (circles) associated with $g_N(z \mid \theta)$ for 100 different values of θ from $\pi(\theta)$ (left). The corresponding estimates of the third (squares) and fourth (circles) moments are displayed.

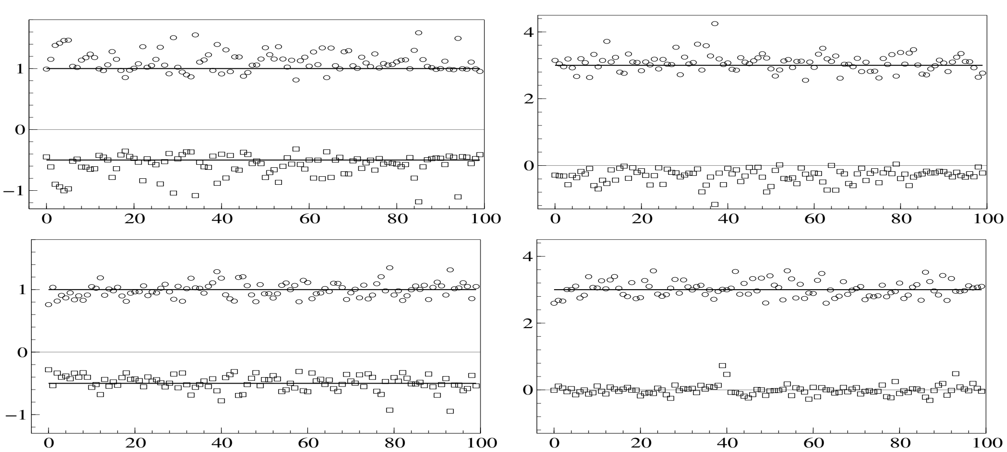

Figure 7: Huang and Tauchen two factor model experiment for S&P 500 data. Top to bottom: $T = 300$, $N = 80$ (top), $T = 2700$, $N = 700$ (bottom). Left to right: mean (squares) and variance (circles) associated with $g_N(z \mid \theta)$ for 100 different values of θ from $\pi(\theta)$ (left). The corresponding estimates of the third (squares) and fourth (circles) moments are displayed.

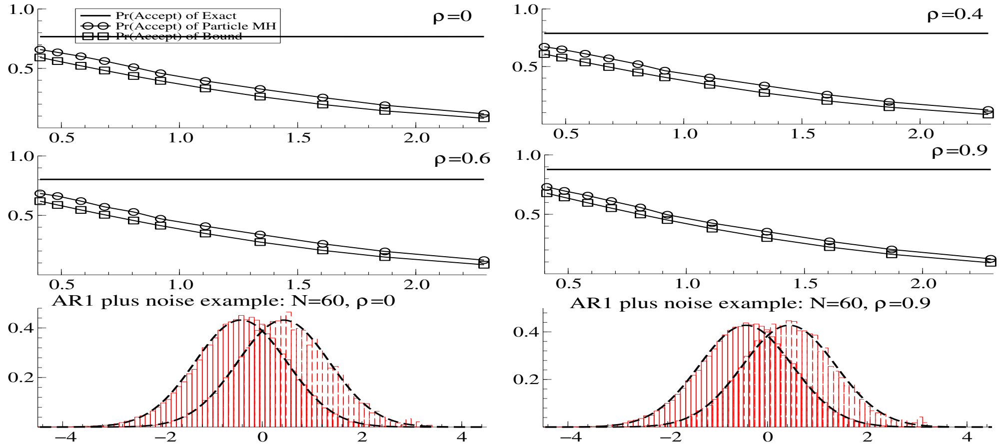

Figure 8: AR(1) plus noise example with $T = 300, \phi = 0.8, \mu = 0.5, \sigma_x^2 = 1, \sigma_\varepsilon^2$ fixed at 0.5. Marginal acceptance probabilities displayed against $\sigma(\overline{\theta}; N)$. The estimated (constant) marginal probabilities of acceptance for Q_{EX} are shown (solid line) together with the estimated probabilities (circles) from Q. The lower bound (squares) is given as the probability from the exact scheme times $2\Phi(-\sigma/\sqrt{2})$. The proposal autocorrelations are $\rho = 0, 0.4, 0.6$ and 0.9. See Tables 2 and 3. Bottom: Histograms for the accepted and proposed values of Z, the log-likelihood error for $\rho = 0$ (left) and for $\rho = 0.9$ (right). The theoretical Gaussian densities for the proposal $\varphi(-\sigma^2/2, \sigma^2)$ and the accepted values $\varphi(\sigma^2/2, \sigma^2)$ are overlaid where $\sigma = \sigma(\overline{\theta}; N)$.

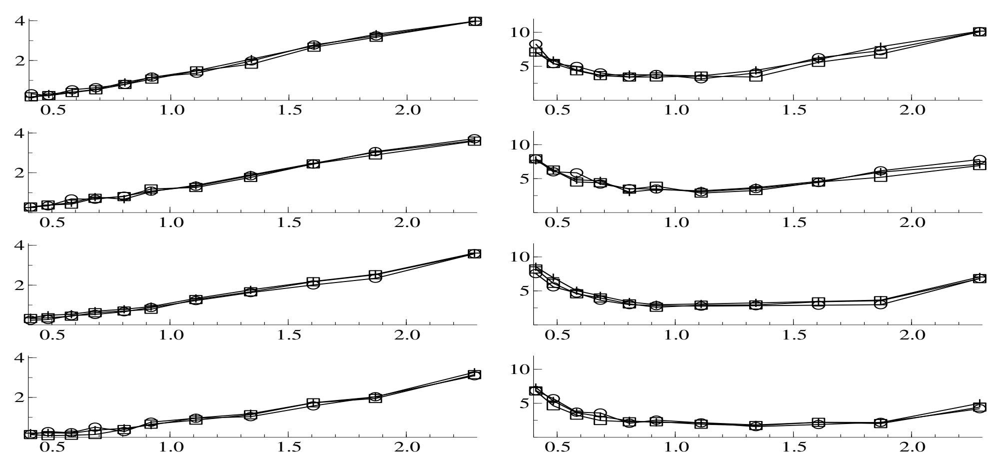

Figure 9: AR(1) plus noise example with $T = 300$, $\phi = 0.8, \mu = 0.5$, $\sigma_x^2 = 1$ and σ_ε^2 fixed at 0.5. Left: Logarithm RIF against $\sigma(\overline{\theta}; N)$. Right: RCT = RIF$/\sigma^2(\overline{\theta}; N)$ against $\sigma(\overline{\theta}; N)$. The three plots on all graphs are for ϕ (square), μ (circle) and σ_x (cross). From Top to bottom: $\rho = 0$, $0.4, 0.6$ and 0.9. Here $\sigma(\overline{\theta}; N)$ is the standard deviation of the log-likelihood estimator evaluated at the posterior mean $\overline{\theta}$. See Tables 2 and 3.