Adv. Appl. Probab. **56,** 1440–1466 (2024)
*doi:*10.1017/apr.2024.16

CONVERGENCE OF HYBRID SLICE SAMPLING VIA SPECTRAL GAP

KRZYSZTOF ŁATUSZYŃSKI,* *University Warwick*
DANIEL RUDOLF,** *Universität Passau*

Abstract

It is known that the simple slice sampler has robust convergence properties; however, the class of problems where it can be implemented is limited. In contrast, we consider hybrid slice samplers which are easily implementable and where another Markov chain approximately samples the uniform distribution on each slice. Under appropriate assumptions on the Markov chain on the slice, we give a lower bound and an upper bound of the spectral gap of the hybrid slice sampler in terms of the spectral gap of the simple slice sampler. An immediate consequence of this is that the spectral gap and geometric ergodicity of the hybrid slice sampler can be concluded from the spectral gap and geometric ergodicity of the simple version, which is very well understood. These results indicate that robustness properties of the simple slice sampler are inherited by (appropriately designed) easily implementable hybrid versions. We apply the developed theory and analyze a number of specific algorithms, such as the stepping-out shrinkage slice sampling, hit-and-run slice sampling on a class of multivariate targets, and an easily implementable combination of both procedures on multidimensional bimodal densities.

Keywords: Slice sampler; spectral gap; geometric ergodicity

2020 Mathematics Subject Classification: Primary 60J22
Secondary 65C05; 60J05

1. Introduction

Slice sampling algorithms are designed for Markov chain Monte Carlo (MCMC) sampling from a distribution given by a possibly unnormalized density. They belong to the class of auxiliary-variable algorithms that define a suitable Markov chain on an extended state space. Following [6] and [41], a number of different versions have been discussed and proposed in [4, 9, 20–22, 25, 29, 30, 37]. We refer to these papers for details of algorithmic design and applications in Bayesian inference and statistical physics. Here let us first focus on the appealing *simple slice sampler* setting, in which no further algorithmic tuning or design by the user is necessary: assume that $K \subseteq \mathbb{R}^d$ and let the unnormalized density be $\varrho\colon K \to (0, \infty)$. The goal is to sample approximately with respect to the distribution π determined by ϱ, i.e.

Received 4 February 2023; accepted 9 February 2024.
* Postal address: Department of Statistics, CV47AL Coventry, United Kingdom. Email address: K.G.Latuszynski@warwick.ac.uk
** Postal address: Universität Passau, Innstraße 33, 94032 Passau, Germany. Email address: daniel.rudolf@uni-passau.de

$$\pi(A) = \frac{\int_A \varrho(x)\,\mathrm{d}x}{\int_K \varrho(x)\,\mathrm{d}x}, \qquad A \in \mathcal{B}(K),$$

where $\mathcal{B}(K)$ denotes the Borel σ-algebra. Given the current state $X_n = x \in K$, the *simple slice sampling* algorithm generates the next Markov chain instance X_{n+1} by the following two steps:

1. Choose t uniformly at random from $(0, \varrho(x))$, i.e. $t \sim \mathcal{U}(0, \varrho(x))$.
2. Choose X_{t+1} uniformly at random from

$$K(t) := \{x \in K \mid \varrho(x) > t\},$$

   the level set of ϱ determined by t.

The above-defined *simple slice sampler* transition mechanism is known to be reversible with respect to π and possesses very robust convergence properties that have been observed empirically and established formally. For example Mira and Tierney [21] proved that if ϱ is bounded and the support of ϱ has finite Lebesgue measure, then the simple slice sampler is *uniformly ergodic*. Roberts and Rosenthal in [29] provide criteria for *geometric ergodicity*. Moreover, in [29, 30] the authors prove explicit estimates for the total variation distance of the distribution of X_n to π. In the recent work [24], depending on the volume of the level sets, an explicit lower bound of the *spectral gap* of simple slice sampling is derived.

Unfortunately, the applicability of the simple slice sampler is limited. In high dimensions it is in general infeasible to sample uniformly from the level set of ϱ, and thus the second step of the algorithm above cannot be performed. Consequently, the second step is replaced by sampling a Markov chain on the level set, which has the uniform distribution as the invariant one. Following the terminology of [28] we call such algorithms *hybrid slice samplers*. We refer to [26], where various procedures and designs for the Markov chain on the slice are suggested and insightful expert advice is given.

Despite being easy to implement, hybrid slice sampling in general has not been analyzed theoretically, and little is known about its convergence properties. (A notable exception is elliptical slice sampling [22], which has recently been investigated in [23], where a geometric ergodicity statement is provided.) The present paper is aimed at closing this gap by providing statements about the inheritance of convergence from the simple to the hybrid setting.

To this end we study the absolute spectral gap of hybrid slice samplers. The absolute spectral gap of a Markov operator P or a corresponding Markov chain $(X_n)_{n\in\mathbb{N}}$ is given by

$$\mathrm{gap}(P) = 1 - \|P\|_{L^0_{2,\pi} \to L^0_{2,\pi}},$$

where $L^0_{2,\pi}$ is the space of functions $f\colon K \to \mathbb{R}$ with zero mean and finite variance (i.e. $\int_K f(x)\mathrm{d}\pi(x) = 0$; $\|f\|_2^2 = \int_K |f(x)|^2\,\mathrm{d}\pi(x) < \infty$) and $\|P\|_{L^0_{2,\pi} \to L^0_{2,\pi}}$ denotes the operator norm. We refer to [32] for the functional-analytic background. From the computational point of view, the existence of the spectral gap (i.e. $\mathrm{gap}(P) > 0$) implies a number of desirable and well studied robustness properties, in particular the following:

- The spectral gap implies geometric ergodicity [15, 28] and the variance bounding property [31].
- For reversible Markov chains, the spectral gap implies that a CLT holds for all functions $f \in L_{2,\pi}$ (cf. [8, 14]).

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

- Furthermore, consistent estimation of the CLT asymptotic variance is well established for geometrically ergodic chains (cf. [2, 7, 10, 11]).

Additionally, quantitative information on the spectral gap allows the formulation of precise non-asymptotic statements. In particular, it is well known (see e.g. [27, Lemma 2]) that if ν is the initial distribution of the reversible Markov chain in question, i.e. $\nu = \mathbb{P}_{X_1}$, then

$$\left\| \nu P^n - \pi \right\|_{\text{tv}} \le (1 - \text{gap}(P))^n \left\| \frac{d\nu}{d\pi} - 1 \right\|_2 ,$$

where $\nu P^n = \mathbb{P}_{X_{n+1}}$. See [1, Section 6] for a related $L_{2,\pi}$ convergence result. Moreover, when considering the sample average, one obtains

$$\mathbb{E} \left| \frac{1}{n} \sum_{j=1}^{n} f(X_j) - \int_K f(x) \mathrm{d}\pi(x) \right|^2 \le \frac{2}{n \cdot \text{gap}(P)} + \frac{c_p \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty}{n^2 \cdot \text{gap}(P)},$$

for any $p > 2$ and any function $f\colon K \to \mathbb{R}$ with $\|f\|_p^p = \int_K |f(x)|^p \, \pi(\mathrm{d}x) \le 1$, where c_p is an explicit constant which depends only on p. One can also take a burn-in into account; for further details see [33, Theorem 3.41]. This indicates that the spectral gap of a Markov chain is central to robustness and a crucial quantity in both asymptotic and non-asymptotic analysis of MCMC estimators.

The route we undertake is to conclude the spectral gap of the hybrid slice sampler from the more tractable spectral gap of the simple slice sampler. So what is known about the spectral gap of the simple slice sampler? To say more on this, we require the following notation. Define $v_\varrho\colon [0, \infty) \to [0, \infty]$ by $v_\varrho(t) := \text{vol}_d(K(t))$, which for level t returns the volume of the level set. We say for $m \in \mathbb{N}$ that $v_\varrho \in \Lambda_m$ if

- v_ϱ is continuously differentiable and $v'_\varrho(t) < 0$ for any $t \ge 0$, and
- the mapping $t \mapsto t v'_\varrho(t)/v_\varrho(t)^{1-1/m}$ is decreasing on the support of v_ϱ.

Recently, in [24, Theorem 3.10], it has been shown that if $v_\varrho \in \Lambda_m$, then $\text{gap}(U) \ge 1/(m+1)$. This provides a criterion for the existence of a spectral gap as well as a quantitative lower bound, essentially depending on whether $t \mapsto t v'_\varrho(t)/v_\varrho(t)^{1-1/m}$ is decreasing or not.

Now we are in a position to explain our contributions. Let H be the Markov kernel of the hybrid slice sampler determined by a family of transition kernels H_t, where each H_t is a Markov kernel with uniform limit distribution, say U_t, on the level determined by t. Consider

$$\beta_k := \sup_{x \in K} \left(\int_0^{\varrho(x)} \left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2 \frac{\mathrm{d}t}{\varrho(x)} \right)^{1/2},$$

and note that the quantity $\left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2$ measures how quickly H_t gets close to U_t. Thus β_k is the supremum over expectations of a function which measures the speed of convergence of H_t^k to U_t. The main result, stated in Theorem 1, is as follows. Assume that $\beta_k \to 0$ for increasing k and assume H_t induces a positive semidefinite Markov operator for every level t. Then

$$\frac{\text{gap}(U) - \beta_k}{k} \le \text{gap}(H) \le \text{gap}(U), \qquad k \in \mathbb{N}. \tag{1}$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

The first inequality implies that whenever there exists a spectral gap of the simple slice sampler and $\beta_k \to 0$, then there is a spectral gap of the hybrid slice sampler. The second inequality of (1) verifies a very intuitive result, namely that the simple slice sampler is always better than the hybrid one.

We demonstrate how to apply our main theorem in different settings. First, we consider a stepping-out shrinkage slice sampler, suggested in [26], in a simple bimodal 1-dimensional setting. Next we turn to the d-dimensional case and on each slice perform a single step of the hit-and-run algorithm, studied in [3, 16, 38]. Using our main theorem we prove equivalence of the spectral gap (and hence geometric ergodicity) of this hybrid hit-and-run on the slice and the simple slice sampler. Let us also mention here that in [36] the hit-and-run algorithm, hybrid hit-and-run on the slice, and the simple slice sampler are compared, according to covariance ordering [19], to a random-walk Metropolis algorithm. Finally, we combine the stepping-out shrinkage and hit-and-run slice sampler. The resulting algorithm is practical and easily implementable in multidimensional settings. For this version we again show equivalence of the spectral gap and geometric ergodicity with the simple slice sampler for multidimensional bimodal targets.

Further note that we consider single-auxiliary-variable methods to keep the arguments simple. We believe that a similar analysis can also be done if one considers multi-auxiliary-variable methods.

The structure of the paper is as follows. In Section 2 the notation and preliminary results are provided. These include a necessary and sufficient condition for reversibility of hybrid slice sampling in Lemma 1, followed by a useful representation of slice samplers in Section 2.1, which is crucial in the proof of the main result. In Section 3 we state and prove the main result. For example, Corollary 1 provides a lower bound on the spectral gap of a hybrid slice sampler, which performs several steps with respect to H_t on the chosen level. In Section 4 we apply our result to analyze a number of specific hybrid slice sampling algorithms in different settings that include multidimensional bimodal distributions.

2. Notation and basics

Recall that $\varrho : K \to (0, \infty)$ is an unnormalized density on $K \subseteq \mathbb{R}^d$, and denote the level set of ϱ by

$$K(t) = \{x \in K \mid \varrho(x) > t\}.$$

Hence the sequence $(K(t))_{t\geq 0}$ of subsets of $\mathbb{R}^d$ satisfies the following:

1. $K(0) = K$.
2. $K(s) \subseteq K(t)$ for $t < s$.
3. $K(t) = \emptyset$ for $t \geq \|\varrho\|_\infty$.

Let vol_d be the d-dimensional Lebesgue measure, and let $(U_t)_{t\in(0,\|\varrho\|_\infty)}$ be a sequence of distributions, where U_t is the uniform distribution on $K(t)$, i.e.

$$U_t(A) = \frac{\mathrm{vol}_d(A \cap K(t))}{\mathrm{vol}_d(K(t))}, \quad A \in \mathcal{B}(K).$$

Furthermore, let $(H_t)_{t\in(0,\|\varrho\|_\infty)}$ be a sequence of transition kernels, where H_t is a transition kernel on $K(t) \subseteq \mathbb{R}^d$. For convenience we extend the definition of the transition kernel $H_t(\cdot, \cdot)$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

on the measurable space $(K, \mathcal{B}(K))$. We set

$$\bar{H}_t(x, A) = \begin{cases} 0, & x \notin K(t), \\ H_t(x, A \cap K(t)), & x \in K(t). \end{cases} \tag{2}$$

In the following we write H_t for $\bar{H}_t$ and consider H_t as extension on $(K, \mathcal{B}(K))$. The transition kernel of the hybrid slice sampler is given by

$$H(x, A) = \frac{1}{\varrho(x)} \int_0^{\varrho(x)} H_t(x, A)\, \mathrm{d}t, \qquad x \in K,\ A \in \mathcal{B}(K).$$

If $H_t = U_t$ we have the simple slice sampler studied in [21, 24, 29, 30]. The transition kernel of this important special case is given by

$$U(x, A) = \frac{1}{\varrho(x)} \int_0^{\varrho(x)} U_t(A)\, \mathrm{d}t, \qquad x \in K,\ A \in \mathcal{B}(K).$$

We provide a criterion for reversibility of H with respect to π. Therefore let us define the density

$$\ell(s) = \frac{\mathrm{vol}_d(K(s))}{\int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(r))\, \mathrm{d}r}, \qquad s \in (0, \|\varrho\|_\infty),$$

of the distribution of the level sets on $((0, \|\varrho\|_\infty), \mathcal{B}((0, \|\varrho\|_\infty)))$.

Lemma 1. *The transition kernel H is reversible with respect to π if and only if*

$$\int_0^{\|\varrho\|_\infty} \int_B H_t(x, A)\, U_t(\mathrm{d}x)\, \ell(t)\mathrm{d}t = \int_0^{\|\varrho\|_\infty} \int_A H_t(x, B)\, U_t(\mathrm{d}x)\, \ell(t)\mathrm{d}t, \qquad A, B \in \mathcal{B}(K). \tag{3}$$

In particular, if H_t is reversible with respect to U_t for almost all t (concerning ℓ), then H is reversible with respect to π.

Equation (3) is the detailed balance condition of H_t with respect to U_t in the average sense, i.e.

$$\mathbb{E}_\ell[H_\cdot(x, \mathrm{d}y)U_\cdot(\mathrm{d}x)] = \mathbb{E}_\ell[H_\cdot(y, \mathrm{d}x)U_\cdot(\mathrm{d}y)], \qquad x, y \in K.$$

Now we prove Lemma 1.

Proof. First, note that

$$\begin{aligned} \int_K \varrho(x)\, \mathrm{d}x &= \int_0^{\|\varrho\|_\infty} \int_K \mathbf{1}_{(0,\varrho(x))}(s)\, \mathrm{d}x\, \mathrm{d}s \\ &= \int_0^{\|\varrho\|_\infty} \int_K \mathbf{1}_{K(s)}(x)\, \mathrm{d}x\, \mathrm{d}s = \int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s))\, \mathrm{d}s. \end{aligned}$$

From this we obtain for any $A, B \in \mathcal{B}(K)$ that

$$\int_B H(x, A)\, \pi(\mathrm{d}x) = \int_B \int_0^{\varrho(x)} H_t(x, A) \frac{\mathrm{d}t}{\int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s))\mathrm{d}s} \mathrm{d}x$$

$$= \int_B \int_0^{\|\varrho\|_\infty} \mathbf{1}_{K(t)}(x) H_t(x, A) \frac{\ell(t)}{\mathrm{vol}_d(K(t))} \mathrm{d}t\mathrm{d}x = \int_0^{\|\varrho\|_\infty} \int_B H_t(x, A)\, U_t(\mathrm{d}x)\, \ell(t)\mathrm{d}t.$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

As an immediate consequence of the previous equation, we have the claimed equivalence of reversibility and (3). By the definition of the reversibility of H_t according to U_t, we have

$$\int_B H_t(x, A)\, U_t(\mathrm{d}x) = \int_A H_t(x, B)\, U_t(\mathrm{d}x).$$

This, combined with (3), leads to the reversibility of H. □

We always want to have that H is reversible with respect to π. Therefore we formulate the following assumption.

Assumption 1. *Let H_t be reversible with respect to U_t for any $t \in (0, \|\varrho\|_\infty)$.*

Now we define Hilbert spaces of square-integrable functions and Markov operators. Let $L_{2,\pi} = L_2(K, \pi)$ be the space of functions $f: K \to \mathbb{R}$ which satisfy $\|f\|_{2,\pi}^2 := \langle f, f\rangle_\pi < \infty$, where

$$\langle f, g\rangle_\pi := \int_K f(x)\, g(x)\, \pi(\mathrm{d}x)$$

denotes the corresponding inner product of $f, g \in L_{2,\pi}$. For $f \in L_{2,\pi}$ and $t \in (0, \|\varrho\|_\infty)$ define

$$H_t f(x) = \int_{K(t)} f(y)\, H_t(x, \mathrm{d}y), \qquad x \in K. \tag{4}$$

Note that if $x \notin K(t)$, then we have $H_t f(x) = 0$ by the convention on H_t; see (2). The Markov operator $H: L_{2,\pi} \to L_{2,\pi}$ is defined by

$$Hf(x) = \frac{1}{\varrho(x)} \int_0^{\varrho(x)} H_t f(x)\, \mathrm{d}t,$$

and similarly $U: L_{2,\pi} \to L_{2,\pi}$ is defined by

$$Uf(x) = \frac{1}{\varrho(x)} \int_0^{\varrho(x)} U_t(f)\, \mathrm{d}t,$$

where $U_t(f) = \int_{K(t)} f(x)\, U_t(\mathrm{d}x)$ is a special case of (4). Furthermore, for $t \in (0, \|\varrho\|_\infty)$ let $L_{2,t} = L_2(K(t), U_t)$ be the space of functions $f: K(t) \to \mathbb{R}$ with $\|f\|_{2,t}^2 := \langle f, f\rangle_t < \infty$, where

$$\langle f, g\rangle_t := \int_{K(t)} f(x)\, g(x)\, U_t(\mathrm{d}x)$$

denotes the corresponding inner product of $f, g \in L_{2,t}$. Then $H_t: L_{2,t} \to L_{2,t}$ can also be considered as a Markov operator. Define the functional

$$S(f) = \int_K f(x)\, \pi(\mathrm{d}x), \quad f \in L_{2,\pi},$$

as the operator $S: L_{2,\pi} \to L_{2,\pi}$ which maps functions to the constant functions given by their mean value. We say $f \in L_{2,\pi}^0$ if and only if $f \in L_{2,\pi}$ and $S(f) = 0$. Now the absolute spectral gap of a Markov kernel or Markov operator $P: L_{2,\pi} \to L_{2,\pi}$ is given by

$$\mathrm{gap}(P) = 1 - \|P - S\|_{L_{2,\pi} \to L_{2,\pi}} = 1 - \|P\|_{L_{2,\pi}^0 \to L_{2,\pi}^0}.$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

For details of the last equality we refer to [33, Lemma 3.16]. Moreover, for the equivalence of $\text{gap}(P) > 0$ and (almost sure) geometric ergodicity we refer to [15, Proposition 1.2]. For any $t > 0$ the norm $\|f\|_{2,t}$ can also be considered for $f : K \to \mathbb{R}$. With this in mind we have the following relation between $\|f\|_{2,\pi}$ and $\|f\|_{2,t}$.

Lemma 2. *For any $f : K \to \mathbb{R}$, with the notation defined above, we obtain*

$$S(f) = \int_0^{\|\varrho\|_\infty} U_t(f)\ \ell(t)\,\mathrm{d}t. \tag{5}$$

In particular,

$$\|f\|_{2,\pi}^2 = \int_0^{\|\varrho\|_\infty} \|f\|_{2,t}^2\ \ell(t)\,\mathrm{d}t. \tag{6}$$

Proof. The assertion of (6) is a special case of (5), since $S(\,|f|^2\,) = \|f\|_{2,\pi}^2$. By $\int_K \varrho(x)\,\mathrm{d}x = \int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s))\,\mathrm{d}s$ (see the proof of Lemma 1), one obtains

$$\begin{aligned} S(f) &= \frac{\int_K f(x)\,\varrho(x)\,\mathrm{d}x}{\int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s))\,\mathrm{d}s} = \int_K \int_0^{\varrho(x)} f(x) \frac{\mathrm{d}t\,\mathrm{d}x}{\int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s))\,\mathrm{d}s} \\ &= \int_0^{\|\varrho\|_\infty} \int_{K(t)} f(x)\,\frac{\mathrm{d}x}{\mathrm{vol}_d(K(t))}\,\ell(t)\mathrm{d}t = \int_0^{\|\varrho\|_\infty} U_t(f)\,\ell(t)\,\mathrm{d}t, \end{aligned}$$

which proves (5). ☐

2.1. A useful representation

As in [35, Section 3.3], we derive a suitable representation of H and U. We define a $(d+1)$-dimensional auxiliary state space. Let

$$K_\varrho = \{(x, t) \in \mathbb{R}^{d+1} \mid x \in K,\ \ t \in (0, \varrho(x))\}$$

and let μ be the uniform distribution on $(K_\varrho, \mathcal{B}(K_\varrho))$, i.e.

$$\mu(\mathrm{d}(x, t)) = \frac{\mathrm{d}t\,\mathrm{d}x}{\mathrm{vol}_{d+1}(K_\varrho)}.$$

Note that $\mathrm{vol}_{d+1}(K_\varrho) = \int_K \varrho(x)\,\mathrm{d}x$. By $L_{2,\mu} = L_2(K_\varrho, \mu)$ we denote the space of functions $f : K_\varrho \to \mathbb{R}$ that satisfy $\|f\|_{2,\mu}^2 := \langle f, f\rangle_\mu < \infty$, where

$$\langle f, g\rangle_\mu := \int_{K_\varrho} f(x, s)\,g(x, s)\,\mu(\mathrm{d}(x, s))$$

denotes the corresponding inner product for $f, g \in L_{2,\mu}$. Here, similarly to (6), we have

$$\|f\|_{2,\mu}^2 = \int_0^{\|\varrho\|_\infty} \|f(\cdot, s)\|_{2,s}^2\,\ell(s)\mathrm{d}s.$$

Let $T : L_{2,\mu} \to L_{2,\pi}$ and $T^* : L_{2,\pi} \to L_{2,\mu}$ be given by

$$Tf(x) = \frac{1}{\varrho(x)} \int_0^{\varrho(x)} f(x, s)\,\mathrm{d}s, \quad \text{and} \quad T^*f(x, s) = f(x).$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Then T^* is the adjoint operator of T, i.e. for all $f \in L_{2,\pi}$ and $g \in L_{2,\mu}$ we have

$$\langle f, Tg \rangle_\pi = \langle T^* f, g \rangle_\mu .$$

Then, for $f \in L_{2,\mu}$, define

$$\widetilde{H} f(x, s) = \int_{K(s)} f(y, s)\, H_s(x, \mathrm{d}y).$$

By the stationarity of U_s according to H_s it is easily seen that

$$\begin{aligned} \|\widetilde{H} f\|_{2,\mu}^2 &= \int_K \int_0^{\varrho(x)} \left|\widetilde{H} f(x, s)\right|^2 \frac{\mathrm{d}s\, \mathrm{d}x}{\int_K \varrho(y)\mathrm{d}y} = \int_0^{\|\varrho\|_\infty} \int_{K(s)} \left|\widetilde{H} f(x, s)\right|^2 U_s(\mathrm{d}x)\, \ell(s)\mathrm{d}s \\ &\le \int_0^{\|\varrho\|_\infty} \int_{K(s)} \int_{K(s)} |f(y, s)|^2 H_s(x, \mathrm{d}y)\, U_s(\mathrm{d}x)\, \ell(s)\mathrm{d}s \\ &= \int_0^{\|\varrho\|_\infty} \int_{K(s)} |f(x, s)|^2\ U_s(\mathrm{d}x)\, \ell(s)\mathrm{d}s = \|f\|_{2,\mu}^2 . \end{aligned}$$

Furthermore, define

$$\widetilde{U} f(x, s) = \int_{K(s)} f(y, s)\, U_s(\mathrm{d}y).$$

Then $\widetilde{H} \colon L_{2,\mu} \to L_{2,\mu}$, $\widetilde{U} \colon L_{2,\mu} \to L_{2,\mu}$, and

$$\|\widetilde{H}\|_{L_{2,\mu} \to L_{2,\mu}} = 1, \qquad \|\widetilde{U}\|_{L_{2,\mu} \to L_{2,\mu}} = 1.$$

By construction we have the following.

Lemma 3. *Let H, U, T, T^*, $\widetilde{H}$, and $\widetilde{U}$ be as above. Then*

$$H = T\widetilde{H}T^* \quad \textit{and} \quad U = T\widetilde{U}T^*.$$

Here $TT^* \colon L_{2,\pi} \to L_{2,\pi}$ satisfies $TT^* f(x) = f(x)$, i.e. TT^* is the identity operator, and $T^*T \colon L_{2,\mu} \to L_{2,\mu}$ satisfies

$$T^*Tf(x, s) = Tf(x),$$

i.e. it returns the average of the function $f(x, \cdot)$ over the second variable.

3. On the spectral gap of hybrid slice samplers

We start with a relation between the convergence on the slices and the convergence of $T\widetilde{H}^k T^*$ to $T\widetilde{U}T^*$ for increasing k.

Lemma 4. *Let $k \in \mathbb{N}$. Then*

$$\left\| T(\widetilde{H}^k - \widetilde{U})T^* \right\|_{L_{2,\pi} \to L_{2,\pi}} \le \sup_{x \in K} \left(\int_0^{\varrho(x)} \left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2 \frac{\mathrm{d}t}{\varrho(x)} \right)^{1/2} .$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Proof. First note that $\|f\|_{2,\pi} < \infty$ implies $\|f\|_{2,t} < \infty$ for ℓ-almost every t. For any $k \in \mathbb{N}$ and $f \in L_{2,\pi}$ we have

$$(\widetilde{H}^k T^* f)(x, t) = (H_t^k f)(x) \quad \text{and} \quad (\widetilde{U} T^* f)(x, t) = U_t(f),$$

so that

$$T(\widetilde{H}^k - \widetilde{U})T^* f(x) = \int_0^{\|\varrho\|_\infty} (H_t^k - U_t) f(x) \, \frac{\mathbf{1}_{K(t)}(x)}{\varrho(x)} \, \mathrm{d}t.$$

It follows that

$$\begin{aligned}
\left\| T(\widetilde{H}^k - \widetilde{U})T^* f \right\|_{2,\pi}^2 &= \int_K \left| \int_0^{\|\varrho\|_\infty} (H_t^k - U_t) f(x) \, \frac{\mathbf{1}_{K(t)}(x)}{\varrho(x)} \, \mathrm{d}t \right|^2 \pi(\mathrm{d}x) \\
&\leq \int_K \int_0^{\|\varrho\|_\infty} \left| (H_t^k - U_t) f(x) \right|^2 \frac{\mathbf{1}_{K(t)}(x)}{\varrho(x)} \, \mathrm{d}t \, \frac{\varrho(x)}{\int_K \varrho(y) \, \mathrm{d}y} \, \mathrm{d}x \\
&= \int_0^{\|\varrho\|_\infty} \int_{K(t)} \left| (H_t^k - U_t) f(x) \right|^2 \frac{\mathrm{d}x}{\mathrm{vol}_d(K(t))} \, \frac{\mathrm{vol}_d(K(t))}{\int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s)) \, \mathrm{d}s} \, \mathrm{d}t \\
&= \int_0^{\|\varrho\|_\infty} \left\| (H_t^k - U_t) f \right\|_{2,t}^2 \ell(t) \, \mathrm{d}t \\
&\leq \int_0^{\|\varrho\|_\infty} \left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2 \|f\|_{2,t}^2 \, \ell(t) \, \mathrm{d}t \\
&= \int_0^{\|\varrho\|_\infty} \int_{K(t)} \left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2 |f(x)|^2 \frac{\mathrm{d}x}{\mathrm{vol}_d(K(t))} \, \frac{\mathrm{vol}_d(K(t))}{\int_0^{\|\varrho\|_\infty} \mathrm{vol}_d(K(s)) \, \mathrm{d}s} \, \mathrm{d}t \\
&= \int_K \int_0^{\varrho(x)} \left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2 \frac{\mathrm{d}t}{\varrho(x)} \, |f(x)|^2 \, \frac{\varrho(x)}{\int_K \varrho(y) \, \mathrm{d}y} \mathrm{d}x \\
&\leq \|f\|_{2,\pi}^2 \, \sup_{x \in K} \int_0^{\varrho(x)} \left\| H_t^k - U_t \right\|_{L_{2,t} \to L_{2,t}}^2 \frac{\mathrm{d}t}{\varrho(x)}.
\end{aligned}$$

□

Remark 1. If there exists a number $\beta \in [0, 1]$ such that $\|H_t - U_t\|_{L_{2,t} \to L_{2,t}} \leq \beta$ for any $t \in (0, \|\varrho\|_\infty)$, then one obtains (as a consequence of the former lemma) that

$$\left\| T\widetilde{H}^k T^* - S \right\|_{L_{2,\pi} \to L_{2,\pi}} \leq \left\| T\widetilde{U}T^* - S \right\|_{L_{2,\pi} \to L_{2,\pi}} + \beta^k.$$

Here we have employed the triangle inequality and the fact that

$$\|H_t^k - U_t\|_{L_{2,t} \to L_{2,t}} \leq \|H_t - U_t\|_{L_{2,t} \to L_{2,t}}^k \leq \beta^k;$$

see for example [33, Lemma 3.16].

Now a corollary follows which provides a lower bound for $\mathrm{gap}(T\widetilde{H}^k T^*)$.

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Corollary 1. *Let us assume that* $\text{gap}(U) > 0$, *i.e.* $\|U - S\|_{L_{2,\pi}\to L_{2,\pi}} < 1$, *and let*

$$\beta_k = \sup_{x\in K}\left(\int_0^{\varrho(x)} \left\|H_t^k - U_t\right\|^2_{L_{2,t}\to L_{2,t}} \frac{\mathrm{d}t}{\varrho(x)}\right)^{1/2}.$$

Then

$$\text{gap}(T\widetilde{H}^k T^*) \geq \text{gap}(U) - \beta_k. \tag{7}$$

Proof. It is enough to prove

$$\left\|T\widetilde{H}^k T^* - S\right\|_{L_{2,\pi}\to L_{2,\pi}} \leq \|U - S\|_{L_{2,\pi}\to L_{2,\pi}} + \beta_k.$$

By $\widetilde{H}^k = \widetilde{U} + \widetilde{H}^k - \widetilde{U}$ and Lemma 4 we have

$$\begin{aligned}\left\|T\widetilde{H}^k T^* - S\right\|_{L_{2,\pi}\to L_{2,\pi}} &= \left\|T\widetilde{U}T^* - S + T(\widetilde{H}^k - \widetilde{U})T^*\right\|_{L_{2,\pi}\to L_{2,\pi}} \\ &\leq \|U - S\|_{L_{2,\pi}\to L_{2,\pi}} + \beta_k.\end{aligned}$$

□

Remark 2. If one can sample with respect to U_t for every $t \geq 0$, then $H_t = U_t$, and in the estimate of Corollary 1 we obtain $\beta_k = 0$ and equality in (7).

Now let us state the main theorem.

Theorem 1. *Let us assume that for almost all* t *(with respect to* ℓ*)* H_t *is positive semidefinite on* $L_{2,t}$, *and let*

$$\beta_k = \sup_{x\in K}\left(\int_0^{\varrho(x)} \left\|H_t^k - U_t\right\|^2_{L_{2,t}\to L_{2,t}} \frac{\mathrm{d}t}{\varrho(x)}\right)^{1/2}.$$

Further assume that $\lim_{k\to\infty}\beta_k = 0$. *Then*

$$\frac{\text{gap}(U) - \beta_k}{k} \leq \text{gap}(H) \leq \text{gap}(U), \qquad k \in \mathbb{N}. \tag{8}$$

Several conclusions can be drawn from the theorem. First, under the assumption that $\lim_{k\to\infty}\beta_k = 0$, the left-hand side of (8) implies that in the setting of the theorem, whenever the simple slice sampler has a spectral gap, so does the hybrid version. See Section 4 for examples. Second, it also provides a quantitative bound on $\text{gap}(H)$ given appropriate estimates on $\text{gap}(U)$ and β_k. Third, the right-hand side of (8) verifies the intuitive result that the simple slice sampler is better than the hybrid one (in terms of the spectral gap).

To prove the theorem we need some further results.

Lemma 5.

1. *For any* $t \in (0, \|\varrho\|_\infty)$ *assume that* H_t *is reversible with respect to* U_t. *Then* $\widetilde{H}$ *is self-adjoint on* $L_{2,\mu}$.
2. *Assume that for almost all* t *(with respect to* ℓ*)* H_t *is positive semidefinite on* $L_{2,t}$, *i.e. for all* $f \in L_{2,t}$, *it holds that* $\langle H_t f, f\rangle_t \geq 0$. *Then* $\widetilde{H}$ *is positive semidefinite on* $L_{2,\mu}$.

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Proof. Note that $\|f\|_{2,\mu} < \infty$ implies $\|f(\cdot, t)\|_{2,t} < \infty$ for almost all t (with respect to ℓ).

Part 1: Let $f, g \in L_{2,\mu}$; then we have to show that

$$\langle \widetilde{H} f, g\rangle_\mu = \langle f, \widetilde{H} g\rangle_\mu.$$

Note that for $f, g \in L_{2,\mu}$ we have for almost all t, by the reversibility of H_t, that

$$\langle H_t f(\cdot, t), g(\cdot, t)\rangle_t = \langle f(\cdot, t), H_t g(\cdot, t)\rangle_t.$$

Since

$$\begin{aligned}
\langle \widetilde{H} f, g\rangle_\mu &= \int_{K_\varrho} \widetilde{H} f(x, t) g(x, t)\, \mu(\mathrm{d}(x, t)) \\
&= \int_K \int_0^{\varrho(x)} \int_{K(t)} f(y, t)\, H_t(x, \mathrm{d}y) g(x, t)\, \frac{\mathrm{d}t\, \mathrm{d}x}{\mathrm{vol}_{d+1}(K_\varrho)} \\
&= \int_0^{\|\varrho\|_\infty} \int_{K(t)} \int_{K(t)} f(y, t)\, H_t(x, \mathrm{d}y) g(x, t)\, U_t(\mathrm{d}y)\, \ell(t)\, \mathrm{d}t \\
&= \int_0^{\|\varrho\|_\infty} \langle H_t f(\cdot, t), g(\cdot, t)\rangle_t\, \ell(t)\, \mathrm{d}t,
\end{aligned}$$

the assertion of Part 1 holds.

Part 2. We have to prove for all $f \in L_{2,\mu}$ that

$$\langle \widetilde{H} f, f\rangle_\mu = \int_{K_\varrho} \widetilde{H} f(x, t) f(x, t)\, \mu(\mathrm{d}(x, t)) \geq 0.$$

Note that for $f \in L_{2,\mu}$ we have for almost all t that

$$\langle H_t f(\cdot, t), f(\cdot, t)\rangle_t \geq 0.$$

By the same computation as in Part 1 we obtain that the positive semidefiniteness of H_t carries over to $\widetilde{H}$. □

The statements and proofs of the following lemmas closely follow the lines of argument in [39, 40] and make essential use of [40, Lemma 12 and Lemma 13]. We provide alternative proofs of the aforementioned lemmas in Appendix A.

Lemma 6. *Let $\widetilde{H}$ be positive semidefinite on $L_{2,\mu}$. Then*

$$\left\| T\widetilde{H}^{k+1} T^* - S \right\|_{L_{2,\pi} \to L_{2,\pi}} \leq \left\| T\widetilde{H}^{k} T^* - S \right\|_{L_{2,\pi} \to L_{2,\pi}}, \qquad k \in \mathbb{N}. \tag{9}$$

Furthermore, if

$$\beta_k = \sup_{x \in K} \left(\int_0^{\varrho(x)} \left\| H_t^k - U_t \right\|^2_{L_{2,t} \to L_{2,t}} \frac{\mathrm{d}t}{\varrho(x)} \right)^{1/2}$$

and $\lim_{k\to\infty} \beta_k = 0$*, then*

$$\|U - S\|_{L_{2,\pi}\to L_{2,\pi}} \le \|H - S\|_{L_{2,\pi}\to L_{2,\pi}} .$$

Proof. Let $S_1 : L_{2,\mu} \to L_{2,\pi}$ and the adjoint $S_1^* : L_{2,\pi} \to L_{2,\mu}$ be given by

$$S_1(f) = \int_{K_\varrho} f(x, s)\, \mu(\mathrm{d}(x, s)) \quad \text{and} \quad S_1^*(g) = \int_K g(x)\, \pi(\mathrm{d}x).$$

Thus, $\langle S_1 f, g\rangle_\pi = \langle f, S_1^* g\rangle_\mu$. Furthermore, observe that $S_1 S_1^* = S$. Let $R = T - S_1$ and note that $RR^* = I - S$, with identity I, and $RR^* = (RR^*)^2$. Since $RR^* \neq 0$, and by the projection property $RR^* = (RR^*)^2$, one gets $\|RR^*\|_{L_{2,\pi}\to L_{2,\pi}} = 1$. We have

$$\begin{aligned} R\widetilde{H}^k R^* &= (T - S_1)\widetilde{H}^k (T^* - S_1^*) \\ &= T\widetilde{H}^k T^* - T\widetilde{H}^k S_1^* - S_1\widetilde{H}^k T^* + S_1\widetilde{H}^k S_1^* = T\widetilde{H}^k T^* - S. \end{aligned}$$

Furthermore, $\left\| S_1 \widetilde{H} S_1^* \right\|_{L_{2,\mu}\to L_{2,\mu}} \le 1$. By Lemma 12 it follows that

$$\left\| R\widetilde{H}^{k+1} R^* \right\|_{L_{2,\pi}\to L_{2,\pi}} \le \left\| R\widetilde{H}^k R^* \right\|_{L_{2,\pi}\to L_{2,\pi}} ,$$

and the proof of (9) is completed.

By Lemma 4 we obtain $\left\| T(\widetilde{H}^k - \widetilde{U})T^* \right\|_{L_{2,\pi}\to L_{2,\pi}} \le \beta_k$, and by (9) as well as Lemma 3 we obtain

$$\left\| T\widetilde{H}^k T^* - S \right\|_{L_{2,\pi}\to L_{2,\pi}} \le \|H - S\|_{L_{2,\pi}\to L_{2,\pi}} , \quad k \in \mathbb{N}.$$

This implies by the triangle inequality that

$$\lim_{k\to\infty} \left\| T\widetilde{H}^k T^* - S \right\|_{L_{2,\pi}\to L_{2,\pi}} = \|U - S\|_{L_{2,\pi}\to L_{2,\pi}} ,$$

and the assertion is proven. □

Lemma 7. *Let* $\widetilde{H}$ *be positive semidefinite on* $L_{2,\mu}$*. Then*

$$\|H - S\|^k_{L_{2,\pi}\to L_{2,\pi}} \le \left\| T\widetilde{H}^k T^* - S \right\|_{L_{2,\pi}\to L_{2,\pi}} , \tag{10}$$

for any $k \in \mathbb{N}$.

Proof. As in the proof of Lemma 6 we use $R\widetilde{H}^k R^* = T\widetilde{H}^k T^* - S$ to reformulate the assertion. It remains to prove that

$$\left\| R\widetilde{H} R^* \right\|^k_{L_{2,\pi}\to L_{2,\pi}} \le \left\| R\widetilde{H}^k R^* \right\|_{L_{2,\pi}\to L_{2,\pi}} .$$

Recall that RR^* is a projection and satisfies $\|RR^*\|_{L_{2,\pi}\to L_{2,\pi}} = 1$. By Lemma 13 the assertion is proven.

Now we turn to the proof of Theorem 1.

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Proof of Theorem 1. By Lemma 5 we know that $\widetilde{H}: L_{2,\mu} \to L_{2,\mu}$ is self-adjoint and positive semidefinite. By Lemma 6 we have

$$\|U - S\|_{L_{2,\pi}\to L_{2,\pi}} \le \|H - S\|_{L_{2,\pi}\to L_{2,\pi}} .$$

By Theorem 1 we have for any $k \in \mathbb{N}$ that

$$\left\|T\widetilde{H}^k T^* - S\right\|_{L_{2,\pi}\to L_{2,\pi}} \le \|U - S\|_{L_{2,\pi}\to L_{2,\pi}} + \beta_k. \tag{11}$$

Then

$$\begin{aligned}
\|U - S\|_{L_{2,\pi}\to L_{2,\pi}} &\underset{11}{\ge} \left\|T\widetilde{H}^k T^* - S\right\|_{L_{2,\pi}\to L_{2,\pi}} - \beta_k \\
&\underset{10}{\ge} \|H - S\|^k_{L_{2,\pi}\to L_{2,\pi}} - \beta_k \\
&\ge 1 - k\left(1 - \|H - S\|_{L_{2,\pi}\to L_{2,\pi}}\right) - \beta_k \\
&= 1 - k\,\mathrm{gap}(H) - \beta_k,
\end{aligned}$$

where we applied a version of Bernoulli's inequality, i.e. $1 - x^n \le n(1 - x)$ for $x \ge 0$ and $n \in \mathbb{N}$. Thus

$$\frac{\mathrm{gap}(U) - \beta_k}{k} \le \mathrm{gap}(H),$$

and the proof is completed. □

4. Applications

In this section we apply Theorem 1 under different assumptions with different Markov chains on the slices. We provide a criterion for geometric ergodicity of these hybrid slice samplers by showing that there is a spectral gap whenever the simple slice sampler has a spectral gap.

First we consider a class of bimodal densities in a 1-dimensional setting. We study a stepping-out shrinkage slice sampler, suggested in [26], which is explained in Algorithm 1.

Then we consider a hybrid slice sampler which performs a hit-and-run step on the slices in a d-dimensional setting. Here we impose very weak assumptions on the unnormalized densities. The drawback is that an implementation of this algorithm may be difficult.

Motivated by this difficulty, we study a combination of the previous sampling procedures on the slices. The resulting hit-and-run stepping-out shrinkage slice sampler is presented in Algorithm 2. Here we consider a class of bimodal densities in a d-dimensional setting.

4.1. Stepping-out and shrinkage procedure

Let $w > 0$ be a parameter and $\varrho: \mathbb{R} \to (0, \infty)$ be an unnormalized density. We say $\varrho \in \mathcal{R}_w$ if there exist $t_1, t_2 \in (0, \|\varrho\|_\infty)$ with $t_1 \le t_2$ such that the following hold:

1. For all $t \in (0, t_1) \cup [t_2, \|\varrho\|_\infty)$, the level set $K(t)$ is an interval.
2. For all $t \in [t_1, t_2)$, there are disjoint intervals $K_1(t)$, $K_2(t)$ with strictly positive Lebesgue measure such that

$$K(t) = K_1(t) \cup K_2(t),$$

and for all $\varepsilon > 0$ it holds that $K_i(t+\varepsilon) \subseteq K_i(t)$ for $i = 1, 2$. For convenience we set $K_i(t) = \emptyset$ for $t \notin [t_1, t_2)$.

3. For all $t \in (0, \|\varrho\|_\infty)$ we assume $\delta_t < w$, where

$$\delta_t := \begin{cases} \inf_{r \in K_1(t),\, s \in K_2(t)} |r - s|, & t \in [t_1, t_2), \\ 0 & \text{otherwise.} \end{cases}$$

The next result shows that certain bimodal densities belong to $\mathcal{R}_w$.

Lemma 8. *Let* $\varrho_1 \colon \mathbb{R} \to (0, \infty)$ *and* $\varrho_2 \colon \mathbb{R} \to (0, \infty)$ *be unnormalized density functions. Let us assume that* ϱ_1, ϱ_2 *are lower semicontinuous and quasi-concave, i.e. the level sets are open intervals, and*

$$\inf_{r \in \arg\max \varrho_1,\, s \in \arg\max \varrho_2} |r - s| < w.$$

Then $\varrho_{\max} := \max\{\varrho_1, \varrho_2\} \in \mathcal{R}_w$.

Proof. For $t \in (0, \|\varrho_{\max}\|_\infty)$ let $K_{\varrho_{\max}}(t)$, $K_{\varrho_1}(t)$, and $K_{\varrho_2}(t)$ be the level sets of $\varrho_{\max}$, ϱ_1, and ϱ_2 of level t. Note that

$$K_{\varrho_{\max}}(t) = K_{\varrho_1}(t) \cup K_{\varrho_2}(t).$$

With the choice

$$t_1 = \inf\{t \in (0, \|\varrho_{\max}\|_\infty) : K_{\varrho_1}(t) \cap K_{\varrho_2}(t) = \emptyset\},$$
$$t_2 = \min\{\|\varrho_1\|_\infty, \|\varrho_2\|_\infty\},$$

we have the properties 1 and 2. Observe that $\arg\max \varrho_i \subseteq K_{\varrho_i}(t)$ for $i = 1, 2$, which yields

$$\inf_{r \in K_{\varrho_1}(t),\, s \in K_{\varrho_2}(t)} |r - s| \le \inf_{r \in \arg\max \varrho_1,\, s \in \arg\max \varrho_2} |r - s| < w.$$

□

In [26] a stepping-out and shrinkage procedure is suggested for the transitions on the level sets. The procedures are explained in Algorithm 1, where a single transition from the resulting hybrid slice sampler from x to y is presented.

Algorithm 1. *A hybrid slice sampling transition of the stepping-out and shrinkage procedure from x to y, i.e. with input x and output y. The stepping-out procedure has input x (current state), t (chosen level),* $w > 0$ *(step size parameter from* $\mathcal{R}_w$*) and outputs an interval [L, R]. The shrinkage procedure has input [L, R] and output y:*

1. *Choose a level* $t \sim \mathcal{U}(0, \varrho(x))$;
2. *Stepping-out with input x, t, w outputs an interval [L, R]:*
   (a) *Choose* $u \sim \mathcal{U}[0, 1]$. *Set* $L = x - uw$ *and* $R = L + w$;
   (b) *Repeat until* $t \ge \varrho(L)$, *i.e.* $L \notin K(t)$: *Set* $L = L - w$;
   (c) *Repeat until* $t \ge \varrho(R)$, *i.e.* $R \notin K(t)$: *Set* $R = R + w$;

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

3. *Shrinkage procedure with input [L, R] outputs y:*

   (a) *Set* $\bar{L} = L$ *and* $\bar{R} = R$*;*

   (b) *Repeat:*

      i. *Choose* $v \sim \mathcal{U}[0, 1]$ *and set* $y = \bar{L} + v(\bar{R} - \bar{L})$*;*

      ii. *If* $y \in K(t)$ *then return y and exit the loop;*

      iii. *If* $y < x$ *then set* $\bar{L} = y$*, else* $\bar{R} = y$*.*

For short we write $|K(t)| = \mathrm{vol}_1(K(t))$, and for $t \in (0, \|\varrho\|_\infty)$ we set

$$\gamma_t := \frac{(w - \delta_t)}{w} \frac{|K(t)|}{(\,|K(t)| + \delta_t)}.$$

Now we provide useful results for applying Theorem 1.

Lemma 9. *Let* $\varrho \in \mathcal{R}_w$ *with* $t_2 > 0$ *satisfying the properties 1 and 2 of the definition of* $\mathcal{R}_w$*. Moreover, let* $t \in (0, \|\varrho\|_\infty)$*.*

1. *The transition kernel* H_t *of the stepping-out and shrinkage slice sampler from Algorithm 1 takes the form*

$$H_t(x, A) = \gamma_t\, U_t(A) + (1 - \gamma_t) \left[\mathbf{1}_{K_1(t)}(x) U_{t,1}(A) + \mathbf{1}_{K_2(t)}(x) U_{t,2}(A)\right],$$

   with $x \in \mathbb{R}$*,* $A \in \mathcal{B}(\mathbb{R})$*, and*

$$U_{t,i}(A) = \begin{cases} \frac{|K_i(t) \cap A|}{|K_i(t)|}, & t \in [t_1, t_2), \\ 0, & t \in (0, t_1) \cup [t_2, \|\varrho\|_\infty), \end{cases}$$

   for $i = 1, 2$*; i.e. in the case* $t \in [t_1, t_2)$ *we have that* $U_{t,i}$ *denotes the uniform distribution in* $K_i(t)$*. (For* $t \in (0, t_1) \cup [t_2, \|\varrho\|_\infty$ *) we have* $H_t = U_t$ *since* $\delta_t = 0$ *yields* $\gamma_t = 1$*.)*

2. *The transition kernel* H_t *is reversible and induces a positive semidefinite operator, i.e. for any* $f \in L_{2,t}$*, it holds that* $\langle H_t f, f \rangle_t \geq 0$*.*

3. *We have* $\|H_t - U_t\|_{L_{t,2} \to L_{t,2}} = 1 - \gamma_t$ *and*

$$\beta_k \leq \left(\frac{1}{t_2} \int_0^{t_2} (1 - \gamma_t)^{2k}\, \mathrm{d}t \right)^{1/2}, \quad k \in \mathbb{N}. \tag{12}$$

Proof. **Part 1.** For $t \in (0, t_1) \cup [t_2, \|\varrho\|_\infty)$ the stepping-out procedure returns an interval that contains $K(t)$ entirely. Then, since $K(t)$ is also an interval, the shrinkage scheme returns a sample with respect to U_t in $K(t)$.

For $t \in [t_1, t_2)$, $i \in \{1, 2\}$, and $x \in K_i(t)$, within the stepping-out procedure, with probability $(w - \delta_t)/w$ an interval that contains $K(t) = K_1(t) \cup K_2(t)$ is returned, and with probability

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

$1-(w-\delta_t)/w$ an interval that contains $K_i(t)$ but not $K(t)\setminus K_i(t)$ is returned. We distinguish these cases:

Case 1: K(t) contained in the stepping-out output:

Then with probability $|K(t)|/(|K(t)|+\delta_t)$ the shrinkage scheme returns a sample with respect to U_t, and with probability $1-|K(t)|/(|K(t)|+\delta_t)$ it returns a sample with respect to $U_{t,i}$.

Case 2: $K_i(t)$, but not $K(t)\setminus K_i(t)$, *contained in the stepping-out output:*

Then with probability 1 the shrinkage scheme returns a sample with respect to $U_{t,i}$. In total, for $x\in K_i(t)$ we obtain

$$\begin{aligned} H_t(x,A) &= \frac{(w-\delta_t)}{w}\Big[\frac{|K(t)|}{|K(t)|+\delta_t}\,U_t(A)+\big(1-\frac{|K(t)|}{|K(t)|+\delta_t}\big)U_{t,i}(A)\Big] \\ &\quad+\big(1-\frac{(w-\delta_t)}{w}\big)U_{t,i}(A) \\ &= \gamma_t U_t(A)+(1-\gamma_t)U_{t,i}(A), \end{aligned}$$

where we emphasize that for $t\in(0,t_1)\cup[t_2,\|\varrho\|_\infty)$ it follows that $\gamma_t=1$ (since $\delta_t=0$), so that $H_t(x,A)$ coincides with $U_t(A)$.

Part 2. For $A,B\in\mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned} \int_A H_t(x,A)\,U_t(\mathrm{d}x) &= \gamma_t\,U_t(B)U_t(A) \\ &\quad+(1-\gamma_t)\int_A\big[\mathbf{1}_{K_1(t)}(x)U_{t,1}(B)+\mathbf{1}_{K_2(t)}(x)U_{t,2}(B)\big]U_t(\mathrm{d}x) \\ &= \gamma_t\,U_t(B)U_t(A)+(1-\gamma_t)\Big[\frac{|K_1(t)|}{|K(t)|}U_{t,1}(A)U_{t,1}(B)+\frac{|K_2(t)|}{|K(t)|}U_{t,2}(A)U_{t,2}(B)\Big], \end{aligned}$$

which is symmetric in A,B and therefore implies the claimed reversibility with respect to U_t. Similarly, we have

$$\langle H_t f,f\rangle_t=\gamma_t\,U_t(f)^2+(1-\gamma_t)\left[\frac{|K_1(t)|}{|K(t)|}\,U_{t,1}(f)^2+\frac{|K_2(t)|}{|K(t)|}\,U_{t,2}(f)^2\right]\ge 0, \tag{13}$$

where $U_{t,i}(f)$ denotes the integral of f with respect to $U_{t,i}$ for $i=1,2$, which proves the positive semidefiniteness.

Part 3. By virtue of [33, Lemma 3.16], the reversibility (or equivalently self-adjointness) of H_t, and [42, Theorem V.5.7], we have

$$\|H_t-U_t\|_{L_{2,t}\to L_{2,t}}=\sup_{\|f\|_{2,t}\le 1,\,U_t(f)=0}|\langle H_t f,f\rangle|=\sup_{\|f\|_{2,t}\le 1,\,U_t(f)=0}\langle H_t f,f\rangle, \tag{14}$$

where the last equality follows by the positive semidefiniteness. Observe that for any $f\in L_{2,s}$ with $s\in[t_1,t_2)$ we have by $U_{s,i}(f)^2\le U_{s,i}(f^2)$ for $i=1,2$ that

$$\frac{|K_1(s)|}{|K(s)|}\,U_{s,1}(f)^2+\frac{|K_2(s)|}{|K(s)|}\,U_{s,2}(f)^2\le\frac{|K_1(s)|}{|K(s)|}\,U_{s,1}(f^2)+\frac{|K_2(s)|}{|K(s)|}\,U_{s,2}(f^2)=\|f\|_{2,s}^2.$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Therefore, the equality in (13) yields

$$\|H_t - U_t\|_{L_{2,t}\to L_{2,t}} \le \sup_{\|f\|_{2,t}\le 1,\, U_t(f)=0} (1-\gamma_t)\|f\|_{2,t}^2 = 1-\gamma_t. \tag{15}$$

For $t\in(0, t_1)\cup[t_2, \|\varrho\|_\infty)$, by $H_t = U_t$ and $1-\gamma_t = 0$ we have an equality. For $t\in[t_1, t_2)$ with

$$h(x) = \frac{|K(t)|}{|K_1(t)|}\,\mathbf{1}_{K_1(t)}(x) - \frac{|K(t)|}{|K_2(t)|}\,\mathbf{1}_{K_2(t)}(x)$$

the upper bound of (15) is attained for $f = h/\|h\|_{2,t}$ in the supremum expression of (14).

We turn to the verification of (12). For $t\in(t_2, \|\varrho\|_\infty]$ we have $1-\gamma_t = 0$, and for $t\in(0, t_2)$ the function $1-\gamma_t$ is increasing, which also yields that $t\mapsto(1-\gamma_t)^{2k}$ is increasing on $(0, t_2)$ for any $k\in\mathbb{N}$. By [33, Lemma 3.16] we obtain

$$\|H_t^k - U_t\|_{L_{2,t}\to L_{2,t}} \le (1-\gamma_t)^k.$$

Consequently, we have

$$\beta_k \le \sup_{r\in(0,t_2)} \frac{1}{r}\int_0^r (1-\gamma_t)^{2k}\mathrm{d}t. \tag{16}$$

Furthermore, note that for $a\in(0,\infty)$, any increasing function $g\colon(0,a)\to\mathbb{R}$, and $p, q\in(0,a)$ with $p\le q$, it holds that

$$\frac{1}{p}\int_0^p g(t)\,\mathrm{d}t \le \frac{1}{q}\int_0^q g(t)\,\mathrm{d}t. \tag{17}$$

The former inequality can be verified by showing that the function $p\mapsto G(p)$ for $p\ge 0$ with $G(p) = \frac{1}{p}\int_0^p g(t)\,\mathrm{d}t$ satisfies $G'(p)\ge 0$. Applying (17) with $g(t) = (1-\gamma_t)^{2k}$ in combination with (16) yields (12). □

By Theorem 1 and the previous lemma we have the following result.

Corollary 2. *For any $\varrho\in\mathcal{R}_w$ the stepping-out and shrinkage slice sampler has a spectral gap if and only if the simple slice sampler has a spectral gap.*

Remark 3. We want to discuss two extremal situations:

- Consider densities $\varrho\colon\mathbb{R}\to(0,\infty)$ that are lower semicontinuous and quasi-concave, i.e. the level sets are open intervals. Loosely speaking, we assume we have unimodal densities. Then for any $w>0$ we have $\varrho\in\mathcal{R}_w$ (just take $t_1 = t_2$ arbitrarily) and $\delta_t = 0$ for all $t\in(0,\|\varrho\|_\infty)$. Hence $H_t = U_t$ for all $t\in(0,\|\varrho\|_\infty)$, and Algorithm 1 provides an effective implementation of simple slice sampling.
- Assume that $\varrho\colon\mathbb{R}\to(0,\infty)$ satisfies the properties 1 and 2 from the definition of $\mathcal{R}_w$ for some $w>0$, but for any $t\in(0,\|\varrho\|_\infty)$ we have $\delta_t\ge w$. In this setting our theory is not applicable and it is clear that the corresponding Markov chain does not work well, since it is not possible to get from one part of the support to the other part.

4.2. Hit-and-run slice sampler

The idea is to combine the hit-and-run algorithm with slice sampling. We ask whether a spectral gap of simple slice sampling implies a spectral gap of this combination. The hit-and-run algorithm was proposed by Smith [38]. It is well studied (see for example [3, 5, 12, 13,

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

16, 17, 35]) and used for numerical integration (see [33, 34]). We define the setting and the transition kernel of hit-and-run.

We consider a d-dimensional state space $K \subseteq \mathbb{R}^d$ and suppose $\varrho\colon K \to (0, \infty)$ is an unnormalized density. We denote the diameter of a level set by

$$\operatorname{diam}(K(t)) = \sup_{x,y \in K(t)} |x - y|$$

with the Euclidean norm $|\cdot|$. We impose the following assumption.

Assumption 2. *The limit* $\kappa := \lim_{t \downarrow 0} \frac{\operatorname{vol}_d(K(t))}{\operatorname{diam}(K(t))^d}$ *exists, and there are numbers* $c, \varepsilon \in (0, 1]$ *such that*

$$\inf_{t \in (0,\varepsilon)} \frac{\operatorname{vol}_d(K(t))}{\operatorname{diam}(K(t))^d} = c > 0. \tag{18}$$

Note that under Assumption 2 we always have $\kappa \geq c$. If K is bounded, has positive Lebesgue measure, and satisfies $\inf_{x \in K} \varrho(x) > 0$, then Assumption 2 is satisfied with $\kappa = c$. Moreover, for instance, the density of a standard normal distribution satisfies Assumption 2 with unbounded K, where again $c = \kappa$. However, the following example indicates that this is not always the case.

Example 1. *Let* $K = (0, 1)^2$ *and* $\varrho(x_1, x_2) = 2 - x_1 - x_2$. *Then for* $t \in (0, 1]$ *we have*

$$K(t) = \{(x_1, x_2) \in (0, 1)^2 \colon x_2 \in (0, \min\{1, 2 - t - x_1\})\},$$

so that $\operatorname{vol}_2(K(t)) = 1 - t^2/2$. *Moreover, the fact that* $\{(\alpha, 1 - \alpha)\colon \alpha \in (0, 1)\} \subseteq K(t)$ *yields* $\operatorname{diam}(K(t)) = \sqrt{2}$, *so that for* $\varepsilon = 1$ *we have* $c = 1/4$ *and* $\kappa = 1/2$.

In general, we consider Assumption 2 a weak regularity requirement, since there is no condition on the level sets and also no condition on the modality.

Let S_{d-1} be the Euclidean unit sphere and $\sigma_d = \operatorname{vol}_{d-1}(S_{d-1})$. A transition from x to y by hit-and-run on the level set $K(t)$ works as follows:

1. Choose $\theta \in S_{d-1}$ uniformly distributed.
2. Choose y according to the uniform distribution on the line $x + r\theta$ intersected with $K(t)$.

This leads to

$$H_t(x, A) = \int_{S_{d-1}} \int_{L_t(x,\theta)} \mathbf{1}_A(x + s\theta) \frac{\mathrm{d}s}{\operatorname{vol}_1(L_t(x, \theta))} \frac{\mathrm{d}\theta}{\sigma_d}$$

with

$$L_t(x, \theta) = \{r \in \mathbb{R} \mid x + r\theta \in K(t)\}.$$

The hit-and-run algorithm is reversible and induces a positive semidefinite operator on $L_{2,t}$; see [35]. The following property is well known; see for example [5].

Proposition 1. *For* $t \in (0, \|\varrho\|_\infty)$, $x \in K(t)$ *and* $A \in \mathcal{B}(K)$ *we have*

$$H_t(x, A) = \frac{2}{\sigma_d} \int_A \frac{\mathrm{d}y}{|x - y|^{d-1} \operatorname{vol}_1(L(x, \frac{x-y}{|x-y|}))} \tag{19}$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

and

$$\|H_t - U_t\|_{L_{2,t}\to L_{2,t}} \leq 1 - \frac{2}{\sigma_d}\frac{\mathrm{vol}_d(K(t))}{\mathrm{diam}(K(t))^d}. \tag{20}$$

Proof. The representation of H_t stated in (19) is well known; see for example [5]. From this we have for any $x \in K(t)$ that

$$H_t(x, A) \geq \frac{2}{\sigma_d}\frac{\mathrm{vol}_d(K(t))}{\mathrm{diam}(K(t))^d} \cdot \frac{\mathrm{vol}_d(K(t)\cap A)}{\mathrm{vol}_d(K(t))},$$

which means that the whole state space $K(t)$ is a small set. By [18] we have uniform ergodicity, and by [33, Proposition 3.24] we obtain (20). □

Furthermore, we obtain the following helpful result.

Lemma 10. *Under Assumption 2, with*

$$\beta_k = \sup_{x\in K}\left(\int_0^{\varrho(x)} \left\|H_t^k - U_t\right\|^2_{L_{2,t}\to L_{2,t}} \frac{\mathrm{d}t}{\varrho(x)}\right)^{1/2},$$

we have that $\lim_{k\to\infty}\beta_k = 0$.

Proof. By (20) and [33, Lemma 3.16], in combination with reversibility (or equivalently self-adjointness) of H_t, it holds that

$$\left\|H_t^k - U_t\right\|^2_{L_{2,t}\to L_{2,t}} \leq \left(1 - \frac{2}{\sigma_d}\frac{\mathrm{vol}_d(K(t))}{\mathrm{diam}(K(t))^d}\right)^{2k}. \tag{21}$$

Let $g_k\colon [0, \|\varrho\|_\infty) \to [0, 1]$ be given by

$$g_k(t) = \begin{cases} \left(1 - \frac{2}{\sigma_d}\frac{\mathrm{vol}_d(K(t))}{\mathrm{diam}(K(t))^d}\right)^{2k}, & t \in (0, \|\varrho\|_\infty), \\ \left(1 - \frac{2\kappa}{\sigma_d}\right)^{2k}, & t = 0, \end{cases}$$

which is the continuous extension at zero of the upper bound of (21) with $\kappa \geq c \in (0, 1]$ from Assumption 2. Note that $\lim_{k\to\infty} g_k(t) = 0$ for all $t \in [0, \|\varrho\|_\infty)$ and $\beta_k \leq \sup_{r\in(0,\|\varrho\|_\infty]}\left(\frac{1}{r}\int_0^r g_k(t)\,\mathrm{d}t\right)^{1/2}$. Considering the continuous function

$$h_k(r) = \begin{cases} \frac{1}{r}\int_0^r g_k(t)\mathrm{d}t, & r \in (0, \|\varrho\|_\infty], \\ g_k(0), & r = 0, \end{cases}$$

the supremum can be replaced by a maximum over $r \in [0, \|\varrho\|_\infty]$ which is attained, say for $r^{(k)} \in [0, \|\varrho\|_\infty]$, i.e. $\beta_k \leq h_k(r^{(k)})^{1/2}$. Define

$$(r_0^{(k)})_{k\in\mathbb{N}} := \{r^{(k)} \mid r^{(k)} = 0,\ k \in \mathbb{N}\} \subseteq (r^{(k)})_{k\in\mathbb{N}},$$

$$(r_1^{(k)})_{k\in\mathbb{N}} := \{r^{(k)} \mid r^{(k)} \in (0, \varepsilon),\ k \in \mathbb{N}\} \subseteq (r^{(k)})_{k\in\mathbb{N}},$$

$$(r_2^{(k)})_{k\in\mathbb{N}} := \{r^{(k)} \mid r^{(k)} \geq \varepsilon,\ k \in \mathbb{N}\} \subseteq (r^{(k)})_{k\in\mathbb{N}}.$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Without loss of generality we assume that $(r_0^{(k)})_{k\in\mathbb{N}} \neq \emptyset$, $(r_1^{(k)})_{k\in\mathbb{N}} \neq \emptyset$ and $(r_2^{(k)})_{k\in\mathbb{N}} \neq \emptyset$. Then $\lim_{k\to\infty} h_k(r_0^{(k)}) = 0$, and using Assumption 2 we have

$$0 \leq \lim_{k\to\infty} h_k(r_1^{(k)}) \leq \lim_{k\to\infty} \sup_{s\in(0,\varepsilon)} g_k(s) \leq \lim_{k\to\infty} \left(1 - \frac{2c}{\sigma_d}\right)^{2k} = 0.$$

Moreover, by the definition of $(r_2^{(k)})_{k\in\mathbb{N}}$, note that $1/r_2^{(k)} \cdot \mathbf{1}_{(0,r_2^{(k)})}(t) \leq \varepsilon^{-1}$ for $t \in (0, \infty)$, so that

$$\lim_{k\to\infty} h_k(r_1^{(k)}) = \lim_{k\to\infty} \int_0^{\|\varrho\|_\infty} \frac{\mathbf{1}_{(0,r_1^{(k)})}(t)}{r_1^{(k)}} g_k(t)\mathrm{d}t = \int_0^{\|\varrho\|_\infty} \lim_{k\to\infty} \frac{\mathbf{1}_{(0,r_1^{(k)})}(t)}{r_1^{(k)}} g_k(t)\mathrm{d}t = 0.$$

Consequently $\lim_{k\to\infty} h_k(r^{(k)}) = 0$, so that $\lim_{k\to\infty} \beta_k \leq \lim_{k\to\infty} h_k(r^{(k)})^{1/2} = 0$. □

By Theorem 1, this observation leads to the following result.

Corollary 3. *Let $\varrho\colon K \to (0, \infty)$ and let Assumption 2 be satisfied. Then the hit-and-run slice sampler has an absolute spectral gap if and only if the simple slice sampler has an absolute spectral gap.*

We stress that we do not know whether the level sets of ϱ are convex or star-shaped or have any additional structure. In this sense the assumptions imposed on ϱ can be considered weak. This also means that it may be difficult to implement hit-and-run in this generality. In the next section we consider a combination of hit-and-run, stepping-out, and the shrinkage procedure, for which we provide a concrete implementable algorithm.

4.3. Hit-and-run, stepping-out, shrinkage slice sampler

We combine hit-and-run, stepping-out, and the shrinkage procedure. Let $w > 0$, let $K \subseteq \mathbb{R}^d$, and assume that $\varrho\colon K \to (0, \infty)$. We say $\varrho \in \mathcal{R}_{d,w}$ if the following conditions are satisfied:

1. There are not necessarily normalized lower semicontinuous and quasi-concave densities $\varrho_1, \varrho_2\colon K \to (0, \infty)$, i.e. the level sets are open and convex, with

$$\inf_{y\in\arg\max \varrho_1,\, z\in\arg\max \varrho_2} |z - y| \leq \frac{w}{2},$$

   such that $\varrho(x) = \max\{\varrho_1(x), \varrho_2(x)\}$.

2. The limit $\kappa := \lim_{t\downarrow 0} \frac{\mathrm{vol}_d(K(t))}{\mathrm{diam}(K(t))^d}$ exists, and there are numbers $c, \varepsilon \in (0, 1]$ such that

$$\inf_{t\in(0,\varepsilon)} \frac{\mathrm{vol}_d(K(t))}{\mathrm{diam}(K(t))^d} = c.$$

For $i = 1, 2$, let the level set of ϱ_i be denoted by $K_i(t)$ for $t \in [0, \|\varrho_i\|_\infty)$, and set $K_i(t) = \emptyset$ for $t \geq \|\varrho_i\|_\infty$. Then, since $\varrho = \max\{\varrho_1, \varrho_2\}$, it follows that $K(t) = K_1(t) \cup K_2(t)$. If K is bounded and has positive Lebesgue measure, then the condition 2 is always satisfied. For $K = \mathbb{R}^d$ one has to check the condition 2. For example, $\varrho\colon \mathbb{R}^d \to (0, \infty)$ with

$$\varrho(x) = \max\{\exp(-\alpha\, |x|^2\,), \exp(-\beta\, |x - x_0|^2\,)\},$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

and $2\beta > \alpha$ satisfies the conditions 1 and 2 for $w = 2\,|x_0|$. The rough idea for a transition from x to y of the combination of the different methods on the level set $K(t)$ is as follows. Consider a line/segment of the form

$$L_t(x, \theta) = \{r \in \mathbb{R} \mid x + r\theta \in K(t)\}.$$

Then run the stepping-out and shrinkage procedure on $L_t(x, \theta)$ and return y. We present in detail a single transition from x to y of the hit-and-run, stepping-out, and shrinkage slice sampler in Algorithm 2.

Algorithm 2. *A hybrid slice sampling transition of the hit-and-run, stepping-out, and shrinkage procedure from x to y, i.e. with input x and output y. The stepping-out procedure on $L_t(x, \theta)$ (line of hit-and-run on level set) has inputs x, $w > 0$ (step size parameter from $\mathcal{R}_{d,w}$) and outputs an interval [L, R]. The shrinkage procedure has input [L, R] and output $y = x + s\theta$:*

1. *Choose a level $t \sim \mathcal{U}(0, \varrho(x))$;*
2. *Choose a direction $\theta \in S_{d-1}$ uniformly distributed;*
3. *Stepping-out on $L_t(x, \theta)$ with $w > 0$ outputs an interval [L, R]:*
   (a) *Choose $u \sim \mathcal{U}[0, 1]$. Set $L = uw$ and $R = L + w$;*
   (b) *Repeat until $t \geq \varrho(x + L\theta)$, i.e. $L \notin L_t(x, \theta)$: Set $L = L - w$;*
   (c) *Repeat until $t \geq \varrho(x + R\theta)$, i.e. $R \notin L_t(x, \theta)$: Set $R = R + w$;*
4. *Shrinkage procedure with input [L, R] outputs y:*
   (a) *Set $\bar{L} = L$ and $\bar{R} = R$;*
   (b) *Repeat:*
      i. *Choose $v \sim \mathcal{U}[0, 1]$ and set $s = \bar{L} + v(\bar{R} - \bar{L})$;*
      ii. *If $s \in L_t(x, \theta)$ return $y = x + s\theta$ and exit the loop;*
      iii. *If $s < 0$ then set $\bar{L} = s$, else $\bar{R} = s$.*

Now we present the corresponding transition kernel on $K(t)$. Since $\varrho \in \mathcal{R}_{d,w}$, we can define for $i = 1, 2$ the open intervals

$$L_{t,i}(x, \theta) = \{s \in \mathbb{R} \mid x + s\theta \in K_i(t)\}$$

and have $L_t(x, \theta) = L_{t,1}(x, \theta) \cup L_{t,2}(x, \theta)$. Let

$$\delta_{t,\theta,x} = \inf_{r \in L_{t,1}(x,\theta),\, s \in L_{t,2}(x,\theta)} |r - s|\,,$$

and note that if $\delta_{t,\theta,x} > 0$ then $L_{t,1}(x, \theta) \cap L_{t,2}(x, \theta) = \emptyset$.

We also write for short $|L_t(x, \theta)| = \mathrm{vol}_1(L_t(x, \theta))$, and for $A \in \mathcal{B}(K)$, $x \in K$, $\theta \in S_{d-1}$, let $A_{x,\theta} = \{s \in \mathbb{R} \mid x + s\theta \in A\}$. With this notation, for $t > 0$, the transition kernel H_t on $K(t)$ is given by

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

$$H_t(x, A) = \int_{S_{d-1}} \left[\gamma_t(x, \theta) \frac{\left| L_t(x, \theta) \cap A_{x,\theta} \right|}{|L_t(x, \theta)|} \right.$$

$$\left. + (1 - \gamma_t(x, \theta)) \sum_{i=1}^{2} \mathbf{1}_{K_i(t)}(x) \frac{\left| L_{t,i}(x, \theta) \cap A_{x,\theta} \right|}{\left| L_{t,i}(x, \theta) \right|} \right] \frac{\mathrm{d}\theta}{\sigma_d},$$

with

$$\gamma_t(x, \theta) = \frac{(w - \delta_{t,x,\theta})}{w} \cdot \frac{|L_t(x, \theta)|}{|L_t(x, \theta)| + \delta_{t,x,\theta}}.$$

The following result is helpful.

Lemma 11. *For $\varrho \in \mathcal{R}_{d,w}$ and for any $t \in (0, \|\varrho\|_\infty)$, the following hold:*

1. *The transition kernel H_t is reversible and induces a positive semidefinite operator on $L_{2,t}$, i.e. for $f \in L_{2,t}$, it holds that $\langle H_t f, f \rangle_t \geq 0$.*
2. *We have*

$$\|H_t - U_t\|_{L_{2,t} \to L_{2,t}} \leq 1 - \frac{\mathrm{vol}_d(K(t))}{\sigma_d \, \mathrm{diam}(K(t))^d}; \tag{22}$$

in particular $\lim_{k \to \infty} \beta_k = 0$ with β_k defined in Theorem 1.

Proof. First, note that $L_t(x + s\theta, \theta) = L_t(x, \theta) - s$, $|L_t(x + s\theta, \theta)| = |L_t(x, \theta)|$, and $\gamma_t(x + s\theta, \theta) = \gamma_t(x, \theta)$ for any $x \in \mathbb{R}^d$, $\theta \in S_{d-1}$, and $s \in \mathbb{R}$.

Part 1. The reversibility of H_t with respect to U_t (in the setting of $\varrho \in \mathcal{R}_{d,w}$) is inherited by the reversibility of hit-and-run and the reversibility of the combination of the stepping-out and shrinkage procedure; see Lemma 9.

We turn to the positive semidefiniteness. Let $C_t = \mathrm{vol}_d(K(t))$. We have

$$\langle f, H_t f \rangle_t = \int_{S_{d-1}} \int_{K(t)} \gamma_t(x, \theta) f(x) \int_{L_t(x,\theta)} f(x + r\theta) \frac{\mathrm{d}r}{|L_t(x, \theta)|} \frac{\mathrm{d}x}{C_t} \frac{\mathrm{d}\theta}{\sigma_d}$$

$$+ \sum_{i=1}^{2} \int_{S_{d-1}} \int_{K_i(t)} (1 - \gamma_t(x, \theta)) f(x) \int_{L_{t,i}(x,\theta)} f(x + r\theta) \frac{\mathrm{d}r}{\left| L_{t,i}(x, \theta) \right|} \frac{\mathrm{d}x}{C_t} \frac{\mathrm{d}\theta}{\sigma_d}.$$

We prove positivity of the first summand. The positivity of the other two summands follows by the same arguments. For $\theta \in S_{d-1}$ let us define the projected set

$$P_{\theta^\perp}(K(t)) = \{\tilde{x} \in \mathbb{R}^d \mid \tilde{x} \perp \theta, \ \exists s \in \mathbb{R} \text{ s.t. } \tilde{x} + \theta s \in K(t)\}.$$

Then

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

$$\int_{S_{d-1}}\int_{K(t)}\gamma_t(x,\theta)f(x)\int_{L_t(x,\theta)}f(x+r\theta)\,\frac{\mathrm{d}r}{|L_t(x,\theta)|}\,\frac{\mathrm{d}x}{C_t}\,\frac{\mathrm{d}\theta}{\sigma_d}$$

$$=\int_{S_{d-1}}\int_{P_{\theta^\perp}(K(t))}\int_{L_t(\tilde{x},\theta)}\gamma_t(\tilde{x}+s\theta,\theta)f(\tilde{x}+s\theta)\times$$

$$\int_{L_t(\tilde{x}+s\theta,\theta)}f(\tilde{x}+(r+s)\theta)\,\frac{\mathrm{d}r}{|L_t(\tilde{x}+s\theta,\theta)|}\,\frac{\mathrm{d}s\,\mathrm{d}\tilde{x}}{C_t}\,\frac{\mathrm{d}\theta}{\sigma_d}$$

$$=\int_{S_{d-1}}\int_{P_{\theta^\perp}(K(t))}\int_{L_t(\tilde{x},\theta)}\gamma_t(\tilde{x},\theta)f(\tilde{x}+s\theta)\times$$

$$\int_{L_t(\tilde{x},\theta)-s}f(\tilde{x}+(r+s)\theta)\,\frac{\mathrm{d}r}{|L_t(\tilde{x},\theta)-s|}\,\frac{\mathrm{d}s\,\mathrm{d}\tilde{x}}{C_t}\,\frac{\mathrm{d}\theta}{\sigma_d}$$

$$=\int_{S_{d-1}}\int_{P_{\theta^\perp}(K(t))}\frac{\gamma_t(\tilde{x},\theta)}{|L_t(\tilde{x},\theta)|}\left(\int_{L_t(\tilde{x},\theta)}f(\tilde{x}+u\theta)\mathrm{d}u\right)^2\frac{\mathrm{d}\tilde{x}}{C_t}\,\frac{\mathrm{d}\theta}{\sigma_d}\geq 0.$$

This yields that H_t is positive semidefinite.

Part 2. For any $x\in K(t)$ and measurable $A\subseteq K(t)$ we have

$$H_t(x,A)\geq\int_{S_{d-1}}\gamma_t(x,\theta)\int_{L_t(x,\theta)}\mathbf{1}_A(x+s\theta)\,\frac{\mathrm{d}s}{|L_t(x,\theta)|}\,\frac{\mathrm{d}\theta}{\sigma_d}$$

$$=\int_{S_{d-1}}\int_0^\infty\gamma_t(x,\theta)\mathbf{1}_A(x-s\theta)\,\frac{\mathrm{d}s}{|L_t(x,\theta)|}\,\frac{\mathrm{d}\theta}{\sigma_d}$$

$$+\int_{S_{d-1}}\int_0^\infty\gamma_t(x,\theta)\mathbf{1}_A(x+s\theta)\,\frac{\mathrm{d}s}{|L_t(x,\theta)|}\,\frac{\mathrm{d}\theta}{\sigma_d}$$

$$=\int_{\mathbb{R}^d}\frac{\gamma_t(x,\frac{y}{|y|})}{\sigma_d\cdot\left|L_t(x,\frac{y}{|y|})\right|}\,\frac{\mathbf{1}_A(x-y)}{|y|^{d-1}}\,\mathrm{d}y+\int_{\mathbb{R}^d}\frac{\gamma_t(x,\frac{y}{|y|})}{\sigma_d\cdot\left|L_t(x,\frac{y}{|y|})\right|}\,\frac{\mathbf{1}_A(x+y)}{|y|^{d-1}}\,\mathrm{d}y$$

$$=\frac{2}{\sigma_d}\int_A\frac{\gamma_t(x,\frac{x-y}{|x-y|})}{|x-y|^{d-1}\left|L_t(x,\frac{x-y}{|x-y|})\right|}\,\mathrm{d}y\geq\frac{\mathrm{vol}_d(K(t))}{\sigma_d\,\mathrm{diam}(K(t))^d}\cdot\frac{\mathrm{vol}_d(A)}{\mathrm{vol}_d(K(t))}.$$

Here the last inequality follows from the fact that $\delta_{t,x,\theta}\leq w/2$ and $|L_t(x,\theta)|+\delta_{t,x,\theta}\leq\mathrm{diam}(K(t))$. Thus, by [18] we have uniform ergodicity and by [33, Proposition 3.24] we obtain (22). Finally, $\lim_{k\to\infty}\beta_k=0$ follows by the same arguments as in Lemma 10. □

By Theorem 1, this observation leads to the following result.

Corollary 4. *Let $\varrho\in\mathcal{R}_{d,w}$. Then the hit-and-run, stepping-out, shrinkage slice sampler has an absolute spectral gap if and only if the simple slice sampler has an absolute spectral gap.*

5. Concluding remarks

We provide a general framework to prove convergence results for hybrid slice sampling via spectral gap arguments. More precisely, we state sufficient conditions for the spectral gap of an

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

appropriately designed hybrid slice sampler to be equivalent to the spectral gap of the simple slice sampler. Since all Markov chains we are considering are reversible, this also provides a criterion for geometric ergodicity; see [28].

To illustrate how our analysis can be applied to specific hybrid slice sampling implementations, we analyze the hit-and-run algorithm on the slice on multidimensional targets under weak conditions, as well as the easily implementable stepping-out shrinkage hit-and-run on the slice for bimodal d-dimensional distributions. The latter analysis can in principle be extended to settings with more than two modes, at the price of further notational and computational complexity.

These examples demonstrate that the robustness of the simple slice sampler is inherited by appropriately designed hybrid versions of it in realistic computational settings, providing theoretical underpinnings for their use in applications.

Appendix A. Technical lemmas

Lemma 12. *Let H_1 and H_2 be two Hilbert spaces. Furthermore, let $R\colon H_2 \to H_1$ be a bounded linear operator with adjoint $R^*\colon H_1 \to H_2$, and let $Q\colon H_2 \to H_2$ be a bounded linear operator which is self-adjoint. Then*

$$\left\| RQ^{k+1}R^* \right\|_{H_1\to H_1} \le \|Q\|_{H_2\to H_2} \left\| R\,|Q|^k\,R^* \right\|_{H_1\to H_1}.$$

Let us additionally assume that Q is positive semidefinite. Then

$$\left\| RQ^{k+1}R^* \right\|_{H_1\to H_1} \le \|Q\|_{H_2\to H_2} \left\| RQ^kR^* \right\|_{H_1\to H_1}.$$

Proof. Let us denote the inner products of H_1 and H_2 by $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$, respectively. By the spectral theorem for the bounded and self-adjoint operator $Q\colon H_2 \to H_2$ we obtain

$$\frac{\langle QR^*f, R^*f\rangle_2}{\langle R^*f, R^*f\rangle_2} = \int_{\mathrm{spec}(Q)} \lambda \,\mathrm{d}\nu_{Q,R^*f}(\lambda),$$

where $\mathrm{spec}(Q)$ denotes the spectrum of Q and ν_{Q,R^*f} denotes the normalized spectral measure. Thus,

$$\begin{aligned}
\left\| RQ^{k+1}R^* \right\|_{H_1\to H_1} &= \sup_{\langle f,f\rangle_1\neq 0} \frac{\left|\langle Q^{k+1}R^*f, R^*f\rangle_2\right|}{\langle f,f\rangle_1} \\
&= \sup_{\langle f,f\rangle_1\neq 0} \frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1} \frac{\left|\langle Q^{k+1}R^*f, R^*f\rangle_2\right|}{\langle R^*f, R^*f\rangle_2} \\
&= \sup_{\langle f,f\rangle_1\neq 0} \frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1} \left| \int_{\mathrm{spec}(Q)} \lambda^{k+1}\,\mathrm{d}\nu_{Q,R^*f}(\lambda) \right| \\
&\le \|Q\|_{H_2\to H_2} \sup_{\langle f,f\rangle_1\neq 0} \frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1} \int_{\mathrm{spec}(Q)} |\lambda|^k\,\mathrm{d}\nu_{Q,R^*f}(\lambda) \\
&= \|Q\|_{H_2\to H_2} \sup_{\langle f,f\rangle_1\neq 0} \frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1} \frac{\langle |Q|^k R^*f, R^*f\rangle_2}{\langle R^*f, R^*f\rangle_2} \\
&= \|Q\|_{H_2\to H_2} \left\| R\,|Q|^k\,R^* \right\|_{H_1\to H_1}.
\end{aligned}$$

https://doi.org/10.1017/apr.2024.16 Published online by Cambridge University Press

Here we used that the operator norm of $Q\colon H_2 \to H_2$ is the same as the operator norm of $|Q| : H_2 \to H_2$. If Q is positive semidefinite, then $Q = |Q|$. □

Lemma 13. *Let us assume that the conditions of Lemme 12 are satisfied. Furthermore, let* $\|R\|^2_{H_2\to H_1} = \|RR^*\|_{H_1\to H_1} \le 1$. *Then*

$$\|RQR^*\|^k_{H_1\to H_1} \le \left\|R\,|Q|^k\,R^*\right\|_{H_1\to H_1}.$$

Let us additionally assume that Q is positive semidefinite. Then

$$\|RQR^*\|^k_{H_1\to H_1} \le \left\|RQ^kR^*\right\|_{H_1\to H_1}.$$

Proof. We use the same notation as in the proof of Lemma 12. Thus

$$\begin{aligned}
\|RQR^*\|^k_{H_1\to H_1} &= \sup_{\langle f,f\rangle_1\neq 0}\left(\frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1}\frac{|\langle QR^*f, R^*f\rangle_2|}{\langle R^*f, R^*f\rangle_2}\right)^k \\
&= \sup_{\langle f,f\rangle_1\neq 0}\left(\frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1}\right)^k\left|\int_{\mathrm{spec}(Q)}\lambda\,\mathrm{d}\nu_{Q,R^*f}(\lambda)\right|^k \\
&\le \sup_{\langle f,f\rangle_1\neq 0}\left(\frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1}\right)^k\int_{\mathrm{spec}(Q)}|\lambda|^k\,\mathrm{d}\nu_{Q,R^*f}(\lambda) \\
&= \sup_{\langle f,f\rangle_1\neq 0}\left(\frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1}\right)^k\frac{\langle |Q|^k\,R^*f, R^*f\rangle_2}{\langle R^*f, R^*f\rangle_2} \\
&= \sup_{\langle f,f\rangle_1\neq 0}\left(\frac{\langle R^*f, R^*f\rangle_2}{\langle f,f\rangle_1}\right)^{k-1}\frac{\langle |Q|^k\,R^*f, R^*f\rangle_2}{\langle f,f\rangle_1} \\
&\le \|RR^*\|^{k-1}_{H_1\to H_1}\left\|R\,|Q|^k\,R^*\right\|_{H_1\to H_1} \le \left\|R\,|Q|^k\,R^*\right\|_{H_1\to H_1}.
\end{aligned}$$

Note that we applied Jensen's inequality here. Furthermore, if Q is positive semidefinite, then $Q = |Q|$, which finishes the proof. □

Acknowledgements

We are extremely grateful for the referees' careful reading and comments on the paper. In particular, we thank one of the referees, who pointed us to Example 1.

Funding information

K. L. was supported by the Royal Society through the Royal Society University Research Fellowship. D. R. gratefully acknowledges support from the DFG, within project 522337282.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BAXENDALE, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15**, 700–738.

[2] BEDNORZ, W. AND ŁATUSZYŃSKI, K. (2007). A few remarks on 'Fixed-width output analysis for Markov chain Monte Carlo' by Jones *et al. J. Amer. Statist. Assoc.* 102, 1485–1486.

[3] BÉLISLE, C., ROMEIJN, H. AND SMITH, R. (1993). Hit-and-run algorithms for generating multivariate distributions. *Math. Operat. Res.* **18**, 255–266.

[4] BESAG, J. AND GREEN, P. (1993). Spatial statistics and Bayesian computation. *J. R. Statist. Soc. B [Statist. Methodology]* **55**, 25–37.

[5] DIACONIS, P. AND FREEDMAN, D. (1997). On the hit and run process. Tech. rep. 497, Department of Statistics, University of California, Berkeley, 16 pp.

[6] EDWARDS, R. AND SOKAL, A. (1988). Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *Phys. Rev. D* 38, article no. 2009.

[7] FLEGAL, J. AND JONES, G. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38**, 1034–1070.

[8] GEYER, C. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7**, 473–483.

[9] HIGDON, D. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Statist. Assoc.* **93**, 585–595.

[10] HOBERT, J., JONES, G., PRESNELL, B. AND ROSENTHAL, J. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika* **89**, 731–743.

[11] JONES, G., HARAN, M., CAFFO, B. AND NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101**, 1537–1547.

[12] KAUFMAN, D. AND SMITH, R. (1998). Direction choice for accelerated convergence in hit-and-run sampling. *Operat. Res.* **46**, 84–95.

[13] KIATSUPAIBUL, S., SMITH, R. AND ZABINSKY, Z. (2011). An analysis of a variation of hit-and-run for uniform sampling from general regions. *ACM Trans. Model. Comput. Simul.* **21**, 16:1–16:11.

[14] KIPNIS, C. AND S, V. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1–19.

[15] KONTOYIANNIS, I. AND MEYN, S. (2012). Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Prob. Theory Relat. Fields* **154**, 327–339.

[16] LOVÁSZ, L. (1999). Hit-and-run mixes fast. *Math. Program.* **86**, 443–461.

[17] LOVÁSZ, L. AND VEMPALA, S. (2006). Hit-and-run from a corner. *SIAM J. Comput.* **35**, 985–1005.

[18] MEYN, S. AND TWEEDIE, R. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.

[19] MIRA, A. (2001). Ordering and improving the performance of Monte Carlo Markov chains. *Statist. Sci.* **16**, 340–350.

[20] MIRA, A., MØLLER, J. AND ROBERTS, G. (2001). Perfect slice samplers. *J. R. Statist. Soc. B [Statist. Methodology]* **63**, 593–606.

[21] MIRA, A. AND TIERNEY, L. (2002). Efficiency and convergence properties of slice samplers. *Scand. J. Statist.* **29**, 1–12.

[22] MURRAY, I., ADAMS, R. AND MACKAY, D. (2010). Elliptical slice sampling. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, Journal of Machine Learning Research, pp. 541–548.

[23] NATAROVSKII, V., RUDOLF, D. AND SPRUNGK, B. (2021). Geometric convergence of elliptical slice sampling. In *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research 139, pp. 7969–7978.

[24] NATAROVSKII, V., RUDOLF, D. AND SPRUNGK, B. (2021). Quantitative spectral gap estimate and Wasserstein contraction of simple slice sampling. *Ann. Appl. Prob.* **31**, 806–825.

[25] NEAL, R. (1993). *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto.

[26] NEAL, R. (2003). Slice sampling. *Ann. Statist.* **31**, 705–767.

[27] NOVAK, E. AND RUDOLF, D. (2014). Computation of expectations by Markov chain Monte Carlo methods. In *Extraction of Quantifiable Information from Complex Systems*, Springer, Cham, pp. 397–411.

[28] ROBERTS, G. AND ROSENTHAL, J. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Prob.* **2**, 13–25.

[29] ROBERTS, G. AND ROSENTHAL, J. (1999). Convergence of slice sampler Markov chains. *J. R. Statist. Soc. B [Statist. Methodology]* **61**, 643–660.

[30] ROBERTS, G. AND ROSENTHAL, J. (2002). The polar slice sampler. *Stoch. Models* **18**, 257–280.

[31] ROBERTS, G. AND ROSENTHAL, J. (2008). Variance bounding Markov chains. *Ann. Appl. Prob.* **18**, 1201–1214.
[32] RUDIN, W. (1991). *Functional Analysis*, 2nd edn. McGraw-Hill, New York.
[33] RUDOLF, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math.* **485**, 93 pp.
[34] RUDOLF, D. (2013). Hit-and-run for numerical integration. In *Monte Carlo and Quasi-Monte Carlo Methods 2012,* eds J. Dick, F. Kuo, G. Peters, and I. Sloan, Springer, Berlin, pp. 565–579.
[35] RUDOLF, D. AND ULLRICH, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Prob.* **18**, 1–8.
[36] RUDOLF, D. AND ULLRICH, M. (2018). Comparison of hit-and-run, slice sampler and random walk Metropolis. *J. Appl. Prob.* **55**, 1186–1202.
[37] SCHÄR, P., HABECK, M. AND RUDOLF, D. (2023). Gibbsian polar slice sampling. In *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research 202, pp. 30204–30223.
[38] SMITH, R. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operat. Res.* **32**, 1296–1308.
[39] ULLRICH, M. (2012). *Rapid mixing of Swendsen–Wang and single-bond dynamics in two dimensions*. Preprint. Available at https://arxiv.org/abs/1202.6321.
[40] ULLRICH, M. (2014). Rapid mixing of Swendsen–Wang dynamics in two dimensions. *Dissertationes Math.* **502**, 64 pp.
[41] WANG, J. AND SWENDSEN, R. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* 58, article no. 86.
[42] WERNER, D. (2005). *Funktionalanalysis*. Springer, Berlin.