Check for
updates

# Curve Based Approximation of Measures on Manifolds by Discrepancy Minimization

Martin Ehler[1] · Manuel Gräf[2] · Sebastian Neumayer[3] · Gabriele Steidl[3]

## Abstract

The approximation of probability measures on compact metric spaces and in particular on Riemannian manifolds by atomic or empirical ones is a classical task in approximation and complexity theory with a wide range of applications. Instead of point measures we are concerned with the approximation by measures supported on Lipschitz curves. Special attention is paid to push-forward measures of Lebesgue measures on the unit interval by such curves. Using the discrepancy as distance between measures, we prove optimal approximation rates in terms of the curve's length and Lipschitz constant. Having established the theoretical convergence rates, we are interested in the numerical minimization of the discrepancy between a given probability measure and the set of push-forward measures of Lebesgue measures on the unit interval by Lipschitz curves. We present numerical examples for measures on the 2- and 3-dimensional torus, the 2-sphere, the rotation group on $\mathbb{R}^3$ and the Grassmannian of all 2-dimensional linear subspaces of $\mathbb{R}^4$. Our algorithm of choice is a conjugate gradient method on these manifolds, which incorporates second-order information. For efficient gradient and Hessian evaluations within the algorithm, we approximate the given measures by truncated Fourier series and use fast Fourier transform techniques on these manifolds.

**Keywords** Approximation of measures · Curves · Discrepancies · Fourier methods · Manifolds · Non-convex optimization · Quadrature rules · Sampling theory

**Mathematics Subject Classification** 28E99 · 49Q99 · 65D99

## 1 Introduction

The approximation of probability measures by atomic or empirical ones based on their discrepancies is a well examined problem in approximation and complexity theory

Communicated by Alan Edelman.

Extended author information available on the last page of the article

Springer

[59,62,67] with a wide range of applications, e.g., in the derivation of quadrature rules and in the construction of designs. Recently, discrepancies were also used in image processing for dithering [46,72,77], i.e., for representing a gray-value image by a finite number of black dots, and in generative adversarial networks [28].

Besides discrepancies, Optimal Transport (OT) and in particular Wasserstein distances have emerged as powerful tools to compare probability measures in recent years, see [24,81] and the references therein. In fact, so-called Sinkhorn divergences, which are computationally much easier to handle than OT, are known to interpolate between OT and discrepancies [30]. For the sample complexity of Sinkhorn divergences we refer to [37]. The rates for approximating probability measures by atomic or empirical ones with respect to Wasserstein distances depend on the dimension of the underlying spaces, see [21,58]. In contrast, approximation rates based on discrepancies can be given independently of the dimension [67], i.e., they do not suffer from the curse of dimensionality. Additionally, we should keep in mind that the computation of discrepancies does not involve a minimization problem, which is a major drawback of OT and Sinkhorn divergences. Moreover, discrepancies admit a simple description in Fourier domain and hence the use of fast Fourier transforms is possible, leading to better scalability than the aforementioned methods.

Instead of point measures, we are interested in approximations with respect to measures supported on curves. More precisely, we consider push-forward measures of probability measures $\omega \in \mathcal{P}([0, 1])$ by Lipschitz curves of bounded speed, with special focus on absolutely continuous measures $\omega = \rho\lambda$ and the Lebesgue measure $\omega = \lambda$. In this paper, we focus on approximation with respect to discrepancies. For related results on quadrature and approximation on manifolds, we refer to [31,47,64, 65] and the references therein. An approximation model based on the 2-Wasserstein distance was proposed in [61]. That work exploits completely different techniques than ours both in the theoretical and numerical part. Finally, we want to point out a relation to principal curves which are used in computer science and graphics for approximating distributions approximately supported on curves [49,50,50,55,57]. For the interested reader, we further comment on this direction of research in Remark 3 and in the conclusions. Next, we want to motivate our framework by numerous potential applications:

– In MRI sampling [11,17], it is desirable to construct sampling curves with short sampling times (short curve) and high reconstruction quality. Unfortunately, these requirements usually contradict each other and finding a good trade-off is necessary. Experiments demonstrating the power of this novel approach on a real-world scanner are presented in [60].
– For laser engraving [61] and 3D printing [20], we require nozzle trajectories based on our (continuous) input densities. Compared to the approach in [20], where points given by Llyod's method are connected as a solution of the TSP (traveling salesman problem), our method jointly selects the points and the corresponding curve. This avoids the necessity of solving a TSP, which can be quite costly, although efficient approximations exist. Further, it is not obvious that the fixed initial point approximation is a good starting point for constructing a curve.

– The model can be used for wire sculpture creation [2]. In view of this, our numerical experiment presented in Fig. 5 can be interpreted as a building plan for a wire sculpture of the Spock head, namely of a 2D surface. Clearly, the approach can be also used to create images similar to TSP Art [54], where images are created from points by solving the corresponding TSP.

– In a more manifold related setting, the approach can be used for grand tour computation on $\mathcal{G}_{2,4}$ [5], see also our numerical experiment in Fig. 11. More technical details are provided in the corresponding section.

Our contribution is two-fold. On the theoretical side, we provide estimates of the approximation rates in terms of the maximal speed of the curve. First, we prove approximation rates for general probability measures on compact Ahlfors $d$-regular length spaces $\mathbb{X}$. These spaces include many compact sets in the Euclidean space $\mathbb{R}^d$, e.g., the unit ball or the unit cube as well as $d$-dimensional compact Riemannian manifolds without boundary. The basic idea consists in combining the known convergence rates for approximation by atomic measures with cost estimates for the traveling salesman problem. As for point measures, the approximation rate $L^{d/(2d-2)} \leq L^{-1/2}$ for general $\omega \in \mathcal{P}([0,1])$ and $L^{d/(3d-2)} \leq L^{-1/3}$ for $\omega = \lambda$ in terms of the maximal Lipschitz constant (speed) $L$ of the curves does not crucially depend on the dimension of $\mathbb{X}$. In particular, the second estimate improves a result given in [18] for the torus.

If the measures fulfill additional smoothness properties, these estimates can be improved on compact, connected, $d$-dimensional Riemannian manifolds without boundary. Our results are formulated for absolutely continuous measures (with respect to the Riemannian measure) having densities in the Sobolev space $H^s(\mathbb{X})$, $s > d/2$. In this setting, the optimal approximation rate becomes roughly speaking $L^{-s/(d-1)}$. Our proofs rely on a general result of Brandolini et al. [13] on the quadrature error achievable by integration with respect to a measure that exactly integrates all eigenfunctions of the Laplace–Beltrami with eigenvalues smaller than a fixed number. Hence, we need to construct measures supported on curves that fulfill the above exactness criterion. More precisely, we construct such curves for the $d$ dimensional torus $\mathbb{T}^d$, the spheres $\mathbb{S}^d$, the rotation group SO(3) and the Grassmannian $\mathcal{G}_{2,4}$.

On the numerical side, we are interested in finding (local) minimizers of discrepancies between a given continuous measure and those from the set of push-forward measures of the Lebesgue measure by bounded Lipschitz curves. This problem is tackled numerically on $\mathbb{T}^2$, $\mathbb{T}^3$, $\mathbb{S}^2$ as well as SO(3) and $\mathcal{G}_{2,4}$ by switching to the Fourier domain. The minimizers are computed using the method of conjugate gradients (CG) on manifolds, which incorporates second order information in form of a multiplication by the Hessian. Thanks to the approach in the Fourier domain, the required gradients and the calculations involving the Hessian can be performed efficiently by fast Fourier transform techniques at arbitrary nodes on the respective manifolds. Note that in contrast to our approach, semi-continuous OT minimization relies on Laguerre tessellations [41], which are not available in the required form on the 2-sphere, SO(3) or $\mathcal{G}_{2,4}$.

This paper is organized as follows: In Sect. 2 we give the necessary preliminaries on probability measures. In particular, we introduce the different sets of measures supported on Lipschitz curves that are used for the approximation. Note that measures

supported on continuous curves of finite length can be equivalently characterized by push-forward measures of probability measures by Lipschitz curves. Section 3 provides the notation on reproducing kernel Hilbert spaces and discrepancies including their representation in the Fourier domain. Section 4 contains our estimates of the approximation rates for general given measures and different approximation spaces of measures supported on curves. Following the usual lines in approximation theory, we are then concerned with the approximation of absolutely continuous measures with density functions lying in Sobolev spaces. Our main results on the approximation rates of smoother measures are contained in Sect. 5, where we distinguish between the approximation with respect to the push-forward of general measures $\omega \in \mathcal{P}[0, 1]$, absolute continuous measures and the Lebesgue measure on [0, 1]. In Sect. 6 we formulate our numerical minimization problem. Our numerical algorithms of choice are briefly described in Sect. 7. For a comprehensive description of the algorithms on the different manifolds, we refer to respective papers. Section 8 contains numerical results demonstrating the practical feasibility of our findings. Conclusions are drawn in Sect. 9. Finally, Appendix A briefly introduces the different manifolds $\mathbb{X}$ used in our numerical examples together with the Fourier representation of probability measures on $\mathbb{X}$.

## 2 Probability Measures and Curves

In this section, the basic notation on measure spaces is provided, see [3,32], with focus on probability measures supported on curves. At this point, let us assume that

> $\mathbb{X}$ is a compact metric space endowed with a bounded non-negative Borel mea-
> sure $\sigma_{\mathbb{X}} \in \mathcal{M}(\mathbb{X})$ such that $\mathrm{supp}(\sigma_{\mathbb{X}}) = \mathbb{X}$. Further, we denote the metric by
> $\mathrm{dist}_{\mathbb{X}}$.

Additional requirements on $\mathbb{X}$ are added along the way and notations are explained below. By $\mathcal{B}(\mathbb{X})$ we denote the Borel $\sigma$-algebra on $\mathbb{X}$ and by $\mathcal{M}(\mathbb{X})$ the linear space of all finite signed Borel measures on $\mathbb{X}$, i.e., the space of all $\mu \colon \mathcal{B}(\mathbb{X}) \to \mathbb{R}$ satisfying $\mu(\mathbb{X}) < \infty$ and for any sequence $(B_k)_{k \in \mathbb{N}} \subset \mathcal{B}(\mathbb{X})$ of pairwise disjoint sets the relation $\mu(\bigcup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} \mu(B_k)$. The *support of a measure* $\mu$ is the closed set

$$\mathrm{supp}(\mu) := \big\{ x \in \mathbb{X} : B \subset \mathbb{X} \text{ open}, \ x \in B \implies \mu(B) > 0 \big\}.$$

For $\mu \in \mathcal{M}(\mathbb{X})$ the total variation measure is defined by

$$|\mu|(B) := \sup\bigg\{ \sum_{k=1}^{\infty} |\mu(B_k)| : \bigcup_{k=1}^{\infty} B_k = B, \ B_k \text{ pairwise disjoint} \bigg\}.$$

With the norm $\|\mu\|_{\mathcal{M}} = |\mu|(\mathbb{X})$ the space $\mathcal{M}(\mathbb{X})$ becomes a Banach space. By $\mathcal{C}(\mathbb{X})$ we denote the Banach space of continuous real-valued functions on $\mathbb{X}$ equipped with the norm $\|\varphi\|_{\mathcal{C}(\mathbb{X})} := \max_{x \in \mathbb{X}} |\varphi(x)|$. The space $\mathcal{M}(\mathbb{X})$ can be identified via Riesz' theorem with the dual space of $\mathcal{C}(\mathbb{X})$ and the weak-* topology on $\mathcal{M}(\mathbb{X})$ gives rise to

the *weak convergence of measures*, i.e., a sequence $(\mu_k)_k \subset \mathcal{M}(\mathbb{X})$ converges *weakly* to $\mu$ and we write $\mu_k \rightharpoonup \mu$, if

$$\lim_{k \to \infty} \int_{\mathbb{X}} \varphi \, \mathrm{d}\mu_k = \int_{\mathbb{X}} \varphi \, \mathrm{d}\mu \quad \text{for all } \varphi \in \mathcal{C}(\mathbb{X}).$$

For a non-negative, finite measure $\mu$, let $L^p(\mathbb{X}, \mu)$ be the Banach space (of equivalence classes) of complex-valued functions with norm

$$\|f\|_{L^p(\mathbb{X}, \mu)} = \left( \int_{\mathbb{X}} |f|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}} < \infty.$$

By $\mathcal{P}(\mathbb{X})$ we denote the space of Borel probability measures on $\mathbb{X}$, i.e., non-negative Borel measures with $\mu(\mathbb{X}) = 1$. This space is *weakly compact*, i.e., compact with respect to the topology of weak convergence. We are interested in the approximation of measures in $\mathcal{P}(\mathbb{X})$ by probability measures supported on points and curves in $\mathbb{X}$. To this end, we associate with $x \in \mathbb{X}$ a probability measure $\delta_x$ with values $\delta_x(B) = 1$ if $x \in B$ and $\delta_x(B) = 0$ otherwise.

The *atomic probability measures* at $N$ points are defined by

$$\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X}) := \left\{ \sum_{k=1}^{N} w_k \delta_{x_k} : (x_k)_{k=1}^{N} \in \mathbb{X}^N, \ (w_k)_{k=1}^{N} \in [0, 1]^N, \ \sum_{k=1}^{N} w_k = 1 \right\}.$$

In other words, $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$ is the collection of probability measures, whose support consists of at most $N$ points. Further restriction to equal mass distribution leads to the *empirical probability measures* at $N$ points denoted by

$$\mathcal{P}_N^{\mathrm{emp}}(\mathbb{X}) := \left\{ \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} : (x_k)_{k=1}^{N} \in \mathbb{X}^N \right\}.$$

In this work, we are interested in the approximation by measures having their support on curves. Let $\mathcal{C}([a, b], \mathbb{X})$ denote the set of closed, continuous curves $\gamma : [a, b] \to \mathbb{X}$. Although our presented experiments involve solely closed curves, some applications might require open curves. Hence, we want to point out that all of our approximation results still hold without this requirement. Upper bounds would not get worse and we have not used the closedness for the lower bounds on the approximation rates. The *length of* a curve $\gamma \in \mathcal{C}([a, b], \mathbb{X})$ is given by

$$\ell(\gamma) := \sup_{\substack{a \le t_0 \le \dots \le t_n \le b \\ n \in \mathbb{N}}} \sum_{k=1}^{n} \mathrm{dist}_{\mathbb{X}}\big(\gamma(t_k), \gamma(t_{k-1})\big).$$

If $\ell(\gamma) < \infty$, then $\gamma$ is called *rectifiable*. By reparametrization, see [48, Thm. 3.2], the image of any rectifiable curve in $\mathcal{C}([a, b], \mathbb{X})$ can be derived from the set of closed

*Lipschitz continuous curves*

$$\mathrm{Lip}(\mathbb{X}):=\big\{\gamma \in \mathcal{C}([0,1],\mathbb{X}) : \exists L \in \mathbb{R} \text{ with } \mathrm{dist}_{\mathbb{X}}\big(\gamma(s),\gamma(t)\big)\leq L|s-t| \ \forall s,t \in [0,1]\big\}.$$

The *speed of a curve* $\gamma \in \mathrm{Lip}(\mathbb{X})$ is defined a.e. by the metric derivative

$$|\dot{\gamma}|(t) := \lim_{s\to t} \frac{\mathrm{dist}_{\mathbb{X}}\big(\gamma(s),\gamma(t)\big)}{|s-t|}, \qquad t \in [0,1],$$

cf. [4, Sec. 1.1]. The optimal Lipschitz constant $L = L(\gamma)$ of a curve $\gamma$ is given by $L(\gamma) = \||\dot{\gamma}|\|_{\infty([0,1])}$. For a constant speed curve it holds $L(\gamma) = \ell(\gamma)$.

We aim to approximate measures in $\mathcal{P}(\mathbb{X})$ from those of the subset

$$\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X}) := \big\{\nu \in \mathcal{P}(\mathbb{X}) : \exists \gamma \in \mathcal{C}([a,b],\mathbb{X}),\ \mathrm{supp}(\nu) \subset \gamma([a,b]),\ \ell(\gamma) \leq L\big\}. \tag{1}$$

This space is quite large and in order to define further meaningful subsets, we derive an equivalent formulation in terms of push-forward measures. For $\gamma \in \mathcal{C}([0,1],\mathbb{X})$, the *push-forward* $\gamma_*\omega \in \mathcal{P}(\mathbb{X})$ of a probability measure $\omega \in \mathcal{P}([0,1])$ is defined by $\gamma_*\omega(B) := \omega(\gamma^{-1}(B))$ for $B \in \mathcal{B}(\mathbb{X})$. We directly observe $\mathrm{supp}(\gamma_*\omega) = \gamma(\mathrm{supp}(\omega))$. By the following lemma, $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ consists of the push-forward of measures in $\mathcal{P}([0,1])$ by constant speed curves.

**Lemma 1** *The space $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ in (1) is equivalently given by*

$$\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X}) = \big\{\gamma_*\omega : \gamma \in \mathrm{Lip}(\mathbb{X}) \text{ has constant speed } L(\gamma) \leq L,\ \omega \in \mathcal{P}([0,1])\big\}. \tag{2}$$

**Proof** Let $\nu \in \mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ as in (1). If $\mathrm{supp}(\nu)$ consists of a single point $x \in \mathbb{X}$ only, then the constant curve $\gamma \equiv x$ pushes forward an arbitrary $\delta_t$ for $t \in [a,b]$, which shows that $\nu$ is contained in (2).

Suppose that $\mathrm{supp}(\nu)$ contains at least two distinct points and let $\gamma \in \mathcal{C}([a,b],\mathbb{X})$ with $\mathrm{supp}(\nu) \subset \gamma([a,b])$ and $\ell(\gamma) < \infty$. According to [16, Prop. 2.5.9], there exists a continuous curve $\tilde{\gamma} \in \mathrm{Lip}(\mathbb{X})$ with constant speed $\ell(\gamma)$ and a continuous nondecreasing function $\varphi: [a,b] \to [0,1]$ with $\gamma = \tilde{\gamma} \circ \varphi$. Now, define $f: \mathbb{X} \to [0,1]$ by $f(x) := \min\{\tilde{\gamma}^{-1}(x)\}$. This function is measurable, since for every $t \in [0,1]$ it holds that

$$\big\{x \in \mathbb{X} : f(x) \leq t\big\} = \big\{x \in \mathbb{X} : \min\{\tilde{\gamma}^{-1}(x)\} \leq t\big\} = \tilde{\gamma}([0,t])$$

is compact. Due to $\mathrm{supp}(\nu) \subset \tilde{\gamma}([0,1])$, we can define $\omega := f_*\nu \in \mathcal{P}([0,1])$. By construction, $\omega$ satisfies $\tilde{\gamma}_*\omega(B) = \omega(\tilde{\gamma}^{-1}(B)) = \nu(f^{-1} \circ \tilde{\gamma}^{-1}(B)) = \nu(B)$ for all $B \in \mathcal{B}(\mathbb{X})$. This concludes the proof. □

The set $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ contains $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$ if $L$ is sufficiently large compared to $N$ and $\mathbb{X}$ is sufficiently nice, cf. Sect. 4. It is reasonable to ask for more restrictive sets

of approximation measures, e.g., when $\omega \in \mathcal{P}([0, 1])$ is assumed to be absolutely continuous. For the Lebesgue measure $\lambda$ on $[0, 1]$, we consider

$$\mathcal{P}_L^{\text{a-curv}}(\mathbb{X}) := \left\{ \gamma_* \omega : \gamma \in \text{Lip}(\mathbb{X}), \ L(\gamma) \leq L, \ \omega = \rho\lambda \in \mathcal{P}([0, 1]), \ L(\rho) \leq L \right\}.$$

In the literature [18,61], the special case of push-forward of the Lebesgue measure $\omega = \lambda$ on $[0, 1]$ by Lipschitz curves in $\mathbb{T}^d$ was discussed and successfully used in certain applications [11,17]. Therefore, we also consider approximations from

$$\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X}) := \left\{ \gamma_* \lambda : \gamma \in \text{Lip}(\mathbb{X}), \ L(\gamma) \leq L \right\}.$$

It is obvious that our probability spaces related to curves are nested,

$$\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X}) \subset \mathcal{P}_L^{\text{a-curv}}(\mathbb{X}) \subset \mathcal{P}_L^{\text{curv}}(\mathbb{X}).$$

Hence, one may expect that establishing good approximation rates is most difficult for $\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})$ and easier for $\mathcal{P}_L^{\text{curv}}(\mathbb{X})$.

## 3 Discrepancies and RKHS

The aim of this section is to introduce the way we quantify the distance ("discrepancy") between two probability measures. To this end, choose a continuous, symmetric function $K : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ that is positive definite, i.e., for any finite number $n \in \mathbb{N}$ of points $x_j \in \mathbb{X}$, $j = 1, \ldots, n$, the relation

$$\sum_{i,j=1}^{n} a_i a_j K(x_i, x_j) \geq 0$$

is satisfied for all $a_j \in \mathbb{R}$, $j = 1, \ldots, n$. We know by Mercer's theorem [23,63,76] that there exists an orthonormal basis $\{\phi_k : k \in \mathbb{N}\}$ of $L^2(\mathbb{X}, \sigma_{\mathbb{X}})$ and non-negative coefficients $(\alpha_k)_{k \in \mathbb{N}} \in \ell_1$ such that $K$ has the Fourier expansion

$$K(x, y) = \sum_{k=0}^{\infty} \alpha_k \phi_k(x) \overline{\phi_k(y)} \tag{3}$$

with absolute and uniform convergence of the right-hand side. If $\alpha_k > 0$ for some $k \in \mathbb{N}_0$, the corresponding function $\phi_k$ is continuous. Every function $f \in L^2(\mathbb{X}, \sigma_{\mathbb{X}})$ has a Fourier expansion

$$f = \sum_{k=0}^{\infty} \hat{f}_k \phi_k, \quad \hat{f}_k := \int_{\mathbb{X}} f \overline{\phi_k} \, d\sigma_{\mathbb{X}}.$$

The kernel $K$ gives rise to a *reproducing kernel Hilbert space* (RKHS). More precisely, the function space

$$H_K(\mathbb{X}) := \left\{ f \in L^2(\mathbb{X}, \sigma_{\mathbb{X}}) : \sum_{k=0}^{\infty} \alpha_k^{-1} |\hat{f}_k|^2 < \infty \right\}$$

equipped with the inner product and the corresponding norm

$$\langle f, g \rangle_{H_K(\mathbb{X})} = \sum_{k=0}^{\infty} \alpha_k^{-1} \hat{f}_k \overline{\hat{g}_k}, \qquad \|f\|_{H_K(\mathbb{X})} = \sqrt{\langle f, f \rangle}_{H_K(\mathbb{X})} \tag{4}$$

forms a Hilbert space with reproducing kernel, i.e.,

$$K(x, \cdot) \in H_K(\mathbb{X}) \qquad \text{for all } x \in \mathbb{X},$$
$$f(x) = \langle f, K(x, \cdot) \rangle_{H_K(\mathbb{X})} \qquad \text{for all } f \in H_K(\mathbb{X}), \ x \in \mathbb{X}.$$

Note that $f \in H_K(\mathbb{X})$ implies $\hat{f}_k = 0$ if $\alpha_k = 0$, in which case we make the convention $\alpha_k^{-1} \hat{f}_k = 0$ in (4). The space $H_K(\mathbb{X})$ is the closure of the linear span of $\{K(x_j, \cdot) : x_j \in \mathbb{X}\}$ with respect to the norm (4), and $H_K(\mathbb{X})$ is continuously embedded in $C(\mathbb{X})$. In particular, the point evaluations in $H_K(\mathbb{X})$ are continuous.

The *discrepancy* $\mathscr{D}_K(\mu, \nu)$ is defined as the dual norm on $H_K(\mathbb{X})$ of the linear operator $T \colon H_K(\mathbb{X}) \to \mathbb{C}$ with $\varphi \mapsto \int_{\mathbb{X}} \varphi \, \mathrm{d}(\mu - \nu)$:

$$\mathscr{D}_K(\mu, \nu) = \max_{\|\varphi\|_{H_K(\mathbb{X})} \leq 1} \left| \int_{\mathbb{X}} \varphi \, \mathrm{d}(\mu - \nu) \right|, \tag{5}$$

see [40,67]. Note that this looks similar to the 1-Wasserstein distance, where the space of test functions consists of Lipschitz continuous functions and is larger. Since

$$\int_{\mathbb{X}} \varphi \, \mathrm{d}\mu = \int_{\mathbb{X}} \langle \varphi, K(x, \cdot) \rangle_{H_K(\mathbb{X})} \, \mathrm{d}\mu(x) = \left\langle \varphi, \int_{\mathbb{X}} K(x, \cdot) \, \mathrm{d}\mu(x) \right\rangle_{H_K(\mathbb{X})},$$

we obtain by Riesz's representation theorem

$$\max_{\|\varphi\|_{H_K(\mathbb{X})} \leq 1} \int_{\mathbb{X}} \varphi \, \mathrm{d}\mu = \left\| \int_{\mathbb{X}} K(x, \cdot) \, \mathrm{d}\mu(x) \right\|_{H_K(\mathbb{X})},$$

which yields by Fubini's theorem, (3), (4) and symmetry of $K$ that

$$\mathscr{D}_K^2(\mu, \nu) = \iint_{\mathbb{X} \times \mathbb{X}} K \, \mathrm{d}\mu \, \mathrm{d}\mu - 2 \iint_{\mathbb{X} \times \mathbb{X}} K \, \mathrm{d}\mu \, \mathrm{d}\nu + \iint_{\mathbb{X} \times \mathbb{X}} K \, \mathrm{d}\nu \, \mathrm{d}\nu \tag{6}$$

$$= \sum_{k=0}^{\infty} \alpha_k |\hat{\mu}_k - \hat{\nu}_k|^2, \tag{7}$$

where the *Fourier coefficients* of $\mu, \nu \in \mathcal{P}(\mathbb{X})$ are well-defined for $k$ with $\alpha_k \neq 0$ by

$$\hat{\mu}_k := \int_{\mathbb{X}} \overline{\phi_k} \, \mathrm{d}\mu, \qquad \hat{\nu}_k := \int_{\mathbb{X}} \overline{\phi_k} \, \mathrm{d}\nu.$$

**Remark 1** The Fourier coefficients $\hat{\mu}_k$ and $\hat{\nu}_k$ depend on both $K$ and $\sigma_{\mathbb{X}}$, but the identity (6) shows that $\mathscr{D}_K(\mu, \nu)$ only depends on $K$. Thus, our approximation rates do not depend on the choice of $\sigma_{\mathbb{X}}$. On the other hand, our numerical algorithms in Sect. 7 depend on $\phi_k$ and hence on the choice of $\sigma_{\mathbb{X}}$.

If $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$ as $n \to \infty$, then also $\mu_n \otimes \nu_n \rightharpoonup \mu \otimes \nu$. Therefore, the continuity of $K$ implies that $\lim_{n \to \infty} \mathscr{D}_K(\mu_n, \nu_n) = \mathscr{D}_K(\mu, \nu)$, so that $\mathscr{D}_K$ is continuous with respect to weak convergence in both arguments. Thus, for any weakly compact subset $P \subset \mathcal{P}(\mathbb{X})$, the infimum

$$\inf_{\nu \in P} \mathscr{D}_K(\mu, \nu)$$

is actually a minimum. All of the subsets introduced in the previous section are weakly compact.

**Lemma 2** *The sets* $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$, $\mathcal{P}_N^{\mathrm{emp}}(\mathbb{X})$, $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$, $\mathcal{P}_L^{\mathrm{a\text{-}curv}}(\mathbb{X})$, *and* $\mathcal{P}_L^{\mathrm{\check{}}\text{-}curv}(\mathbb{X})$ *are weakly compact.*

**Proof** It is well-known that $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$ and $\mathcal{P}_N^{\mathrm{emp}}(\mathbb{X})$ are weakly compact.

We show that $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ is weakly compact. In view of (2), let $(\gamma_k)_{k \in \mathbb{N}}$ be Lipschitz curves with constant speed $L(\gamma_k) \leq L$ and $(\omega_k)_{k \in \mathbb{N}} \subset \mathcal{P}([0, 1])$. Since $\mathcal{P}([0, 1])$ is weakly compact, we can extract a subsequence $(\omega_{k_j})_{j \in \mathbb{N}}$ with weak limit $\hat{\omega} \in \mathcal{P}([0, 1])$. Now, we observe that $\mathrm{dist}_{\mathbb{X}}(\gamma_{k_j}(s), \gamma_{k_j}(t)) \leq L|s - t|$ for all $j \in \mathbb{N}$. Since $\mathbb{X}$ is compact, the Arzelà–Ascoli theorem implies that there exists a subsequence of $(\gamma_{k_j})_{j \in \mathbb{N}}$ which converges uniformly towards $\hat{\gamma} \in \mathrm{Lip}(\mathbb{X})$ with $L(\hat{\gamma}) \leq L$. Then, $\hat{\nu} := \hat{\gamma}_* \hat{\omega}$ fulfills $\mathrm{supp}(\hat{\nu}) \subset \hat{\gamma}([0, 1])$, so that $\hat{\nu} \in \mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ by (1). Thus, $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ is weakly compact.

The proof for $\mathcal{P}_L^{\mathrm{a\text{-}curv}}(\mathbb{X})$ and $\mathcal{P}_L^{\mathrm{\check{}}\text{-}curv}(\mathbb{X})$ is analogous and hence omitted. □

**Remark 2** (Discrepancies and Convolution Kernels) Let $\mathbb{X} = \mathbb{T}^d := \mathbb{R}^d / \mathbb{Z}^d$ be the torus and $h \in \mathcal{C}(\mathbb{T}^d)$ be a function with Fourier series

$$h(x) = \sum_{k \in \mathbb{Z}^d} \hat{h}_k \mathrm{e}^{2\pi \mathrm{i} \langle k, x \rangle}, \quad \hat{h}_k := \int_{\mathbb{T}^d} h(x) \mathrm{e}^{-2\pi \mathrm{i} \langle k, x \rangle} \, \mathrm{d}\sigma_{\mathbb{T}^d}(x),$$

which converges in $L^2(\mathbb{T}^d)$ so that $\sum_k |\hat{h}_k|^2 < \infty$.

Assume that $\hat{h}_k \neq 0$ for all $k \in \mathbb{Z}^d$. We consider the special Mercer kernel

$$K(x, y) := \sum_{k \in \mathbb{Z}^d} |\hat{h}_k|^2 \mathrm{e}^{2\pi \mathrm{i} \langle k, x - y \rangle} = \sum_{k \in \mathbb{Z}^d} |\hat{h}_k|^2 \cos\big(2\pi \langle k, x - y \rangle\big)$$

with associated discrepancy $\mathscr{D}_h$ via (6), i.e., $\phi_k(x) = \mathrm{e}^{2\pi \mathrm{i} \langle k, x \rangle}$, $\alpha_k = |\hat{h}_k|^2$, $k \in \mathbb{Z}^d$ in (3). The convolution of $h$ with $\mu \in \mathcal{M}(\mathbb{T}^d)$ is the function $h * \mu \in C(\mathbb{T}^d)$ defined by

$$(h * \mu)(x) := \int_{\mathbb{T}^d} h(x - y) \, d\mu(y).$$

By the convolution theorem for Fourier transforms it holds $\widehat{(h * \mu)}_k = \hat{h}_k \hat{\mu}_k, k \in \mathbb{Z}^d$, and we obtain by Parseval's identity for $\mu, \nu \in \mathcal{M}(\mathbb{T}^d)$ and (7) that

$$\|h * (\mu - \nu)\|_{L^2(\mathbb{T}^d)}^2 = \left\| \left( \hat{h}_k \left( \hat{\mu}_k - \hat{\nu}_k \right) \right)_{k \in \mathbb{Z}^d} \right\|_{\ell_2}^2 = \sum_{k \in \mathbb{Z}^d} |\hat{h}_k|^2 |\hat{\mu}_k - \hat{\nu}_k|^2 = \mathscr{D}_h^2(\mu, \nu).$$

In image processing, metrics of this kind were considered in [18,33,77].

**Remark 3** (Relations to Principal Curves)  A similar concept, sharing the common theme of "a curve which passes through the *middle* of a distribution" with the intention of our paper, is that of principle curves. The notion of principal curves has been developed in a statistical framework and was successfully applied in statistics and machine learning, see [38,55,57]. The idea is to generalize the concept of principal components with just one direction to so-called self-consistent (principal) curves. In the seminal paper [49], the authors showed that these principal curves $\gamma$ are critical points of the energy functional

$$E(\gamma, \mu) = \int_{\mathbb{X}} \|x - \text{proj}_\gamma(x)\|_2^2 \, d\mu(x), \tag{8}$$

where $\mu$ is a given probability measure on $\mathbb{X}$ and $\text{proj}_\gamma(x) = \text{argmin}_{y \in \gamma} \|x - y\|_2$ is a projection of a point $x \in \mathbb{X}$ on $\gamma$. This notion has also been generalized to Riemannian manifolds in [50], see also [57] for an application on the sphere. Further investigation of principal curves in the plane, cf. [27], showed that self-consistent curves are not (local) minimizers, but saddle points of (8). Moreover, the existence of such curves is established only for certain classes of measures, such as elliptical ones. By additionally constraining the length of curves minimizing (8), these unfavorable effects were eliminated, cf. [55]. In comparison to the objective (8), the discrepancy (6) averages for fixed $x \in \mathbb{X}$ the distance encoded by $K$ to any point on $\gamma$, instead of averaging over the squared minimal distances to $\gamma$.

## 4 Approximation of General Probability Measures

Given $\mu \in \mathcal{P}(\mathbb{X})$, the estimates[1]

$$\min_{\nu \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_N^{\text{emp}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim N^{-\frac{1}{2}}, \tag{9}$$

---

[1]  We use the symbols $\lesssim$ and $\gtrsim$ to indicate that the corresponding inequalities hold up to a positive constant factor on the respective right-hand side. The notation $\sim$ means that both relations $\lesssim$ and $\gtrsim$ hold. The dependence of the constants on other parameters shall either be explicitly stated or clear from the context.

are well-known, cf. [43, Cor. 2.8]. Here, the constant hidden in $\lesssim$ depends on $\mathbb{X}$ and $K$ but is independent of $\mu$ and $N \in \mathbb{N}$. In this section, we are interested in approximation rates with respect to measures supported on curves.

Our approximation rates for $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ are based on those for $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$ combined with estimates for the traveling salesman problem (TSP). Let $\mathrm{TSP}_{\mathbb{X}}(N)$ denote the worst case minimal cost tour in a fully connected graph $G$ of $N$ arbitrary nodes represented by $x_1, \ldots, x_N \in \mathbb{X}$ and edges with cost $\mathrm{dist}_{\mathbb{X}}(x_i, x_j)$, $i, j = 1, \ldots, N$. Similarly, let $\mathrm{MST}_{\mathbb{X}}(N)$ denote the worst case cost of the minimal spanning tree of $G$. To derive suitable estimates, we require that $\mathbb{X}$ is *Ahlfors d-regular* (sometimes also called Ahlfors-David $d$-regular), i.e., there exists $0 < d < \infty$ such that

$$\sigma_{\mathbb{X}}\big(B_r(x)\big) \sim r^d, \quad \text{for all } x \in \mathbb{X}, \quad 0 < r \le \mathrm{diam}(\mathbb{X}), \tag{10}$$

where $B_r(x) = \{y \in \mathbb{X} : \mathrm{dist}_{\mathbb{X}}(x, y) \le r\}$ and the constants in $\sim$ do not depend on $x$ or $r$. Note that $d$ is not required to be an integer and turns out to be the Hausdorff dimension. For $\mathbb{X}$ being the unit cube the following lemma was proved in [75].

**Lemma 3** *If $\mathbb{X}$ is a compact Ahlfors d-regular metric space, then there is a constant $0 < C_{\mathrm{TSP}} < \infty$ depending on $\mathbb{X}$ such that*

$$\mathrm{TSP}_{\mathbb{X}}(N) \le C_{\mathrm{TSP}} N^{1-\frac{1}{d}}.$$

***Proof*** Using (10) and the same covering argument as in [74, Lem. 3.1], we see that for every choice $x_1, \ldots, x_N \in \mathbb{X}$, there exist $i \ne j$ such that $\mathrm{dist}_{\mathbb{X}}(x_i, x_j) \lesssim N^{-1/d}$, where the constant depends on $\mathbb{X}$.

Let $S = \{x_1, \ldots, x_N\}$ be an arbitrary selection of $N$ points from $\mathbb{X}$. First, we choose $x_i$ and $x_j$ with $\mathrm{dist}_{\mathbb{X}}(x_i, x_j) \le cN^{-1/d}$. Then, we form a minimal spanning tree $T$ of $S \setminus \{x_i\}$ and augment the tree by adding the edge between $x_i$ and $x_j$. This construction provides us with a spanning tree and hence we can estimate $\mathrm{MST}_{\mathbb{X}}(N) \le \mathrm{MST}_{\mathbb{X}}(N-1) + cN^{-1/d}$. Iterating the argument, we deduce

$$\mathrm{MST}_{\mathbb{X}}(N) \lesssim N^{1-\frac{1}{d}},$$

cf. [75]. Finally, the standard relation $\mathrm{TSP}_{\mathbb{X}}(N) \le 2\,\mathrm{MST}_{\mathbb{X}}(N)$ for edge costs satisfying the triangular inequality concludes the proof. $\qquad\square$

To derive a curve in $\mathbb{X}$ from a minimal cost tour in the graph, we require the additional assumption that $\mathbb{X}$ is a *length space*, i.e., a metric space with

$$\mathrm{dist}_{\mathbb{X}}(x, y) = \inf\big\{\ell(\gamma) : \gamma \text{ a continuous curve that connects } x \text{ and } y\big\},$$

cf. [15,16]. Thus, for the rest of this section, we are assuming that

$\mathbb{X}$ is a compact Ahlfors $d$-regular length space.

In this case, Lemma 3 yields the next proposition.

**Proposition 1** *For a compact, Ahlfors d-regular length space $\mathbb{X}$ it holds $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X}) \subset \mathcal{P}_{C_{\mathrm{TSP}} N^{1-1/d}}^{\mathrm{curv}}(\mathbb{X})$.*

**Proof** The Hopf-Rinow Theorem for metric measure spaces, see [15, Chap. I.3] and [16, Thm. 2.5.28], yields that every pair of points $x, y \in \mathbb{X}$ can be connected by a geodesic, i.e., there is $\gamma \in \mathrm{Lip}(\mathbb{X})$ with constant speed and $\ell(\gamma|_{[s,t]}) = \mathrm{dist}_{\mathbb{X}}(\gamma(s), \gamma(t))$ for all $0 \leq s \leq t \leq 1$. Thus, for any pair $x, y \in \mathbb{X}$, there is a constant speed curve $\gamma_{x,y} \in \mathrm{Lip}(\mathbb{X})$ of length $\ell(\gamma_{x,y}) = \mathrm{dist}_{\mathbb{X}}(x, y)$ with $\gamma_{x,y}(0) = x$, $\gamma_{x,y}(1) = y$, cf. [16, Rem. 2.5.29]. For $\mu_N \in \mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$, let $\{x_1, \ldots, x_N\} = \mathrm{supp}(\mu_N)$. The minimal cost tour in Lemma 3 leads to a curve $\gamma \in \mathrm{Lip}(\mathbb{X})$, so that $\mu_N = \gamma_* \omega \in \mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$ for an appropriate measure $\omega \in \mathcal{P}_N^{\mathrm{atom}}([0, 1])$.                     $\square$

By Proposition 1 we can transfer approximation rates from $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})$ to $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$.

**Theorem 1** *For $\mu \in \mathcal{P}(\mathbb{X})$, it holds with a constant depending on $\mathbb{X}$ and $K$ that*

$$\min_{\nu \in \mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{d}{2d-2}}.$$

**Proof** Choose $\alpha = \frac{d-1}{d}$. For $L$ large enough, set $N := \lfloor (L/C_{\mathrm{TSP}})^{\frac{1}{\alpha}} \rfloor \in \mathbb{N}$, so that we observe $\mathcal{P}_N^{\mathrm{atom}}(\mathbb{X}) \subset \mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$. According to (9), we obtain

$$\min_{\nu \in \mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_N^{\mathrm{atom}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim N^{-\frac{1}{2}} \lesssim L^{-\frac{1}{2\alpha}}.$$

$\square$

Next, we derive approximation rates for $\mathcal{P}_L^{\mathrm{a\text{-}curv}}(\mathbb{X})$ and $\mathcal{P}_L^{\check{}\mathrm{-curv}}(\mathbb{X})$.

**Theorem 2** *For $\mu \in \mathcal{P}(\mathbb{X})$, we have with a constant depending on $\mathbb{X}$ and $K$ that*

$$\min_{\nu \in \mathcal{P}_L^{\mathrm{a\text{-}curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_L^{\check{}\mathrm{-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{d}{3d-2}}. \tag{11}$$

**Proof** Let $\alpha = \frac{d-1}{d}$, $d \geq 2$. For $L$ large enough, set $N := \lfloor L^{\frac{2}{2\alpha+1}} / \mathrm{diam}(\mathbb{X}) \rfloor \in \mathbb{N}$. By (9), there is a set of points $\{x_1, \ldots, x_N\} \subset \mathbb{X}$ such that

$$\mathscr{D}_K(\mu, \nu_N) \lesssim N^{-\frac{1}{2}} \lesssim L^{-\frac{1}{2\alpha+1}}, \qquad \nu_N := \frac{1}{N} \sum_{j=1}^N \delta_{x_j}. \tag{12}$$

Let these points be ordered as a solution of the corresponding TSP. Set $x_0 := x_N$ and $\tau_i := \mathrm{dist}_{\mathbb{X}}(x_i, x_{i+1})/L$, $i = 0, \ldots, N - 1$. Note that

$$N \leq L^{\frac{2}{2\alpha+1}} / \mathrm{diam}(\mathbb{X}) \leq L / \mathrm{dist}_{\mathbb{X}}(x_i, x_{i+1}),$$

so that $\tau_i \leq N^{-1}$ for all $i = 0, \ldots, N-1$. We construct a closed curve $\gamma_L : [0, 1] \to \mathbb{X}$ that rests in each $x_i$ for a while and then rushes from $x_i$ to $x_{i+1}$. As in the proof of

Proposition 1, $\mathbb{X}$ being a compact length space enables us to choose $\gamma_i \in \mathrm{Lip}(\mathbb{X})$ with $\gamma_i(0) = x_i$, $\gamma_i(1) = x_{i+1}$ and $L(\gamma_i) = \mathrm{dist}_{\mathbb{X}}(x_i, x_{i+1})$. For $i = 0, \ldots, N_L - 1$, we define

$$\gamma_L(t) := \begin{cases} x_i & \text{for } t \in \left[\frac{i}{N}, \frac{i+1}{N} - \tau_i\right), \\ \gamma_i\left(\frac{1}{\tau_i}\left(t - \frac{i+1}{N} + \tau_i\right)\right) & \text{for } t \in \left[\frac{i+1}{N} - \tau_i, \frac{i+1}{N}\right). \end{cases}$$

By construction, $L(\gamma_L)$ is bounded by $\min_i d(x_i, x_{i+1})\tau_i^{-1} \leq L$. Defining the measure $\nu := (\gamma_L)_* \lambda \in \mathcal{P}_L^{\text{-curv}}(\mathbb{X})$, the related discrepancy can be estimated by

$$\mathscr{D}_K(\mu, \nu) = \sup_{\|\varphi\|_{H_K(\mathbb{X})} \leq 1} \left| \int_{\mathbb{X}} \varphi \, d\mu - \int_0^1 \varphi \circ \gamma_L \, d\lambda \right|$$

$$\leq \mathscr{D}_K(\mu, \nu_N) + \sup_{\|\varphi\|_{H_K(\mathbb{X})} \leq 1} \sum_{i=0}^{N-1} \left( \tau_i |\varphi(x_i)| + \left| \int_{\frac{i+1}{N} - \tau_i}^{\frac{i+1}{N}} \varphi \circ \gamma_L \, d\lambda \right| \right).$$

The relation (12) yields $\mathscr{D}_K(\mu, \nu_N) \leq C L^{-\frac{1}{2\alpha+1}}$ with some constant $C > 0$. Since for $\varphi \in H_K(\mathbb{X})$ it holds $\|\varphi\|_{L^\infty(\mathbb{X})} \leq C_K \|\varphi\|_{H_K(\mathbb{X})}$ with $C_K := \sup_{x \in \mathbb{X}} \sqrt{K(x,x)}$, we finally obtain by Lemma 3

$$\mathscr{D}_K(\mu, \nu) \leq C L^{-\frac{1}{2\alpha+1}} + 2 C_K \sum_{i=0}^{N-1} \tau_i \leq C L^{-\frac{1}{2\alpha+1}} + 2 C_K C_{\text{TSP}} \frac{N^\alpha}{L}$$

$$\leq \left( C + 2 C_K C_{\text{TSP}}/\operatorname{diam}(\mathbb{X}) \right) L^{-\frac{1}{2\alpha+1}}.$$

$\square$

Note that many compact sets in $\mathbb{R}^d$ are compact Ahlfors $d$-regular length spaces with respect to the Euclidean metric and the normalized Lebesgue measure such as the unit ball or the unit cube. Moreover many compact connected manifolds with or without boundary satisfy these conditions. All assumptions in this section are indeed satisfied for $d$-dimensional connected, compact Riemannian manifolds without boundary equipped with the Riemannian metric and the normalized Riemannian measure. The latter setting is studied in the subsequent section to refine our investigations on approximation rates.

**Remark 4** For $\mathbb{X} = \mathbb{T}^d$ with $d \in \mathbb{N}$, the estimate

$$\min_{\nu \in \mathcal{P}_L^{\text{-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{1}{d}}. \tag{13}$$

was derived in [18] provided that $K$ satisfies an additional Lipschitz condition, where the constant in (13) depends on $d$ and $K$. The rate coincides with our rate in (11) for $d = 2$ and is worse for higher dimensions as $\frac{d}{3d-2} > \frac{1}{3}$ for all $d \geq 3$.

## 5 Approximation of Probability Measures Having Sobolev Densities

To study approximation rates in more detail, we follow the standard strategy in approximation theory and take additional smoothness properties into account. We shall therefore consider $\mu$ with a density satisfying smoothness requirements. To define suitable smoothness spaces, we make additional structural assumptions on $\mathbb{X}$. Throughout the remaining part of this work, we suppose that

> $\mathbb{X}$ is a $d$-dimensional connected, compact Riemannian manifold without boundary equipped with the *Riemannian metric* dist$_{\mathbb{X}}$ and the *normalized Riemannian measure* $\sigma_{\mathbb{X}}$.

In the first part of this section, we introduce the necessary background on Sobolev spaces and derive general lower bounds for the approximation rates. Then, we focus on upper bounds in the rest of the section. So far, we only have general upper bounds for $\mathcal{P}_L^{\text{curv}}(\mathbb{X})$. In case of the smaller spaces $\mathcal{P}_L^{\text{a-curv}}(\mathbb{X})$ and $\mathcal{P}_L^{\text{-curv}}(\mathbb{X})$, we have to restrict to special manifolds $\mathbb{X}$ in order to obtain bounds. For a better overview, all theorems related to approximation rates are named accordingly.

### 5.1 Sobolev Spaces and Lower Bounds

In order to define a smoothness class of functions on $\mathbb{X}$, let $-\Delta$ denote the (negative) Laplace–Beltrami operator on $\mathbb{X}$. It is self-adjoint on $L^2(\mathbb{X}, \sigma_{\mathbb{X}})$ and has a sequence of positive, non-decreasing eigenvalues $(\lambda_k)_{k \in \mathbb{N}}$ (with multiplicities) with a corresponding orthonormal complete system of smooth eigenfunctions $\{\phi_k : k \in \mathbb{N}\}$. Every function $f \in L^2(\mathbb{X}, \sigma_{\mathbb{X}})$ has a Fourier expansion

$$f = \sum_{k=0}^{\infty} \hat{f}(k)\phi_k, \qquad \hat{f}(k) := \int_{\mathbb{X}} f\overline{\phi_k} \, d\sigma_{\mathbb{X}}.$$

The *Sobolev space* $H^s(\mathbb{X})$, $s > 0$, is the set of all functions $f \in L^2(\mathbb{X}, \sigma_{\mathbb{X}})$ with distributional derivative $(I - \Delta)^{s/2} f \in L^2(\mathbb{X}, \sigma_{\mathbb{X}})$ and norm

$$\|f\|_{H^s(\mathbb{X})} := \|(I - \Delta)^{s/2} f\|_{L^2(\mathbb{X}, \sigma_{\mathbb{X}})} = \left( \sum_{k=0}^{\infty} (1 + \lambda_k)^s |\hat{f}(k)|^2 \right)^{\frac{1}{2}}.$$

For $s > d/2$, the space $H^s(\mathbb{X})$ is continuously embedded into the space of Hölder continuous functions of degree $s - d/2$, and every function $f \in H^s(\mathbb{X})$ has a uniformly convergent Fourier series, see [70, Thm. 5.7]. Actually, $H^s(\mathbb{X})$, $s > d/2$, is a RKHS with reproducing kernel

$$K(x, y) := \sum_{k=0}^{\infty} (1 + \lambda_k)^{-s} \phi_k(x) \overline{\phi_k(y)}.$$

Hence, the discrepancy $\mathscr{D}_K(\mu, \nu)$ satisfies (5) with $H_K(\mathbb{X}) = H^s(\mathbb{X})$. Clearly, each kernel of the above form with coefficients having the same decay as $(1 + \lambda_k)^{-s}$ for $k \to \infty$ gives rise to a RKHS that coincides with $H^s(\mathbb{X})$ with an equivalent norm. Appendix A contains more details of the above discussion for the torus $\mathbb{T}^d$, the sphere $\mathbb{S}^d$, the special orthogonal group SO(3) and the Grassmannian $\mathcal{G}_{k,d}$.

Now, we are in the position to establish lower bounds on the approximation rates. Again, we want to remark that our results still hold if we drop the requirement that the approximating curves are closed.

**Theorem 3** (Lower bound) *For $s > d/2$ suppose that $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Assume that $\mu$ is absolutely continuous with respect to $\sigma_{\mathbb{X}}$ with a continuous density $\rho$. Then, there are constants depending on $\mathbb{X}$, $K$, and $\rho$ such that*

$$N^{-\frac{s}{d}} \lesssim \min_{\nu \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_N^{\text{emp}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu),$$

$$L^{-\frac{s}{d-1}} \lesssim \min_{\nu \in \mathcal{P}_L^{\text{curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_L^{\text{a-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_L^{\text{c-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu).$$

**Proof** The proof is based on the construction of a suitable fooling function to be used in (5) and follows [13, Thm. 2.16]. There exists a ball $B \subset \mathbb{X}$ with $\rho(x) \geq \epsilon = \epsilon(B, \rho)$ for all $x \in B$ and $\sigma_{\mathbb{X}}(B) > 0$, which is chosen as the support of the constructed fooling functions. We shall verify that for every $\nu \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})$ there exists $\varphi \in H^s(\mathbb{X})$ such that $\varphi$ vanishes on supp($\nu$) but

$$\int_B \varphi \, d\mu \gtrsim \|\varphi\|_{H^s(\mathbb{X})} N^{-\frac{s}{d}}, \tag{14}$$

where the constant depends on $\mathbb{X}$, $K$, and $\rho$. For small enough $\delta$ we can choose $2N$ disjoint balls in $B$ with diameters $\delta N^{-1/d}$, see also [39]. For $\nu \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})$, there are $N$ of these balls that do not intersect with supp($\nu$). By putting together bump functions supported on each of the $N$ balls, we obtain a non-negative function $\varphi$ supported in $B$ that vanishes on supp($\nu$) and satisfies (14), with a constant that depends on $\epsilon$, cf. [13, Thm. 2.16]. This yields

$$\left| \int_{\mathbb{X}} \varphi \, d\mu - \int_{\mathbb{X}} \varphi \, d\nu \right| = \int_B \varphi \, d\mu \gtrsim \|\varphi\|_{H^s(\mathbb{X})} N^{-\frac{s}{d}}.$$

The inequality for $\mathcal{P}_L^{\text{curv}}(\mathbb{X})$ is derived in a similar way. Given a continuous curve $\gamma \colon [0, 1] \to \mathbb{X}$ of length $L$, choose $N$ such that $L \leq \delta N N^{-1/d}$. By taking half of the radius of the above balls, there are $2N$ pairwise disjoint balls of radius $\frac{\delta}{2} N^{-1/d}$ contained in $B$ with pairwise distances at least $\delta N^{-1/d}$. Any curve of length $\delta N N^{-1/d}$ intersects at most $N$ of those balls. Hence, there are $N$ balls of radius $\frac{\delta}{2} N^{-1/d}$ that do not intersect supp($\gamma$). As above, this yields a fooling function $\varphi$ satisfying (14), which ends the proof. $\qquad \square$

## 5.2 Upper Bounds for $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$

In this section, we derive upper bounds that match the lower bounds in Theorem 3 for $\mathcal{P}_L^{\mathrm{curv}}(\mathbb{X})$. Our analysis makes use of the following theorem, which was already proved for $\mathbb{X} = \mathbb{S}^d$ in [51].

**Theorem 4** *[13, Thm. 2.12] Assume that $\nu_r \in \mathcal{P}(\mathbb{X})$ provides an exact quadrature for all eigenfunctions $\varphi_k$ of the Laplace–Beltrami operator with eigenvalues $\lambda_k \leq r^2$, i.e.,*

$$\int_{\mathbb{X}} \varphi_k \, \mathrm{d}\sigma_{\mathbb{X}} = \int_{\mathbb{X}} \varphi_k \, \mathrm{d}\nu_r. \tag{15}$$

*Then, it holds for every function $f \in H^s(\mathbb{X})$, $s > d/2$, that there is a constant depending on $\mathbb{X}$ and $s$ with*

$$\left| \int_{\mathbb{X}} f \, \mathrm{d}\sigma_{\mathbb{X}} - \int_{\mathbb{X}} f \, \mathrm{d}\nu_r \right| \lesssim r^{-s} \|f\|_{H^s(\mathbb{X})}.$$

For our estimates it is important that the number of eigenfunctions of the Laplace–Beltrami operator on $\mathbb{X}$ belonging to eigenvalues with $\lambda_k \leq r^2$ is of order $r^d$, see [19, Chap. 6.4] and [52, Thm. 17.5.3, Cor. 17.5.8]. This is known as Weyl's estimates on the spectrum of an elliptic operator. For some special manifolds, the eigenfunctions are explicitly given in the appendix. In the following lemma, the result from Theorem 4 is rewritten in terms of discrepancies and generalized to absolutely continuous measures with densities $\rho \in H^s(\mathbb{X})$.

**Lemma 4** *For $s > d/2$ suppose that $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms and that $\nu_r \in \mathcal{P}(\mathbb{X})$ satisfies (15). Let $\mu \in \mathcal{P}(\mathbb{X})$ be absolutely continuous with respect to $\sigma_{\mathbb{X}}$ with density $\rho \in H^s(\mathbb{X})$. For sufficiently large $r$, the measures $\tilde{\nu}_r := \frac{\rho}{\beta_r} \nu_r \in \mathcal{P}(\mathbb{X})$ with $\beta_r := \int_{\mathbb{X}} \rho \, \mathrm{d}\nu_r$ are well defined and there is a constant depending on $\mathbb{X}$ and $K$ with*

$$\mathscr{D}_K\big(\mu, \tilde{\nu}_r\big) \lesssim \|\rho\|_{H^s(\mathbb{X})} r^{-s}.$$

**Proof** Note that $H^s(\mathbb{X})$ is a Banach algebra with respect to addition and multiplication [22], in particular, for $f, g \in H^s(\mathbb{X})$ we have $fg \in H^s(\mathbb{X})$ with

$$\|fg\|_{H^s(\mathbb{X})} \leq \|f\|_{H^s(\mathbb{X})} \|g\|_{H^s(\mathbb{X})}. \tag{16}$$

By Theorem 4, we obtain for all $\varphi \in H^s(\mathbb{X})$ that

$$\left| \int_{\mathbb{X}} \varphi \rho \, \mathrm{d}\sigma_{\mathbb{X}} - \int_{\mathbb{X}} \varphi \rho \, \mathrm{d}\nu_r \right| \lesssim r^{-s} \|\varphi \, \rho\|_{H^s(\mathbb{X})} \lesssim r^{-s} \|\varphi\|_{H^s(\mathbb{X})} \|\rho\|_{H^s(\mathbb{X})}. \tag{17}$$

In particular, this implies for $\varphi \equiv 1$ that

$$\left| 1 - \beta_r \right| \lesssim r^{-s} \|\rho\|_{H^s(\mathbb{X})}. \tag{18}$$

Then, application of the triangle inequality results in

$$\left| \int_{\mathbb{X}} \varphi \, d\mu - \int_{\mathbb{X}} \varphi \, d\tilde{\nu}_r \right| \leq \left| \int_{\mathbb{X}} \varphi \, d\mu - \int_{\mathbb{X}} \varphi \rho \, d\nu_r \right| + \left| \int_{\mathbb{X}} \varphi \rho \frac{\beta_r - 1}{\beta_r} \, d\nu_r \right|.$$

According to (17), the first summand is bounded by $\lesssim r^{-s} \|\varphi\|_{H^s(\mathbb{X})} \|\rho\|_{H^s(\mathbb{X})}$. It remains to derive matching bounds on the second term. Hölder's inequality yields

$$\left| \int_{\mathbb{X}} \varphi \rho \frac{\beta_r - 1}{\beta_r} \, d\nu_r \right| \lesssim \|\varphi\|_{L^\infty(\mathbb{X})} |\beta_r - 1| \lesssim \|\varphi\|_{H^s(\mathbb{X})} r^{-s} \|\rho\|_{H^s(\mathbb{X})},$$

where the last inequality is due to $H^s(\mathbb{X}) \hookrightarrow L^\infty(\mathbb{X})$ and (18). □

Using the previous lemma, we derive optimal approximation rates for $\mathcal{P}_N^{\text{atom}}(\mathbb{X})$ and $\mathcal{P}_L^{\text{curv}}(\mathbb{X})$.

**Theorem 5** (Upper bounds) *For $s > d/2$ suppose that $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Assume that $\mu$ is absolutely continuous with respect to $\sigma_{\mathbb{X}}$ with density $\rho \in H^s(\mathbb{X})$. Then, there are constants depending on $\mathbb{X}$ and $K$ such that*

$$\min_{\nu \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})} \mathcal{D}_K(\mu, \nu) \lesssim \|\rho\|_{H^s(\mathbb{X})} N^{-\frac{s}{d}}, \tag{19}$$

$$\min_{\nu \in \mathcal{P}_L^{\text{curv}}(\mathbb{X})} \mathcal{D}_K(\mu, \nu) \lesssim \|\rho\|_{H^s(\mathbb{X})} L^{-\frac{s}{d-1}}. \tag{20}$$

**Proof** By [13, Lem. 2.11] and since the Laplace–Beltrami has $N \sim r^d$ eigenfunctions belonging to eigenvectors $\lambda_k < r^2$, there exists a measure $\nu_r \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})$ that satisfies (15). Hence, (15) is satisfied with $r \sim N^{1/d}$, where the constants depend on $\mathbb{X}$ and $K$. Thus, Lemma 4 with $\tilde{\nu}_r \in \mathcal{P}_N^{\text{atom}}(\mathbb{X})$ leads to (19).

The assumptions of Lemma 3 are satisfied, so that analogous arguments as in the proof of Theorem 1 yield $\mathcal{P}_N^{\text{atom}}(\mathbb{X}) \subset \mathcal{P}_L^{\text{curv}}(\mathbb{X})$ with suitable $N \sim L^{d/(d-1)}$. Hence, (19) implies (20).

□

## 5.3 Upper Bounds for $\mathcal{P}_L^{\text{a-curv}}(\mathbb{X})$ and special manifolds $\mathbb{X}$

To establish upper bounds for the smaller space $\mathcal{P}_L^{\text{a-curv}}(\mathbb{X})$, restriction to special manifolds is necessary. The basic idea consists in the construction of a curve and a related measure $\nu_r$ such that all eigenfunctions of the Laplace–Beltrami operator belonging to eigenvalues smaller than a certain value are exactly integrated by this measure and then applying Lemma 4 for estimating the minimum of discrepancies. We begin with the torus.

**Theorem 6** (Torus) *Let $\mathbb{X} = \mathbb{T}^d$ with $d \in \mathbb{N}$, $s > d/2$ and suppose that $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Then, for any absolutely continuous measure*

$\mu \in \mathcal{P}(\mathbb{X})$ *with Lipschitz continuous density* $\rho \in H^s(\mathbb{X})$*, there exists a constant depending on d, K, and* $\rho$ *such that*

$$\min_{\nu \in \mathcal{P}_L^{\text{a-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{d-1}}.$$

**Proof** 1. First, we construct a closed curve $\gamma_r$ such that the trigonometric polynomials from $\Pi_r(\mathbb{T}^d)$, see (33) in the appendix, are exactly integrated along this curve. Clearly, the polynomials in $\Pi_r(\mathbb{T}^{d-1})$ are exactly integrated at equispaced nodes $x_k = \frac{k}{n}$, $k = (k_1, \dots, k_{d-1}) \in \mathbb{N}_0^{d-1}, 0 \le k_i \le n-1$, with weights $1/n^{d-1}$, where $n := r+1$. Set $z(k) := k_1 + k_2 n + \dots + k_{d-1} n^{d-2}$ and consider the curves

$$\gamma_k \colon I_k := \left[\tfrac{z(k)}{n^{d-1}}, \tfrac{z(k)+1}{n^{d-1}}\right] \to \mathbb{T}^d \quad \text{with} \quad \gamma_k(t) := \begin{pmatrix} x_k \\ n^{d-1}t \end{pmatrix}.$$

Then, each element in $\Pi_r^d$ is exactly integrated along the union of these curves, i.e., using $I := \{0, \dots, n-1\}^{d-1}$, we have

$$\int_{\mathbb{T}^d} p \, d\sigma_{\mathbb{T}^d} = \sum_{k \in I} \int_{I_k} p \circ \gamma_k \, d\lambda, \quad p \in \Pi_r^d.$$

The argument is repeated for every other coordinate direction, so that we end up with $dn^{d-1}$ curves mapping from an interval of length $\frac{1}{dn^{d-1}}$ to $\mathbb{T}^d$. The intersection points of these curves are considered as vertices of a graph, where each vertex has $2d$ edges. Consequently, there exists an Euler path $\gamma_r \colon [0, 1] \to \mathbb{T}^d$ trough the vertices build from all curves. It has constant speed $dn^{d-1}$ and the polynomials $\Pi_r^d$ are exactly integrated along $\gamma_r$, i.e.,

$$\int_{\mathbb{T}^d} p \, d\sigma_{\mathbb{T}^d} = \int_{\mathbb{T}^d} p \, d\gamma_{r*}\lambda, \quad p \in \Pi_r^d.$$

2. Next, we apply Lemma 4 for $\nu_r = \gamma_{r*}\lambda$. We observe $\tilde{\nu}_r = \gamma_{r*}((\rho \circ \gamma_r)/\beta_r \lambda)$ and deduce $L(\rho \circ \gamma_r/\beta_r) \le L(\gamma_r)L(\rho)/\beta_r \lesssim r^{d-1} \sim L$ as $\beta_r \sim 1$. Here, constants depend on $d$, $K$, and $\rho$. □

Now, we provide approximation rates for $\mathbb{X} = \mathbb{S}^d$.

**Theorem 7** (Sphere) *Let* $\mathbb{X} = \mathbb{S}^d$ *with* $d \ge 2$, $s > d/2$ *and suppose that* $H_K(\mathbb{X}) = H^s(\mathbb{X})$ *holds with equivalent norms. Then, we have for any absolutely continuous measure* $\mu \in \mathcal{P}(\mathbb{X})$ *with Lipschitz continuous density* $\rho \in H^s(\mathbb{X})$ *that there is a constant depending on d, K, and* $\rho$ *with*

$$\min_{\nu \in \mathcal{P}_L^{\text{a-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{d-1}}.$$

**Proof** 1. First, we construct a constant speed curve $\gamma_r : [0, 1] \to \mathbb{S}^d$ and a probability measure $\omega_r = \rho_r \lambda$ with Lipschitz continuous density $\rho_r : [0, 1] \to \mathbb{R}_{\geq 0}$ such that for all $p \in \Pi_r(\mathbb{S}^d)$, it holds

$$\int_{\mathbb{S}^d} p \, d\sigma_{\mathbb{S}^d} = \int_0^1 p \circ \gamma_r \, d\omega_r. \tag{21}$$

Utilizing spherical coordinates

$$x_1 = \cos\theta_1, \;\; x_2 = \sin\theta_1 \cos\theta_2, \;\; \ldots, \;\; x_d = \prod_{j=1}^{d-1} \sin\theta_j \cos\phi, \;\; x_{d+1} = \prod_{j=1}^{d-1} \sin\theta_j \sin\phi, \tag{22}$$

where $\theta_k \in [0, \pi]$, $k = 1, \ldots, d-1$, and $\phi \in [0, 2\pi)$, we obtain

$$\int_{\mathbb{S}^d} p \, d\sigma_{\mathbb{S}^d} = \int_0^\pi c_d \sin(\theta_1)^{d-1} \int_{\mathbb{S}^{d-1}} p\big(\cos(\theta_1), \sin(\theta_1)\tilde{x}\big) \, d\sigma_{\mathbb{S}^{d-1}}(\tilde{x}) \, d\theta_1, \tag{23}$$

where $c_d := (\int_0^\pi \sin(\theta)^{d-1} \, d\theta)^{-1}$. There exist nodes $\tilde{x}_i \in \mathbb{S}^{d-1}$ and positive weights $a_i, i = 1, \ldots, n \sim r^{d-1}$, with $\sum_{i=1}^n a_i = 1$, such that for all $p \in \Pi_r(\mathbb{S}^{d-1})$ it holds

$$\int_{\mathbb{S}^{d-1}} p \, d\sigma_{\mathbb{S}^{d-1}} = \sum_{i=1}^n a_i \, p(\tilde{x}_i).$$

To see this, substitute $u_k = \sin\theta_k$, $k = 2, \ldots, d-1$, apply Gaussian quadrature with nodes $\lceil (r+1)/2 \rceil$ and corresponding weights to exactly integrate over $u_k$, and equispaced nodes and weights $1/(2r+1)$ for the integration over $\phi$ as, e.g., in [82]. Then, we define $\gamma_r : [0, 1] \to \mathbb{S}^d$ for $t \in [(i-1)/n, i/n], i = 1, \ldots, n$, by

$$\gamma_r(t) := \gamma_{r,i}(2\pi nt), \qquad \gamma_{r,i}(\alpha) := \big(\cos(\alpha), \sin(\alpha)\tilde{x}_i\big), \quad \alpha \in [0, 2\pi].$$

Since $(1, 0, \ldots, 0) = \gamma_{r,i}(0) = \gamma_{r,i}(2\pi)$ for all $i = 1, \ldots, n$, the curve is closed. Furthermore, $\gamma_r(t)$ has constant speed since for $i = 1, \ldots, n$, i.e.,

$$|\dot{\gamma}_r|(t) = |\dot{\gamma}_{r,i}|(2\pi nt) = 2\pi n \sim r^{d-1}.$$

Next, the density $\rho_r : [0, 1] \to \mathbb{R}$ is defined for $t \in [(i-1)/n, i/n], i = 1, \ldots, n$, by

$$\rho_r(t) := \rho_{r,i}(2\pi nt), \qquad \rho_{r,i}(\alpha) := a_i c_d \pi n |\sin(\alpha)|^{d-1}, \qquad \alpha \in [0, 2\pi].$$

We directly verify that $\rho_r$ is Lipschitz continuous with $L(\rho_r) \lesssim \max_i a_i n^2$. By [34], the quadrature weights fulfill $a_i \lesssim \frac{1}{r^{d-1}}$ so that $L(\rho_r) \lesssim n^2 r^{-(d-1)} \sim r^{d-1}$. By definition of the constant $c_d$ and weights $a_i$, we see that $\rho_r$ is indeed a probability

density

$$\int_0^1 \rho_r \, d\lambda = \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \rho_{r,i}(2\pi n t) \, dt = \frac{1}{2\pi n} \sum_{i=1}^n \int_0^{2\pi} \rho_{r,i}(\alpha) \, d\alpha$$

$$= \frac{c_d}{2} \sum_{i=1}^n a_i \int_0^{2\pi} |\sin(\theta)|^{d-1} \, d\theta = 1.$$

For $p \in \Pi_r(\mathbb{S}^d)$, we obtain

$$\int_0^1 p \circ \gamma_r \, \rho_r \, d\lambda$$

$$= \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} p(\gamma_{r,i}(2\pi n t)) \rho_{r,i}(2\pi M t) \, dt = \int_0^{2\pi} \frac{1}{2\pi n} \sum_{i=1}^n p(\gamma_{r,i}(\alpha)) \rho_{r,i}(\alpha) \, d\alpha$$

$$= \frac{c_d}{2} \int_0^{2\pi} |\sin(\alpha)|^{d-1} \sum_{i=1}^n a_i \, p(\cos(\alpha), \sin(\alpha)\tilde{x}_i) \, d\alpha$$

$$= \frac{c_d}{2} \int_0^{\pi} |\sin(\alpha)|^{d-1} \sum_{i=1}^n a_i \Big( p(\cos(\alpha), \sin(\alpha)\tilde{x}_i) + p(-\cos(\alpha), -\sin(\alpha)\tilde{x}_i) \Big) \, d\alpha.$$

Without loss of generality, $p$ is chosen as a homogeneous polynomial of degree $k \le r$, i.e., $p(tx) = t^k p(x)$. Then,

$$\int_0^1 p \circ \gamma_r \, \rho_r \, d\lambda = \frac{1 + (-1)^k}{2} \int_0^{\pi} c_d |\sin(\alpha)|^{d-1} \sum_{i=1}^n a_i \, p(\cos(\alpha), \sin(\alpha)\tilde{x}_i) \, d\alpha,$$

and regarding that for fixed $\alpha \in [0, 2\pi]$ the function $\tilde{x} \mapsto p(\cos(\alpha), \sin(\alpha)\tilde{x})$ is a polynomial of degree at most $r$ on $\mathbb{S}^{d-1}$, we conclude

$$\int_0^1 p \circ \gamma_r \, \rho_r \, d\lambda = \frac{1 + (-1)^k}{2} \int_0^{\pi} c_d |\sin(\alpha)|^{d-1} \int_{\mathbb{S}^{d-1}} p(\cos(\alpha), \sin(\alpha)\tilde{x}) \, d\sigma_{\mathbb{S}^{d-1}}(\tilde{x}) \, d\alpha.$$

Now, the assertion (21) follows from (23) and since $\int_{\mathbb{S}^d} p \, d\sigma_{\mathbb{S}^d} = 0$ if $k$ is odd.

2. Next, we apply Lemma 4 for $\nu_r = \gamma_{r*}\rho_r \lambda$, from which we obtain that $\tilde{\nu}_r = \gamma_{r*}((\rho \circ \gamma_r)\rho_r/\beta_r \lambda)$. As all $\rho_r$ are uniformly bounded by construction and $\rho$ is bounded due to continuity, we conclude using $L(\rho_r) \lesssim r^{d-1}$ and $L(\gamma_r) \sim r^{d-1}$ that

$$L(\rho \circ \gamma_r \, \rho_r/\beta_r) \le \big( L(\rho \circ \gamma_r)\|\rho_r\|_\infty + L(\rho_r)\|\rho\|_\infty \big)/\beta_r \lesssim \big( L(\rho) + \|\rho\|_\infty \big) r^{d-1},$$

which concludes the proof. □

Finally, we derive approximation rates for $\mathbb{X} = \mathrm{SO}(3)$.

**Corollary 1** *(Special orthogonal group) Let $\mathbb{X} = \mathrm{SO}(3)$, $s > 3/2$ and suppose $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Then, we have for any absolutely continuous measure $\mu \in \mathcal{P}(\mathbb{X})$ with Lipschitz continuous density $\rho \in H^s(\mathbb{X})$ that*

$$\min_{\nu \in \mathcal{P}_L^{\text{a-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{d-1}},$$

*where the constant depends on $K$ and $\rho$.*

**Proof** 1. For fixed $L \sim r^2$, we shall construct a curve $\gamma_r : [0,1] \to \mathrm{SO}(3)$ with $L(\gamma_r) \lesssim L$ and a probability measure $\omega_r = \rho_r \lambda$ with density $\rho_r : [0,1] \to \mathbb{R}_{\geq 0}$ and $L(\rho_r) \lesssim L$, such that

$$\int_{\mathrm{SO}(3)} p \, d\sigma_{\mathrm{SO}(3)} = \int_{\mathrm{SO}(3)} p \, d\gamma_{r*}(\rho_r \lambda).$$

We use the fact that the sphere $\mathbb{S}^3$ is a double covering of $\mathrm{SO}(3)$. That is, there is a surjective two-to-one mapping $a : \mathbb{S}^3 \to \mathrm{SO}(3)$ satisfying $a(x) = a(-x)$, $x \in \mathbb{S}^3$. Moreover, we know that $a : \mathbb{S}^3 \to \mathrm{SO}(3)$ is a local isometry, see [42], i.e., it respects the Riemannian structures, implying the relations $\sigma_{\mathrm{SO}(3)} = a_* \sigma_{\mathbb{S}^3}$ and

$$\mathrm{dist}_{\mathrm{SO}(3)}\big(a(x_1), a(x_2)\big) = \min\big(\mathrm{dist}_{\mathbb{S}^3}(x_1, x_2), \mathrm{dist}_{\mathbb{S}^3}(x_1, -x_2)\big).$$

It also maps $\Pi_r(\mathrm{SO}(3))$ into $\Pi_{2r}(\mathbb{S}^3)$, i.e., $p \in \Pi_r(\mathrm{SO}(3))$ implies $p \circ a \in \Pi_{2r}(\mathbb{S}^3)$. Now, let $\tilde{\gamma}_r : [0,1] \to \mathbb{S}^3$ and $\tilde{\omega}_r$ be given as in the first part of the proof of Theorem 7 for $d = 3$, i.e., $\tilde{\gamma}_{r*}\tilde{\omega}_r$ satisfies (21) with $L(\tilde{\gamma}_r) \lesssim L$ and $\tilde{\omega}_r = \tilde{\rho}_r \lambda$ with $L(\tilde{\rho}_r) \lesssim L$.

We now define a curve $\gamma_r$ in $\mathrm{SO}(3)$ by

$$\gamma_r : [0,1] \to \mathrm{SO}(3), \qquad \gamma_r(t) := a \circ \tilde{\gamma}_{2r}(t),$$

and let $\omega_r := \tilde{\omega}_{2r}$. For $p \in \Pi_r(\mathrm{SO}(3))$, the push-forward measure $\gamma_{r*}\omega_r$ leads to

$$\int_{\mathrm{SO}(3)} p \, d\sigma_{\mathrm{SO}(3)} = \int_{\mathrm{SO}(3)} p \, da_* \sigma_{\mathbb{S}^3} = \int_{\mathbb{S}^3} p \circ a \, d\sigma_{\mathbb{S}^3}$$
$$= \int_{\mathbb{S}^3} p \circ a \, d\tilde{\gamma}_{2r*}\tilde{\omega}_{2r} = \int_{\mathrm{SO}(3)} p \, d\gamma_{r*}\omega_r.$$

Hence, property (15) is satisfied for $\gamma_{r*}\omega_r = \gamma_{r*}(\tilde{\rho}_{2r}\lambda)$.

2. The rest follows along the lines of step 2. in the proof of Theorem 7. $\square$

## 5.4 Upper Bounds for $\mathcal{P}_L^{\text{ˇ-curv}}(\mathbb{X})$ and special manifolds $\mathbb{X}$

To derive upper bounds for the smallest space $\mathcal{P}_L^{\text{ˇ-curv}}(\mathbb{X})$, we need the following specification of Lemma 4.

**Lemma 5** *For $s > d/2$ suppose that $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Let $\mu \in \mathcal{P}(\mathbb{X})$ be absolutely continuous with respect to $\sigma_{\mathbb{X}}$ with positive density $\rho \in H^s(\mathbb{X})$. Suppose that $\nu_r := \gamma_{r*}\lambda$ with $\gamma_r \in \mathrm{Lip}(\mathbb{X})$ satisfies (15) and let $\beta_r := \int_{\mathbb{X}} \rho \, d\nu_r$. Then, for sufficiently large $r$,*

$$g : [0, 1] \to [0, 1], \qquad g(t) := \frac{1}{\beta_r} \int_0^t \rho \circ \gamma_r \, d\lambda$$

*is well-defined and invertible. Moreover, $\tilde{\gamma}_r := \gamma_r \circ g^{-1}$ satisfies $L(\tilde{\gamma}_r) \lesssim L(\gamma_r)$ and*

$$\mathscr{D}_K(\mu, \tilde{\gamma}_{r*}\lambda) \lesssim r^{-s}, \tag{24}$$

*where the constants depend on $\mathbb{X}$, $K$, and $\rho$.*

**Proof** Since $\rho$ is continuous, there is $\epsilon > 0$ with $\rho \geq \epsilon$. To bound the Lipschitz constant $L(\tilde{\gamma}_r)$, we apply the mean value theorem together with the definition of $g$ and the fact that $(g^{-1})'(s) = 1/g'(g^{-1}(s))$ to obtain

$$\left|\tilde{\gamma}_r(s) - \tilde{\gamma}_r(t)\right| \leq L(\gamma_r)\left|g^{-1}(s) - g^{-1}(t)\right| \leq L(\gamma_r)\frac{\beta_r}{\epsilon}|s - t|.$$

Using (18), this can be further estimated for sufficiently large $r$ as

$$\left|\tilde{\gamma}_r(s) - \tilde{\gamma}_r(t)\right| \lesssim L(\gamma_r)\frac{1 + \|\rho\|_{H^s(\mathbb{X})}r^{-s}}{\epsilon}|s - t| \lesssim L(\gamma_r)\frac{2}{\epsilon}|s - t|.$$

To derive (24), we aim to apply Lemma 4 with $\nu_r = \gamma_{r*}\lambda$. We observe

$$\tilde{\nu}_r = \frac{\rho}{\beta_r}\gamma_{r*}\lambda = \gamma_{r*}\left(\frac{\rho \circ \gamma_r}{\beta_r}\lambda\right) = \gamma_{r*}(g'\lambda) = (\gamma_r \circ g^{-1})_*\lambda = \tilde{\gamma}_{r*}\lambda,$$

so that Lemma 4 indeed implies (24). $\qquad\square$

In comparison to Theorem 6, we now trade the Lipschitz condition on $\rho$ with the positivity requirement, which enables us to cover $\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})$.

**Theorem 8** (Torus) *Let $\mathbb{X} = \mathbb{T}^d$ with $d \in \mathbb{N}$, $s > d/2$ and suppose that $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Then, for any absolutely continuous measure $\mu \in \mathcal{P}(\mathbb{X})$ with positive density $\rho \in H^s(\mathbb{X})$, there is a constant depending on $d$, $K$, and $\rho$ with*

$$\min_{\nu \in \mathcal{P}_L^{\text{a-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{d-1}}.$$

**Proof** The first part of the proof is identical to the proof of Theorem 6. Instead of Lemma 4 though, we now apply Lemma 5 for $\gamma_r$ and $\rho_r \equiv 1$. Hence, $\tilde{\gamma}_r = \gamma_r \circ g_r^{-1}$ satisfies $L(\tilde{\gamma}_r) \leq \frac{\beta_r}{\epsilon}d(2r + 1)^{d-1} \lesssim r^{d-1}$, so that $\tilde{\gamma}_{r*}\lambda$ satisfies (24) and is in $\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})$ with $L \sim r^{d-1}$. $\qquad\square$

The construction on $\mathbb{X} = \mathbb{S}^d$ for $\mathcal{P}_L^{\text{a-curv}}(\mathbb{X})$ in the proof of Theorem 7 is not compatible with $\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})$. Thus, the situation is different from the torus, where we have used the same underlying construction and only switched from Lemma 4 to Lemma 5. Now, we present a new construction for $\mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})$, which is tailored to $\mathbb{X} = \mathbb{S}^2$. In this case, we can transfer the ideas of the torus, but with Gauss-Legendre quadrature points.

**Theorem 9** (2-sphere) *Let $\mathbb{X} = \mathbb{S}^2$, $s > 1$ and suppose $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Then, we have for any absolutely continuous measure $\mu \in \mathcal{P}(\mathbb{X})$ with positive density $\rho \in H^s(\mathbb{X})$ that there is a constant depending on $K$ and $\rho$ with*

$$\min_{\nu \in \mathcal{P}_L^{\text{a-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_L^{\check{}\text{-curv}}(\mathbb{X})} \mathscr{D}_K(\mu, \nu) \lesssim L^{-s}.$$

*Proof* 1. We construct closed curves such that the spherical polynomials from $\Pi_r(\mathbb{S}^2)$, see (35) in the appendix, are exactly integrated along this curve. It suffices to show this for the polynomials $p(x) = x_1^{k_1} x_2^{k_2} x_3^{k_3} \in \Pi_r(\mathbb{S}^2)$ with $k_1 + k_2 + k_3 \leq r$ restricted to $\mathbb{S}^2$. We select $n = \lceil (r+1)/2 \rceil$ Gauss-Legendre quadrature points $u_j = \cos(\theta_j) \in [-1, 1]$ and corresponding weights $2\omega_j$, $j = 1, \ldots, n$. Note that $\sum_{j=1}^n \omega_j = 1$. Using spherical coordinates $x_1 = \cos(\theta)$, $x_2 = \sin(\theta)\cos(\phi)$, and $x_3 = \sin(\theta)\sin(\phi)$ with $(\theta, \phi) \in [0, \pi] \times [0, 2\pi]$, we obtain

$$\int_{\mathbb{S}^2} p \, d\sigma_{\mathbb{S}^2} = \frac{1}{4\pi} \int_0^{2\pi} \cos(\phi)^{k_2} \sin(\phi)^{k_3} \int_0^\pi \cos(\theta)^{k_1} \sin(\theta)^{k_2 + k_3} \sin(\phi) \, d\theta \, d\phi$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \cos(\phi)^{k_2} \sin(\phi)^{k_3} \int_{-1}^1 u^{k_1} (1 - u^2)^{\frac{k_2 + k_3}{2}} \, du \, d\phi,$$

see also [83]. If $k_2 + k_3$ is odd, then the integral over $\phi$ becomes zero. If $k_2 + k_3$ is even, the inner integrand is a polynomial of degree $\leq r$. In both cases we get

$$\int_{\mathbb{S}^2} p \, d\sigma_{\mathbb{S}^2} = \frac{1}{2\pi} \sum_{j=1}^n \omega_j \int_0^{2\pi} p\big(\cos(\theta_j), \sin(\theta_j)\cos(\phi), \sin(\theta_j)\sin(\phi)\big) \, d\phi.$$

Substituting in each summand $\phi = 2\pi t / \omega_j$, $j = 1, \ldots, n$, yields

$$\int_{\mathbb{S}^2} p \, d\sigma_{\mathbb{S}^2} = \sum_{j=1}^n \int_0^{\omega_j} p \circ \gamma_j \, d\lambda,$$

where $\gamma_j \colon [0, \omega_j] \to \mathbb{S}^2$ is defined by

$$\gamma_j(t) := \big(\cos(\theta_j), \sin(\theta_j)\cos(2\pi t/\omega_j), \sin(\theta_j)\sin(2\pi t/\omega_j)\big),$$

and has constant speed $L(\gamma_j) = 2\pi \sin(\theta_j)/\omega_j$. The lower bound $\omega_j \gtrsim \frac{1}{n} \sin(\theta_j)$, cf. [34], implies that $L(\gamma_j) \lesssim n$. Defining a curve $\tilde{\gamma} : [0, 1] \to \mathbb{S}^2$ piecewise via

$$\tilde{\gamma}|_{[0,s_1]} = \gamma_1, \quad \tilde{\gamma}|_{[s_1,s_2]} = \gamma_2(\cdot - s_1), \quad \ldots \quad , \quad \tilde{\gamma}|_{[s_{n-1},1]} = \gamma_n(\cdot - s_{n-1}),$$

where $s_j := \omega_1 + \ldots + \omega_j$, we obtain

$$\int_{\mathbb{S}^2} p \, \mathrm{d}\sigma_{\mathbb{S}^2} = \int_0^1 p \, \mathrm{d}\tilde{\gamma}_*\lambda, \quad p \in \Pi_r(\mathbb{S}^2).$$

Further, the curve satisfies $L(\tilde{\gamma}) \lesssim r$.

As with the torus, we now "turn" the sphere (or switch the position of $\phi$) so that we get circles along orthogonal directions. This large collection of circles is indeed connected. As with the torus, each intersection point has an incoming and outgoing part of a circle, so that all this corresponds to a graph, where again each vertex has an even number of "edges". Hence, there is an Euler path inducing our final curve $\gamma_r : [0, 1] \to \mathbb{S}^2$ with piecewise constant speed $L(\gamma_r) \lesssim r$ satisfying

$$\int_{\mathbb{S}^2} p \, \mathrm{d}\sigma_{\mathbb{S}^2} = \int_0^1 p \, \mathrm{d}(\gamma_{r*}\lambda), \quad p \in \Pi_r(\mathbb{S}^2).$$

2. Let $r \sim L$. Analogous to the end of the proof of Theorem 8, Lemma 5 now yields the assertion. $\qquad\Box$

To get the approximation rate for $\mathbb{X} = \mathcal{G}_{2,4}$, we make use of its double covering $\mathbb{X} = \mathbb{S}^2 \times \mathbb{S}^2$, cf. Remark 8.

**Theorem 10** (Grassmannian) *Let $\mathbb{X} = \mathcal{G}_{2,4}$, $s > 2$ and suppose $H_K(\mathbb{X}) = H^s(\mathbb{X})$ holds with equivalent norms. Then, we have for any absolutely continuous measure $\mu \in \mathcal{P}(\mathbb{X})$ with positive density $\rho \in H^s(\mathbb{X})$ that there exists a constant depending on $K$ and $\rho$ with*

$$\min_{\nu \in \mathcal{P}_L^{\mathrm{a\text{-}curv}}(\mathbb{X})} \mathcal{D}_K(\mu, \nu) \leq \min_{\nu \in \mathcal{P}_L^{\mathrm{\text{-}curv}}(\mathbb{X})} \mathcal{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{3}}.$$

*Proof* By Remark 8 in the appendix, we know that $\mathcal{G}_{2,4} \cong \mathbb{S}^2 \times \mathbb{S}^2/\{\pm 1\}$ so that is remains to prove the assertion for $\mathbb{X} = \mathbb{S}^2 \times \mathbb{S}^2$.

There exist pairwise distinct points $\{x_1, \ldots, x_N\} \subset \mathbb{S}^2$ such that $\frac{1}{N}\sum_{j=1}^N \delta_{x_j}$ satisfies (15) on $\mathbb{S}^2$ with $N \sim r^2$, cf. [9,10]. On the other hand, let $\tilde{\gamma}$ be the curve on $\mathbb{S}^2$ constructed in the proof of Theorem 9, so that $\tilde{\gamma}_*\lambda$ satisfies (15) on $\mathbb{S}^2$ with $\ell(\tilde{\gamma}) \leq L(\tilde{\gamma}) \sim r$. Let us introduce the virtual point $x_{N+1} := x_1$.

The curve $\tilde{\gamma}([0, 1])$ contains a great circle. Thus, for each pair $x_j$ and $x_{j+1}$ there is $O_j \in \mathrm{O}(3)$ such that $x_j, x_{j+1} \in \Gamma_j := O_j\tilde{\gamma}([0, 1])$.

It turns out that the set on $\mathbb{S}^2 \times \mathbb{S}^2$ given by $\bigcup_{j=1}^N (\{x_j\} \times \Gamma_j) \cup (\Gamma_j \times \{x_{j+1}\})$ is connected. We now choose $\gamma_j := O_j\tilde{\gamma}$ and know that the union of the trajectories of

the set of curves

$$t \mapsto \big(x_j, \gamma_j(t)\big), \qquad t \mapsto \big(\gamma_j(t), x_{j+1}\big), \quad j = 1, \dots, N,$$

is connected. Combinatorial arguments involving Euler paths, see Theorems 6 and 9, lead to a curve $\gamma$ with $\ell(\gamma) \le L(\gamma) \sim NL(\tilde{\gamma}) \sim r^3$, so that $\gamma_*\lambda$ satisfies (15). The remaining part follows along the lines of the proof of Theorem 7.                                      □

Our approximation results can be extended to diffeomorphic manifolds, e.g., from $\mathbb{S}^2$ to ellipsoids, see also the 3D-torus example in Sect. 8. To this end, recall that we can describe the Sobolev space $H^s(\mathbb{X})$ using local charts, see [78, Sec. 7.2]. The exponential maps $\exp_x \colon T_x\mathbb{X} \to \mathbb{X}$ give rise to local charts $(\mathring{B}_x(r_0), \exp_x^{-1})$, where $\mathring{B}_x(r_0) := \{y \in \mathbb{X} : \mathrm{dist}_{\mathbb{X}}(x, y) < r_0\}$ denotes the geodesic balls around $x$ with the injectivity radius $r_0$. If $\delta < r_0$ is chosen small enough, there exists a uniformly locally finite covering of $\mathbb{X}$ by a sequence of balls $(\mathring{B}_{x_j}(\delta))_j$ with a corresponding smooth resolution of unity $(\psi_j)_j$ with $\mathrm{supp}(\psi_j) \subset \mathring{B}_{x_j}(\delta)$, see [78, Prop. 7.2.1]. Then, an equivalent Sobolev norm is given by

$$\|f\|_{H^s(\mathbb{X})} := \Big( \sum_{j=1}^{\infty} \|(\psi_j f) \circ \exp_{x_j}\|_{H^s(\mathbb{R}^d)}^2 \Big)^{\frac{1}{2}}, \tag{25}$$

where $(\psi_j f) \circ \exp_{x_j}$ is extended to $\mathbb{R}^d$ by zero, see [78, Thm. 7.4.5]. Using Definition (25), we are able to pull over results from the Euclidean setting.

**Proposition 2** *Let $\mathbb{X}_1$, $\mathbb{X}_2$ be two d-dimensional connected, compact Riemannian manifolds without boundary, which are $s + 1$ diffeomorphic with $s > d/2$. Assume that for $H_K(\mathbb{X}_2) = H^s(\mathbb{X}_2)$ and every absolutely continuous measure $\mu$ with positive density $\rho \in H^s(\mathbb{X}_2)$ it holds*

$$\min_{\nu \in \mathcal{P}_L^{\text{-curv}}} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{d-1}},$$

*where the constant depends on $\mathbb{X}_2$, $K$, and $\rho$. Then, the same property holds for $\mathbb{X}_1$, where the constant additionally depends on the diffeomorphism.*

**Proof** Let $f \colon \mathbb{X}_2 \to \mathbb{X}_1$ denote such a diffeomorphism and $\rho \in H^s(\mathbb{X}_1)$ the density of the measure $\mu$ on $\mathbb{X}_1$. Any curve $\tilde{\gamma} \colon [0, 1] \to \mathbb{X}_2$ gives rise to a curve $\gamma \colon [0, 1] \to \mathbb{X}_1$ via $\gamma = f \circ \tilde{\gamma}$, which for every $\varphi \in H^s(\mathbb{X}_1)$ satisfies

$$\Big| \int_{\mathbb{X}_1} \varphi\rho \, \mathrm{d}\sigma_{\mathbb{X}_1} - \int_0^1 \varphi \circ \gamma \, \mathrm{d}\lambda \Big| = \Big| \int_{\mathbb{X}_2} (\varphi\rho) \circ f |\det(J_f)| \, \mathrm{d}\sigma_{\mathbb{X}_2} - \int_0^1 \varphi \circ f \circ \tilde{\gamma} \, \mathrm{d}\lambda \Big|,$$

where $J_f$ denotes the Jacobian of $f$. Now, note that $\varphi \circ f$, $\rho \circ f |\det(J_f)| \in H^s(\mathbb{X}_2)$, see (16) and [78, Thm. 4.3.2], which is lifted to manifolds using (25). Hence, we can define a measure $\tilde{\mu}$ on $\mathbb{X}_2$ through the probability density $\rho \circ f |\det(J_f)|$. Choosing $\tilde{\gamma}_L$ as a

realization for some minimizer of $\inf_{\nu \in \mathcal{P}_L^{\text{-curv}}} \mathscr{D}(\tilde{\mu}, \nu)$, we can apply the approximation result for $\mathbb{X}_2$ and estimate for $\gamma_L = f \circ \tilde{\gamma}_L$ that

$$\left| \int_{\mathbb{X}_1} \varphi \rho \, d\sigma_{X_1} - \int_0^1 \varphi \circ \gamma_L \, d\lambda \right| \lesssim L^{-\frac{s}{d-1}} \|\varphi \circ f\|_{H^s(\mathbb{X}_2)} \lesssim L^{-\frac{s}{d-1}} \|\varphi\|_{H^s(\mathbb{X}_1)},$$

where the second estimate follows from [78, Thm. 4.3.2]. Now, $L(\gamma_L) \leq L(f)L$ implies

$$\inf_{\nu \in \mathcal{P}_L^{\text{-curv}}} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s}{d-1}}.$$

$\square$

**Remark 5** Consider a probability measure $\mu$ on $\mathbb{X}$ such that the dimension $d_\mu$ of its support is smaller than the dimension $d$ of $\mathbb{X}$. Then, $\mu$ does not have any density with respect to $\sigma_{\mathbb{X}}$. If $\text{supp}(\mu)$ is itself a $d_\mu$-dimensional connected, compact Riemannian manifold $\mathbb{Y}$ without boundary, we switch from $\mathbb{X}$ to $\mathbb{Y}$. Sobolev trace theorems and reproducing kernel Hilbert space theory imply that the assumption $H_K(\mathbb{X}) = H^s(\mathbb{X})$ leads to $H_{K'}(\mathbb{Y}) = H^{s'}(\mathbb{Y})$, where $K' := K|_{\mathbb{Y} \times \mathbb{Y}}$ is the restricted kernel and $s' = s - (d - d_\mu)/2$, cf. [36]. If, for instance, $\mathbb{Y}$ is diffeomorphic to $\mathbb{T}^{d_\mu}$ (or $\mathbb{S}^{d_\mu}$ with $d_\mu = 2$), and $\mu$ has a positive density $\rho \in H^{s'}(\mathbb{Y})$ with respect to $\sigma_{\mathbb{Y}}$, then Theorem 8 (or 9) and Proposition 2 eventually yield

$$\min_{\nu \in \mathcal{P}_L^{\text{-curv}}} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s'}{d_\mu - 1}}.$$

If $\text{supp}(\mu)$ is a proper subset of $\mathbb{Y}$, we are able to analyze approximations with $\mathcal{P}_L^{\text{a-curv}}(\mathbb{Y})$. First, we observe that the analogue of Proposition 2 also holds for $\mathcal{P}_L^{\text{a-curv}}(\mathbb{X}_1), \mathcal{P}_L^{\text{a-curv}}(\mathbb{X}_2)$ when the positivity assumption on $\rho$ is replaced with the Lipschitz requirement as in Theorems 6 and 7. If, for instance, $\mathbb{Y}$ is diffeomorphic to $\mathbb{T}^{d_\mu}$ or $\mathbb{S}^{d_\mu}$ and $\mu$ has a Lipschitz continuous density $\rho \in H^{s'}(\mathbb{Y})$ with respect to $\sigma_{\mathbb{Y}}$, then Theorems 6 and 7, and Proposition 2 eventually yield

$$\min_{\nu \in \mathcal{P}_L^{\text{a-curv}}} \mathscr{D}_K(\mu, \nu) \lesssim L^{-\frac{s'}{d_\mu - 1}}.$$

## 6 Discretization

In our numerical experiments, we are interested in determining minimizers of

$$\min_{\nu \in \mathcal{P}_L^{\text{-curv}}(\mathbb{X})} \mathscr{D}_K^2(\mu, \nu). \tag{26}$$

Defining $A_L := \{\gamma \in \text{Lip}(\mathbb{X}) : L(\gamma) \leq L\}$ and using the indicator function

$$\iota_{A_L}(\gamma) := \begin{cases} 0 & \text{if } \gamma \in A_L, \\ +\infty & \text{otherwise,} \end{cases}$$

we can rephrase problem (26) as a minimization problem over curves

$$\min_{\gamma \in \mathcal{C}([0,1],\mathbb{X})} \mathcal{J}_L(\gamma),$$

where $\mathcal{J}_L(\gamma) := \mathscr{D}_K^2(\mu, \gamma_*\lambda) + \iota_{A_L}(\gamma)$. As $\mathbb{X}$ is a connected Riemannian manifold, we can approximate curves in $A_L$ by piecewise shortest geodesics with $N$ parts, i.e., by curves from

$$A_{L,N} := \left\{ \gamma \in A_L : \gamma|_{[(i-1)/N, i/N]} \text{ is a shortest geodesic for } i = 1, \ldots, N \right\}.$$

Next, we approximate the Lebesgue measure on $[0, 1]$ by $e_N := \frac{1}{N} \sum_{i=1}^N \delta_{i/N}$ and consider the minimization problems

$$\min_{\gamma \in \mathcal{C}([0,1],\mathbb{X})} \mathcal{J}_{L,N}(\gamma), \qquad (27)$$

where $\mathcal{J}_{L,N}(\gamma) := \mathscr{D}_K^2(\mu, \gamma_*e_N) + \iota_{A_{L,N}}(\gamma)$. Since $\text{ess sup}_{t \in [0,1]} |\dot{\gamma}|(t) = L(\gamma)$, the constraint $L(\gamma) \leq L$ can be reformulated as $\int_0^1 (|\dot{\gamma}|(t) - L)_+^2 \, dt = 0$.[1] Hence, using $x_i = \gamma(i/N)$, $i = 1, \ldots, N$, $x_0 = x_N$ and regarding that $|\dot{\gamma}|(t) = N \, \text{dist}_{\mathbb{X}}(x_{i-1}, x_i)$ for $t \in \left( \frac{i-1}{N}, \frac{i}{N} \right)$, problem (27) is rewritten in the computationally more suitable form

$$\min_{(x_1,\ldots,x_N) \in \mathbb{X}^N} \mathscr{D}_K^2 \left( \mu, \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \right) \quad \text{s.t.} \quad \frac{1}{N} \sum_{i=1}^N \left( N \, \text{dist}_{\mathbb{X}}(x_{i-1}, x_i) - L \right)_+^2 = 0. \quad (28)$$

This discretization is motivated by the next proposition. To this end, recall that a sequence $(f_N)_{N \in \mathbb{N}}$ of functions $f_N \colon \mathbb{X} \to (-\infty, +\infty]$ is said to $\Gamma$-converge to $f \colon \mathbb{X} \to (-\infty, +\infty]$ if the following two conditions are fulfilled for each $x \in \mathbb{X}$, see [12]:

(i) $f(x) \leq \liminf_{N \to \infty} f_N(x_N)$ whenever $x_N \to x$,
(ii) there is a sequence $(y_N)_{N \in \mathbb{N}}$ with $y_N \to x$ and $\limsup_{N \to \infty} f_N(y_N) \leq f(x)$.

The importance of $\Gamma$-convergence relies in the fact that every cluster point of minimizers of $(f_N)_{N \in \mathbb{N}}$ is a minimizer of $f$. Note that for non-compact manifolds $\mathbb{X}$ an additional equi-coercivity condition would be required.

**Proposition 3** *The sequence* $(\mathcal{J}_{L,N})_{N \in \mathbb{N}}$ *is* $\Gamma$*-convergent with limit* $\mathcal{J}_L$.

*Proof* 1. First, we verify the lim inf-inequality. Let $(\gamma_N)_{N \in \mathbb{N}}$ with $\lim_{N \to \infty} \gamma_N = \gamma$, i.e., the sequence satisfies $\sup_{t \in [0,1]} \text{dist}_{\mathbb{X}}(\gamma(t), \gamma_N(t)) \to 0$. By excluding the trivial

---

[1] For $r \in \mathbb{R}$, we use the notation $r_+ = \begin{cases} r, & r \geq 0, \\ 0, & \text{otherwise.} \end{cases}$

case $\liminf_{N \to \infty} \mathcal{J}_{L,N}(\gamma_N) = \infty$ and restricting to a subsequence $(\gamma_{N_k})_{k \in \mathbb{N}}$, we may assume $\gamma_{N_k} \in A_{L,N_k} \subset A_L$. Since $A_L$ is closed, we directly infer $\gamma \in A_L$. It holds $e_N \rightharpoonup \lambda$, which is equivalent to the convergence of Riemann sums for $f \in C[0, 1]$, and hence also $\gamma_{N*} e_N \rightharpoonup \gamma_* dr$. By the weak continuity of $\mathscr{D}_K^2$, we obtain

$$\mathcal{J}_L(\gamma) = \mathscr{D}_K^2(\mu, \gamma_* \lambda) = \lim_{N \to \infty} \mathscr{D}_K^2(\mu, \gamma_{N*} e_N) = \liminf_{N \to \infty} \mathcal{J}_{L,N}(\gamma_N). \qquad (29)$$

2. Next, we prove the lim sup-inequality, i.e., we are searching for a sequence $(\gamma_N)_{N \in \mathbb{N}}$ with $\gamma_N \to \gamma$ and $\limsup_{N \to \infty} \mathcal{J}_{L,N}(\gamma_N) \leq \mathcal{J}_L(\gamma)$. First, we may exclude the trivial case $\mathcal{J}_L(\gamma) = \infty$. Then, $\gamma_N$ is defined on every interval $[(i-1)/N, i/N]$, $i = 1, \ldots, N$, as a shortest geodesic from $\gamma((i-1)/N)$ to $\gamma(i/N)$. By construction we have $\gamma_N \in A_{L,N}$. From $\gamma, \gamma_N \in A_L$ we conclude

$$\sup_{t \in [0,1]} \operatorname{dist}_{\mathbb{X}}\big(\gamma(t), \gamma_N(t)\big) = \max_{i=1,\ldots N} \sup_{t \in [(i-1)/N, i/N]} \operatorname{dist}_{\mathbb{X}}\big(\gamma(t), \gamma_N(t)\big)$$

$$\leq \max_{i=1,\ldots N} \sup_{t \in [(i-1)/N, i/N]} \operatorname{dist}_{\mathbb{X}}\big(\gamma(t), \gamma(i/N)\big) + \operatorname{dist}_{\mathbb{X}}\big(\gamma_N(i/N), \gamma_N(t)\big) \leq \frac{2L}{N},$$

implying $\gamma_N \to \gamma$. Similarly as in (29), we infer $\limsup_{N \to \infty} \mathcal{J}_{L,N}(\gamma_N) \leq \mathcal{J}_L(\gamma)$. $\qquad \square$

In the numerical part, we use the penalized form of (28) and minimize

$$\min_{(x_1,\ldots,x_N) \in \mathbb{X}^N} \mathscr{D}_K^2\Big(\mu, \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}\Big) + \frac{\lambda}{N} \sum_{i=1}^{N} \big(N \operatorname{dist}_{\mathbb{X}}(x_{i-1}, x_i) - L\big)_+^2, \qquad \lambda > 0. \tag{30}$$

## 7 Numerical Algorithm

For a detailed overview on Riemannian optimization we refer to [69] and the books [1,79]. In order to minimize (30), we have a closer look at the discrepancy term. By (6) and (7), the discrepancy can be represented as follows

$$\mathscr{D}_K^2\Big(\mu, \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}\Big) = \frac{1}{N^2} \sum_{i,j=1}^{N} K(x_i, x_j) - 2 \sum_{i=1}^{N} \int_{\mathbb{X}} K(x_i, x) \, d\mu(x) + \iint_{\mathbb{X} \times \mathbb{X}} K \, d\mu \, d\mu$$

$$= \sum_{k=0}^{\infty} \alpha_k \Big| \hat{\mu}_k - \frac{1}{N} \sum_{i=1}^{N} \varphi_k(x_i) \Big|^2.$$

Both formulas have pros and cons: The first formula allows for an exact evaluation only if the expressions $\Phi(x) := \int_{\mathbb{X}} K(x, y) \, d\mu(y)$ and $\int_{\mathbb{X}} \Phi \, d\mu$ can be written in closed forms. In this case the complexity scales quadratically in the number of points $N$. The

second formula allows for exact evaluation only if the kernel has a finite expansion (3). In that case the complexity scales linearly in $N$.

Our approach is to use kernels fulfilling $H_K(\mathbb{X}) = H^s(\mathbb{X})$, $s > d/2$, and approximating them by their truncated representation with respect to the eigenfunctions of the Laplace–Beltrami operator

$$K_r(x, y) := \sum_{k \in \mathcal{I}_r} \alpha_k \varphi_k(x) \overline{\varphi_k(y)}, \quad \mathcal{I}_r := \{k : \varphi_k \in \Pi_r(\mathbb{X})\}.$$

Then, we finally aim to minimize

$$\min_{x \in \mathbb{X}^N} F(x) := \sum_{k \in \mathcal{I}_r} \alpha_k \left( \hat{\mu}_k - \frac{1}{N} \sum_{i=1}^N \varphi_k(x_i) \right)^2 + \frac{\lambda}{N} \sum_{i=1}^N \left( N \operatorname{dist}_{\mathbb{X}}(x_{i-1}, x_i) - L \right)_+^2, \tag{31}$$

---

### Algorithm 1 (CG Method with Restarts)

**Parameters:** maximal iterations $k_{\max} \in \mathbb{N}$
**Input:** twice differentiable function $F : \mathbb{X}^N \to [0, \infty)$, initial point $x^{(0)} \in \mathbb{X}^N$
**Initialization:** $g^{(0)} := \nabla_{\mathbb{X}^N} F(x^{(0)})$, $d^{(0)} := -g^{(0)}$, $r := 0$
**for** $k := 0, \ldots, k_{\max}$ **do**
$\quad x^{(k+1)} := \gamma_{x^{(k)}, d^{(k)}}(\tau^{(k)})$ where $\tau^{(k)}$ is determined by Algorithm 2
$\quad \tilde{d}^{(k)} := \dot{\gamma}_{x^{(k)}, d^{(k)}}(\tau^{(k)})$
$\quad g^{(k+1)} := \nabla_{\mathbb{X}^N} F(x^{(k+1)})$
$\quad \beta^{(k)} := \begin{cases} \dfrac{\langle \tilde{d}^{(k)}, H_{\mathbb{X}^N} F(x^{(k+1)}) g^{(k+1)} \rangle}{\langle \tilde{d}^{(k)}, H_{\mathbb{X}^N} F(x^{(k+1)}) \tilde{d}^{(k)} \rangle}, & \langle \tilde{d}^{(k)}, H_{\mathbb{X}^N} F(x^{(k+1)}) \tilde{d}^{(k)} \rangle \neq 0, \\ 0, & \text{else} \end{cases}$
$\quad d^{(k+1)} := -g^{(k+1)} + \beta^{(k)} \tilde{d}^{(k)}$
$\quad$ **if** $\langle d^{(k+1)}, g^{(k+1)} \rangle > 0$ or $(k+1) \equiv r \mod N \dim(\mathbb{X})$ **then**
$\quad\quad d^{(k+1)} = -g^{(k+1)}$
$\quad\quad r := k + 1$
**Output:** iteration sequence $x^{(0)}, x^{(1)}, \cdots \in \mathbb{X}^N$

---

### Algorithm 2 (Armijo Line Search)

**Parameters:** $0 < a < \frac{1}{2}$, $0 < b < 1$, maximal iterations $k_{\max} \in \mathbb{N}$
**Input:** smooth function $F : \mathbb{X}^N \to [0, \infty)$, start point $x \in \mathbb{X}^N$, descent direction $d \in T_x \mathbb{X}^N$
**Initialization:** $k := 0$,

$$\tau^{(0)} := \begin{cases} \left| \dfrac{\langle d, \nabla_{\mathbb{X}^N} F(x) \rangle}{\langle d, H_{\mathbb{X}^N} F(x) d \rangle} \right|, & \langle d, H_{\mathbb{X}^N} F(x) d \rangle \neq 0, \\ 1, & \text{else} \end{cases}$$

**while** $f \circ \gamma_{x,d}(\tau^{(k)}) - F(x) \geq a \tau^{(k)} \langle \nabla_{\mathbb{X}^N} F(x), d \rangle$ and $k < k_{\max}$ **do**
$\quad \tau^{(k+1)} := b \tau^{(k)}$
$\quad k := k + 1$
**Output:** $\tau^{(k)}$ (success if $k \leq k_{\max}$)

where $\lambda > 0$. Our algorithm of choice is the nonlinear conjugate gradient (CG) method with Armijo line search as outlined in Algorithm 1 with notation and implementation details described in the comments after Remark 6, see [25] for Euclidean spaces. Note that the notation is independent of the special choice of $\mathbb{X}$ in our comments. The proposed method is of "exact conjugacy" and uses the second order derivative information provided by the Hessian. For the Armijo line search itself, the sophisticated initialization in Algorithm 2 is used, which also incorporates second order information via the Hessian. The main advantage of the CG method is its simplicity together with fast convergence at low computational cost. Indeed, Algorithm 1, together with Algorithm 2 replaced by an exact line search, converges under suitable assumptions superlinearly, more precisely $dN$-step quadratically towards a local minimum, cf. [73, Thm. 5.3] and [43, Sec. 3.3.2, Thm. 3.27].

**Remark 6** The objective in (31) violates the smoothness requirements whenever $x_{k-1} = x_k$ or $\mathrm{dist}_{\mathbb{X}}(x_{k-1}, x_k) = L/N$. However, we observe numerically that local minimizers of (31) do not belong to this set of measure zero. This means in turn, if a local minimizer has a positive definite Hessian, then there is a local neighborhood where the CG method (with exact line search) permits a superlinear convergence rate. We do indeed observe this behavior in our numerical experiments.

Let us briefly comment on Algorithm 1 for $\mathbb{X} \in \{\mathbb{T}^2, \mathbb{T}^3, \mathbb{S}^2, \mathrm{SO}(3), \mathcal{G}_{2,4}\}$ which are considered in our numerical examples. For additional implementation details we refer to [43]. By $\gamma_{x,d}$ we denote the geodesic with $\gamma_{x,d}(0) = x$ and $\dot{\gamma}_{x,d}(0) = d$. Besides evaluating the geodesics $\gamma_{x^{(k)},d^{(k)}}(\tau^{(k)})$ in the first iteration step, we have to compute the parallel transport of $d^{(k)}$ along the geodesics in the second step. Furthermore, we need to compute the Riemannian gradient $\nabla_{\mathbb{X}^N} F$ and products of the Hessian $H_{\mathbb{X}^N} F$ with vectors $d$, which are approximated by the finite difference

$$H_{\mathbb{X}^N} F(x)d \approx \tfrac{\|d\|}{h}\Big(\nabla_{\mathbb{X}^N} F\big(\gamma_{x,hd/\|d\|}\big) - \nabla_{\mathbb{X}^N} F(x)\Big), \qquad h := 10^{-8}.$$

The computation of the gradient of the *penalty term* in (30) is done by applying the chain rule and noting that for $x \mapsto \mathrm{dist}_{\mathbb{X}}(x, y)$, we have $\nabla_{\mathbb{X}} \mathrm{dist}_{\mathbb{X}}(x, y) = \log_x y / \mathrm{dist}_{\mathbb{X}}(x, y)$, $x \neq y$ with the logarithmic map log on $\mathbb{X}$, while the distance is not differentiable for $x = y$. Concerning the later point, see Remark 5. The evaluation of the gradient of the penalty term at a point in $\mathbb{X}^N$ requires only $\mathcal{O}(N)$ arithmetic operations. The computation of the Riemannian gradient of the *data term* in (30) is done analytically via the gradient of the eigenfunctions $\varphi_k$ of the Laplace–Beltrami operator. Then, the evaluation of the gradient of the whole data term at given points can be done efficiently by *fast Fourier transform* (FFT) techniques at non-equispaced nodes using the NFFT software package of Potts et al. [56]. The overall complexity of the algorithm and references for the computation details for the above manifolds are given in Table 1.

**Table 1** References for implementation details of Algorithm 1 (left) and arithmetic complexity for the evaluations per iteration for the different manifolds (right)

| $\mathbb{X}$ | References | Complexity |
|---|---|---|
| $\mathbb{T}^d$ | [46], [43, Sect. 5.2.1] | $\mathcal{O}(r^d \log(r) + N)$ |
| $\mathbb{S}^2$ | [45,46], [43, Sect. 5.2.2] | $\mathcal{O}(r^2 \log^2(r) + N)$ |
| SO(3) | [42,44], [43, Sect. 5.2.3] | $\mathcal{O}(r^3 \log^2(r) + N)$ |
| $\mathcal{G}_{2,4}$ | [26] | $\mathcal{O}(r^4 \log^2(r) + N)$ |

## 8 Numerical Results

In this section, we underline our theoretical results by numerical examples. We start by studying the parameter choice in our numerical model. Then, we provide examples for the approximation of absolutely continuous measures with densities in $H^s(\mathbb{X})$, $s > d/2$, by push-forward measures of the Lebesgue measure on $[0, 1]$ by Lipschitz curves for the manifolds $\mathbb{X} \in \{\mathbb{T}^2, \mathbb{T}^3, \mathbb{S}^2, \text{SO}(3), G_{2,4}\}$. Supplementary material can be found on our webpage.

### 8.1 Parameter Choice

We like to emphasize that the optimization problem (31) is highly nonlinear and the objective function has a large number of local minimizers, which appear to increase exponentially in N. In order to find for fixed $L$ reasonable (local) solutions of (26), we carefully adjust the parameters in problem (31), namely the number of points $N$, the polynomial degree $r$ in the kernel truncation, and the penalty parameter $\lambda$. In the following, we suppose that $\dim(\text{supp}(\mu)) = d \geq 2$.

(i) **Number of points** $N$ Clearly, $N$ should not be too small compared to $L$. However, from a computational perspective it should also be not too large since the optimization procedure is hampered by the vast number of local minimizers. From the asymptotic of the path lengths of TSP in Lemma 3, we conclude that $N \gtrsim \vdash(\gamma)^{d/(d-1)}$ is a reasonable choice, where $\vdash(\gamma) \leq L$ is the length of the resulting curve $\gamma$ going through the points.

(ii) **Polynomial degree** $r$ Based on the proofs of the theorems in Sect. 5.4 it is reasonable to choose

$$r \sim L^{\frac{1}{d-1}} \sim N^{\frac{1}{d}}.$$

(iii) **Penalty parameter** $\lambda$ If $\lambda$ is too small, we cannot enforce that the points approximate a regular curve, i.e., $L/N \gtrsim \text{dist}_{\mathbb{X}}(x_{k-1}, x_k)$. Otherwise, if $\lambda$ is too large the optimization procedure is hampered by the rigid constraints. Hence, to find a reasonable choice for $\lambda$ in dependence on $L$, we assume that the minimizers of (31) treat both terms proportionally, i.e., for $N \to \infty$ both terms are of the same order. Therefore, our heuristic is to choose the parameter $\lambda$ such that

$$\min_{x_1,\dots,x_N} \mathscr{D}_K^2\left(\mu, \frac{1}{N}\sum_{k=1}^N \delta_{x_k}\right) \sim N^{-\frac{2s}{d}} \sim \frac{\lambda}{N}\sum_{k=1}^N \left(N\,\text{dist}_{\mathbb{X}}(x_{k-1}, x_k) - L\right)_+^2.$$

On the other hand, assuming that for the length $\ell(\gamma) = \sum_{k=1}^{N} \mathrm{dist}_{\mathbb{X}}(x_{k-1}, x_k)$ of a minimizer $\gamma$ we have $\ell(\gamma) \sim L \sim N^{(d-1)/d}$, so that $N \, \mathrm{dist}_{\mathbb{X}}(x_{k-1}, x_k) \sim L$, the value of the penalty term behaves like

$$\frac{\lambda}{N} \sum_{k=1}^{N} \left( N \, \mathrm{dist}_{\mathbb{X}}(x_{k-1}, x_k) - L \right)_{+}^{2} \sim \lambda L^2 \sim \lambda N^{\frac{2d-2}{d}}.$$

Hence, a reasonable choice is

$$\lambda \sim L^{\frac{-2s-2(d-1)}{d-1}} \sim N^{\frac{-2s-2(d-1)}{d}}. \tag{32}$$

**Remark 7** In view of Remark 5 the relations in i)-iii) become

$$N \sim L^{\frac{d_\mu}{d_\mu - 1}}, \quad r \sim N^{\frac{1}{d_\mu}} \sim L^{\frac{1}{d_\mu - 1}}, \quad \lambda \sim L^{\frac{-2s - 3d_\mu + d + 2}{d_\mu - 1}} \sim N^{\frac{-2s - 3d_\mu + d + 2}{d_\mu}}.$$

In the rest of this subsection, we aim to provide some numerical evidence for the parameter choice above. We restrict our attention to the torus $\mathbb{X} = \mathbb{T}^2$ and the kernel $K$ given in (34) with $d = 2$ and $s = 3/2$. Choose $\mu$ as the Lebesgue measure on $\mathbb{T}^2$. From (32), we should keep in mind $\lambda \sim N^{-5/2} \sim L^{-5}$.

**Influence of $N$ and $\lambda$** We fix $L = 4$ and a large polynomial degree $r = 128$ for truncating the kernel. For any $\lambda_i = 0.1 \cdot 2^{-5i/2}$, $i = 1, \ldots, 4$, we compute local minimizers with $N_j = 10 \cdot 2^j$, $j = 1, \ldots, 4$. More precisely, keeping $\lambda_i$ fixed we start with $N_1 = 20$ and refine successively the curves by inserting the midpoints of the line segments connecting consecutive points and applying a local minimization with this initialization. The results are depicted in Fig. 1. For fixed $\lambda$ (fixed row) we can clearly notice that the local minimizers converge towards a smooth curve for increasing $N$. Moreover, the diagonal images correspond to the choice $\lambda = 0.1(N/10)^{-5/2}$, where we can already observe good approximation of the curves emerging to the right of it. This should provide some evidence that the choice of the penalty parameter $\lambda$ and the number of points $N$ discussed above is reasonable. Indeed, for $\lambda \to \infty$ we observe $L(\gamma) \to \ell(\gamma) \to L = 4$.

**Influence of the polynomial degree $r$** In Fig. 2 we illustrate the local minimizers of (31) for fixed Lipschitz parameters $L_i = 2^i$ and corresponding regularization weights $\lambda_i = 0.2 \cdot L_i^{-5}$, $i = 1, \ldots, 4$, (rows) in dependence on the polynomial degrees $r_j = 8 \cdot 2^j$, $j = 1, \ldots, 5$ (columns). According to the previous experiments, it seems reasonable to choose $N = 20L^2$. Note, that the (numerical) choice of $\lambda$ leads to curves with length $\ell(\gamma) \approx 2L$. In Fig. 2 we observe that for $r = cL$ the corresponding local minimizers have common features. For instance, if $c = 4$ (i.e., $r \approx \ell(\gamma)$) the minimizers have mostly vertical and horizontal line segments. Furthermore, for fixed $r$ it appears that the length of the curves increases linearly with $L$ until $L$ exceeds $2r$, from where it remains unchanged. This observation can be explained by the fact that there are curves of bounded length $cr$ which provide exact quadratures for degree $r$.
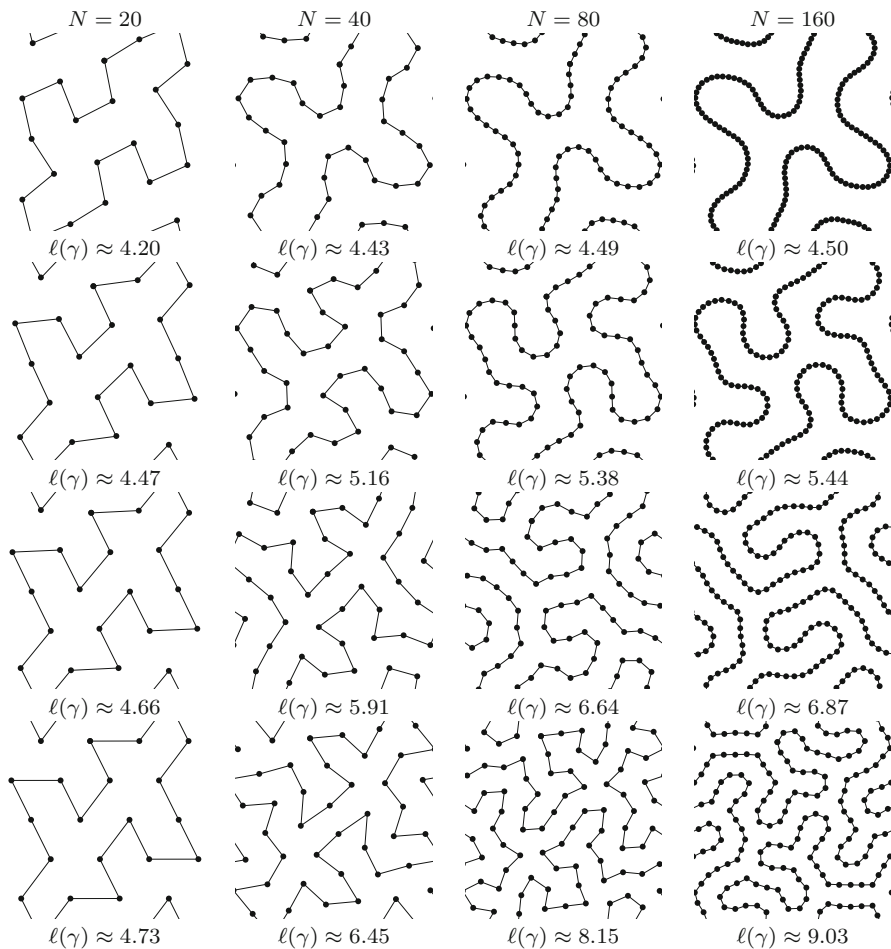
**Fig. 1** Influence of $N$ and $\lambda$ on local minimizers of (31) for the Lebesgue measure on $\mathbb{T}^2$, $L = 4$ and $r = 128$. Results for increasing $N$ (column-wise) and decreasing $\lambda = 0.1 \cdot 2^{-5i/2}$, $i = 1, \ldots, 4$, (row-wise). Here, the curve length increases for decreasing $\lambda$ or increasing $N$, until stagnation for sufficient small $\lambda$ or large $N$. For all minimizer the distance between consecutive points is around $\ell(\gamma)/N$

## 8.2 Quasi-Optimal Curves on Special Manifolds

In this subsection, we give numerical examples for $\mathbb{X} \in \{\mathbb{T}^2, \mathbb{T}^3, \mathbb{S}^2, \text{SO}(3), \mathcal{G}_{2,4}\}$. Since the objective function in (31) is highly non-convex, the main problem is to find nearly optimal curves $\gamma_L \in \mathcal{P}_L^{\vee\text{-curv}}(\mathbb{X})$ for increasing $L$. Our heuristic is as follows:

(i) We start with a curve $\gamma_{L_0} \colon [0, 1] \to \mathbb{X}$ of small length $\ell(\gamma) \approx L_0$ and solve the problem (31) for increasing $L_i = cL_{i-1}$, $c > 1$, where we choose the parameters $N_i$, $\lambda_i$ and $r_i$ in dependence on $L_i$ as described in the previous subsection. In each step a local minimizer is computed using the CG method with 100 iterations.
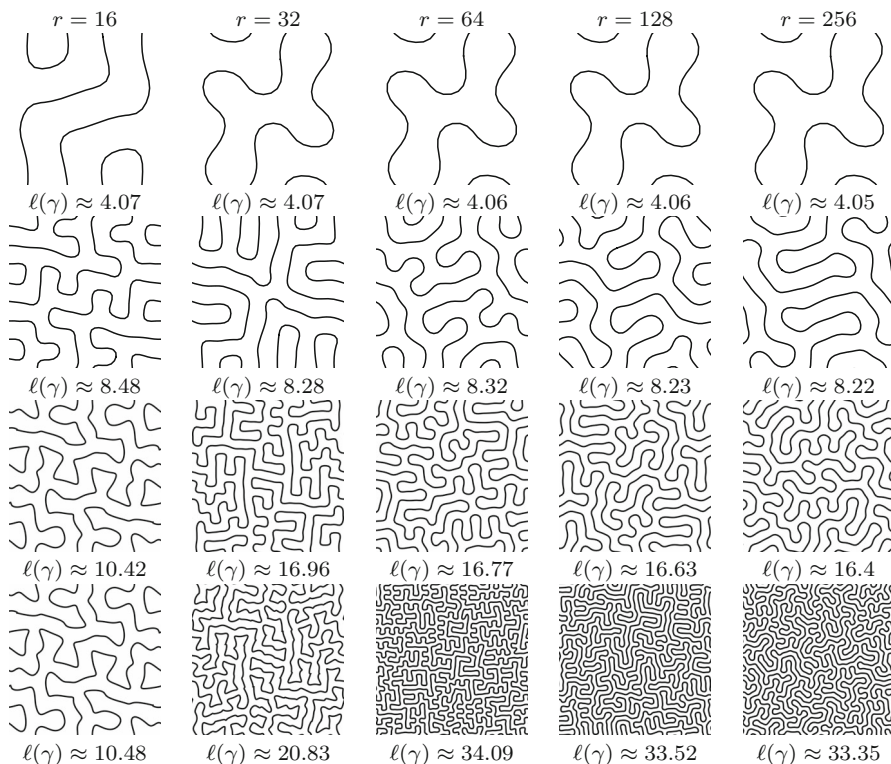
**Fig. 2** Influence of $r$ on the local minimizer of (31) for the Lebesgue measure on $\mathbb{T}^2$. Column-wise we increase $r = 16, 32, 64, 128, 256$ and row-wise we increase $L = 2, 4, 8, 16$, where $\lambda = 0.2L^{-5}$ and $N = 20L^2$. Note that the degree $r$ steers the resolution of the curves. It appears that the spacing of the curves is bounded by $r^{-1}$

Then, the obtained minimizer $\gamma_i$ serves as the initial guess in the next step, which is obtained by inserting the midpoints.

(ii) In case that the resulting curves $\gamma_i$ have non-constant speed, each is refined by increasing $\lambda_i$ and $N_i$. Then, the resulting problem is solved with the CG method and $\gamma_i$ as initialization. Details on the parameter choice are given in the according examples.

The following examples show that this recipe indeed enables us to compute "quasi-optimal" curves, meaning that the obtained minimizers have optimal decay in the discrepancy.

**2d-Torus** $\mathbb{T}^2$ In this example we illustrate how well a gray-valued image (considered as probability density) may be approximated by an almost constant speed curve. The original image of size $170 \times 170$ is depicted in the bottom-right corner of Fig. 3. Its Fourier coefficients $\hat{\mu}_{k_1,k_2}$ are computed by a discrete Fourier transform (DFT) using the FFT algorithm and normalized appropriately. The kernel $K$ is given by (34) with $d = 2$ and $s = 3/2$.

We start with $N_0 = 96$ points on a circle given by the formula

$$x_{0,k} = \left( \tfrac{1}{5} \cos(2\pi k/N_0), \tfrac{1}{5} \sin(2\pi k/N_0) \right), \qquad k = 0, \ldots, N_0.$$

Then, we apply our procedure for $i = 0, \ldots, 11$ with parameters

$$L_i = 0.97 \cdot 2^{\frac{i+5}{2}}, \quad \lambda_i = 100 \cdot L_i^{-5}, \quad N_i = 96 \cdot 2^i \sim L_i^2 \quad r_i = \lfloor 2^{\frac{i+11}{2}} \rfloor \sim L_i,$$

chosen such that the length of the local minimizer $\gamma_i$ satisfies $\ell(\gamma_i) \approx 2^{(i+5)/2}$ and the maximal speed is close to $L_i$.

To get nearly constant speed curves $\gamma_i$, see ii), we increase $\lambda_i$ by a factor of 100, $N_i$ by a factor of 2 and set $L_i := 2^{(i+5)/2}$. Then, we apply the CG method with maximal 100 iterations and $i$ restarts. The results are depicted in Fig. 3. Note that the complexity for the evaluation of the function in (31) scales roughly as $N \sim L^2$. In Fig. 4 we observe that the decay-rate of the squared discrepancy $\mathscr{D}_K^2(\mu, \nu)$ in dependence on the Lipschitz constant $L$ matches indeed the theoretical findings of Theorem 8.

**3D-Torus $\mathbb{T}^3$** The aim of this example is two-fold. First, it shows that the algorithm works pretty well in three dimensions. Second, we are able to approximate any compact surface in the three-dimensional space by a curve. We construct a measure $\mu$ supported around a two-dimensional surface by taking samples from Spock's head[2] and placing small Gaussian peaks at the sampling points, i.e., the density is given for $x \in [-\tfrac{1}{2}, \tfrac{1}{2}]$ by

$$\rho(x) := c^{-1} \sum_{p \in S} e^{-30000\|p-x\|_2^2}, \qquad c := \int_{[-\frac{1}{2}, \frac{1}{2}]^3} \sum_{p \in S} e^{-30000\|p-x\|_2^2} \, dx,$$

where $S \subset [-\tfrac{1}{2}, \tfrac{1}{2}]^3$ is the discrete sampling set. From a numerical point of view it holds $\dim(\mathrm{supp}(\mu)) = 2$. The Fourier coefficients are again computed by a DFT and the kernel $K$ is given by (34) with $d = 3$ and $s = 2$ so that $H_K = H^2(\mathbb{T}^3)$.

We start with $N_0 = 100$ points on a smooth curve given by the formula

$$x_{0,k} = \left( \tfrac{3}{10} \cos(2\pi k/N_0), \tfrac{3}{10} \sin(2\pi k/N_0), \tfrac{3}{10} \sin(4\pi k/N_0) \right), \qquad k = 0, \ldots, N_0.$$

Then, we apply our procedure for $i = 0, \ldots, 8$ with parameters, cf. Remark 7,

$$L_i = 2^{\frac{i+5}{2}}, \quad \lambda_i = 10 \cdot L_i^{-5}, \quad N_i = 100 \cdot 2^i \sim L_i^2, \quad r_i = \lfloor 2^{\frac{i+5}{2}} \rfloor \sim L_i.$$

To get nearly constant speed curves $\gamma_i$, we increase $\lambda_i$ by a factor of 100, $N_i$ by a factor of 2 and set $L_i := 2^{(i+6)/2}$. Then, we apply the CG method with maximal 100 iterations and one restart to the previously found curve $\gamma_i$. The results are illustrated in Fig. 5. Note that the complexity of the function evaluation in (31) scales roughly as $N^{3/2} \sim L^3$. In Fig. 6 we depict the squared discrepancy $\mathscr{D}_K^2(\mu, \nu)$ of the computed
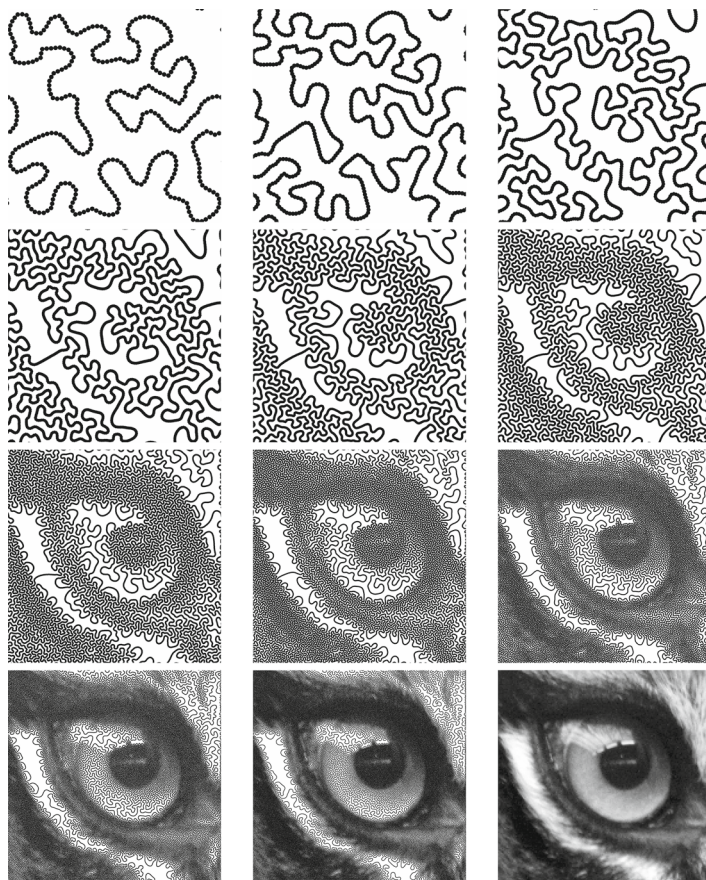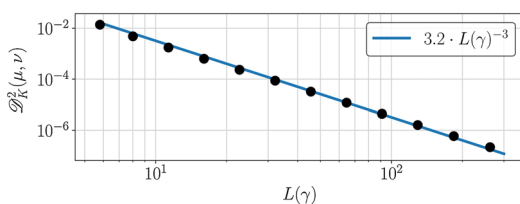
---

Springer

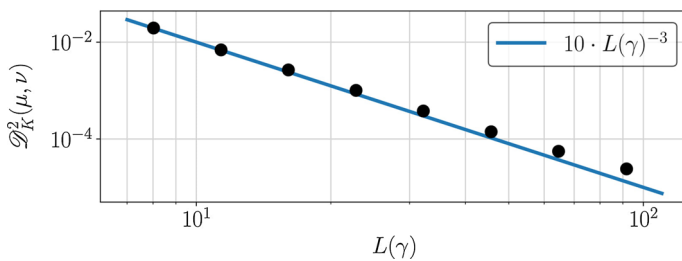**Fig. 3** Local minimizers of (31) for the image at bottom right

**Fig. 4** Squared discrepancy
between the measure $\mu$ given by
the image in Fig. 3 and the
computed local minimizers
(black dots) on $\mathbb{T}^2$ in log-scale.
The blue line corresponds to the
optimal decay-rate in Theorem 8



curves. For small Lipschitz constants, say $L(\gamma) \leq 50$, we observe a decrease of
approximately $L(\gamma)^{-3}$, which matches the optimal decay-rate for measures supported
on surfaces as discussed in Remark 5.

**2-Sphere** $\mathbb{S}^2$ Next, we approximate a gray-valued image on the sphere $\mathbb{S}^2$ by an
almost constant speed curve. The image represents the earth's elevation data provided
by MATLAB, given by samples $\rho_{i,j}, i = 1, \ldots, 180, \; j = 1, \ldots, 360$, on the grid

$$x_{i,j} := \left( \sin\left(i\,\tfrac{\pi}{180}\right) \sin\left(j\,\tfrac{\pi}{180}\right), \sin\left(i\,\tfrac{\pi}{180}\right) \cos\left(j\,\tfrac{\pi}{180}\right), \cos\left(i\,\tfrac{\pi}{180}\right) \right).$$
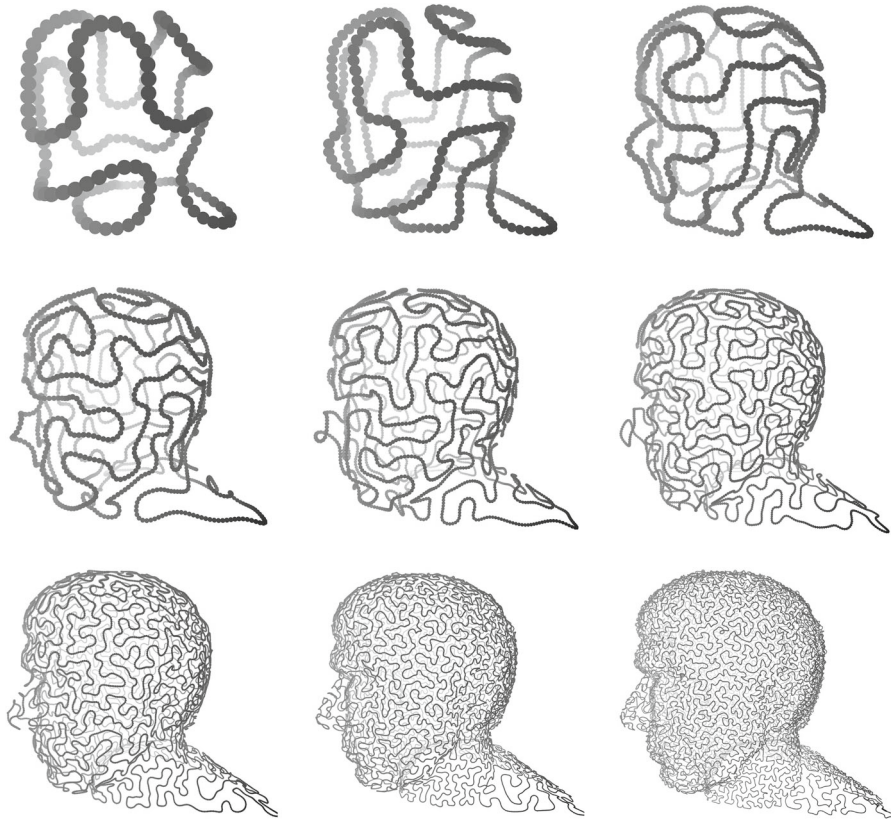
**Fig. 5** Local minimizers of (31) for a measure $\mu$ concentrated on a surface (head of Spock) in $\mathbb{T}^3$



**Fig. 6** Squared discrepancy between the measure $\mu$ given by the surface in Fig. 5 and the computed local minimizers (black dots) on $\mathbb{T}^3$ in log-scale. The blue line corresponds to the optimal decay-rate in Theorem 8

The Fourier coefficients are computed by discretizing the Fourier integrals, i.e.,

$$\hat{\mu}_k^m := \begin{cases} \frac{1}{180 \cdot 360} \sum_{i=1}^{180} \sum_{j=1}^{360} \rho_{i,j} \overline{Y_k^m(x_{i,j})} \sin\left(i \frac{\pi}{180}\right), & 1 \le k \le 2m+1, m \le 180, \\ 0, & \text{else}, \end{cases}$$
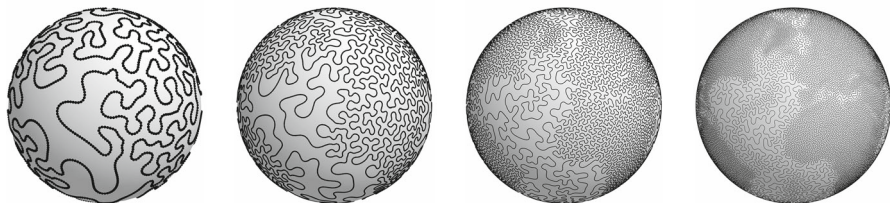
**Fig. 7** Local minimizers of (31) for $\mu$ given by the earth's elevation data on the sphere $\mathbb{S}^2$
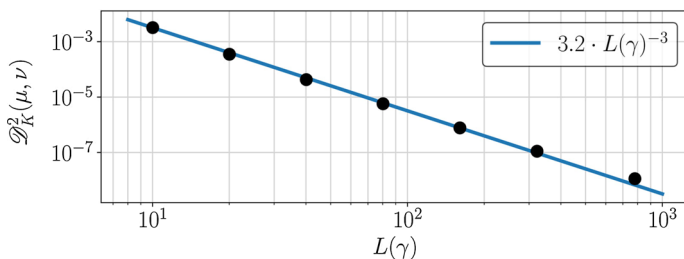


**Fig. 8** Squared discrepancy between the measure $\mu$ and the computed local minimizers (black dots) in log-scale. The blue line corresponds to the optimal decay-rate in Theorem 9

followed by a suitable normalization such that $\hat{\mu}_0^0 = 1$. The corresponding sums are efficiently computed by an adjoint non-equispaced fast spherical Fourier transform (NFSFT), see [68]. The kernel $K$ is given by (36). Similar to the previous examples, we apply our procedure for $i = 0, \ldots, 12$ with parameters

$$L_i = 9.7 \cdot 2^{\frac{i}{2}}, \quad \lambda_i = 100 \cdot L_i^{-5}, \quad N_i = 100 \cdot 2^i \sim L_i^2, \quad r_i = \lfloor L_i \rfloor \sim L_i.$$

To get nearly constant speed curves, we increase $\lambda_i$ by a factor of 100, $N_i$ by a factor of 2 and set $L_i := L_0 2^{i/2}$. Then, we apply the CG method with maximal 100 iterations and one restart to the previously constructed curves $\gamma_i$. The results for $i = 6, 8, 10, 12$ are depicted in Fig. 7. Note that the complexity of the function evaluation in (31) scales roughly as $N \sim L^2$. In Fig. 8 we observe that the decay-rate of the squared discrepancy $\mathscr{D}_K^2(\mu, \nu)$ in dependence on the Lipschitz constant matches indeed the theoretical findings in Theorem 9.

**3D-Rotations** SO(3) There are several possibilities to parameterize the rotation group SO(3). We apply those by Euler angles and an axis-angle representation for visualization. Euler angles $(\varphi_1, \theta, \varphi_2) \in [0, 2\pi) \times [0, \pi] \times [0, 2\pi)$ correspond to rotations $\text{Rot}(\varphi_1, \theta, \varphi_2)$ in SO(3) that are the successive rotations around the axes $e_3, e_2, e_3$ by the respective angles. Then, the Haar measure of SO(3) is determined by

$$\mathrm{d}\mu_{\mathrm{SO}(3)}(\varphi_1, \theta, \varphi_2) = \tfrac{1}{8\pi^2} \sin(\theta)\, \mathrm{d}\varphi_1\, \mathrm{d}\theta\, \mathrm{d}\varphi_2.$$

We are interested in the full three-dimensional doughnut

$$D = \left\{ \text{Rot}(\varphi_1, \theta, \varphi_2) \; : \; 0 \leq \theta \leq \tfrac{\pi}{2}, \; 0 \leq \varphi_1, \varphi_2 \leq 2\pi \right\} \subset \text{SO}(3).$$

Next, we want to approximate the Haar measure $\mu = \mu_D$ restricted to $D$, i.e., with normalization we consider the measure defined for $f \in C(SO(3))$ by

$$\int_{SO(3)} f \, d\mu_D = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \int_0^{2\pi} f(\varphi_1, \theta, \varphi_2) \sin(\theta) \, d\varphi_1 \, d\theta \, d\varphi_2.$$

The Fourier coefficients of $\mu_D$ can be explicitly computed by

$$\hat{\mu}_{l,l'}^k = \begin{cases} P_{k-1}(0) - P_{k+1}(0), & l, l' = 0, \ k \geq 0, \\ 0, & l, l' \neq 0, \end{cases}$$

where $P_k$ are the Legendre polynomials. The kernel $K$ is given by (37) with $d = 3$ and $s = 2$. For $i = 0, \ldots, 8$ the parameters are chosen as

$$L_i = 0.93 \cdot 2^{\frac{2i+12}{3}}, \quad \lambda_i = 10 \cdot L_i^{-4}, \quad N_i = 64 \cdot 2^i \sim L_i^2, \quad r_i = \lfloor 2^{\frac{i+9}{3}} \rfloor \sim L_i^{\frac{1}{2}}.$$

Here, we use a CG method with 100 iterations and one restart. Step ii) appears to be not necessary. Note that the complexity for the function evaluations in (31) scales roughly as $N \sim L^{3/2}$.

The constructed curves are illustrated in Fig. 9, where we utilized the following visualization: Every rotation $R(\alpha, r) \in SO(3)$ is determined by a rotation axis $r = (r_1, r_2, r_3) \in \mathbb{S}^2$ and a rotation angle $\alpha \in [0, \pi]$, i.e.,

$$R(\alpha, r)x = r(r^\mathsf{T}x) + \cos(\alpha)\,((r \times x) \times r) + \sin(\alpha)(r \times x).$$

Setting $q := (\cos(\frac{\alpha}{2}), \sin(\frac{\alpha}{2})r) \in \mathbb{S}^3$ with $r \in \mathbb{S}^2$ and $\alpha \in [0, 2\pi]$, see (22), we observe that the same rotation is generated by $-q = (\cos(\frac{2\pi - \alpha}{2}), \sin(\frac{2\pi - \alpha}{2}(-r)) \in \mathbb{S}^3$, in other words $SO(3) \cong \mathbb{S}^3/\{\pm 1\}$. Then, by applying the stereographic projection $\pi(q) = (q_2, q_3, q_4)/(1 + q_1)$, we map the upper hemisphere onto the three dimensional unit ball. Note that the equatorial plane of $\mathbb{S}^3$ is mapped onto the sphere $\mathbb{S}^2$, hence on the surface of the ball antipodal points have to be identified. In other words, the rotation $R(\alpha, r)$ is plotted as the point

$$\pi(q) = \frac{\sin(\frac{\alpha}{2})}{1 + \cos(\frac{\alpha}{2})} r = \tan(\frac{\alpha}{4})r \in \mathbb{R}^3.$$

In Fig. 10 we observe that the decay-rate of $\mathscr{D}_K^2(\mu, \nu)$ in dependence on the Lipschitz constant $L$ matches the theoretical findings in Corollary 1.

**The 4-dimensional Grassmannian $\mathcal{G}_{2,4}$** Here, we aim to approximate the Haar measure of the Grassmannian $\mathcal{G}_{2,4}$ by a curve of almost constant speed. As this curve samples the space $\mathcal{G}_{2,4}$ quite evenly, it could be used for the grand tour, a technique to analyze high-dimensional data by their projections onto two-dimensional subspaces, cf. [5].
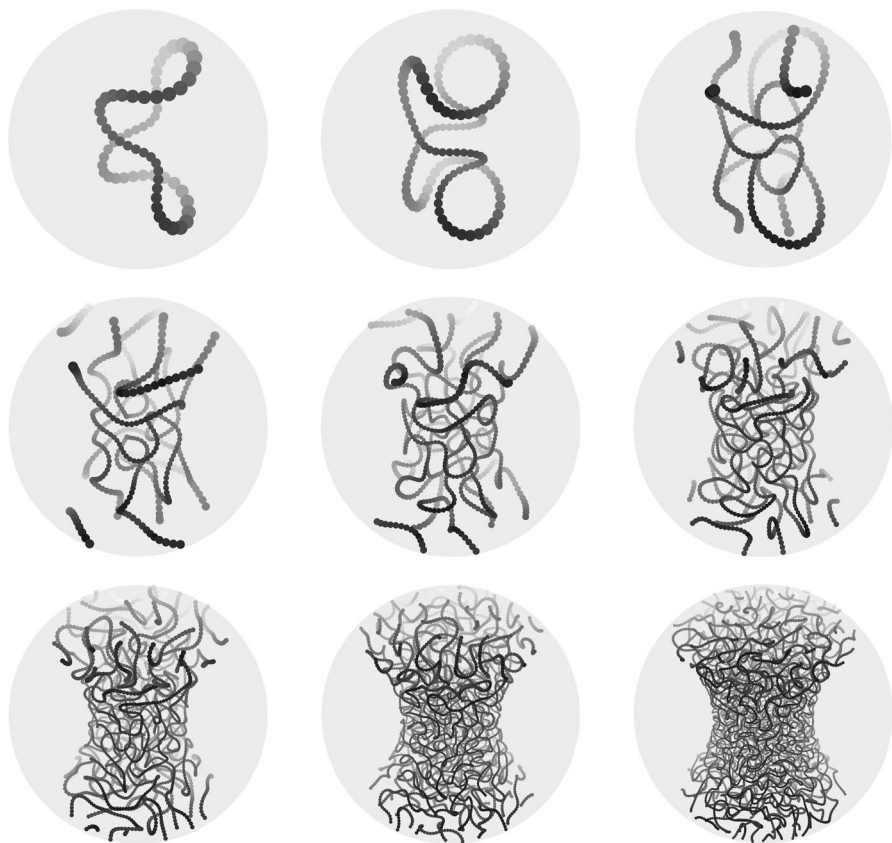
**Fig. 9** Local minimizers of (31) for the Haar measure $\mu_D$ of three-dimensional doughnut $D$ in the rotation group SO(3) with a color scheme for better visibility of the 3D structure
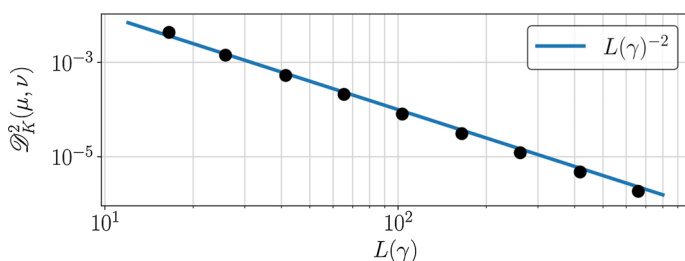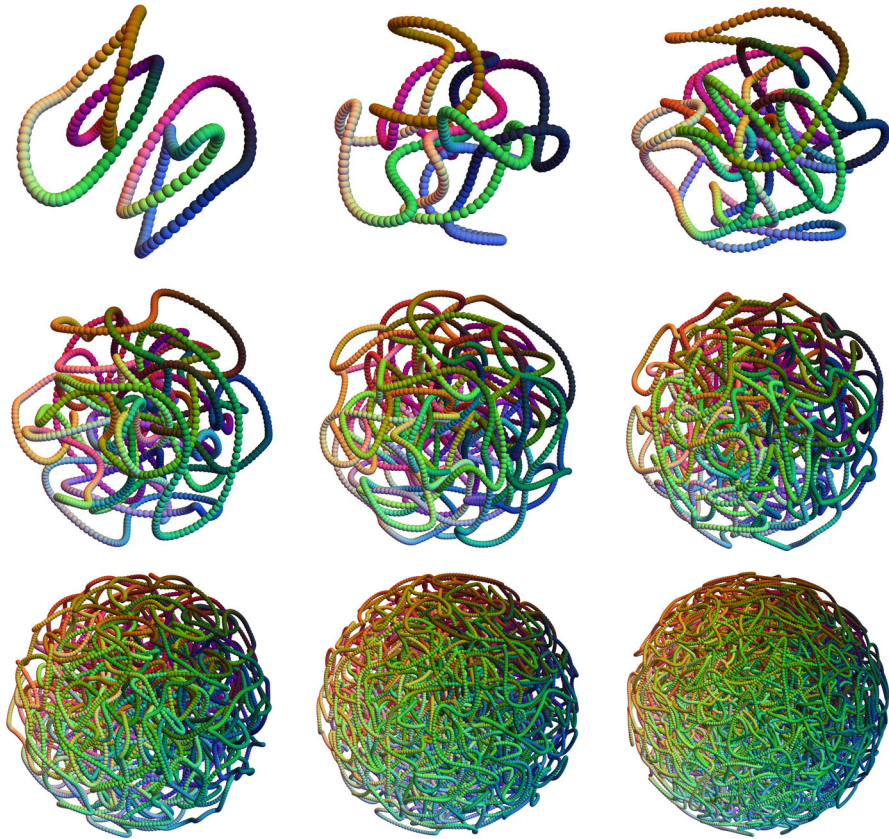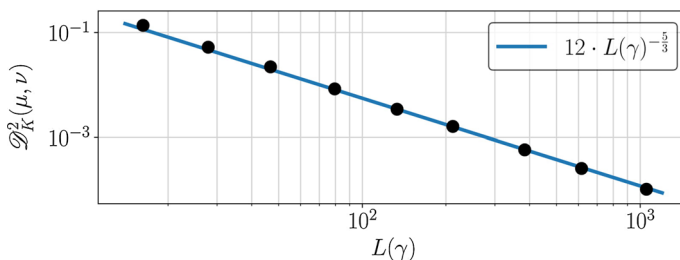


**Fig. 10** Squared discrepancy between the measure $\mu_D$ and the computed local minimizers (black dots) in log-scale. The blue line corresponds to the optimal decay-rate in Corollary 1

The kernel $K$ of the Haar measure is given by (38) and the Fourier coefficients are given by $\hat{\mu}_{m,m'}^{k,k'} = \delta_{m,0}\delta_{m',0}\delta_{k,0}\delta_{k',0}$. For $i = 0, \ldots, 8$ the parameters are chosen as

$$L_i = 0.91 \cdot 2^{\frac{3i+16}{4}}, \quad \lambda_i = 100 \cdot L_i^{-\frac{11}{3}}, \quad N_i = 128 \cdot 2^i \sim L_i^2, \quad r_i = \lfloor 2^{\frac{3i+16}{12}} \rfloor + 1 \sim L_i^{\frac{1}{3}}.$$

**Fig. 11** Local minimizers of (31) for the Haar measure of the Grassmannian $\mathcal{G}_{2,4}$



**Fig. 12** The squared discrepancy between the Haar measure $\mu$ and the computed local minimizers (black dots) in log-scale. Here, the blue line corresponds to the optimal decay-rate, cf. Theorem 10

Here, we use a CG method with 100 iterations and one restart. Our experiments suggest that step ii) is not necessary. Note that the complexity for the function evaluation in (31) scales roughly as $N \sim L^{3/2}$.

The computed curves are illustrated in Fig. 11, where we use the following visualization. By Remark 8, there exists an isometric one-to-one mapping $P : \mathbb{S}^2 \times \mathbb{S}^2 / \{\pm 1\} \rightarrow$

$\mathcal{G}_{2,4}$. Using this relation, we plot the point $P(u, v) \in \mathcal{G}_{2,4}$ by two antipodal points $z_1 = u + v$, $z_2 = -u - v \in \mathbb{R}^3$ together with the RGB color-coded vectors $\pm u$.[3] More precisely, $R = (1 \mp u_1)/2$, $G = (1 \mp u_2)/2$, $B = (1 \mp u_3)/2$. This means a curve $\gamma(t) \in \mathcal{G}_{2,4}$ only intersects itself if the corresponding curve $z(t) \in \mathbb{R}^3$ intersects and has the same colors at the intersection point. In Fig. 12 we observe that the decay-rate of the squared discrepancy $\mathscr{D}_K^2(\mu, \nu)$ in dependence on the Lipschitz constant $L$ matches indeed the theoretical findings in Theorem 10.

## 9 Conclusions

In this paper, we provided approximation results for general probability measures on compact Ahlfors $d$-regular metric spaces $\mathbb{X}$ by

  (i) measures supported on continuous curves of finite length, which are actually push-forward measures of probability measures on [0, 1] by Lipschitz curves;
 (ii) push-forward measures of absolutely continuous probability measures on [0, 1] by Lipschitz curves;
(iii) push-forward measures of the Lebesgue measure on [0, 1] by Lipschitz curves.

Our estimates rely on discrepancies between measures. In contrast to Wasserstein distances, these estimates do not reflect the curse of dimensionality.

In approximation theory, a natural question is how the approximation rates improve as the "measures become smoother". Therefore, we considered absolutely continuous probability measures with densities in Sobolev spaces, where we have to restrict ourselves to compact Riemannian manifolds $\mathbb{X}$. We proved lower estimates for all three approximation spaces i)-iii). Concerning upper estimates, we gave a result for the approximation space i). Unfortunately, we were not able to show similar results for the smaller approximation spaces ii) and iii). Nevertheless, for these cases, we could provide results for the $d$-dimensional torus, the $d$-sphere, the three-dimensional rotation group and the Grassmannian $\mathcal{G}_{2,4}$, which are all of interest on their own. Numerical examples on these manifolds underline our theoretical findings.

Our results can be seen as starting point for future research. Clearly, we want to have more general results also for the approximation spaces ii) and iii). We hope that our research leads to further practical applications. It would be also interesting to consider approximation spaces of measures supported on higher dimensional submanifolds as, e.g., surfaces.

Recently, results on the principal component analysis (PCA) on manifolds were obtained. It may be interesting to see if some of our approximation results can be also modified for the setting of principal curves, cf. Remark 3. In contrast to [55, Thm. 1] that bounds the discretization error for fixed length, we were able to provide precise error bounds for the discrepancy in dependence on the Lipschitz constant $L$ of $\gamma$ and the smoothness of the density $\mathrm{d}\mu$.

---

[3] Note that the decomposition of $z \in \mathbb{R}^3$ with $0 < \|z\| < 2$ into $u$ and $v$ is not unique. There is a one-parameter family of points $u_s, v_s \in \mathbb{S}^2$ such $z = u_s + v_s$. The point $z = 0$ has a two-dimensional ambiguity $v = -u$, $u \in \mathbb{S}^2$ and the point $z \in 2\mathbb{S}^2$ has a unique pre-image $v = u = \frac{1}{2}z$.

# A Special Manifolds

Here, we introduce the main examples that are addressed in the numerical part. The measure $\sigma_{\mathbb{X}}$ is always the normalized Riemannian measure on the manifold $\mathbb{X}$. Note that for simplicity of notation all eigenspaces are complex in this section. We are interested in the following special manifolds.

*Example 1:* $\mathbb{X} = \mathbb{T}^d$. For $\boldsymbol{k} \in \mathbb{Z}^d$, set $|\boldsymbol{k}|^2 := k_1^2 + \ldots + k_d^2$ and $|\boldsymbol{k}|_\infty := \max\{|k_1|, \ldots, |k_d|\}$. Then $-\Delta$ has eigenvalues $\{4\pi^2 |\boldsymbol{k}|^2\}_{\boldsymbol{k} \in \mathbb{Z}^d}$ with eigenfunctions $\{e^{2\pi i \langle \boldsymbol{k}, \cdot \rangle}\}_{\boldsymbol{k} \in \mathbb{Z}^d}$. The *space of $d$-variate trigonometric polynomials of degree $r$*,

$$\Pi_r(\mathbb{T}^d) := \mathrm{span}\{e^{2\pi i \langle \boldsymbol{k}, x \rangle} : |\boldsymbol{k}|_\infty \le r\} \tag{33}$$

has dimension $(2r+1)^d$ and contains the eigenspaces belonging to eigenvalues smaller than $4\pi^2 r^2$. As kernel for $H^s$, $s = (d+1)/2$, we use in our numerical examples

$$K(x, y) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} (1 + |\boldsymbol{k}|_2^2)^{-\frac{d+1}{2}} e^{2\pi i \langle \boldsymbol{k}, x-y \rangle} = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} (1 + |\boldsymbol{k}|_2^2)^{-\frac{d+1}{2}} \cos(2\pi \langle \boldsymbol{k}, x-y \rangle).$$
$$\tag{34}$$

*Example 2:* $\mathbb{X} = \mathbb{S}^d \subset \mathbb{R}^{d+1}$, $d \ge 1$. We use distance $\mathrm{dist}_{\mathbb{S}^d}(x, z) = \arccos(\langle x, z \rangle)$. The Laplace–Beltrami operator $-\Delta$ on $\mathbb{S}^d$ has the eigenvalues $\{k(k + d - 1)\}_{k \in \mathbb{N}}$ with the *spherical harmonics of degree $k$*,

$$\{Y_l^k : l = 1, \ldots, Z(d, k)\}, \quad Z(d, k) := (2k + d - 1)\frac{\Gamma(k+d-1)}{\Gamma(d)\Gamma(k+1)}$$

as corresponding orthonormal eigenfunctions [66]. The span of eigenfunctions with eigenvalues smaller than $r(r + d - 1)$ is given by

$$\Pi_r(\mathbb{S}^d) := \mathrm{span}\{Y_l^k : k = 0, \ldots, r, \, l = 1, \ldots, Z(d, k)\}. \tag{35}$$

It has dimension $\sum_{k=0}^{r} Z(d,k) = \frac{(d+2r)\Gamma(d+r)}{\Gamma(d+1)\Gamma(r+1)} \sim r^d$ and coincides with the space of polynomials of total degree $r$ in $d$ variables restricted to the sphere. As kernel for $H^s(\mathbb{S}^2)$, $s = 3/2$, we use

$$
K(x,y) = \frac{1}{3} + \sum_{k=1}^{\infty} \frac{2}{(2k-1)(2k+1)(2k+3)} \sum_{l=1}^{2k+1} Y_l^k(x)\overline{Y_l^k(y)}
$$

$$
= \frac{1}{3} + \sum_{k=1}^{\infty} \frac{2}{(2k-1)(2k+3)} P_k(\langle x,y\rangle) = 1 - \frac{1}{2}\|x - y\|_2 \quad (36)
$$

with the Legendre polynomials $P_k$. Note that the coefficients have the decay as $(k(k+1))^{-3/2}$.

*Example 3:* $\mathbb{X} = \mathrm{SO}(3)$. This 3-dimensional manifold is equipped with the distance function $\mathrm{dist}_{\mathrm{SO}(3)}(x,y) = \arccos((\mathrm{trace}(x^{\top}y) - 1)/2)/2$. The eigenvalues of $-\Delta$ are $\{k(k+1)\}_{k=0}^{\infty}$ and the (normalized) *Wigner-$\mathcal{D}$ functions* $\{\mathcal{D}_{l,l'}^k : l,l' = -k, \ldots, k\}$ provide an orthonormal basis for $L^2(\mathrm{SO}(3))$, cf. [80]. The span of eigenspaces belonging to eigenvalues smaller than $r(r+1)$ is

$$
\Pi_r(\mathrm{SO}(3)) := \mathrm{span}\{\mathcal{D}_{l,l'}^k : k = 0, \ldots, r, \ l, l' = -k, \ldots, k\}
$$

and has dimension $(r+1)(2r+1)(2r+3)/3$. In the numerical part, we use the following kernel for $H^s(\mathrm{SO}(3))$, $s = 2$,

$$
K(x,y) = \frac{\pi}{8} - \frac{1}{3} + \sum_{k=1}^{\infty} \frac{1}{(2k-1)(2k+1)^2(2k+3)} \sum_{l=-k}^{k}\sum_{l'=-k}^{k} \mathcal{D}_{l,l'}^k(x)\overline{\mathcal{D}_{l,l'}^k(y)}
$$

$$
= \frac{\pi}{8} - \frac{1}{3} + \sum_{k=1}^{\infty} \frac{1}{(2k-1)(2k+1)(2k+3)} U_{2k}\left(\tfrac{1}{2}\sqrt{\mathrm{tr}(x^{\top}y) + 1}\right)
$$

$$
= \frac{\pi}{8} - \pi \frac{\sqrt{2}}{16}\|x - y\|_{\mathrm{F}}, \quad (37)
$$

where $U_k$ are the Chebyshev polynomials of the second kind.

*Example 4:* $\mathbb{X} = \mathcal{G}_{2,4}$. For integers $1 \leq s < r$, the $(s,r)$-Grassmannian is the collection of all $s$-dimensional linear subspaces of $\mathbb{R}^r$ and carries the structure of a closed Riemannian manifold. By identifying a subspace with the orthogonal projector onto this subspace, the Grassmannian becomes

$$
\mathcal{G}_{s,r} := \{x \in \mathbb{R}^{r \times r} : x^{\top} = x, \ x^2 = x, \ \mathrm{rank}(x) = s\}.
$$

In our context, the cases $\mathcal{G}_{1,2}$, $\mathcal{G}_{1,3}$, and $\mathcal{G}_{2,3}$ can essentially be treated by the spheres $\mathbb{S}^1$ and $\mathbb{S}^2$. The simplest Grassmannian that is algebraically different is $\mathcal{G}_{2,4}$. It is a 4-dimensional manifold and the geodesic distance between $x, y \in \mathcal{G}_{2,4}$ is given by

$$
\mathrm{dist}_{\mathcal{G}_{2,4}}(x,y) = \sqrt{2}\sqrt{\theta_1^2(x,y) + \theta_2^2(x,y)},
$$

where $\theta_1(x, y)$ and $\theta_2(x, y)$ are the principal angles between the subspaces associated to $x$ and $y$, respectively. The terms $\cos(\theta_1(x, y))^2$ and $\cos(\theta_2(x, y))^2$ correspond to the two largest singular values of the product $xy$. The eigenvalues of $-\Delta$ on $\mathcal{G}_{2,4}$ are $4(\lambda_1^2 + \lambda_2^2 + \lambda_1)$, where $\lambda_1$ and $\lambda_2$ run through all integers with $\lambda_1 \geq \lambda_2 \geq 0$, cf. [6–8,29,53,71]. The associated eigenfunctions are denoted by $\varphi_l^\lambda$ with $l = 1, \ldots, Z(\lambda)$, where $Z(\lambda) = (1 + \lambda_1 + \lambda_2)\eta(\lambda_2)$ and $\eta(\lambda_2) = 1$ if $\lambda_2 = 0$ and $2$ if $\lambda_2 > 0$ cf. [35, (24.29) and (24.41)] as well as [7,8].

The space of polynomials of total degree $r$ on $\mathbb{R}^{16} \cong \mathbb{R}^{4 \times 4}$ restricted to $\mathcal{G}_{2,4}$ is

$$\Pi_r(\mathcal{G}_{2,4}) := \text{span}\big\{\varphi_l^\lambda : \lambda_1 + \lambda_2 \leq r, \; l = 1, \ldots, Z(\lambda)\big\}.$$

It contains all eigenfunctions $\varphi_l^\lambda$ with $4(\lambda_1^2 + \lambda_2^2 + \lambda_1) < 2(r + 1)(r + 2)$, cf. [14, Thm. 5].

For $H^s(\mathcal{G}_{2,4})$ with $s = 5/2$, we chose the kernel

$$K(x, y) = \sum_{\lambda_1 \geq \lambda_2 \geq 0} \left(1 + \lambda_1^2 + \lambda_2^2\right)^{-\frac{5}{2}} \sum_{l=1}^{Z(\lambda)} \varphi_l^\lambda(x)\overline{\varphi_l^\lambda(y)}. \tag{38}$$

**Remark 8** It is well-known that $\mathbb{S}^2 \times \mathbb{S}^2$ is a double covering of $\mathcal{G}_{2,4}$. More precisely, there is an isometric one-to-one mapping $P : \mathbb{S}^2 \times \mathbb{S}^2/\{\pm 1\} \to \mathcal{G}_{2,4}$ given by

$$P(u, v) = P(-u, -v) := \frac{1}{2} \begin{pmatrix} 1 + u^{\mathrm{T}}v & -(u \times v)^{\mathrm{T}} \\ -u \times v & uv^{\mathrm{T}} + vu^{\mathrm{T}} + (1 - u^{\mathrm{T}}v)I_3 \end{pmatrix},$$

cf. [26]. Moreover, the $\varphi_l^\lambda$ are essentially tensor products of spherical harmonics, which enables transferring the non-equispaced fast Fourier transform from $\mathbb{S}^2 \times \mathbb{S}^2$ to $\mathcal{G}_{2,4}$, see [26] for details.

## References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008)
2. Akleman, E., Xing, Q., Garigipati, P., Taubin, G., Chen, J., Hu, S.: Hamiltonian cycle art: Surface covering wire sculptures and duotone surfaces. Comput. Graph. **37**(5), 316–332 (2013)
3. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, New York (2000)
4. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhäuser, Basel (2005)
5. Asimov, D.: The Grand Tour: A tool for viewing multidimensional data. SIAM J. Sci. Stat. Comput. **6**(1), 28–143 (1985)
6. Bachoc, C.: Linear programming bounds for codes in Grassmannian spaces. IEEE Trans. Inf. Th. **52**(5), 2111–2125 (2006)
7. Bachoc, C., Bannai, E., Coulangeon, R.: Codes and designs in Grassmannian spaces. Discrete Math. **277**(1-3), 15–28 (2004)
8. Bachoc, C., Coulangeon, R., Nebe, G.: Designs in Grassmannian spaces and lattices. J. Algebr. Comb. **16**(1), 5–19 (2002)

9. Bondarenko, A., Radchenko, D., Viazovska, M.: Optimal asymptotic bounds for spherical designs. Ann. Math. **178**(2), 443–452 (2013)

10. Bondarenko, A., Radchenko, D., Viazovska, M.: Well-separated spherical designs. Constr. Approx. **41**(1), 93–112 (2015)

11. Boyer, C., Chauffert, N., Ciuciu, P., Kahn, J., Weiss, P.: On the generation of sampling schemes for magnetic resonance imaging. SIAM J. Imaging Sci. **9**(4), 2039–2072 (2016)

12. Braides, A.: $\Gamma$-Convergence for Beginners. Oxford University Press, Oxford (2002)

13. Brandolini, L., Choirat, C., Colzani, L., Gigante, G., Seri, R., Travaglini, G.: Quadrature rules and distribution of points on manifolds. Ann. Scuola Norm.-Sci. **13**(4), 889–923 (2014)

14. Breger, A., Ehler, M., Gräf, M.: Quasi Monte Carlo integration and kernel-based function approximation on Grassmannians. In: Frames and Other Bases in Abstract and Function Spaces: Novel Methods in Harmonic Analysis, pp. 333–353. Birkhäuser, Basel (2017)

15. Bridson, M., Häfliger, A.: Metric Spaces of Non-Positive Curvature, *A Series of Comprehensive Studies in Mathematics*, vol. 319. Springer, Berlin (1999)

16. Burago, D., Burago, Y., Ivanov, S.: A Course in Metric Geometry, *Graduate Studies in Mathematics*, vol. 33. Amer. Math. Soc., Providence (2001)

17. Chauffert, N., Ciuciu, P., Kahn, J., Weiss, P.: Variable density sampling with continuous trajectories. SIAM J. Imaging Sci. **7**(4), 1962–1992 (2014)

18. Chauffert, N., Ciuciu, P., Kahn, J., Weiss, P.: A projection method on measures sets. Constr. Approx. **45**(1), 83–111 (2017)

19. Chavel, I.: Eigenvalues in Riemannian Geometry. Academic Press, Orlando (1984)

20. Chen, Z., Shen, Z., Guo, J., Cao, J., Zeng, X.: Line drawing for 3D printing. Comput. Graph. **66**, 85–92 (2017)

21. Chevallier, J.: Uniform decomposition of probability measures: Quantization, clustering and rate of convergence. J. Appl. Probab. **55**(4), 1037–1045 (2018)

22. Coulhon, T., Russ, E., Tardivel-Nachef, V.: Sobolev algebras on Lie groups and Riemannian manifolds. Amer. J. Math. **123**(2), 283–342 (2001)

23. Cucker, F., Smale, S.: On the mathematical foundations of learning. Bull. Amer. Math. Soc. **39**(1), 1–49 (2002)

24. Cuturi, M., Peyré, G.: Computational optimal transport. Found. Trends Mach. Learn. **11**(5-6), 355–607 (2019)

25. Daniel, J.W.: The conjugate gradient method for linear and nonlinear operator equations. SIAM J. Numer. Anal. **4**(1), 10–26 (1967)

26. Dick, J., Ehler, M., Gräf, M., Krattenthaler, C.: Spectral decomposition of discrepancy kernels on the Euclidean ball, the special orthogonal group, and the Grassmannian manifold. arXiv:1909.12334 (2019)

27. Duchamp, T., Stuetzle, W.: Extremal properties of principal curves in the plane. Ann. Stat. **24**(4), 1511–1520 (1996)

28. Dziugaite, G.K., Roy, D.M., Ghahramani, Z.: Training generative neural networks via maximum mean discrepancy optimization. In: Proc. of the 31st Conference on Uncertainty in Artificial Intelligence, pp. 258–267 (2015)

29. Ehler, M., Gräf, M.: Reproducing kernels for the irreducible components of polynomial spaces on unions of Grassmannians. Constr. Approx. **49**(1), 29–58 (2018)

30. Feydy, J., Séjourné, T., Vialard, F.X., Amari, S., Trouvé, A., Peyré, G.: Interpolating between optimal transport and MMD using Sinkhorn divergences. In: Proc. of Machine Learning Research, vol. 89, pp. 2681–2690. PMLR (2019)

31. Filbir, F., Mhaskar, H.N.: Marcinkiewicz–Zygmund measures on manifolds. J. Complex. **27**(6), 568–596 (2011)

32. Fonseca, I., Leoni, G.: Modern Methods in the Calculus of Variations: $L^p$ Spaces. Springer, New York (2007)

33. Fornasier, M., Haskovec, J., Steidl, G.: Consistency of variational continuous-domain quantization via kinetic theory. Appl. Anal. **92**(6), 1283–1298 (2013)

34. Förster, K.J., Petras, K.: On estimates for the weights in Gaussian quadrature in the ultraspherical case. Math. Comp. **55**(191), 243–264 (1990)

35. Fulton, W., Harris, J.: Representation Theory: A First Course. Springer, New York (1991)

36. Fuselier, E., Wright, G.B.: Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. SIAM J. Numer. Anal. **50**(3), 1753–1776 (2012)

37. Genevay, A., Chizat, L., Bach, F., Cuturi, M., Peyré, G.: Sample complexity of Sinkhorn divergences. In: Proc. of Machine Learning Research, vol. 89, pp. 1574–1583. PMLR (2019)
38. Gerber, S., Whitaker, R.: Regularization-free principal curve estimation. J. Mach. Learn. Res. **14**(1), 1285–1302 (2013)
39. Gigante, G., Leopardi, P.: Diameter bounded equal measure partitions of Ahlfors regular metric measure spaces. Discrete Comput. Geom. **57**(2), 419–430 (2017)
40. Gnewuch, M.: Weighted geometric discrepancies and numerical integration on reproducing kernel Hilbert spaces. J. Complex. **28**(1), 2–17 (2012)
41. de Gournay, F., Kahn, J., Lebrat, L.: Differentiation and regularity of semi-discrete optimal transport with respect to the parameters of the discrete measure. Numer. Math. **141**(2), 429–453 (2019)
42. Gräf, M.: A unified approach to scattered data approximation on $\mathbb{S}^3$ and SO(3). Adv. Comput. Math. **37**(3), 379–392 (2012)
43. Gräf, M.: Efficient algorithms for the computation of optimal quadrature points on Riemannian manifolds. PhD thesis, TU Chemnitz (2013)
44. Gräf, M., Potts, D.: Sampling sets and quadrature formulae on the rotation group. Numer. Funct. Anal. Optim. **30**(7-8), 665–688 (2009)
45. Gräf, M., Potts, D.: On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms. Numer. Math. **119**(4), 699–724 (2011)
46. Gräf, M., Potts, D., Steidl, G.: Quadrature errors, discrepancies and their relations to halftoning on the torus and the sphere. SIAM J. Sci. Comput. **34**(5), 2760–2791 (2013)
47. Gröchenig, K., Romero, J.L., Unnikrishnan, J., Vetterli, M.: On minimal trajectories for mobile sampling of bandlimited fields. Appl. Comput. Harmon. Anal. **39**(3), 487–510 (2015)
48. Hajlasz, P.: Sobolev spaces on metric-measure spaces. In: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, *Contemp. Math.*, vol. 338, pp. 173–218. Amer. Math. Soc., Providence (2003)
49. Hastie, T., Stuetzle, W.: Principal curves. J. Am. Stat. Assoc. **84**(406), 502–516 (1989)
50. Hauberg, S.: Principal curves on Riemannian manifolds. IEEE Trans. Pattern Anal. Mach. Intell. **38**(9), 1915–1921 (2015)
51. Hesse, K., Mhaskar, H.N., Sloan, I.H.: Quadrature in Besov spaces on the Euclidean sphere. J. Complex. **23**(4-6), 528–552 (2007)
52. Hörmander, L.: The Analysis of Linear Partial Differential Operators I. Springer, Berlin (1983)
53. James, A.T., Constantine, A.G.: Generalized Jacobi polynomials as spherical functions of the Grassmann manifold. Proc. London Math. Soc. **29**(3), 174–192 (1974)
54. Kaplan, C.S., Bosch, R.: TSP art. In: Renaissance Banff: Mathematics, Music, Art, Culture, pp. 301–308. Bridges Conference (2005)
55. Kégl, B., Krzyzak, A., Linder, T., Zeger, K.: Learning and design of principal curves. IEEE Trans. Pattern Anal. Mach. Intell. **22**(3), 281–297 (2000)
56. Keiner, J., Kunis, S., Potts, D.: Using NFFT3 – a software library for various nonequispaced fast Fourier transforms. ACM Trans. Math. Software **36**(4), 1–30 (2009)
57. Kim, J.H., Lee, J., Oh, H.S.: Spherical principal curves. arXiv:2003.02578 (2020)
58. Kloeckner, B.: Approximation by finitely supported measures. ESAIM Control Opt. Calc. Var. **18**(2), 343–359 (2012)
59. Kuipers, L., Niederreiter, H.: Uniform Distribution of Sequences. Wiley, New York (1974)
60. Lazarus, C., Weiss, P., Chauffert, N., Mauconduit, F., El Gueddari, L., Destrieux, C., Zemmoura, I., Vignaud, A., Ciuciu, P.: SPARKLING: Variable-density k-space filling curves for accelerated $T_2^*$-weighted MRI. Magn. Reson. Med. **81**(6), 3643–3661 (2019)
61. Lebrat, L., de Gournay, F., Kahn, J., Weiss, P.: Optimal transport approximation of 2-dimensional measures. SIAM J. Imaging Sci. **12**(2), 762–787 (2019)
62. Matousek, J.: Geometric Discrepancy, *Algorithms and Combinatorics*, vol. 18. Springer, Berlin (2010)
63. Mercer, J.: Functions of positive and negative type and their connection with the theory of integral equations. Philos. Trans. Roy. Soc. London Ser. A **209**(441-458), 415–446 (1909)
64. Mhaskar, H.N.: Eignets for function approximation on manifolds. Appl. Comput. Harmon. Anal. **29**(1), 63–87 (2010)
65. Mhaskar, H.N.: Approximate quadrature measures on data-defined spaces. In: Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan. Springer, Cham (2018)
66. Müller, C.: Spherical Harmonics, *Lecture Notes in Mathematics*, vol. 17. Springer, Berlin (1992)

67. Novak, E., Wozniakowski, H.: Tractability of Multivariate Problems. Volume II, *EMS Tracts in Mathematics*, vol. 12. EMS Publishing House, Zürich (2010)
68. Plonka, G., Potts, D., Steidl, G., Tasche, M.: Numerical Fourier Analysis. Birkhäuser, Basel (2019)
69. Ring, W., Wirth, B.: Optimization methods on Riemannian manifolds and their application to shape space. SIAM J. Optim. **22**(2), 596–627 (2012)
70. Roe, J.: Elliptic Operators, Topology and Asymptotic Methods, 2nd edn. Longman, Harlow (1998)
71. Roy, A.: Bounds for codes and designs in complex subspaces. J. Algebr. Comb. **31**(1), 1–32 (2010)
72. Schmaltz, C., Gwosdek, P., Bruhn, A., Weickert, J.: Electrostatic halftoning. Comp. Graph. For. **29**(8), 2313–2327 (2010)
73. Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Hamiltonian and Gradient Flows, Algorithms and Control, *Fields Inst. Commun.*, vol. 3, pp. 113–136. Amer. Math. Soc., Providence (1994)
74. Steele, J.M.: Growth rates of Euclidean minimum spanning trees with power weighted edges. Ann. Probab. **16**(4), 1767–1787 (1988)
75. Steele, J.M., Snyder, T.L.: Worst-case growth rates of some classical problems of combinatorial optimization. SIAM J. Comput. **18**(2), 278–287 (1989)
76. Steinwart, I., Scovel, C.: Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. Constr. Approx. **35**(3), 363–417 (2011)
77. Teuber, T., Steidl, G., Gwosdek, P., Schmaltz, C., Weickert, J.: Dithering by differences of convex functions. SIAM J. Imaging Sci. **4**(1), 79–108 (2011)
78. Triebel, H.: Theory of Function Spaces II. Birkhäuser, Basel (1992)
79. Udrişte, C.: Convex Functions and Optimization Methods on Riemannian Manifolds, *Mathematics and its Applications*, vol. 297. Springer, Dordrecht (1994)
80. Varshalovich, D., Moskalev, A., Khersonskii, V.: Quantum Theory of Angular Momentum. World Scientific, Singapore (1988)
81. Villani, C.: Topics in Optimal Transportation. Amer. Math. Soc., Providence (2003)
82. Wagner, G.: On means of distances on the surface of a sphere II (upper bounds). Pacific J. Math. **154**(2), 381–396 (1992)
83. Wagner, G., Volkmann, B.: On averaging sets. Monatsh. Math. **111**(1), 69–78 (1991)

## Affiliations

**Martin Ehler**[1] · **Manuel Gräf**[2] · **Sebastian Neumayer**[3] · **Gabriele Steidl**[3]

✉ Sebastian Neumayer
  neumayer@math.tu-berlin.de

  Martin Ehler
  martin.ehler@univie.ac.at

  Manuel Gräf
  mgraef@kfs.oeaw.ac.at

  Gabriele Steidl
  steidl@math.tu-berlin.de

[1] Department of Mathematics, University of Vienna, Vienna, Austria

[2] Austrian Academy of Sciences, Acoustics Research Institute, Vienna, Austria

[3] Institute of Mathematics, TU Berlin, 10623 Berlin, Germany