Wasserstein contraction and spectral gap of slice sampling revisited

Philip Schär*

Abstract

We propose a new class of Markov chain Monte Carlo methods, called k-*polar slice sampling* (k-PSS), as a technical tool that interpolates between and extrapolates beyond uniform and polar slice sampling. By examining Wasserstein contraction rates and spectral gaps of k-PSS, we obtain strong quantitative results regarding its performance for different kinds of target distributions. Because k-PSS contains uniform and polar slice sampling as special cases, our results significantly advance the theoretical understanding of both of these methods. In particular, we prove realistic estimates of the convergence rates of uniform slice sampling for arbitrary multivariate Gaussian distributions on the one hand, and near-arbitrary multivariate t-distributions on the other. Furthermore, our results suggest that for heavy-tailed distributions, polar slice sampling performs dimension-independently well, whereas uniform slice sampling suffers a rather strong curse of dimensionality.

1 Introduction

Sampling from intractable distributions is a ubiquitous problem in all fields that rely on probabilistic modeling. One of the standard approaches to solve this problem is Markov chain Monte Carlo (MCMC), which works by constructing a Markov chain whose iterate distribution converges to the target distribution as the chain progresses. Samples from such a chain are then used as approximate samples from the target. Although there are many heuristics for assessing the quality of such samples, there is no general way to definitively determine how well they are suited for estimating an unknown target quantity. It is therefore of vital importance to gain a good understanding of the theoretical properties of MCMC methods, as these provide more reliable, general insights than heuristics and numerical results.

---

*Friedrich Schiller University Jena, Germany. E-mail: philip.schaer@uni-jena.de

In this paper, we focus on a class of MCMC methods called *slice sampling*, which originated in physics research and was popularized in statistics in [1]. We view slice sampling as being defined for distributions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(G)$ denotes the Borel-sigma algebra of a set G, though we note that recent advances have also extended the concept to the d-sphere [4]. More specifically, the setting we assume is the following. Let ν be a probability measure (in the following often just called *distribution*) on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Suppose that ν has a potentially unnormalized Lebesgue density, that is, a measurable function $\eta : \mathbb{R}^d \to \mathbb{R}_+ := [0, \infty[$ such that

$$\nu(A) = \frac{\int_A \eta(x)\mathrm{d}x}{\int_{\mathbb{R}^d} \eta(x)\mathrm{d}x}, \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

We assume that we can evaluate η at any given point $x \in \mathbb{R}^d$ but that its normalization constant (the denominator in the above expression) is unknown. Slice sampling for ν further assumes η to be factorized into measurable functions $\eta_i : \mathbb{R}^d \to \mathbb{R}_+$, $i = 0, 1$, such that

$$\eta(x) = \eta_0(x)\eta_1(x), \qquad x \in \mathbb{R}^d. \tag{1.1}$$

Given an initial state $x_0 \in \mathbb{R}^d$, slice sampling for ν then generates a realization $(x_n)_{n \in \mathbb{N}_0}$ of a Markov chain $(X_n)_{n \in \mathbb{N}_0}$ with invariant distribution ν by setting $X_0 := x_0$ and using the following steps to perform the transition from $X_{n-1} = x_{n-1}$ to $X_n = x_n$:

1. Sample $T_n \sim \mathcal{U}(\,]0, \eta_1(x_{n-1})[\,)$, call the result t_n.

2. Sample $X_n \sim \mu_{t_n}$, where μ_{t_n} is defined by[1].

$$\mu_t(A) := \frac{\int_A \eta_0(x)\mathbb{I}(\eta_1(x) > t)\mathrm{d}x}{\int_{\mathbb{R}^d} \eta_0(x)\mathbb{I}(\eta_1(x) > t)\mathrm{d}x}, \qquad A \in \mathcal{B}(\mathbb{R}^d), t > 0,$$

call the result x_n.

Following [17], we refer to step 2 as the *X-update* of the slice sampling transition. Intuitively, the X-update samples the new state from the distribution whose (unnormalized) density is the first factor η_0 restricted to the super-level set (or *slice*) w.r.t. the second factor η_1 at the current threshold t_n.

The above definition is frequently restricted by specifying a mapping that determines for each target density η the factorization (1.1), thereby defining what we call a *slice sampling variant*. In our view, only three such variants are currently practically relevant. First there is *uniform slice sampling* (USS), which uses the factorization

$$\eta_0(x) := 1, \qquad \eta_1(x) := \eta(x).$$

Then there is *polar slice sampling* (PSS), proposed in [13], which uses

$$\eta_0(x) := \|x\|^{1-d}, \qquad \eta_1(x) := \|x\|^{d-1}\eta(x)$$

for $x \neq 0$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$. Finally, there is *Gaussian slice sampling* (GSS), which sets η_0 to the density of a multivariate Gaussian and η_1 to η divided by it.

As all three variants lead to an X-update that is typically infeasible to implement, in practice one uses methods that mimic these "ideal" formulations while enabling computationally efficient implementation. For USS, one can use the stepping-out and

---

[1]We denote by $\mathbb{I}(cond)$ an indicator that takes the value 1 whenever the condition $cond$ is satisfied and the value 0 otherwise.

shrinkage procedures proposed in [9], for PSS the Gibbsian polar slice sampling framework recently developed in [17], and for GSS the elliptical slice sampling framework proposed in [7]. Of course theoretical results for the ideal methods do not readily yield results for their efficient approximations. However, for USS it has been shown in [6] that some theoretical results, even quantitative ones, can be explicitly transferred from the ideal to the efficient version. We are cautiously optimistic that analogous theories can also be developed for PSS and GSS, making it all the more worthwhile to investigate the theoretical properties of the ideal methods.

Though we are of course also interested in GSS, we focus our efforts in this paper exclusively on the other two methods, USS and PSS. Moreover, we observe that these two can often be analyzed simultaneously. To this end, we propose a new class of slice sampling variants, which we term k-*polar slice sampling* (k-PSS). For any fixed $k \in \mathbb{R}_{>0} := \,]0, \infty[$, we define k-PSS as slice sampling with factorization

$$\eta_0(x) := \eta_{k,0}(x) := \|x\|^{k-d},$$
$$\eta_1(x) := \eta_{k,1}(x) := \|x\|^{d-k}\eta(x)$$

for $x \neq 0$. Clearly, k-PSS coincides with PSS for $k = 1$ and with USS for $k = d$. For $1 \leq k \leq d$, it continuously interpolates between USS and PSS, and for $k < 1$ and $k > d$ it extrapolates beyond them in both directions. As we will see in the following, the quantitative theoretical properties of k-PSS often also interpolate between those of USS and PSS. However, we emphasize right away that we do not deem the k-PSS methods apart from USS and PSS to be useful in actual practical use as samplers. Rather, we view them as technical tools that we can use to prove theoretical results, and, as we shall see later on (in Theorem 3.3), their use in this regard goes far beyond unifying the analysis of USS and PSS.

Let us now briefly mention some related work. It is well-known that slice sampling is geometrically ergodic (essentially meaning that the total variation distance between iterate distribution and target distribution converges geometrically to zero) under extremely weak assumptions [12]. However, this is a purely qualitative result, as the convergence can in practice still be arbitrarily slow, and – unsurprisingly – the known quantitative results are much less extensive. Some quantitative convergence guarantees were already established in [12], but only under fairly restrictive conditions, including on the initial value, and the guarantees given there were, in retrospect, far from sharp. Realistic quantitative results on slice sampling are a comparatively recent achievement. In [8], the authors examine Wasserstein contraction rates and spectral gaps (see below) of USS and in [15] some of their methodology is extended to general slice sampling and a dimension-independent lower bound on the spectral gap of PSS, for rotationally invariant targets that are log-concave along rays emanating from the origin, is proven.

Following the example of [8] and [15], we also focus on Wasserstein contraction rates and spectral gaps as quantities of interest in order to arrive at realistic quantitative results. We now briefly introduce and motivate both of these quantities.

Let $(G, \mathcal{G})$ be a measurable space and let P be a Markov kernel on $G \times \mathcal{G}$ with invariant distribution $\nu \in \mathcal{P}(G)$, where $\mathcal{P}(G)$ denotes the set of probability measures on $(G, \mathcal{G})$. Define the L_2-space belonging to ν as

$$L_2(\nu) := \{g : G \to \mathbb{R} \,:\, \|g\|_{2,\nu} < \infty\},$$

where

$$\|g\|_{2,\nu} := \left(\int_G g(x)^2 \nu(\mathrm{d}x) \right)^{1/2}.$$

Observe that P acting on measurable functions $g : G \to \mathbb{R}$ via $g \mapsto Pg$ with

$$Pg(x) := \int_G g(y)P(x, \mathrm{d}y), \qquad x \in G$$

defines a linear operator mapping $L_2(\nu) \to L_2(\nu)$. Note further that by interpreting ν as a Markov kernel that is constant in its first argument, ν also defines a linear operator $L_2(\nu) \to L_2(\nu)$. This framing allows us to define the *spectral gap* of P as

$$\mathsf{gap}_\nu(P) := 1 - \|P - \nu\|_{L_2(\nu) \to L_2(\nu)},$$

where $\|\cdot\|_{L_2(\nu) \to L_2(\nu)}$ is the operator norm w.r.t. $\|\cdot\|_{2,\nu}$.

The result regarding spectral gaps that one is typically interested in is a lower bound, i.e. a statement of the form $\mathsf{gap}_\nu(P) \geq \delta$ for some $\delta > 0$ (in the following simply called *spectral gap estimate*). The reason for this is that there are various useful implications of such a bound. Suppose $\mathsf{gap}_\nu(P) \geq 1 - \rho$ for some $\rho \in\ ]0, 1[$, then it is well-known, see e.g. [10, Lemma 2], that for any initial distribution $\xi \in \mathcal{P}(G)$ with $\xi \ll \nu$ one has

$$d_{\mathrm{tv}}(\xi P^n, \nu) \leq \rho^n \left\| \frac{\mathrm{d}\xi}{\mathrm{d}\nu} - 1 \right\|_{2,\nu}, \qquad n \in \mathbb{N}_0, \tag{1.2}$$

where $\frac{\mathrm{d}\xi}{\mathrm{d}\nu}$ denotes the Radon-Nikodym derivative of ξ w.r.t. ν and d_{tv} the total variation distance, which for $\xi_1, \xi_2 \in \mathcal{P}(G)$ is defined as

$$d_{\mathrm{tv}}(\xi_1, \xi_2) := \sup_{A \in \mathcal{G}} |\xi_1(A) - \xi_2(A)|.$$

Note that ξP^n is the distribution of X_n when $(X_n)_{n \in \mathbb{N}_0}$ is a Markov chain with transition kernel P and initial distribution $X_0 \sim \xi$, see [2, Theorem 1.3.4]. Aside from this geometric convergence result, consequences of a spectral gap estimate include a quantitative bound on the error of Monte Carlo integration [14, Theorem 3.41], a central limit theorem [5] and an estimate of the central limit theorem's asymptotic variance [3].

The second kind of quantity we examine is the Wasserstein contraction rate. To express it, denote by $\mathcal{W}$ the *Wasserstein 1-distance*, which for $\xi_1, \xi_2 \in \mathcal{P}(\mathbb{R}^d)$ is defined as

$$\mathcal{W}(\xi_1, \xi_2) := \inf_{\gamma \in \Gamma(\xi_1, \xi_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\| \gamma(\mathrm{d}x, \mathrm{d}y),$$

where $\Gamma(\xi_1, \xi_2)$ denotes the set of couplings of ξ_1 and ξ_2,

$$\Gamma(\xi_1, \xi_2) := \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \mid \forall A \in \mathcal{B}(\mathbb{R}^d) :\ \gamma(A \times \mathbb{R}^d) = \xi_1(A),\ \gamma(\mathbb{R}^d \times A) = \xi_2(A)\}.$$

Suppose now that $G \subseteq \mathbb{R}^d$ is a non-empty, closed set and let P be a Markov kernel on $G \times \mathcal{B}(G)$. Then the *Dobrushin coefficient* $\mathsf{Dob}(P)$ of P is given by

$$\mathsf{Dob}(P) = \sup_{x,y \in G,\ x \neq y} \frac{\mathcal{W}(P(x, \cdot), P(y, \cdot))}{\|x - y\|},$$

see [2, Definition 20.3.1, Lemma 20.3.2]. Given this definition, P is called *Wasserstein contractive* if $\mathsf{Dob}(P) < 1$. Moreover, it is called *Wasserstein contractive with rate ρ* if $\mathsf{Dob}(P) \leq \rho < 1$.

There are two main reasons why Wasserstein contraction is an extremely desirable property. The first is that it implies a very intuitive convergence result w.r.t. the Wasserstein distance: Suppose that P is Wasserstein contractive with rate ρ and has $\nu \in \mathcal{P}(G)$ as an invariant distribution, then for any initial distribution $\xi \in \mathcal{P}(G)$ one has

$$\mathcal{W}(\xi P^n, \nu) \leq \rho^n \mathcal{W}(\xi, \nu), \qquad n \in \mathbb{N}_0,$$

see [2, Theorem 20.3.4]. In other words, one has a bound for the error expressed as the Wasserstein distance between iterate and target distribution, and that error bound decreases by the factor ρ with each step of the Markov chain, with its coefficient explicitly given by $\mathcal{W}(\xi, \nu)$, the Wasserstein distance between initial and target distribution.

The second reason why Wasserstein contraction is particularly desirable is that it is a strictly more powerful result than a spectral gap estimate. That is, if P is reversible w.r.t. ν, and if additionally ν has finite second moment, meaning $\int_{\mathbb{R}^d} \|x\|^2 \nu(\mathrm{d}x) < \infty$, then one has

$$\mathsf{Dob}(P) \leq \rho \qquad \Rightarrow \qquad \mathsf{gap}_\nu(P) \geq 1 - \rho, \tag{1.3}$$

for any $\rho < 1$, see [11, Proposition 30]. Hence all the implications of a spectral gap estimate that we mentioned above are also implications of a kernel being Wasserstein contractive.

Given that Wasserstein contraction is a strictly more powerful result than a spectral gap, it is unsurprising that the former is also much harder to establish than the latter. Accordingly, we begin our investigation by proving Wasserstein contraction of slice sampling wherever feasible. First, we generalize the contraction result of [8] from USS to k-PSS, proving that k-PSS is Wasserstein contractive for all target densities η for which the factors $\eta_{k,1}$ are rotationally invariant and log-concave. Afterwards, we investigate the Wasserstein contraction rates of slice sampling for heavy-tailed distributions. Here we end up computing precise rates for, on the one hand, USS applied to standard multivariate t-distributions and, on the other hand, k-PSS applied to a class of distributions that mimics the aforementioned one, while being much simpler to analyze.

We then augment our contraction results by proving spectral gap estimates for broader classes of distributions, even some that are rotationally asymmetric (unlike any of those for which contraction results on slice sampling are available). To enable more elegant and often also further-reaching results, we begin this part of the paper by improving a theorem that can provide realistic estimates of the spectral gap of slice sampling. We note here that the best previously known version of this theorem was proven in [15], whereas the original theorem, applying just to USS, was developed in [8]. Given this improved theorem, we first consider k-PSS for classes of rotationally invariant target densities η with factors $\eta_{k,1}$ of varying degrees of log-concavity, even reaching into log-convexity. We then examine k-PSS for a class of rotationally asymmetric targets, which allows us in particular to prove a realistic spectral gap estimate for USS applied to arbitrary multivariate Gaussian distributions. Finally, we consider USS in isolation once more to augment the aforementioned contraction result for standard multivariate t-distributions by a spectral gap estimate for near-arbitrary multivariate t-distributions.

We take this opportunity to emphasize that our interest in the quantitative properties of slice sampling for standard distributions (such as Gaussians or multivariate t-distributions) does not reflect a desire to actually apply slice sampling to such targets in practice. Even if strong guarantees regarding the performance of slice sampling are available, standard methods to directly generate samples from these tractable targets (i.e. ways to suitably transform samples from the uniform distribution) would still be substantially more efficient. Rather, we view the standard distributions we often consider in this paper as proxies for intractable target distributions that share some of the standard distributions' properties (e.g. their tail behavior). We then expect our results on the performance of slice sampling for standard targets to be indicative, to some extent, of their performance for these somehow similar intractable targets. Of course this then further needs to carry over to the performance of the efficiently implementable approximations of the slice samplers we analyze here, see our remarks on this above.

The paper is structured as follows. Section 2 details our investigation of the Wasserstein contraction rates of slice sampling for different classes of target distributions.

Section 3 establishes a theorem that allows estimating the spectral gap of slice sampling, and then discusses several applications of said theorem. We conclude each subsection with a detailed discussion of our results. Some auxiliary results are stated and (if necessary) proven in Appendix A. Finally, Appendix B gives some numerical results that strongly suggest the sharpness of one of our spectral gap results, for which we were unable to establish this sharpness theoretically.

Throughout the paper, we use a few non-notational conventions: When talking about densities, we mean densities with respect to the Lebesgue measure, unless stated otherwise. We do not require densities to be normalized[2] and will instead explicitly emphasize whenever a density is required or constructed to be normalized. Finally, it should be noted that we view a probability distribution and its density as a fixed unit, ignoring the ambiguous normalization and the fact that non-negative functions which coincide with the density almost everywhere are also densities for the same distribution. Hence, whenever we introduce a distribution with a specific density and talk about a sampler for this distribution, we actually mean the sampler for the density we specified.

2 Wasserstein contraction results

2.1 Targets with log-concave factors

Our first result establishes, for any $k \in \mathbb{R}_{>0}$, Wasserstein contraction of k-PSS for a class of distributions $\mathcal{D}_k(\mathbb{R}^d)$ that is defined as follows. By this we generalize a result of [8] that proved Wasserstein contraction of USS for all rotationally invariant, log-concave target densities.

Definition 2.1. *For any $k \in \mathbb{R}_{>0}$ we define a class $\mathcal{D}_k(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$ of distributions by the constraint that each $\nu \in \mathcal{D}_k(\mathbb{R}^d)$ must have a density η of the form*

$$\eta(x) = \|x\|^{k-d} \exp(-\phi(\|x\|)) \mathbb{1}_{]0,\kappa[}(\|x\|), \tag{2.1}$$

where $\kappa \in]0,\infty]$ and $\phi :]0,\kappa[\to \mathbb{R}$ is strictly increasing and convex.

We state our Wasserstein contraction result and prove it using auxiliary results that we provide in Appendix A.

Theorem 2.2. *For any $k > 0$ and $\nu \in \mathcal{D}_k(\mathbb{R}^d)$, k-PSS for ν is Wasserstein contractive with rate $\rho = k/(k+1)$.*

Proof. For the entirety of the proof, arbitrarily fix $k > 0$ and $\nu \in \mathcal{D}_k(\mathbb{R}^d)$ with density η as in (2.1). To estimate the Wasserstein distance between two instances of k-PSS currently positioned at different points $x, y \in \mathbb{R}^d$, we construct a coupling between the respective transition kernels. To this end, we begin by analyzing how the slice sampling procedure simplifies for k-PSS on $\mathcal{D}_k(\mathbb{R}^d)$. First of all, note that in this setting we have

$$\eta_{k,1}(x) = \exp(-\phi(\|x\|)) \mathbb{1}_{]0,\kappa[}(\|x\|). \tag{2.2}$$

Now observe that[3]

$$\sup \eta_{k,1} := \sup_{0 \neq x \in \mathbb{R}^d} \eta_{k,1}(x) = \exp(-\inf \phi).$$

Hence we only need to consider values smaller than $\sup \eta_{k,1}$ for the variable t standing in for the auxiliary variable T_n in the following. Next, observe that

$$\inf \eta_{k,1} := \inf_{x \in \mathbb{R}^d,\, \|x\| \in]0,\kappa[} \eta_{k,1}(x) = \begin{cases} \exp(-\sup \phi) & \kappa < \infty, \\ 0 & \kappa = \infty. \end{cases}$$

---

[2] In our view, this admittedly somewhat unconventional nomenclature is justified by our using the term *density*, rather than *probability density function*, as only the latter explicitly implies normalization.

[3] See Lemma A.3 for notation.

We now characterize the slices of k-PSS for ν. For $t \in\;]0, \inf \eta_{k,1}]$ and $x \in \mathbb{R}^d$, one has

$$\eta_{k,1}(x) > t \quad \Leftrightarrow \quad \|x\| \in\;]0, \kappa[.$$

For $t \in\;]\inf \eta_{k,1}, \sup \eta_{k,1}[$ and $0 \neq x \in \mathbb{R}^d$, we get[4]

$$
\begin{aligned}
\eta_{k,1}(x) > t \quad &\Leftrightarrow \quad \exp(-\phi(\|x\|)) > t \quad \text{and} \quad \|x\| < \kappa \\
&\Leftrightarrow \quad \phi(\|x\|) < -\log t \quad \text{and} \quad \|x\| < \kappa \\
&\Leftrightarrow \quad \|x\| < \phi^{-1}(-\log t) \quad \text{and} \quad \|x\| < \kappa \\
&\Leftrightarrow \quad \|x\| < \phi^{-1}(-\log t),
\end{aligned}
$$

where the third equivalence relies on ϕ^{-1} being strictly increasing and the fourth on it mapping to $D_\phi =\;]0, \kappa[$, s.t. in particular $\phi^{-1} < \kappa$. Using the notation introduced in Lemma A.4, we can summarize these two cases as

$$\eta_{k,1}(x) > t \quad \Leftrightarrow \quad \|x\| \in\;]0, \widehat{\phi}^{-1}(-\log t)[\tag{2.3}$$

for any $t \in\;]0, \sup \eta_{k,1}[$. With this we can write

$$\eta_{k,0}(x)\mathbb{I}(\eta_{k,1}(x) > t) = \|x\|^{k-d}\mathbb{1}_{]0,\widehat{\phi}^{-1}(-\log t)[}(\|x\|) = h_{1,t}(\|x\|) \tag{2.4}$$

for $x \neq 0$, where we define $h_{1,t} : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$h_{1,t}(r) := r^{k-d}\mathbb{1}_{]0,\widehat{\phi}^{-1}(-\log t)[}(r).$$

Denote by λ_d the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, by $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$ the $(d-1)$-sphere and by σ_d the surface measure on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$, that is, the unique rotationally invariant measure whose total mass $\omega_d := \sigma_d(\mathbb{S}^{d-1})$ equals the surface area of $\mathbb{S}^{d-1}$.

Then, by (2.4) and Corollary A.2, the X-update of k-PSS, given $T_n = t$, can be implemented in polar coordinates as $X_n := R_n \Theta_n$ by sampling (R_n, Θ_n) from the joint distribution with density

$$(r, \theta) \mapsto h_{2,t}(r) := r^{d-1}h_{1,t}(r)\mathbb{1}_{\mathbb{R}_+}(r) = r^{k-1}\mathbb{1}_{]0,\widehat{\phi}^{-1}(-\log t)[}(r)$$

w.r.t. $\lambda_1 \otimes \sigma_d$. As this density is constant in θ, it can in turn be implemented by independently sampling Θ_n from $\mathcal{U}(\mathbb{S}^{d-1})$ and R_n from the distribution with density $h_{2,t}$. To couple the sampling of R_n, we apply the inversion method (also called *inverse transform sampling*). For that we first need to compute the c.d.f. F_t corresponding to density $h_{2,t}$. Let $s \in [0, \widehat{\phi}^{-1}(-\log t)]$, then

$$F_t(s) = \frac{\int_{-\infty}^s h_{2,t}(r)\mathrm{d}r}{\int_{-\infty}^\infty h_{2,t}(r)\mathrm{d}r} = \frac{\int_0^s r^{k-1}\mathrm{d}r}{\int_0^{\widehat{\phi}^{-1}(-\log t)} r^{k-1}\mathrm{d}r} = \frac{\left[\frac{1}{k}r^k\right]_0^s}{\left[\frac{1}{k}r^k\right]_0^{\widehat{\phi}^{-1}(-\log t)}} = \left(\frac{s}{\widehat{\phi}^{-1}(-\log t)}\right)^k.$$

From this it is easy to see that F_t restricted to $[0, \widehat{\phi}^{-1}(-\log t)]$ maps bijectively onto $[0, 1]$ and that the inverse $F_t^{-1} : [0, 1] \to [0, \widehat{\phi}^{-1}(-\log t)]$ of this restriction is given by

$$F_t^{-1}(u) = \widehat{\phi}^{-1}(-\log t)u^{1/k}, \qquad u \in [0, 1].$$

Denote by P the transition kernel of k-PSS for ν and fix $x, y \in \mathbb{R}^d$ with $\|x\|, \|y\| \in\;]0, \kappa[$. Intuitively, we can now couple the sampling from $P(x, \cdot)$ and $P(y, \cdot)$ as follows:

---

[4]See Lemma A.3 for context on ϕ^{-1}.

1. Sample $U_1 \sim \mathcal{U}(\,]0,1[\,)$, call the result u_1 and set $t_x := u_1 \eta_{k,1}(x)$ and $t_y := u_1 \eta_{k,1}(y)$.

2. Sample $U_2 \sim \mathcal{U}(\,]0,1[\,)$, call the result u_2 and set $r_x := F_{t_x}^{-1}(u_2)$ and $r_y := F_{t_y}^{-1}(u_2)$.

3. Sample $\Theta \sim \mathcal{U}(\mathbb{S}^{d-1})$, call the result θ and return $r_x \theta$ and $r_y \theta$ as the new samples.

This corresponds to the coupling $\gamma_{x,y}$ determined by

$$
\gamma_{x,y}(A \times B) := \omega_d^{-1} \int_0^1 \int_0^1 \int_{\mathbb{S}^{d-1}} \mathbb{1}_A\left(F_{u_1 \eta_{k,1}(x)}^{-1}(u_2)\theta\right) \\
\cdot \mathbb{1}_B\left(F_{u_1 \eta_{k,1}(y)}^{-1}(u_2)\theta\right) \sigma_d(\mathrm{d}\theta)\mathrm{d}u_2\mathrm{d}u_1
$$

for $A, B \in \mathcal{B}(\mathbb{R}^d)$. Using this coupling, we can estimate the Wasserstein distance as

$$
\begin{aligned}
&\mathcal{W}(P(x,\cdot), P(y,\cdot)) \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\tilde{x} - \tilde{y}\| \gamma_{x,y}(\mathrm{d}\tilde{x} \times \mathrm{d}\tilde{y}) \\
&= \omega_d^{-1} \int_0^1 \int_0^1 \int_{\mathbb{S}^{d-1}} \left\| F_{u_1 \eta_{k,1}(x)}^{-1}(u_2)\theta - F_{u_1 \eta_{k,1}(y)}^{-1}(u_2)\theta \right\| \sigma_d(\mathrm{d}\theta)\mathrm{d}u_2\mathrm{d}u_1 \\
&= \int_0^1 \int_0^1 \left| F_{u_1 \eta_{k,1}(x)}^{-1}(u_2) - F_{u_1 \eta_{k,1}(y)}^{-1}(u_2) \right| \mathrm{d}u_2\mathrm{d}u_1 \\
&= \int_0^1 \int_0^1 \left| \widehat{\phi}^{-1}\big(-\log(u_1 \eta_{k,1}(x))\big) - \widehat{\phi}^{-1}\big(-\log(u_1 \eta_{k,1}(y))\big) \right| u_2^{1/k} \mathrm{d}u_2\mathrm{d}u_1 \\
&= \frac{k}{k+1} \int_0^1 \left| \widehat{\phi}^{-1}\big(\phi(\|x\|) - \log u_1\big) - \widehat{\phi}^{-1}\big(\phi(\|y\|) - \log u_1\big) \right| \mathrm{d}u_1 \\
&\leq \frac{k}{k+1} |\|x\| - \|y\||,
\end{aligned}
$$

where we use

$$
\int_0^1 u^{1/p}\mathrm{d}u = \left[\frac{1}{1+1/p} u^{1+1/p}\right]_0^1 = \frac{1}{1+1/p} = \frac{p}{p+1}, \qquad p \in \mathbb{R} \setminus \{-1\} \qquad (2.5)
$$

to compute the u_2-integral and apply Lemma A.4 and monotonicity of the Lebesgue integral to obtain the inequality in the last line. Furthermore, by triangle inequality of the Euclidean norm, one has $|\|x\| - \|y\|| \leq \|x - y\|$, which is sharp whenever x and y lie on the same ray emanating from the origin. Thus we now have

$$
\mathsf{Dob}(P) = \sup_{x \neq y} \frac{\mathcal{W}(P(x,\cdot), P(y,\cdot))}{\|x-y\|} \leq \sup_{x \neq y} \frac{\frac{k}{k+1}|\|x\| - \|y\||}{\|x-y\|} = \frac{k}{k+1},
$$

proving that P is Wasserstein contractive with rate $\rho = k/(k+1)$. $\qquad \square$

Discussion 2.3.

(a) We can alternatively state Theorem 2.2 as follows. If $\nu \in \mathcal{P}(\mathbb{R}^d)$ is a distribution with density η and $k > 0$ a constant such that $\eta_{k,1}$ is rotationally invariant and log-concave (meaning that it takes the shape (2.2)), then k-PSS for ν is Wasserstein contractive with rate $\rho = k/(k+1)$. This interpretation would become particularly relevant if one were to view the function $\eta_{k,0}$ as an (improper) prior and ask about the performance of the corresponding slice sampler for rotationally invariant, log-concave likelihood functions $\eta_{k,1}$. An analogous interpretation also makes sense for two of our later results, Theorems 3.5 and 3.7.

(b) Letting $k = d$ in Theorem 2.2, we retrieve precisely [8, Theorem 2.1], the result we are generalizing, which established Wasserstein contraction with rate $\rho = d/(d+1)$ of USS for rotationally invariant, log-concave densities.

(c) With $k = 1$, Theorem 2.2 establishes Wasserstein contraction of PSS, for which, to our knowledge, no such result had been proven yet, with rate $\rho = 1/2$. Though the property is only shown for densities that have a pole of order $(d-1)$ at zero, we still deem it noteworthy that the contraction rate is $1/2$, regardless of the dimension d of the underlying space $\mathbb{R}^d$ (in contrast to the result for USS).

(d) Furthermore, we make note of the fact that the densities for which Wasserstein contraction of PSS is proven are not necessarily log-concave, not even along rays emanating from the origin. For example, with $\kappa = \infty$ and $\phi(r) := \alpha\,r$ for an $\alpha > 0$, we get $\log \eta(r\theta) = (1-d)\log r - \alpha\,r$ for $r \in \mathbb{R}_{>0}$, $\theta \in \mathbb{S}^{d-1}$, which for $d > 1$ is actually strictly convex in r. This is particularly noteworthy because even the broadest previously known quantitative convergence result for PSS [13, Theorem 7] required log-concavity (at least along rays emanating from the origin). In this sense, our result broadens the range of target distributions in which quantitative convergence guarantees for PSS are available. We will further extend this range in the following subsection.

(e) Finally, but perhaps most importantly, as k-PSS continuously interpolates between USS and PSS through the parameter k, the contraction rate $\rho = k/(k+1)$ provided by Theorem 2.2 interpolates accordingly, which shows the interpolation between the samplers to be meaningful in the context of convergence rates.

2.2 Heavy-tailed targets

Some numerical experiments recently published in [17] suggest that PSS performs extremely well for heavy-tailed target distributions in high dimension (provided that they are centered on the origin), whereas USS performs very poorly for the same targets. However, as far as we can tell, no quantitative theoretical results are available to substantiate this hypothesis. Even beyond these specifics, the quantitative theoretical properties of slice sampling for heavy-tailed targets appear to be entirely unexplored.

We therefore deem it worthwhile to detail our investigation of the theoretical behavior of PSS and USS for certain – admittedly somewhat inflexible – classes of heavy-tailed targets. The results we state and prove in this section, Theorems 2.4 and 2.6, as well as Propositions 2.5 and 2.7, should serve to provide some first answers about the samplers' properties and may well inspire future work on the topic.

Arguably the textbook example of multivariate heavy-tailed distributions are the *multivariate t-distributions*, which in their standard form have unnormalized densities of the form (2.6) on $\mathbb{R}^d$, where $m > 0$ is called the *degrees of freedom* parameter. Considering only USS, it is feasible to analyze the behavior for (2.6), and we obtain the following result.

Theorem 2.4. *Let $\nu \in \mathcal{P}(\mathbb{R}^d)$ be a distribution with density η given by*

$$\eta(x) = \left(1 + \frac{1}{m}\|x\|^2\right)^{-(d+m)/2}, \tag{2.6}$$

where $m \in]1, \infty[$. Then USS for ν is Wasserstein contractive with rate

$$\rho = \frac{d(d+m)}{(d+1)(d+m-1)}.$$

Proof. We begin by observing for $x \in \mathbb{R}^d$, $t > 0$, that

$$
\eta(x) > t \quad \Leftrightarrow \quad 1 + \frac{1}{m}\|x\|^2 < t^{-2/(d+m)}
$$
$$
\Leftrightarrow \quad \|x\| < \sqrt{m\,t^{-2/(d+m)} - m}.
$$

Thus the level set at threshold t is

$$
\{x \in \mathbb{R}^d \mid \eta(x) > t\} = \mathfrak{B}(\sqrt{m\,t^{-2/(d+m)} - m}),
$$

where $\mathfrak{B}(\kappa)$ denotes the open zero-centered Euclidean ball of radius κ, for any $\kappa > 0$. Note that for USS the X-update consists of uniform sampling from the slice. Moreover, it is well-known (and also a straightforward consequence of Corollary A.2 and inversion method) that if $\kappa > 0$, $U \sim \mathcal{U}([0,1[)$ and $\Theta \sim \mathcal{U}(\mathbb{S}^{d-1})$, then $U^{1/d}\kappa\Theta \sim \mathcal{U}(\mathfrak{B}(\kappa))$. We make use of this in the construction of our coupling.

Denote by P the transition kernel of USS for ν. For any $x, y \in \mathbb{R}^d$, we intuitively couple $P(x, \cdot)$ and $P(y, \cdot)$ as follows:

1. Sample $U_1 \sim \mathcal{U}(]0,1[)$, call the result u_1 and set

$$
t_x := u_1\eta(x) = u_1\left(1 + \frac{1}{m}\|x\|^2\right)^{-(d+m)/2},
$$
$$
t_y := u_1\eta(y) = u_1\left(1 + \frac{1}{m}\|y\|^2\right)^{-(d+m)/2}.
$$

2. Sample $U_2 \sim \mathcal{U}([0,1[)$, call the result u_2 and set

$$
r_x(u_1, u_2) := u_2^{1/d}\sqrt{mt_x^{-2/(d+m)} - m} = u_2^{1/d}\sqrt{mu_1^{-2/(d+m)}\left(1 + \frac{1}{m}\|x\|^2\right) - m}
$$
$$
= u_2^{1/d}\sqrt{\left(u_1^{-1/(d+m)}\|x\|\right)^2 + mu_1^{-2/(d+m)} - m},
$$
$$
r_y(u_1, u_2) := u_2^{1/d}\sqrt{\left(u_1^{-1/(d+m)}\|y\|\right)^2 + mu_1^{-2/(d+m)} - m}.
$$

3. Sample $\Theta \sim \mathcal{U}(\mathbb{S}^{d-1})$, call the result θ and set $\tilde{x} := r_x(u_1, u_2)\theta$ and $\tilde{y} := r_y(u_1, u_2)\theta$.

4. Return $\tilde{x}$ and $\tilde{y}$ as the new states

This corresponds to the coupling $\gamma_{x,y}$ determined by

$$
\gamma_{x,y}(A \times B) = \omega_d^{-1}\int_0^1\int_0^1\int_{\mathbb{S}^{d-1}} \mathbb{1}_A(r_x(u_1, u_2)\theta)\mathbb{1}_B(r_y(u_1, u_2)\theta)\sigma_d(\mathrm{d}\theta)\mathrm{d}u_2\mathrm{d}u_1 \qquad (2.7)
$$

for $A, B \in \mathcal{B}(\mathbb{R}^d)$. Hence we obtain

$$
\mathcal{W}(P(x, \cdot), P(y, \cdot)) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\tilde{x} - \tilde{y}\|\gamma_{x,y}(\mathrm{d}\tilde{x} \times \mathrm{d}\tilde{y})
$$
$$
= \omega_d^{-1}\int_0^1\int_0^1\int_{\mathbb{S}^{d-1}} \|r_x(u_1, u_2)\theta - r_y(u_1, u_2)\theta\|\sigma_d(\mathrm{d}\theta)\mathrm{d}u_2\mathrm{d}u_1
$$
$$
= \int_0^1\int_0^1 |r_x(u_1, u_2) - r_y(u_1, u_2)|\mathrm{d}u_2\mathrm{d}u_1
$$
$$
= \int_0^1 u_2^{1/d}\mathrm{d}u_2 \cdot \int_0^1 \left|\sqrt{\left(u_1^{-1/(d+m)}\|x\|\right)^2 + mu_1^{-2/(d+m)} - m}\right.
$$

$$-\sqrt{\left(u_1^{-1/(d+m)}\|y\|\right)^2 + mu_1^{-2/(d+m)} - m}\ \Bigg|\mathrm{d}u_1$$

$$\le \int_0^1 u_2^{1/d}\mathrm{d}u_2 \cdot \int_0^1 \left|u_1^{-1/(d+m)}\|x\| - u_1^{-1/(d+m)}\|y\|\right|\mathrm{d}u_1$$

$$= \frac{d}{d+1} \cdot \frac{d+m}{d+m-1} \cdot \left|\|x\| - \|y\|\right|.$$

Note that, somewhat analogous to the proof of Theorem 2.2, the second inequality follows by Lemma A.4, here applied with $\phi(s) := s^2$, and monotonicity of the Lebesgue integral. The equality in the last line follows by (2.5), applied with $p = d$ and $p = -(d+m)$. By the same arguments as in Theorem 2.2, the above estimation yields our claim. $\quad\square$

To demonstrate the sensitivity of our estimations in Theorem 2.4, we complement them by the following sharpness result.

Proposition 2.5. *The contraction rate estimate provided by Theorem 2.4 is sharp. That is, for any $m > 1$ the transition kernel P of USS for (2.6) satisfies*

$$\mathsf{Dob}(P) = \frac{d(d+m)}{(d+1)(d+m-1)}. \tag{2.8}$$

Proof. Theorem 2.4 shows (2.8) with "$\le$" in place of equality. Thus it only remains to show "$\ge$". For this we make use of the Kantorovich-Rubinstein theorem [18, Section 1.2], which states that

$$\mathcal{W}(P(x,\cdot), P(y,\cdot)) = \sup_{\|g\|_{\mathrm{Lip}}\le 1}\left|\int_{\mathbb{R}^d} g(z)P(x,\mathrm{d}z) - \int_{\mathbb{R}^d} g(z)P(y,\mathrm{d}z)\right|,$$

where

$$\|g\|_{\mathrm{Lip}} := \sup_{z_1\ne z_2\in\mathbb{R}^d} \frac{|g(z_1) - g(z_2)|}{\|z_1 - z_2\|}.$$

Note that, by triangle inequality of $\|\cdot\|$, for $g := \|\cdot\|$ one has $\|g\|_{\mathrm{Lip}} = 1$.

The analysis in the proof of Theorem 2.4 shows

$$P(x, A) = \omega_d^{-1}\int_0^1\int_0^1\int_{\mathbb{S}^{d-1}} \mathbb{1}_A(r_x(u_1, u_2)\theta)\sigma_d(\mathrm{d}\theta)\mathrm{d}u_2\mathrm{d}u_1,$$

so that

$$\int_{\mathbb{R}^d}\|z\|P(x,\mathrm{d}z) = \omega_d^{-1}\int_0^1\int_0^1\int_{\mathbb{S}^{d-1}}\|r_x(u_1, u_2)\theta\|\sigma_d(\mathrm{d}\theta)\mathrm{d}u_2\mathrm{d}u_1$$

$$= \int_0^1\int_0^1 r_x(u_1, u_2)\mathrm{d}u_2\mathrm{d}u_1$$

$$= \frac{d}{d+1}\int_0^1 \sqrt{\left(u_1^{-1/(d+m)}\|x\|\right)^2 + mu_1^{-2/(d+m)} - m}\ \mathrm{d}u_1.$$

We now consider the points $x = r\theta_0$, $y = \frac{1}{2}r\theta_0$ for varying $r \in \mathbb{R}_+$ and arbitrary but fixed $\theta_0 \in \mathbb{S}^{d-1}$. By the Kantorovich-Rubinstein theorem, we obtain

$$\mathsf{Dob}(P) \ge \sup_{x\ne y} \frac{\left|\int_{\mathbb{R}^d}\|z\|P(x,\mathrm{d}z) - \int_{\mathbb{R}^d}\|z\|P(y,\mathrm{d}z)\right|}{\|x - y\|}$$

$$\ge \lim_{r\to\infty} \frac{\left|\int_{\mathbb{R}^d}\|z\|P(r\theta_0,\mathrm{d}z) - \int_{\mathbb{R}^d}\|z\|P(\frac{1}{2}r\theta_0,\mathrm{d}z)\right|}{\left\|r\theta_0 - \frac{1}{2}r\theta_0\right\|}$$

$$\begin{aligned}
&= \frac{2d}{d+1} \lim_{r \to \infty} \left| \int_0^1 \sqrt{\left(u_1^{-1/(d+m)} \right)^2 + \frac{m u_1^{-2/(d+m)} - m}{r^2}} \, \mathrm{d}u_1 \right. \\
&\qquad\qquad\qquad \left. - \int_0^1 \sqrt{\left(\frac{1}{2} u_1^{-1/(d+m)} \right)^2 + \frac{m u_1^{-2/(d+m)} - m}{r^2}} \, \mathrm{d}u_1 \right| \\
&= \frac{2d}{d+1} \left| \int_0^1 \left(u_1^{-1/(d+m)} - \frac{1}{2} u_1^{-1/(d+m)} \right) \mathrm{d}u_1 \right| = \frac{d}{d+1} \int_0^1 u_1^{-1/(d+m)} \mathrm{d}u_1 \\
&= \frac{d}{d+1} \cdot \frac{d+m}{d+m-1},
\end{aligned}$$

which is precisely what we set out to prove. Note that the limes can be pulled into the integrals by the monotone convergence theorem. $\qquad\square$

Unfortunately, for k-PSS besides USS, so in particular for PSS, the density factors $\eta_{k,1}$ for η as in (2.6) have intractable level sets, which prevents us from extending the analysis of Theorem 2.4 to these cases. However, we may instead consider a class of distributions that have a simpler structure, while possessing tails asymptotically equivalent to those of the standard multivariate t-distributions (2.6). Namely, we consider the target densities (2.9), which also have a parameter $m > 0$ that serves the same role as in (2.6). We obtain the following result.

Theorem 2.6. *For an $\varepsilon > 0$ let $\nu \in \mathcal{P}(\mathbb{R}^d \setminus \mathfrak{B}(\varepsilon)) \subset \mathcal{P}(\mathbb{R}^d)$ be a distribution with density η given by*

$$\eta(x) = \|x\|^{-(d+m)} \mathbb{I}(\|x\| \geq \varepsilon) \tag{2.9}$$

for some $m > 1$. Then k-PSS for ν is Wasserstein contractive with rate

$$\rho = \frac{k(k+m)}{(k+1)(k+m-1)}$$

for any $k \geq 1$.

Proof. Arbitrarily fix $k \geq 1$. Note that for any $r \geq \varepsilon$ and $\theta \in \mathbb{S}^{d-1}$, the distribution-dependent factor of the k-PSS factorization becomes $\eta_{k,1}(r\theta) = r^{-(k+m)}$. Thus, for any $\theta \in \mathbb{S}^{d-1}$ and $t \in \,]0, \varepsilon^{-(k+m)}[$, we get

$$\{ r \in \mathbb{R}_+ \mid \eta_{k,1}(r\theta) > t \} = [\varepsilon, t^{-1/(k+m)}[.$$

As in the proof of Theorem 2.2, we decompose k-PSS into polar coordinates and couple the sampling of the radii via inversion method. To this end, we note that by Corollary A.2, given threshold t, radius and direction need to be drawn from the joint distribution with density

$$(r, \theta) \mapsto r^{k-1} \mathbb{I}(\eta_{k,1}(r\theta) > t) \mathbb{1}_{\mathbb{R}_+}(r) = r^{k-1} \mathbb{1}_{[\varepsilon, t^{-1/(k+m)}[}(r).$$

This corresponds to the c.d.f. F_t that for $s \in [\varepsilon, t^{-1/(k+m)}]$ is given by

$$F_t(s) := \frac{\int_\varepsilon^s r^{k-1} \mathrm{d}r}{\int_\varepsilon^{t^{-1/(k+m)}} r^{k-1} \mathrm{d}r} = \frac{\left[\frac{1}{k} r^k \right]_\varepsilon^s}{\left[\frac{1}{k} r^k \right]_\varepsilon^{t^{-1/(k+m)}}} = \frac{s^k - \varepsilon^k}{t^{-k/(k+m)} - \varepsilon^k}.$$

Observe that F_t bijectively maps $[\varepsilon, t^{-1/(k+m)}]$ onto $[0, 1]$ and that the inverse F_t^{-1} of this restriction is given by

$$F_t^{-1}(u) = ((t^{-k/(k+m)} - \varepsilon^k) u + \varepsilon^k)^{1/k} = (u \, t^{-k/(k+m)} + (1-u)\varepsilon^k)^{1/k}, \quad u \in [0, 1].$$

By introducing the trivially invertible auxiliary function $h : \mathbb{R}_+ \to \mathbb{R}_+, \ r \mapsto r^k$, we may alternatively write

$$F_t^{-1}(u) = h^{-1}(h(u^{1/k} t^{-1/(k+m)}) + (1-u)\varepsilon^k).$$

Denote by P the transition kernel of k-PSS for ν. For any $x, y \in \mathbb{R}^d \setminus \mathfrak{B}(\varepsilon)$, we intuitively construct our coupling of $P(x, \cdot)$ and $P(y, \cdot)$ as follows:

1. Sample $U_1 \sim \mathcal{U}(\,]0, 1[\,)$, call the result u_1 and set

$$t_x := u_1 \eta_{k,1}(x) = u_1 \|x\|^{-(k+m)} \quad \text{and} \quad t_y := u_1 \eta_{k,1}(y) = u_1 \|y\|^{-(k+m)}.$$

2. Sample $U_2 \sim \mathcal{U}(\,[0, 1[\,)$, call the result u_2 and set

$$r_x(u_1, u_2) := F_{t_x}^{-1}(u_2) = h^{-1}(h(u_2^{1/k} u_1^{-1/(k+m)} \|x\|) + (1 - u_2)\varepsilon^k),$$
$$r_y(u_1, u_2) := F_{t_y}^{-1}(u_2) = h^{-1}(h(u_2^{1/k} u_1^{-1/(k+m)} \|y\|) + (1 - u_2)\varepsilon^k).$$

3. Sample $\Theta \sim \mathcal{U}(\mathbb{S}^{d-1})$, call the result θ and set $\tilde{x} := r_x(u_1, u_2)\theta$ and $\tilde{y} := r_y(u_1, u_2)\theta$.

4. Return $\tilde{x}$ and $\tilde{y}$ as the new states.

The resulting coupling again takes the shape (2.7). Thus, omitting some steps that coincide with ones in Theorem 2.4, we get

$$
\begin{aligned}
\mathcal{W}(P(x, \cdot), P(y, \cdot)) &\leq \int_0^1 \int_0^1 |r_x(u_1, u_2) - r_y(u_1, u_2)| \mathrm{d}u_2 \mathrm{d}u_1 \\
&= \int_0^1 \int_0^1 |h^{-1}(h(u_2^{1/k} u_1^{-1/(k+m)} \|x\|) + (1 - u_2)\varepsilon^k) \\
&\qquad\qquad - h^{-1}(h(u_2^{1/k} u_1^{-1/(k+m)} \|y\|) + (1 - u_2)\varepsilon^k)| \mathrm{d}u_2 \mathrm{d}u_1 \\
&\leq \int_0^1 \int_0^1 \left| u_2^{1/k} u_1^{-1/(k+m)} \|x\| - u_2^{1/k} u_1^{-1/(k+m)} \|x\| \right| \mathrm{d}u_2 \mathrm{d}u_1 \\
&= \int_0^1 u_2^{1/k} \mathrm{d}u_2 \cdot \int_0^1 u_1^{-1/(k+m)} \mathrm{d}u_1 \cdot |\|x\| - \|y\|| \\
&= \frac{k}{k+1} \cdot \frac{k+m}{k+m-1} \cdot |\|x\| - \|y\||.
\end{aligned}
$$

Note that, as in Theorem 2.4, the second inequality follows by Lemma A.4, here applied with $\phi := h$, which is convex thanks to our assumption that $k \geq 1$. The final equality again follows by (2.5), here applied with $p = k$ and $p = -(k+m)$. By the same arguments as in the proof of Theorem 2.2, the above estimation yields our claim. $\qquad\square$

Analogous to what we did for Theorem 2.4 with Proposition 2.5, we can complement the estimates of Theorem 2.6 by a sharpness guarantee.

Proposition 2.7. *The contraction rate estimate provided by Theorem 2.6 is sharp. That is, for any $m > 1$ and $k \geq 1$, the transition kernel P of k-PSS for (2.9) satisfies*

$$\mathsf{Dob}(P) = \frac{k(k+m)}{(k+1)(k+m-1)}.$$

Proof. This follows by the same proof strategy as Proposition 2.5, though of course the differing transition kernel causes some of the details to change. For the reader's convenience we provide the part of the estimation affected by this (using notation and auxiliary results from the proof of Theorem 2.6).

$$
\begin{aligned}
\mathsf{Dob}(P) &\geq \lim_{r \to \infty} \frac{\left| \int_{\mathbb{R}^d} \|z\| P(r\theta_0, \mathrm{d}z) - \int_{\mathbb{R}^d} \|z\| P(\frac{1}{2}r\theta_0, \mathrm{d}z) \right|}{\left\| r\theta_0 - \frac{1}{2}r\theta_0 \right\|} \\
&= 2 \lim_{r \to \infty} \frac{1}{r} \left| \int_0^1 \int_0^1 \left(h^{-1}(h(u_2^{1/k} u_1^{-1/(k+m)} r) + (1 - u_2)\varepsilon^k) \right. \right.
\end{aligned}
$$

$$- h^{-1}(h(\tfrac{1}{2}u_2^{1/k}u_1^{-1/(k+m)}r) + (1 - u_2)\varepsilon^k)\Big)\mathrm{d}u_2\mathrm{d}u_1 \bigg|$$

$$= 2\lim_{r\to\infty}\bigg|\int_0^1\int_0^1 \left(h^{-1}\left(h(u_2^{1/k}u_1^{-1/(k+m)}) + \frac{(1 - u_2)\varepsilon^k}{r^k}\right)\right.$$

$$\left. - h^{-1}\left(h(\tfrac{1}{2}u_2^{1/k}u_1^{-1/(k+m)}) + \frac{(1 - u_2)\varepsilon^k}{r^k}\right)\right)\mathrm{d}u_2\mathrm{d}u_1\bigg|$$

$$= 2\bigg|\int_0^1\int_0^1\left(u_2^{1/k}u_1^{-1/(k+m)} - \tfrac{1}{2}u_2^{1/k}u_1^{-1/(k+m)}\right)\mathrm{d}u_2\mathrm{d}u_1\bigg|$$

$$= \int_0^1 u_2^{1/k}\mathrm{d}u_2 \cdot \int_0^1 u_1^{-1/(k+m)}\mathrm{d}u_1 = \frac{k}{k+1}\cdot\frac{k+m}{k+m-1} \qquad \square$$

Discussion 2.8.

(a) We note that, much like in Theorem 2.2, the contraction rate shown in Theorem 2.6 only depends on the sample space dimension d if k depends on it. Consequently the rate obtained for PSS (i.e. $k = 1$), which is

$$\rho = \frac{m+1}{2m},$$

is dimension-independent. We note that it tends to $1/2$ as $m \to \infty$.

(b) For USS (i.e. $k = d$) on the other hand, the rate obtained in both Theorems 2.4 and 2.6 is the dimension-dependent

$$\rho = \frac{d(d+m)}{(d+1)(d+m-1)}. \tag{2.10}$$

(c) That the contraction rates of USS for (2.6) and (2.9) coincide suggests that the relevant properties of the former are retained when approximating it by the latter – especially because, by the sharpness results of Propositions 2.5 and 2.7, we do not merely have coinciding *rate estimates*, but actually coinciding *rates*. We also note here that the weight the density (2.9) places on its tails can be freely controlled by the parameter $\varepsilon > 0$ (which does not affect the contraction rates), so that in particular this tail weight can be matched precisely with that of the corresponding density (2.6). This further justifies using (2.9) as a proxy for (2.6).

(d) We note that (2.10) tends to $d/(d+1)$ for any fixed d as $m \to \infty$. This is perhaps not too surprising, since in the same limit the density (2.6) approaches that of the multivariate standard normal distribution, which in turn fits (with $\kappa = \infty$ and $\phi(r) = r^2/2$) into Theorem 2.2, restricted to USS, where we also obtain Wasserstein contraction with rate $\rho = d/(d+1)$.

(e) On the other hand, for any fixed m, as $d \to \infty$, the rate (2.10) tends to 1. This is particularly noteworthy to us because we are interested in the behavior of USS in high dimension. We emphasize that (2.10) deteriorates fairly quickly with increasing dimension. For example, for target density (2.9) with $m = 2$ in dimension $d = 100$, the contraction rate of PSS is $\rho = 3/4$, whereas that of USS is $\rho \approx 0.9999$.

(f) Our result has the caveat that it does not explain the empirically excellent performance (cf. [17, Section 6.1]) of PSS even for the standard multivariate Cauchy distribution, which results from (2.6) with $m = 1$, and thus corresponds in the simplified setting to (2.9) with $m = 1$. We emphasize that the issue here is not an insufficiently detailed analysis: When plugging $m = 1$ into the proofs of both

Theorem 2.6 and Proposition 2.7, they yield $\mathsf{Dob}(P) = 1$, meaning that PSS simply is not Wasserstein contractive for this target.

This discrepancy between empirical and theoretical results gives reason to believe that the rates we obtain w.r.t. the Wasserstein distance may not reflect those that hold (albeit so far unproven) w.r.t. the total variation distance (i.e. corresponding to spectral gaps).

3 Spectral gap estimates

3.1 General tool

In [8], the authors developed a method to prove quantitative spectral gap estimates for USS. Roughly speaking, their method yields a spectral gap estimate for USS applied to a target density ϱ whenever one is able to verify that the level set function $\ell : \mathbb{R}_{>0} \to \mathbb{R}_+$ given by

$$\ell(t) := \lambda_d(\{x \in \mathbb{R}^d \mid \varrho(x) > t\}), \qquad t > 0$$

has certain properties. In [15] this method was adapted to general slice samplers, using arbitrary factorization $\varrho = \varrho_0 \, \varrho_1$, which necessitated replacing ℓ by the *generalized level set function* $\ell_{\varrho_0,\varrho_1} : \mathbb{R}_{>0} \to \mathbb{R}_+$ given by

$$\ell_{\varrho_0,\varrho_1}(t) := \int_{\mathbb{R}^d} \varrho_0(x) \mathbb{I}(\varrho_1(x) > t) \mathrm{d}x, \qquad t > 0. \tag{3.1}$$

Using the first of our Wasserstein contraction results, Theorem 2.2, we further generalize the method to allow for more fine-grained and in some cases even further-reaching results. For this we require the following auxiliary result from [15].

Lemma 3.1. *Let $d_1, d_2 \in \mathbb{N}$ and let $\pi \in \mathcal{P}(\mathbb{R}^{d_1})$ and $\nu \in \mathcal{P}(\mathbb{R}^{d_2})$ be distributions with densities ϱ and η arbitrarily factorized as $\varrho = \varrho_0 \, \varrho_1$ and $\eta = \eta_0 \, \eta_1$, like in (1.1). Provided that $\ell_{\varrho_0,\varrho_1} \equiv \ell_{\eta_0,\eta_1}$, i.e. if $\ell_{\varrho_0,\varrho_1}(t) = \ell_{\eta_0,\eta_1}(t)$ for all $t \in \mathbb{R}_{>0}$, one has*

$$\mathit{gap}_\pi(P_\pi) = \mathit{gap}_\nu(P_\nu),$$

where P_π and P_ν are the transition kernels of slice sampling for π and ν, respectively.

Proof. See [15, Theorem 3.5]. $\qquad\square$

Following both [8] and [15], we conveniently package the conditions one needs to verify to obtain a spectral gap estimate by defining classes Λ_k of all functions ℓ that satisfy them.

Definition 3.2. *For any fixed $k \in \mathbb{R}_{>0}$, we define Λ_k as the class of continuous functions $\ell : \mathbb{R}_{>0} \to \mathbb{R}_+$ for which all of the following hold:*

(i) $\lim_{t \to \infty} \ell(t) = 0$ and $\sup \ell := \lim_{t \searrow 0} \ell(t) \in \,]0, \infty]$

(ii) ℓ restricted to $\mathsf{supp}(\ell) := \,]0, \sup\{t \in \mathbb{R}_{>0} \mid \ell(t) > 0\}[$ is strictly decreasing

(iii) the function $]0, (\sup \ell)^{1/k}[\to \mathbb{R}_{>0}, \; r \mapsto \ell^{-1}(r^k)$ is log-concave.

Note that for functions $\ell_{\varrho_0,\varrho_1}$ as in (3.1), it is easy to see that condition (i) is always satisfied, so we ignore it when trying to show statements of the form $\ell_{\varrho_0,\varrho_1} \in \Lambda_k$.

Following [15], we note that the assumed continuity and conditions (i) and (ii) altogether imply that ℓ restricted to $\mathsf{supp}(\ell)$ maps bijectively onto $]0, \sup \ell[$, which in turn implies the existence of the inverse function $\ell^{-1} : \,]0, \sup \ell[\to \mathsf{supp}(\ell)$ used in condition (iii). It is easy to see that ℓ^{-1} is also strictly decreasing.

We are now prepared to state and prove our generalization of the method for establishing quantitative spectral gap estimates.

Theorem 3.3. *Let $\pi \in \mathcal{P}(\mathbb{R}^d)$ be a distribution with density ϱ factorized as $\varrho = \varrho_0\, \varrho_1$ like in (1.1). Suppose that $\ell_{\varrho_0,\varrho_1} \in \Lambda_k$ for some $k \in \mathbb{R}_{>0}$, then*

$$gap_\pi(P_\pi) \geq \frac{1}{k+1},$$

where P_π is the transition kernel of slice sampling for π.

Proof. Our proof strategy is to construct a distribution $\nu \in \mathcal{D}_k(\mathbb{R})$ with density η, factorized according to k-PSS, such that $\ell_{\eta_{k,0},\eta_{k,1}} \equiv \ell_{\varrho_0,\varrho_1}$. Afterwards we just need to apply a number of earlier results to obtain the claimed estimate. We note that our proof strategy would also work if we instead constructed a $\nu \in \mathcal{D}_k(\mathbb{R}^m)$ for some $m > 1$ and that we simply chose $m = 1$ for simplicity.

Fix any $k \in \mathbb{R}_{>0}$ such that $\ell_{\varrho_0,\varrho_1} \in \Lambda_k$, set

$$\kappa := \left(\frac{k}{2} \sup \ell_{\varrho_0,\varrho_1} \right)^{1/k} \in\ ]0, \infty]$$

and define $\phi :\ ]0, \kappa[\ \to \mathbb{R}$ by

$$\phi(r) := -\log\left(\ell_{\varrho_0,\varrho_1}^{-1}\left(\frac{2}{k} r^k \right) \right).$$

Note that since ϕ is a composition of the strictly decreasing functions $-\log$ and $\ell_{\varrho_0,\varrho_1}^{-1}$ and the strictly increasing function $r \mapsto \frac{2}{k} r^k$, it is strictly increasing. Furthermore, since ϕ can also be viewed as a composition of the (by condition (iii) of Definition 3.2) convex function $r \mapsto -\log \ell_{\varrho_0,\varrho_1}^{-1}(r^k)$ and the linear function $r \mapsto (2/k)^{1/k} r$, it is convex as well. Hence it satisfies both assumptions of Definition 2.1 and therefore defines a distribution $\nu \in \mathcal{D}_k(\mathbb{R})$ with density η as in (2.1).

It is easy to verify that the inverse[5] $\phi^{-1} : I_\phi \to D_\phi$ of ϕ is given by

$$\phi^{-1}(s) = \left(\frac{k}{2} \ell_{\varrho_0,\varrho_1}\left(\exp(-s) \right) \right)^{1/k}.$$

Furthermore, one has

$$\sup \phi = \lim_{r \nearrow \kappa} \phi(r) = \lim_{r' \nearrow \sup \ell_{\varrho_0,\varrho_1}} -\log \ell_{\varrho_0,\varrho_1}^{-1}(r') = \lim_{r'' \searrow 0} -\log r'' = \infty,$$

which yields $\widehat{\phi}^{-1} \equiv \phi^{-1}$ (see Lemma A.4). Plugging this into (2.3) and the result into the definition of $\ell_{\eta_{k,0},\eta_{k,1}}$, we get

$$\ell_{\eta_{k,0},\eta_{k,1}}(t) = \int_{\mathbb{R}} |x|^{k-1} \mathbb{1}_{]0,\phi^{-1}(-\log t)[}(|x|)\mathrm{d}x = 2 \int_0^{\phi^{-1}(-\log t)} r^{k-1}\mathrm{d}r$$

$$= 2 \left[\frac{1}{k} r^k \right]_0^{\phi^{-1}(-\log t)} = \frac{2}{k} \phi^{-1}(-\log t)^k = \ell_{\varrho_0,\varrho_1}(t)$$

for any $t \in \mathsf{supp}(\ell_{\eta_{k,0},\eta_{k,1}})$. Note that

$$\sup\{t \in \mathbb{R}_{>0} \mid \ell_{\eta_{k,0},\eta_{k,1}}(t) > 0\} = \exp(-\inf \phi) = \lim_{r \searrow 0} \exp(-\phi(r))$$

$$= \lim_{r \searrow 0} \ell_{\varrho_0,\varrho_1}^{-1}\left(\frac{2}{k} r^k \right) = \lim_{r' \searrow 0} \ell_{\varrho_0,\varrho_1}^{-1}(r') = \sup\{t \in \mathbb{R}_{>0} \mid \ell_{\varrho_0,\varrho_1}(t) > 0\}$$

---

[5]See Lemma A.3 for nomenclature.

and therefore $\mathsf{supp}(\ell_{\eta_{k,0},\eta_{k,1}}) = \mathsf{supp}(\ell_{\varrho_0,\varrho_1})$. Of course we also have

$$\ell_{\eta_{k,0},\eta_{k,1}}(t) = 0 = \ell_{\varrho_0,\varrho_1}(t)$$

for thresholds t too large to be contained in the supports. Overall, we have now shown $\ell_{\eta_{k,0},\eta_{k,1}} \equiv \ell_{\varrho_0,\varrho_1}$, so that Lemma 3.1 applies.

Denote by P_ν the transition kernel of k-PSS for ν. From Theorem 2.2 we know that P_ν is Wasserstein contractive with rate $\rho = k/(k+1)$. Moreover, it is easy to see that all distributions in $\mathcal{D}_k(\mathbb{R}^d)$, for any $k > 0$ and $d \in \mathbb{N}$, have finite second moment, so in particular this must be true of $\nu \in \mathcal{D}_k(\mathbb{R})$. Hence (1.3) applies and we can conclude the proof by

$$\mathsf{gap}_\pi(P_\pi) \overset{3.1}{=} \mathsf{gap}_\nu(P_\nu) \overset{(1.3)}{\geq} \frac{1}{k+1}. \qquad\qquad \square$$

Discussion 3.4.

(a) We emphasize that Definition 3.2 is virtually identical to [15, Definition 3.7], which underlies the result we generalize here. The only difference is that the parameter k, which was restricted to $\mathbb{N}$ in [15], is now allowed to take any value in $\mathbb{R}_{>0}$. Hence we generalize the method from a discrete sequence of classes $(\Lambda_k)_{k \in \mathbb{N}}$ to a continuous spectrum $(\Lambda_k)_{k \in \mathbb{R}_{>0}}$ that contains the sequence as a subset.

(b) This generalization is made possible by a subtle change in the proof strategy: Where the corresponding proof in [15] constructed for each $k \in \mathbb{N}$ a ν that was defined on $\mathbb{R}^k$ and then used the Wasserstein contraction result of [8] for USS in that dimension, we instead stay in dimension one and use our own contraction result for k-PSS. Hence the increased flexibility of Theorem 3.3 (w.r.t. the available choices of k) over the theorem's counterpart in [15] is essentially due to the fact that k-PSS is well-defined and Wasserstein contractive for any $k \in \mathbb{R}_{>0}$, not just for integer values $k \in \mathbb{N}$.

(c) We will see in the following that our generalization is useful in the sense that we can find for any $k \in \mathbb{R}_{>0}$ a target distribution π with density ϱ, factorized as in (1.1), satisfying $\ell_{\varrho_0,\varrho_1} \in \Lambda_k$ and $\ell_{\varrho_0,\varrho_1} \notin \Lambda_{k'}$ for all $k' < k$.

(d) A particularly valuable aspect of the continuous spectrum in our result is that we allow the cases $k \in {]0,1[}$, for which the theorem yields lower bounds on the spectral gap that lie in ${]1/2,1[}$, whereas the result from [8] could at best give a lower bound of $1/2$ (with $k = 1$). This is of importance to both of the applications of Theorem 3.3 that we develop in the subsequent subsections.

In order to ease application of Theorem 3.3, we provide in Appendix A two more results from [15] (Lemma A.5 and Proposition A.6). In the remainder of this section, we present two non-trivial applications of Theorem 3.3 to k-PSS and a result for USS that follows by an analogous approach.

3.2 Rotationally invariant targets

Our first application of Theorem 3.3 concerns k-PSS for certain types of rotationally invariant target densities. For target densities ϱ we obtain spectral gap estimates that depend on the degree to which their factor $\varrho_{k,1}$ is log-concave.

Theorem 3.5. *Let $k \in \mathbb{R}_{>0}$ and let $\pi \in \mathcal{P}(\mathbb{R}^d)$ be a distribution with density ϱ given by*

$$\varrho(r\theta) = r^{k-d} \exp(-\phi(r^m)) \mathbb{1}_{]0,\kappa[}(r^m), \qquad r \in \mathbb{R}_+, \theta \in \mathbb{S}^{d-1}, \tag{3.2}$$

where $m \in \mathbb{R}_{>0}$, $\kappa \in]0, \infty]$ and $\phi :]0, \kappa[\to \mathbb{R}$ is strictly increasing and convex and satisfies $\lim_{r \nearrow \kappa} \phi(r) = \infty$. Then one has

$$\mathsf{gap}_\pi(P) \geq \frac{m}{k+m},$$

where P is the transition kernel of k-PSS for π.

Proof. Observe that for $t \in]0, \exp(-\inf \phi)[$, $r > 0$ and $\theta \in \mathbb{S}^{d-1}$ one has

$$
\begin{aligned}
\varrho_{k,1}(r\theta) > t &\quad \Leftrightarrow \quad \exp(-\phi(r^m)) > t \quad \text{and} \quad r^m < \kappa \\
&\quad \Leftrightarrow \quad r^m < \phi^{-1}(-\log t) \quad \text{and} \quad r^m < \kappa \\
&\quad \Leftrightarrow \quad r^m < \phi^{-1}(-\log t) \\
&\quad \Leftrightarrow \quad r < \phi^{-1}(-\log t)^{1/m},
\end{aligned}
$$

where the second and fourth equivalence rely on ϕ^{-1} and $r \mapsto r^{1/m}$ being strictly increasing, and the third equivalence on ϕ^{-1} mapping to $]0, \kappa[$, s.t. in particular $\phi^{-1} < \kappa$. Thus, by the polar coordinates formula (Proposition A.1), for $t \in]0, \exp(-\inf \phi)[$ we get

$$
\begin{aligned}
\ell_{\varrho_{k,0}, \varrho_{k,1}}(t) &= \int_{\mathbb{R}^d} \|x\|^{k-d} \mathbb{I}(\varrho_{k,1}(x) > t) \mathrm{d}x = \omega_d \int_0^\infty r^{k-1} \mathbb{I}(r < \phi^{-1}(-\log t)^{1/m}) \mathrm{d}r \\
&= \omega_d \left[\frac{1}{k} r^k \right]_0^{\phi^{-1}(-\log t)^{1/m}} = \frac{\omega_d}{k} \phi^{-1}(-\log t)^{k/m}.
\end{aligned}
$$

From this it is easy to see that $\ell_{\varrho_{k,0}, \varrho_{k,1}}$ is continuous and strictly decreasing on its support, so only condition (iii) of Definition 3.2 remains to be verified in order to enable application of Theorem 3.3. To this end, observe that the above computation yields

$$\ell_{\varrho_{k,0}, \varrho_{k,1}}(\exp(-s))^{m/k} = \left(\frac{\omega_d}{k} \phi^{-1}(s)^{k/m} \right)^{m/k} = \left(\frac{\omega_d}{k} \right)^{m/k} \phi^{-1}(s),$$

which is concave in s by the assumptions on ϕ and Lemma A.5. By Proposition A.6, this shows $\ell_{\varrho_{k,0}, \varrho_{k,1}}$ to satisfy condition (iii) of Definition 3.2 with parameter k/m. Therefore Theorem 3.3 applies and yields

$$\mathsf{gap}_\pi(P) \geq \frac{1}{k/m + 1} = \frac{m}{k+m}. \qquad \square$$

Discussion 3.6.

(a) We point out that there is a very intuitive relationship between the result of Theorem 3.5 and our Wasserstein contraction result, Theorem 2.2: If we set $m = 1$ in Theorem 3.5, we obtain $\pi \in \mathcal{D}_k(\mathbb{R}^d)$, so that Theorem 2.2 applies and yields that k-PSS for π is Wasserstein contractive with rate $\rho = k/(k+1)$. By (1.3), this in turn yields a spectral gap estimate of $1/(k+1)$, which is the same as that provided by Theorem 3.5 (due to our setting $m = 1$). Hence, if one interprets the contraction result purely as a spectral gap estimate, then Theorem 3.5 generalizes it by introducing the parameter $m \in \mathbb{R}_{>0}$.

(b) Furthermore, in the cases $m > 1$, the value of m intuitively provides information about the 'degree of log-concavity' of $\varrho_{k,1}$, which controls the target density's rate of decay along rays emanating from the origin. This additional information enables a more sensitive estimation of the spectral gap, leading to a lower bound of $m/(k+m)$, which is strictly larger than the lower bound of $1/(k+1)$ that could be obtained without it.

(c) The cases $m < 1$ complement this by requiring *less* information, as the resulting exp-term in ϱ does no longer need to be log-concave (e.g. for $\phi(r) = r$, $m = 1/2$ we get $\phi(r^m) = \sqrt{r}$, which is strictly concave rather than convex). Correspondingly, the resulting lower bounds of $m/(k+m)$ are strictly smaller than the lower bound of $1/(k+1)$ one could obtain in the case $m = 1$.

(d) Though we are unable to formally show that the spectral gap estimates provided by Theorem 3.5 are sharp, there is substantial numerical evidence suggesting that they are. We refer to Appendix B for details.

(e) We deem it worthwhile to consider the implications of Theorem 3.5 for the spectral gap of USS, which thus far was most effectively examined in [8]. The theorem yields for USS applied to

$$\varrho(r\theta) = \exp(-\phi(r^m))\mathbb{1}_{]0,\kappa[}(r^m)$$

the estimate

$$\mathsf{gap}_\pi(P) \geq \frac{m}{d+m}.$$

Despite being a restriction of the theorem, this can still be viewed as a vast generalization of the most closely related previous result: With $\kappa = \infty$, $\phi(r) = \alpha r$ and $m \leq d$ we obtain the setting of [8, Example 3.15]. Aside from generalizing this to potentially bounded support, non-linear ϕ and larger values of m, we also replace the gap estimate $1/(\lceil d/m \rceil + 1)$ shown in [8] for this setting by the much nicer (and for almost all m also larger) estimate $m/(d+m) = 1/(d/m + 1)$.

(f) Note that in the limit $m \to \infty$, Theorem 3.5 yields a spectral gap of 1, corresponding to instantaneous convergence. We can give some intuition for this: Suppose we were using k-PSS for a target density ϱ of the form

$$\varrho(x) = \|x\|^{k-d} c \, \mathbb{1}_A(x)$$

for some $c > 0$ and $A \in \mathcal{B}(\mathbb{R}^d)$ with $\lambda_d(A) < \infty$. Then we would get for $t \in \,]0, c[$ that

$$\varrho_{k,1}(x) > t \quad \Leftrightarrow \quad c\mathbb{1}_A(x) > t \quad \Leftrightarrow \quad x \in A$$

and thus

$$\varrho_{k,0}(x)\mathbb{I}(\varrho_{k,1}(x) > t) = \|x\|^{k-d}\mathbb{1}_A(x) \propto \varrho(x),$$

so that each X-update would sample directly from the target distribution. We say that the sampler is *exact* in this setting. Furthermore, since for $r \geq 0$ one has

$$\lim_{m\to\infty} r^m = \begin{cases} 0 & r < 1 \\ 1 & r = 1 \\ \infty & r > 1 \end{cases}$$

the density (3.2) converges for $m \to \infty$ almost everywhere to

$$r\theta \mapsto r^{k-d}\exp(-\inf\phi)\mathbb{1}_{]0,1[}(r),$$

which fits into the exact sampling setting described above via $c := \exp(-\inf\phi)$ and $A = \mathfrak{B}(1) \setminus \{0\}$. In other words, k-PSS performs increasingly well for the target densities (3.2) with increasing m because they become closer and closer approximations of a target density for which the sampler is exact. Note that the same dynamic is also at play in our next result, Theorem 3.7.

3.3 Rotationally asymmetric targets

In our second application of Theorem 3.3, we consider what happens to the spectral gap of k-PSS when the target density is not rotationally invariant. Here it turns out to be very helpful to first restrict the setting of Theorem 3.5 by eliminating the domain parameter κ and the functional parameter ϕ, and then generalize it by introducing a different functional parameter χ that allows for rotational asymmetry. In the resulting setting, the asymmetry does not have any effect on the spectral gap estimates we obtain, i.e. it does not appear to hamper the samplers' performance.

Theorem 3.7. *Let $k \in \mathbb{R}_{>0}$ and let $\pi \in \mathcal{P}(\mathbb{R}^d)$ be a distribution with density ϱ given by*

$$\varrho(r\theta) = r^{k-d} \exp(-\chi(\theta)r^m), \qquad r \in \mathbb{R}_+, \theta \in \mathbb{S}^{d-1},$$

where $\chi : \mathbb{S}^{d-1} \to \mathbb{R}_{>0}$ is a measurable function[6] and $m \in \mathbb{R}_{>0}$. Then

$$\mathit{gap}_\pi(P) \geq \frac{m}{k+m},$$

where P is the transition kernel of k-PSS for π.

Proof. Note firstly that

$$\varrho_{k,1}(r\theta) = \exp(-\chi(\theta)r^m), \qquad r \in \mathbb{R}_+, \theta \in \mathbb{S}^{d-1},$$

and secondly that by $\chi > 0$ we have $\varrho_{k,1} \leq \varrho_{k,1}(0) = \exp(0) = 1$. For $t \in \,]0,1[$ one has

$$
\begin{aligned}
\varrho_{k,1}(r\theta) > t \quad &\Leftrightarrow \quad r^m < -\log(t)/\chi(\theta) \\
&\Leftrightarrow \quad r < (-\log(t)/\chi(\theta))^{1/m},
\end{aligned}
$$

where the first equivalence again relies on $\chi > 0$ and the second on $r \mapsto r^{1/m}$ being strictly increasing. Thus, by the polar coordinates formula (Proposition A.1),

$$
\begin{aligned}
\ell_{\varrho_{k,0},\varrho_{k,1}}(t) &= \int_{\mathbb{R}^d} \|x\|^{k-d} \mathbb{I}(\varrho_{k,1}(x) > t)\mathrm{d}x = \int_{\mathbb{S}^{d-1}} \int_0^\infty r^{k-1} \mathbb{I}(\varrho_{k,1}(r\theta) > t)\mathrm{d}r\sigma_d(\mathrm{d}\theta) \\
&= \int_{\mathbb{S}^{d-1}} \left[\frac{1}{k}r^k\right]_0^{(-\log(t)/\chi(\theta))^{1/m}} \sigma_d(\mathrm{d}\theta) = \frac{1}{k}\int_{\mathbb{S}^{d-1}} (-\log(t)/\chi(\theta))^{k/m}\sigma_d(\mathrm{d}\theta) \\
&= \frac{1}{k}\int_{\mathbb{S}^{d-1}} \chi(\theta)^{-k/m}\sigma_d(\mathrm{d}\theta) \cdot (-\log(t))^{k/m} =: C(k,\chi,m) \cdot (-\log(t))^{k/m}.
\end{aligned}
$$

From this it is clear that $\ell_{\varrho_{k,0},\varrho_{k,1}}$ is continuous and strictly decreasing on its support $]0,1[$, so it only remains to verify condition (iii) of Definition 3.2 in order to apply Theorem 3.3. To this end, we observe that the above computation shows in particular

$$\ell_{\varrho_{k,0},\varrho_{k,1}}(\exp(-s))^{m/k} = C(k,\chi,m)^{m/k}s,$$

which is linear, and therefore concave, in s. Thus, again by Proposition A.6, we have shown $\ell_{\varrho_{k,0},\varrho_{k,1}} \in \Lambda_{k/m}$, so that Theorem 3.3 applies and yields

$$\mathsf{gap}_\pi(P) \geq \frac{1}{k/m+1} = \frac{m}{k+m}. \qquad \square$$

---

[6]Note that some implicit assumptions are made on χ by assuming ϱ to be integrable. These assumptions are rather weak, however. For example, assuming χ to be uniformly lower-bounded straightforwardly yields the required integrability.

Discussion 3.8.

(a) As far as we can tell, the result in Theorem 3.7 is not closely related to any result for slice sampling from any previously published work. It appears to be a major advance in understanding the spectral gap of slice sampling, particularly USS, in rotationally asymmetric settings, as we will elaborate in the following.

(b) We note, however, that there is some overlap between the classes of distributions for which Theorems 3.5 and 3.7 apply. Namely, when setting $\chi(\theta) := c$ for some $c > 0$ and all $\theta \in \mathbb{S}^{d-1}$, as well as $\kappa := \infty$, $\phi(r) := c\,r$, and choosing $k \in \mathbb{R}_{>0}$ and $m \in \mathbb{R}_{>0}$ the same in both theorems, the target densities they consider coincide. Notably also the spectral gap estimates they provide in these cases coincide.

(c) Aside from allowing us to prove the desired result, the computation of $\ell_{\varrho_{k,0},\varrho_{k,1}}$ in the proof of Theorem 3.7 also shows that

$$\ell_{\varrho_{k,0},\varrho_{k,1}}(\exp(-s))^{1/p} = C(k,\chi,m)^{1/p} s^{k/(mp)}$$

for any $p > 0$. This is concave in s if and only if $k/(pm) \leq 1$, i.e. $p \geq k/m$. By Proposition A.6, this yields $\ell_{\varrho_{k,0},\varrho_{k,1}} \in \Lambda_p$ for $p \geq k/m$ and $\ell_{\varrho_{k,0},\varrho_{k,1}} \notin \Lambda_p$ for $p < k/m$. As both k and m are allowed to take arbitrary values in $\mathbb{R}_{>0}$, this is an example of our generalization of the statement of Theorem 3.3 from a discrete sequence of classes $(\Lambda_k)_{k\in\mathbb{N}}$ (as in [15]) to a continuous spectrum $(\Lambda_k)_{k\in\mathbb{R}_{>0}}$ (as in our Theorem 3.3) leading to a quantifiable improvement in the resulting spectral gap estimates.

(d) Again it is worthwhile to contemplate the theorem's restriction to USS, i.e. the case $k = d$. Here the theorem yields for USS applied to

$$\varrho(r\theta) = \exp(-\chi(\theta)r^m)$$

the estimate

$$\mathsf{gap}_\pi(P) \geq \frac{m}{d+m}. \tag{3.3}$$

(e) As it may not be obvious how the theorem could be applied in practice, we point out that, for USS, setting $m := 2$ and $\chi(\theta) := \frac{1}{2}\theta^T\Sigma^{-1}\theta$ for a symmetric and positive definite $\Sigma \in \mathbb{R}^{d\times d}$, one obtains

$$\varrho(r\theta) = \exp\left(-\frac{1}{2}\theta^T\Sigma^{-1}\theta \cdot r^2\right) = \exp\left(-\frac{1}{2}(r\theta)^T\Sigma^{-1}(r\theta)\right) \propto \mathcal{N}_d(r\theta; 0, \Sigma),$$

so that $\pi = \mathcal{N}_d(0,\Sigma)$. The lower bound on the spectral gap proven in this case is $2/(d+2)$. By location invariance of USS[7], the same lower bound holds for the sampler applied to any multivariate normal target $\mathcal{N}_d(a,\Sigma)$.

(f) Furthermore, by leaving $m \in \mathbb{R}_{>0}$ arbitrary, setting $\chi(\theta) := \frac{1}{2}\left(\theta^T\Sigma^{-1}\theta\right)^{m/2}$ and again using the location invariance of USS, one can cover a large class of generalized Gaussian distributions, namely all those with densities of the shape

$$\varrho(x) = \exp\left(-\frac{1}{2}\left((x-a)^T\Sigma^{-1}(x-a)\right)^{m/2}\right)$$

for an $a \in \mathbb{R}^d$ and a positive definite $\Sigma \in \mathbb{R}^{d\times d}$. The estimate obtained in these cases is (3.3).

---

[7]That is, the fact that the transition mechanism of USS does not depend on where the target distribution's high probably regions are, only how they are located relative to one another.

(g) Regarding the theorem's restriction to PSS, we note that, much like in Theorem 3.5, it is unfortunate that we only obtain a result for target densities with a pole of order $(d-1)$ at the origin. Nevertheless, we find it noteworthy that the only previously available quantitative theoretical result regarding PSS for rotationally asymmetric targets was the broadly applicable but comparatively unsharp convergence estimate in [13], which additionally does not seem to imply a spectral gap. In other words, the restriction of Theorem 3.7 to PSS is, as far as we can tell, both the first result to provide "realistic" (i.e. potentially sharp) estimates of the convergence rate (cf. (1.2)) of PSS for a class of rotationally asymmetric targets, and the first to estimate the sampler's spectral gap in such a setting, which, as explained in the introduction, has many useful implications besides convergence.

3.4 USS for multivariate t-distributions

By combining the strategy of Theorem 3.7 with the results of Theorems 2.4, Lemma 3.1 and (1.3), we can ultimately estimate the spectral gap of USS for virtually arbitrary multivariate t-distributions.

Theorem 3.9. *Let $\pi \in \mathcal{P}(\mathbb{R}^d)$ with density ϱ given by*

$$\varrho(r\theta) = \left(1 + \frac{1}{m}\chi(\theta)r^2\right)^{-(d+m)/2}, \qquad r \in \mathbb{R}_+, \theta \in \mathbb{S}^{d-1}, \tag{3.4}$$

where $m > 2$ and $\chi : \mathbb{S}^{d-1} \to \mathbb{R}_{>0}$ is measurable. Then

$$\mathsf{gap}_\pi(P_\pi) \geq 1 - \frac{d(d+m)}{(d+1)(d+m-1)}, \tag{3.5}$$

where P_π denotes the transition kernel of USS for π.

Proof. First of all, it is easily seen that the transition kernel of USS for any density η coincides, for any $c > 0$, with that of slice sampling for $\zeta := c \cdot \eta$ with factorization $\zeta_0 := c \cdot \mathbf{1}$, $\zeta_1 := \eta$. Of course the transition kernels coinciding means that the respective spectral gaps must also coincide. Furthermore, we trivially get $c \cdot \ell_{\mathbf{1},\eta} \equiv \ell_{c \cdot \mathbf{1},\eta} \equiv \ell_{\zeta_0,\zeta_1}$. Therefore, if $\ell_{\mathbf{1},\varrho} \equiv c \cdot \ell_{\mathbf{1},\eta}$, then $\ell_{\mathbf{1},\varrho} \equiv \ell_{\zeta_0,\zeta_1}$, so that by Lemma 3.1 the spectral gap of USS for ϱ coincides with that of slice sampling for ζ. Since this gap in turn coincides with that of USS for η, we ultimately obtain coinciding spectral gaps of USS for ϱ and USS for η. For proving Theorem 3.9, the relevant consequence of all this is

$$\ell_{\mathbf{1},\varrho} \propto \ell_{\mathbf{1},\eta} \quad \Rightarrow \quad \mathsf{gap}_\pi(P_\pi) = \mathsf{gap}_\nu(P_\nu), \tag{3.6}$$

where ν denotes the distribution with yet-to-be-specified density η and P_ν the transition kernel of USS for ν.

Next, note that for $r \in \mathbb{R}_+$, $\theta \in \mathbb{S}^{d-1}$, $t \in]0,1[$, we get

$$\varrho(r\theta) > t \quad \Leftrightarrow \quad 1 + \frac{1}{m}\chi(\theta)r^2 < t^{-2/(d+m)}$$
$$\Leftrightarrow \quad r < \sqrt{\frac{m}{\chi(\theta)}(t^{-2/(d+m)} - 1)}.$$

Thus, by the polar coordinates formula (Proposition A.1), and omitting some very elementary integration steps already performed in detail in other proofs,

$$\ell_{\mathbf{1},\varrho}(t) = \int_{\mathbb{R}^d} \mathbb{I}(\varrho(x) > t)\mathrm{d}x$$
$$= \int_{\mathbb{S}^{d-1}} \int_0^\infty \mathbb{I}\left(r < \sqrt{\frac{m}{\chi(\theta)}(t^{-2/(d+m)} - 1)}\right) r^{d-1}\mathrm{d}r\sigma_d(\mathrm{d}\theta)$$

$$= \frac{1}{d} \int_{\mathbb{S}^{d-1}} \left(\frac{m}{\chi(\theta)} \left(t^{-2/(d+m)} - 1\right) \right)^{d/2} \sigma_d(\mathrm{d}\theta)$$

$$= \frac{m^{d/2}}{d} \int_{\mathbb{S}^{d-1}} \chi(\theta)^{-d/2} \sigma_d(\mathrm{d}\theta) \cdot \left(t^{-2/(d+m)} - 1 \right)^{d/2}.$$

This shows that $\ell_{\mathbf{1},\varrho} \propto \ell_{\mathbf{1},\eta}$, where η is given by (3.4) with $\chi \equiv \mathbf{1}$, which is simply (2.6). Hence, (3.6) yields $\mathsf{gap}_\pi(P_\pi) = \mathsf{gap}_\nu(P_\nu)$. Finally, by Theorem 2.4 and (1.3) (which can be applied here because the constraint $m > 2$ ensures that π has finite second moment), we get

$$\mathsf{gap}_\nu(P_\nu) \geq 1 - \frac{d(d+m)}{(d+1)(d+m-1)}.$$

Combining the last two partial results yields the claim. $\qquad\square$

Discussion 3.10.

(a) By choosing $\chi(\theta) := \theta^T \Sigma^{-1} \theta$ for a positive definite matrix $\Sigma \in \mathbb{R}^{d \times d}$, choosing any $a \in \mathbb{R}^d$ and again using the location invariance of USS (cf. footnote in Discussion 3.8 (e)), one obtains the gap estimate (3.5) for USS applied to the target density

$$\varrho(x) = \left(1 + \frac{1}{m}(x-a)^T \Sigma^{-1}(x-a) \right)^{-(d+m)/2}.$$

Setting aside their normalization, the densities of this form are precisely those of arbitrary multivariate t-distributions, except that we require $m > 2$, whereas the distribution is also well-defined for $0 < m \leq 2$.

(b) Both here and in our result on USS for arbitrary multivariate Gaussians $\mathcal{N}_d(a, \Sigma)$ (cf. Discussion 3.8 (e)), the matrix Σ that causes the target density to be rotationally asymmetric does not affect the spectral gap estimate we obtain. Our analysis (specifically the computations of $\ell_{\mathbf{1},\varrho}$, combined with (3.6)) shows that this is actually a strong point of our technique, as the spectral gap of USS in these scenarios is indeed unaffected by Σ.

A Auxiliary results

A.1 Polar coordinates

We frequently make use of the following well-known result, which we refer to as the *polar coordinates formula*.

Proposition A.1. *For any function $g : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that is measurable and integrable, one has*

$$\int_{\mathbb{R}^d} g(x)\mathrm{d}x = \int_{\mathbb{S}^{d-1}} \int_0^\infty g(r\theta) r^{d-1} \mathrm{d}r \, \sigma_d(\mathrm{d}\theta),$$

where σ_d denotes the surface measure on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$.

Proof. See [16, Theorem 15.13]. $\qquad\square$

At some points, we require the following consequence of the polar coordinates formula, which, following [17], we refer to as *sampling in polar coordinates*.

Corollary A.2. *Let $\xi \in \mathcal{P}(\mathbb{R}^d)$ with density $h_\xi : \mathbb{R}^d \to \mathbb{R}_+$, then by Proposition A.1 we have*

$$\xi(A) = \int_{\mathbb{R}^d} \mathbb{1}_A(x) h_\xi(x) \mathrm{d}x = \int_{\mathbb{S}^{d-1}} \int_0^\infty \mathbb{1}_A(r\theta) h_\xi(r\theta) r^{d-1} \mathrm{d}r \, \sigma_d(\mathrm{d}\theta), \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

Thus $X \sim \xi$ can be sampled in polar coordinates as $X := R \cdot \Theta$ by sampling (R, Θ) from the joint distribution with density

$$(r, \theta) \mapsto h_\xi(r\theta)r^{d-1}\mathbb{1}_{\mathbb{R}_+}(r)$$

w.r.t. $\lambda_1 \otimes \sigma_d$, where λ_1 denotes the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

A.2 Some elementary results

Lemma A.3. *Let $\phi :]0, \kappa[\to \mathbb{R}$ be as in Definition 2.1. Denote the domain of ϕ by $D_\phi :=]0, \kappa[$ and the image by $I_\phi := \phi(D_\phi)$. Then one has $I_\phi =]\inf \phi, \sup \phi[$, where*

$$\inf \phi := \inf_{r \in D_\phi} \phi(r) = \lim_{r \searrow 0} \phi(r)$$

$$\sup \phi := \sup_{r \in D_\phi} \phi(r) = \lim_{r \nearrow \kappa} \phi(r).$$

Furthermore, ϕ maps bijectively onto I_ϕ and its inverse $\phi^{-1} : I_\phi \to D_\phi$ is again strictly increasing.

Proof. Because ϕ is a convex function defined on an open interval, it is continuous, which shows that its image is again an open interval. That this interval must have the claimed boundaries is clear from the fact that ϕ is strictly increasing.

Observe that the strict monotonicity of ϕ also guarantees its injectivity and note that every function is surjective onto its image. Hence ϕ maps bijectively $D_\phi \to I_\phi$ and there exists an inverse function $\phi^{-1} : I_\phi \to D_\phi$.

Let now $s_1, s_2 \in I_\phi$ with $s_1 > s_2$. Set $r_i := \phi^{-1}(s_i)$, $i = 1, 2$, so that in particular $\phi(r_i) = s_i$. Then, since ϕ is strictly increasing and $\phi(r_1) = s_1 > s_2 = \phi(r_2)$ by assumption, we must have $r_1 > r_2$. This yields $\phi^{-1}(s_1) = r_1 > r_2 = \phi^{-1}(s_2)$, so ϕ^{-1} is again strictly increasing. $\qquad\square$

Lemma A.4. *Let $\phi : D_\phi \to I_\phi$ be as specified in Definition 2.1 and analyzed in Lemma A.3. Define the extension $\widehat{\phi}^{-1} : \mathbb{R}_{>0} \to \mathbb{R}$ of its inverse ϕ^{-1} to $\mathbb{R}_{>0}$ by*

$$\widehat{\phi}^{-1}(s) := \begin{cases} \phi^{-1}(s) & s \in I_\phi =]\inf \phi, \sup \phi[, \\ \kappa & s \in [\sup \phi, \infty[. \end{cases}$$

Note that assuming $\kappa = \infty$ leads to $\sup \phi = \infty$ (because a non-decreasing convex function on $\mathbb{R}_{>0}$ must either be constant or unbounded and we require ϕ to be strictly increasing, thus preventing it from being constant) and consequently $\widehat{\phi}^{-1} \equiv \phi^{-1}$. In particular, $\widehat{\phi}^{-1}$ cannot attain the value ∞.

For this function $\widehat{\phi}^{-1}$ and all $r, \tilde{r} \in D_\phi$, $v \in \mathbb{R}_+$ one has

$$\left| \widehat{\phi}^{-1}(\phi(r) + v) - \widehat{\phi}^{-1}(\phi(\tilde{r}) + v) \right| \le |r - \tilde{r}|.$$

An illustration of this property can be found in Figure 1.

Proof. We use arguments very similar to those in [8]. Fix $r, \tilde{r} \in D_\phi$ and $v \in \mathbb{R}_+$. Assume w.l.o.g. that $r > \tilde{r}$ and set

$$r' := \widehat{\phi}^{-1}(\phi(r) + v),$$
$$\tilde{r}' := \widehat{\phi}^{-1}(\phi(\tilde{r}) + v),$$

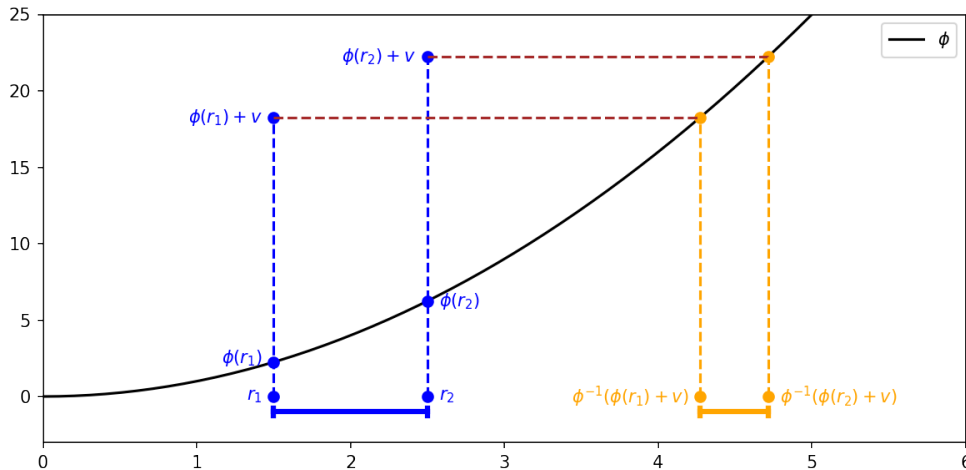s.t. what we need to prove becomes $|r' - \tilde{r}'| \le |r - \tilde{r}|$.

Figure 1: Illustration of the property shown in Lemma A.4 for $\kappa = \infty$. Two points $r_1, r_2 \in \,]0, \kappa[$ are fixed, ϕ is evaluated at each of them and the resulting function values are shifted upwards by the same value v, resulting in $\phi(r_1) + v$ and $\phi(r_2) + v$. The expressions $r_1' := \phi^{-1}(\phi(r_1) + v)$ and $r_2' := \phi^{-1}(\phi(r_2) + v)$ then correspond to the locations where ϕ reaches the shifted values. The lemma's claim is simply that the thick orange line, corresponding to $|r_1' - r_2'|$, is at most as long as the thick blue line, corresponding to $|r_1 - r_2|$.

Since ϕ^{-1} is strictly increasing, $\widehat{\phi}^{-1}$ is still non-decreasing, so $r' \geq \tilde{r}'$ and

$$r' = \widehat{\phi}^{-1}(\phi(r) + v) \geq \widehat{\phi}^{-1}(\phi(r)) = \phi^{-1}(\phi(r)) = r, \tag{A.1}$$

$$\tilde{r}' = \widehat{\phi}^{-1}(\phi(\tilde{r}) + v) \geq \widehat{\phi}^{-1}(\phi(\tilde{r})) = \phi^{-1}(\phi(\tilde{r})) = \tilde{r}. \tag{A.2}$$

We now prove the claim by distinguishing four cases w.r.t. the values of $r', \tilde{r}'$ and $\sup \phi$, given that $r' \geq \tilde{r}'$ (as observed above) and $r', \tilde{r}' \leq \kappa$ (by definition of $\widehat{\phi}^{-1}$):

1. Suppose $r' = \tilde{r}'$. Then $|r' - \tilde{r}'| \leq |r - \tilde{r}|$ holds trivially.

2. Suppose $\tilde{r}' < r' < \kappa$. Then $r', \tilde{r}' \in D_\phi$ and therefore $\phi(r') = \phi(r) + v$ as well as $\phi(\tilde{r}') = \phi(\tilde{r}) + v$. Furthermore, it is well-known that ϕ being convex implies that

$$S_\phi(r_1, r_2) := \frac{\phi(r_1) - \phi(r_2)}{r_1 - r_2}$$

is non-decreasing in r_1 for fixed r_2 and vice-versa. Thus

$$\frac{\phi(r') - \phi(\tilde{r}')}{r' - \tilde{r}'} = S_\phi(r', \tilde{r}') \overset{(A.1)}{\geq} S_\phi(r, \tilde{r}') \overset{(A.2)}{\geq} S_\phi(r, \tilde{r}) = \frac{\phi(r) - \phi(\tilde{r})}{r - \tilde{r}}$$

$$= \frac{(\phi(r) + v) - (\phi(\tilde{r}) + v)}{r - \tilde{r}} = \frac{\phi(r') - \phi(\tilde{r}')}{r - \tilde{r}},$$

which, because all occurring numerators and denominators are positive, implies $r' - \tilde{r}' \leq r - \tilde{r}$. For the general case, i.e. without assuming $r > \tilde{r}$, this translates precisely to the desired relation $|r' - \tilde{r}'| \leq |r - \tilde{r}|$.

3. Suppose $\tilde{r}' < \kappa = r'$ and $\sup \phi = \infty$. Then $I_\phi = \,]\inf \phi, \infty[$, meaning that $\phi(r) + v$ and $\phi(\tilde{r}) + v$ are contained in I_ϕ and thus we again have $r', \tilde{r}' \in D_\phi$. Since the analysis of the second case relied solely on this property, it applies to this case as well.

4. Suppose $\tilde{r}' < \kappa = r'$ and $\sup \phi < \infty$. Note that r' attaining the value κ must mean $\kappa < \infty$. Hence we may continuously extend ϕ to the domain $]0, \kappa] = D_\phi \cup \{\kappa\} \subset \mathbb{R}_{>0}$, denoting the extension $\widehat{\phi} : \ ]0, \kappa] \to \mathbb{R}$, i.e. we set

$$\widehat{\phi}(\hat{r}) := \begin{cases} \phi(\hat{r}) & \hat{r} \in D_\phi, \\ \sup \phi & \hat{r} = \kappa. \end{cases}$$

By construction, $\widehat{\phi}$ is still convex and inverts the restriction of $\widehat{\phi}^{-1}$ to $]\inf \phi, \sup \phi]$. Observe that $\tilde{r}' < \kappa$ again yields $\tilde{r}' \in D_\phi$ and thus $\phi(\tilde{r}') = \phi(\tilde{r}) + v$ and that, by definition of the involved quantities, $r' = \kappa$ implies $\phi(r) + v \geq \sup \phi$. Equipped with all of these relations, we can employ the same approach as in the second case and obtain

$$\frac{\widehat{\phi}(r') - \widehat{\phi}(\tilde{r}')}{r' - \tilde{r}'} = S_{\widehat{\phi}}(r', \tilde{r}') \overset{(A.1)}{\geq} S_{\widehat{\phi}}(r, \tilde{r}') \overset{(A.2)}{\geq} S_{\widehat{\phi}}(r, \tilde{r}) = \frac{\widehat{\phi}(r) - \widehat{\phi}(\tilde{r})}{r - \tilde{r}}$$

$$= \frac{\phi(r) - \phi(\tilde{r})}{r - \tilde{r}} = \frac{(\phi(r) + v) - (\phi(\tilde{r}) + v)}{r - \tilde{r}} = \frac{(\phi(r) + v) - \phi(\tilde{r}')}{r - \tilde{r}}$$

$$\geq \frac{\sup \phi - \phi(\tilde{r}')}{r - \tilde{r}} = \frac{\widehat{\phi}(r') - \widehat{\phi}(\tilde{r}')}{r - \tilde{r}},$$

again showing $r' - \tilde{r}' \leq r - \tilde{r}$, so in the general case $|r' - \tilde{r}'| \leq |r - \tilde{r}|$. $\quad\square$

A.3 Auxiliary results for proving spectral gap estimates

Lemma A.5. *Suppose $h : I_1 \to I_2$ is a strictly monotone, continuous function mapping between open intervals $I_1, I_2 \subseteq \mathbb{R}$. Then h maps bijectively onto its image, w.l.o.g. I_2, and its inverse $h^{-1} : I_2 \to I_1$ has the same monotonicity property. Furthermore,*

- *if h is increasing, then it is convex if and only if h^{-1} is concave*
- *if h is decreasing, then it is convex if and only if h^{-1} is convex.*

Proof. See [15, Lemma A.4]. $\quad\square$

Proposition A.6. *For any $k \in \mathbb{R}_{>0}$, given all the other conditions, condition (iii) of Definition 3.2 is equivalent to concavity of $s \mapsto \ell(\exp(-s))^{1/k}$ restricted to the domain*

$$D := \ ]-\log \sup\{t \in \mathbb{R}_{>0} \mid \ell(t) > 0\}, \infty[.$$

Proof. For $k \in \mathbb{N}$ this was shown in [15, Proposition A.5]. As the proof only requires k to be positive, it does not need to be modified to cover all the cases $k \in \mathbb{R}_{>0}$. $\quad\square$

B Numerical support

We can give some numerical support for the sharpness of our spectral gap estimates in Theorem 3.5 by comparing them with heuristic approximations of the gaps from experimental data. The heuristic we consider is derived by identifying two different notions of effective sample sizes.

Let P be a transition kernel with invariant distribution π and $\mathsf{gap}_\pi(P) > 0$. Denote by $(X_m)_{m \in \mathbb{N}_0}$ a Markov chain with transition kernel P that is initialized with its invariant distribution, i.e. $X_0 \sim \pi$.

Further denoting by $\mathsf{IAT}_g(P)$ the integrated autocorrelation time of P w.r.t. some reasonable summary function $g : \mathbb{R}^d \to \mathbb{R}$, one would commonly define the effective sample size as
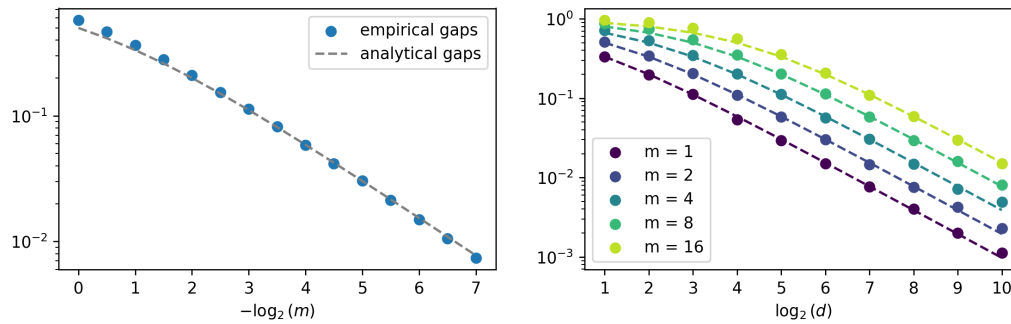
$$n_{\text{eff}} := \frac{n}{\mathsf{IAT}_g(P)}.$$

Figure 2: Comparisons of empirical spectral gaps (scattered points), computed according to (B.1), and spectral gap estimates provided by Theorem 3.5 (dashed lines). The figure on the left shows these values for PSS applied to $\varrho(r\theta) = r^{1-d}\exp(-r^m)$ with parameter choices $m = 2^0, 2^{-0.5}, \ldots, 2^{-7}$. The IATs for the empirical gaps were computed from the sample log radii, i.e. with $g(x) := \log\|x\|$, which renders the results independent of the sample space dimension d. The right figure similarly shows empirical gaps and gap estimates for USS applied to $\varrho(r\theta) = \exp(-r^m)$ with parameter values $m = 2^0, \ldots, 2^4$ in dimensions $d = 2^1, \ldots, 2^{10}$. Here the IATs for the empirical gaps were computed from the sample radii, i.e. with $g(x) := \|x\|$.

On the other hand it is known from [14, Corollary 3.37] that, for a suitably normalized integrand, the mean squared Monte Carlo integration error of the truncated chain $(X_m)_{m \leq n}$ asymptotically (i.e. as $n \to \infty$) behaves exactly as

$$\frac{2 - \mathsf{gap}_\pi(P)}{n \cdot \mathsf{gap}_\pi(P)} = \frac{1}{n}\left(\frac{2}{\mathsf{gap}_\pi(P)} - 1\right).$$

As this quantity would be $1/n$ with i.i.d. sampling, we may view its reciprocal as another notion of effective sample size. By informally identifying the two, we obtain the heuristic

$$\frac{\mathsf{IAT}_g(P)}{n} \approx \frac{1}{n}\left(\frac{2}{\mathsf{gap}_\pi(P)} - 1\right) \quad \Leftrightarrow \quad \mathsf{gap}_\pi(P) \approx \frac{2}{\mathsf{IAT}_g(P) + 1} \tag{B.1}$$

for the approximate computation of $\mathsf{gap}_\pi(P)$. This enables us to estimate spectral gaps from sample runs through their (also approximated) IATs.

The results of some example applications of this technique are shown in Figure 2. They show the empirical gaps to match our gap estimates very well, which, if one believes the heuristic to be good, suggests the gap estimates to be very sharp. We note however that, in our experience, the approach only works this well if the spectral gaps in question are small – if the proven gap estimates are large, the empirical gaps are consistently even larger, which causes discrepancies between the two.

References

[1] Julian Besag and Peter J. Green, *Spatial statistics and Bayesian computation*, Journal of the Royal Statistical Society Series B **55** (1993), no. 1, 25–37. MR1210422

[2] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier, *Markov Chains*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, 2018. MR3889011

[3] James M. Flegal and Galin L. Jones, *Batch means and spectral variance estimators in Markov chain Monte Carlo*, The Annals of Statistics **38** (2010), no. 2, 1034–1070. MR2604704

[4] Michael Habeck, Mareike Hasenpflug, Shantanu Kodgirwar, and Daniel Rudolf, *Geodesic slice sampling on the sphere*, arXiv preprint arXiv:2301.08056, 2023.

[5] Claude P. Kipnis and Srinivasa R. S. Varadhan, *Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions*, Communications in Mathematical Physics **104** (1986), no. 1, 1–19. MR0834478

[6] Krzysztof Latuszyński and Daniel Rudolf, *Convergence of hybrid slice sampling via spectral gap*, arXiv preprint arXiv:1409.2709, 2014.

[7] Iain Murray, Ryan P. Adams, and David MacKay, *Elliptical slice sampling*, in: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 9, 2010, pp. 541–548.

[8] Viacheslav Natarovskii, Daniel Rudolf, and Björn Sprungk, *Quantitative spectral gap estimate and Wasserstein contraction of simple slice sampling*, The Annals of Applied Probability **31** (2021), no. 2, 806–825. MR4254496

[9] Radford M. Neal, *Slice sampling*, The Annals of Statistics **31** (2003), no. 3, 705–767. MR1994729

[10] Erich Novak and Daniel Rudolf, *Computation of expectations by Markov chain Monte Carlo methods*, Lecture Notes in Computational Science and Engineering **102** (2014). MR3329348

[11] Yann Ollivier, *Ricci curvature of Markov chains on metric spaces*, Journal of Functional Analysis **256** (2009), no. 3, 810–864. MR2484937

[12] Gareth O. Roberts and Jeffrey S. Rosenthal, *Convergence of slice sampler Markov chains*, Journal of the Royal Statistical Society Series B **61** (1999), no. 3, 643–660. MR1707866

[13] Gareth O. Roberts and Jeffrey S. Rosenthal, *The polar slice sampler*, Stochastic Models **18** (2002), no. 2, 257–280. MR1904830

[14] Daniel Rudolf, *Explicit error bounds for Markov chain Monte Carlo*, Dissertationes Mathematicae **485** (2012), 1–93. MR2977521

[15] Daniel Rudolf and Philip Schär, *Dimension-independent spectral gap of polar slice sampling*, Statistics and Computing **34** (2024), no. 1, 20. MR4662468

[16] René L. Schilling, *Measures, Integrals and Martingales*, Cambridge University Press, 2005. MR2200059

[17] Philip Schär, Michael Habeck, and Daniel Rudolf, *Gibbsian polar slice sampling*, in: Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 202, 2023, pp. 30204–30223.

[18] Cédric Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, American Mathematical Society, 2003. MR1964483