



# Optimal Monte Carlo integration on closed manifolds

Martin Ehler<sup>1</sup> · Manuel Gräf<sup>2</sup> · Chris. J. Oates<sup>3</sup>

Published online: 30 October 2019  
© The Author(s) 2019

## Abstract

The worst case integration error in reproducing kernel Hilbert spaces of standard Monte Carlo methods with  $n$  random points decays as  $n^{-1/2}$ . However, the re-weighting of random points, as exemplified in the Bayesian Monte Carlo method, can sometimes be used to improve the convergence order. This paper contributes general theoretical results for Sobolev spaces on closed Riemannian manifolds, where we verify that such re-weighting yields optimal approximation rates up to a logarithmic factor. We also provide numerical experiments matching the theoretical results for some Sobolev spaces on the sphere  $\mathbb{S}^2$  and on the Grassmannian manifold  $\mathcal{G}_{2,4}$ . Our theoretical findings also cover function spaces on more general sets such as the unit ball, the cube, and the simplex.

**Keywords** Bayesian cubature · Covering radius · Reproducing kernel

## 1 Introduction

Many problems in statistics and the applied sciences require the numerical integration of one or more functions belonging to a particular class. Given  $\mathcal{M} \subset \mathbb{R}^D$ , endowed with some probability measure  $\mu$ , and an integrable function  $f : \mathcal{M} \rightarrow \mathbb{R}$ , standard Monte Carlo methods approximate the integral  $\int_{\mathcal{M}} f(x) d\mu(x)$  by the finite sum

$$\frac{1}{n} \sum_{j=1}^n f(x_j), \quad (1)$$

where  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  are independent samples from  $\mu$ . On the one hand, Monte Carlo integration is widely used in many

numerical and statistical applications (Robert and Casella 2013). It is well-known, however, that the expected worst case integration error for  $n$  random points using (1) in reproducing kernel Hilbert spaces does not decay faster than  $n^{-1/2}$ , cf. Brauchart et al. (2014), Breger et al. (2018), Hinrichs (2010), Novak and Wozniakowski (2010), Plaskota et al. (2009) and Gräf (2013), proof of Corollary 2.8. To improve the approximation, it has been proposed to re-weight the random points (Briol et al. 2018; Oettershagen 2017; Rasmussen and Ghahramani 2003; Sommariva and Vianello 2006; Ullrich 2017), which is of particular importance when  $\mu$  can only be sampled (Oates et al. 2017) and evaluating  $f$  is rather expensive.

That re-weighting of *deterministic* points can lead to optimal convergence order has been known since the pioneering work of Bakhvalov (1959). For Sobolev spaces on the sphere and more generally on compact Riemannian manifolds, there are numerically feasible strategies to select deterministic points and weights matching optimal worst case error rates, cf. Brandolini et al. (2014), Brauchart et al. (2014), Breger et al. (2017), see also Hellekalek et al. (2016), Hinrichs et al. (2016) and Niederreiter (2003).

The use of *random* points avoids the need to manually specify a point set and can potentially lead to simpler algorithms if the geometry of the manifold  $\mathcal{M}$  is complicated. For random points, it was derived in Briol et al. (2018) that the optimal rate for  $[0, 1]^d$ , the sphere, and quite general domains in  $\mathbb{R}^d$  can be matched up to a logarithmic factor

Handling Editor T. J. Sullivan.

✉ Martin Ehler  
martin.ehler@univie.ac.at

Manuel Gräf  
mgrae@kfs.oeaw.ac.at

Chris. J. Oates  
chris.oates@ncl.ac.uk

- <sup>1</sup> Department of Mathematics, University of Vienna, Vienna, Austria
- <sup>2</sup> Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
- <sup>3</sup> School of Mathematics, Statistics and Physics, Newcastle University, Newcastle-Upon-Tyne, UK

if the weights are optimized with respect to the underlying reproducing kernel. Decay rates of the worst case integration error for Sobolev spaces of dominating mixed smoothness on the torus and the unit cube were studied in Oettershagen (2017). Gaussian kernel quadrature is studied in Karvonen and Särkkä (2019). Numerical experiments on the Grassmannian manifold were provided in Ehler and Gräf (2017). We refer to Trefethen (2017a, b), for further related results.

The present paper is dedicated to verify that, for Sobolev spaces on closed Riemannian manifolds, random points with optimized weights yield optimal decay rates of the worst case error up to a logarithmic factor. We should point out that we additionally allow for the restriction to nonnegative weights, a desirable property not considered in Briol et al. (2018). Our findings also transfer to functions defined on more general sets such as the  $d$ -dimensional unit ball and the simplex.

The paper is structured as follows: First, we bound the worst case error by the covering radius of the underlying points. Second, we use estimates on the covering radius of random points from Reznikov and Saff (2015), see also Brauchart et al. (2018) for the sphere, to establish the optimal approximation rate up to a logarithmic factor. Some consequences for the Bayesian Monte Carlo method are then presented. Numerical experiments for the sphere and the Grassmannian manifold are provided that support our theoretical findings. We also discuss the extension to the unit ball, the cube, and the simplex.

## 2 Preliminaries

Let  $\mathcal{M} \subset \mathbb{R}^D$  be a smooth, connected, closed Riemannian manifold of dimension  $d$ , endowed with the normalized Riemannian measure  $\mu$  throughout the manuscript. Prototypical examples for  $\mathcal{M}$  are the sphere and the Grassmannian

$$\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} : \|x\| = 1\},$$

$$\mathcal{G}_{k,m} = \{x \in \mathbb{R}^{m \times m} : x^\top = x, x^2 = x, \text{rank}(x) = k\},$$

respectively, where  $d = k(m - k)$  with  $D = m^2$  in case of the Grassmannian.

Let  $\mathcal{H}$  be any normed space of continuous functions  $f : \mathcal{M} \rightarrow \mathbb{R}$ . For points  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  and weights  $\{w_j\}_{j=1}^n \subset \mathbb{R}$ , the worst case error of integration is defined by

$$\begin{aligned} & \text{wce}(\{(x_j, w_j)\}_{j=1}^n, \mathcal{H}) \\ & := \sup_{\substack{f \in \mathcal{H} \\ \|f\|_{\mathcal{H}} \leq 1}} \left| \int_{\mathcal{M}} f(x) d\mu(x) - \sum_{j=1}^n w_j f(x_j) \right|. \end{aligned} \tag{2}$$

Suppose now that  $\mathcal{H}$  is a reproducing kernel Hilbert space, denoted by  $\mathcal{H}_K$ . Then the squared worst case error can be expressed in terms of the reproducing kernel  $K$  as

$$\begin{aligned} & \text{wce}(\{(x_j, w_j)\}_{j=1}^n, \mathcal{H}_K)^2 \\ & = \sum_{i,j=1}^n w_i w_j K(x_i, x_j) - 2 \sum_{j=1}^n w_j \int_{\mathcal{M}} K(x_j, y) d\mu(y) \\ & \quad + \int_{\mathcal{M}} \int_{\mathcal{M}} K(x, y) d\mu(x) d\mu(y). \end{aligned} \tag{3}$$

If  $x_1, \dots, x_n \in \mathcal{M}$  are random points, independently distributed according to  $\mu$ , then it holds

$$\sqrt{\mathbb{E} \left[ \text{wce}(\{(x_j, \frac{1}{n})\}_{j=1}^n, \mathcal{H}_K)^2 \right]} \asymp n^{-\frac{1}{2}}, \tag{4}$$

cf. Brauchart et al. (2014), Breger et al. (2018), Novak and Wozniakowski (2010) and Gräf (2013), Proof of Corollary 2.8. Hence, even if  $\mathcal{H}_K$  consists of arbitrarily smooth functions, the left hand side of (4) decays only like  $n^{-1/2}$ .

The present paper is dedicated to the question if and, as the case may be, how much one can actually improve the error rate in (4) when replacing the equal weights  $\frac{1}{n}$  with weights  $\{w_j\}_{j=1}^n$  that are customized to the random points  $\{x_j\}_{j=1}^n$ . From a practical perspective, the methods studied in this paper require that the integrals appearing in (3) can be analytically evaluated.

**Remark 1** The restriction to integrals with respect to the normalized Riemannian measure  $\mu$  is without too much loss of generality, for if  $\nu$  is any  $\sigma$ -finite measure that is absolutely continuous with respect to  $\mu$ , then

$$\int_{\mathcal{M}} f(x) d\nu(x) = \int_{\mathcal{M}} g(x) d\mu(x),$$

where  $g = f \frac{d\nu}{d\mu}$  and  $\frac{d\nu}{d\mu}$  is the Radon–Nikodym derivative of  $\nu$  with respect to  $\mu$ . Often,  $\mathcal{H}$  is an algebra, so that  $f, \frac{d\nu}{d\mu} \in \mathcal{H}$  also yield  $g \in \mathcal{H}$  and our considerations for  $\mu$  do apply.

## 3 Bounding the worst case error by the covering radius

To define appropriate smoothness spaces, let  $\Delta$  denote the Laplace–Beltrami operator on  $\mathcal{M}$  and let  $\{\varphi_\ell\}_{\ell=0}^\infty$  be the collection of its orthonormal eigenfunctions with eigenvalues  $\{-\lambda_\ell\}_{\ell=0}^\infty$  arranged by  $0 = \lambda_0 \leq \lambda_1 \leq \dots$ . We choose each  $\varphi_\ell, \ell = 0, 1, 2, \dots$ , to be real-valued with  $\varphi_0 \equiv 1$ . Given  $f \in L_p(\mathcal{M})$  with  $1 \leq p \leq \infty$ , the Fourier transform is defined by

$$\hat{f}(\ell) := \int_{\mathcal{M}} f(x) \varphi_\ell(x) d\mu(x), \quad \ell = 0, 1, 2, \dots,$$

with the usual extension to distributions on  $\mathcal{M}$ . For  $1 \leq p \leq \infty$  and  $s > 0$ , the Sobolev space  $H_p^s(\mathcal{M})$  is the collection

of all distributions on  $\mathcal{M}$  with  $(I - \Delta)^{s/2} f \in L_p(\mathcal{M})$ , i.e., with

$$\|f\|_{H_p^s} := \|(I - \Delta)^{s/2} f\|_{L_p} = \left\| \sum_{\ell=0}^{\infty} (1 + \lambda_{\ell})^{s/2} \hat{f}(\ell) \varphi_{\ell} \right\|_{L_p} < \infty.$$

For  $s > d/p$ , each function in  $H_p^s(\mathcal{M})$  is continuous, cf. Brandolini et al. (2014) and Triebel (1992), Theorem 7.4.5, Section 7.4.2, so that point evaluation is well-defined.

For  $s > d/p$  and any set of points  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  with arbitrary weights  $\{w_j\}_{j=1}^n \subset \mathbb{R}$ , we have

$$n^{-s/d} \lesssim \text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M})), \tag{5}$$

see Brauchart et al. (2014) for the sphere and Brandolini et al. (2014) for the general case. Note that the constant in (5) may depend on  $s, \mathcal{M}$ , and  $p$ .

Another lower bound involves the covering radius,

$$\rho_n := \max_{x \in \mathcal{M}} \min_{j=1, \dots, n} \text{dist}_{\mathcal{M}}(x, x_j),$$

where  $\text{dist}_{\mathcal{M}}$  denotes the geodesic distance. According to Breger et al. (2018), it also holds

$$\rho_n^{s+d/q} \lesssim \text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M})), \tag{6}$$

where  $1/p + 1/q = 1$ .

Attempting to match this lower bound, we shall optimize the weights. Given points  $\{x_j\}_{j=1}^n \subset \mathcal{M}$ , we define optimal weights with nonnegativity constraints by

$$\{\hat{w}_j^{\geq 0; p}\}_{j=1}^n := \arg \min_{w_1, \dots, w_n \geq 0} \text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M})). \tag{7}$$

It should be mentioned that similar weight optimization is suggested in Liu and Lee (2016), where the additional constraint  $\sum_{j=1}^n w_j = 1$  is used for stabilization purposes.

The worst case error for the optimized weights is upper bounded by the covering radius:

**Theorem 1** *Let  $1 \leq p \leq \infty$ , suppose  $s > d/p$ , and let  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  be any set of points with covering radius  $\rho_n$ . Then the optimized weights  $\{\hat{w}_j^{\geq 0; p}\}_{j=1}^n$  in (7) satisfy*

$$\text{wce}(\{(x_j, \hat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \lesssim \rho_n^s. \tag{8}$$

Note that the constant in (8) may depend on  $\mathcal{M}, s$ , and  $p$ .

**Remark 2** If we fix a constant  $c > 0$ , independent of  $n$  and  $\{x_j\}_{j=1}^n$ , then any weights  $\{\tilde{w}_j^p\}_{j=1}^n \subset \mathbb{R}$  with

$$\begin{aligned} & \text{wce}(\{(x_j, \tilde{w}_j^p)\}_{j=1}^n, H_p^s(\mathcal{M})) \\ & \leq c \cdot \text{wce}(\{(x_j, \hat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \end{aligned}$$

satisfy the estimate

$$\text{wce}(\{(x_j, \tilde{w}_j^p)\}_{j=1}^n, H_p^s(\mathcal{M})) \lesssim \rho_n^s.$$

This fact is beneficial when we compute weights numerically.

**Proof of Theorem 1** Let  $X := \{x_j\}_{j=1}^n$  and  $\rho(X) := \rho_n$ . There is a subset  $Y = \{y_j\}_{j=1}^m \subset X$  with covering radius  $\rho(Y) \leq 2\rho(X)$  and minimal separation

$$\delta(Y) := \min_{\substack{a, b \in Y \\ a \neq b}} \text{dist}_{\mathcal{M}}(a, b)$$

such that  $\rho(Y) \leq 2\delta(Y)$ , cf. Filbir and Mhaskar (2010), Section 3. We observe that our present setting satisfies the technical requirements of Filbir and Mhaskar (2010), cf. Hsu (1999) and Chavel (1984), p. 159. We deduce from Brandolini et al. (2014), Lemma 2.14 and Filbir and Mhaskar (2010), Theorem 3.1, see also Mhaskar et al. (2001) for  $\mathcal{M} = \mathbb{S}^d$ , with Brandolini et al. (2014), Corollary 2.15 that there exist  $w_1, \dots, w_m \gtrsim \rho_n^d$ , such that

$$\text{wce}(\{(y_j, w_j)\}_{j=1}^m, H_p^s(\mathcal{M})) \lesssim m^{-s/d}.$$

Since the bound

$$\begin{aligned} & \text{wce}(\{(x_j, \hat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \\ & \leq \text{wce}(\{(y_j, w_j)\}_{j=1}^m, H_p^s(\mathcal{M})), \end{aligned}$$

holds, we also obtain

$$\text{wce}(\{(x_j, \hat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \lesssim m^{-s/d}.$$

The general lower bound on the covering radius  $m^{-1/d} \lesssim \rho(Y)$ , cf. Breger et al. (2018), implies

$$m^{-s/d} \lesssim \rho(Y)^s \lesssim \rho(X)^s,$$

which concludes the proof. □

Combining (6) with Theorem 1 for  $p = 1$  yields

$$\text{wce}(\{(x_j, \hat{w}_j^{\geq 0; 1})\}_{j=1}^n, H_1^s(\mathcal{M})) \asymp \rho_n^s,$$

so that the worst case error's asymptotic behavior is governed by the covering radius.

**Remark 3** The above proof reveals that in the setting of Theorem 1 there exist  $\{w_j\}_{j=1}^n$  with either  $w_j \gtrsim \rho_n^d$  or  $w_j = 0$ , for  $j = 1, \dots, n$ , such that

$$\text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M})) \lesssim \rho_n^s.$$

The covering radius  $\rho_n$  of any  $n$  points in  $\mathcal{M}$  is lower bounded by  $\gtrsim n^{-1/d}$ , which follows from standard volume arguments. If  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  are points with asymptotically optimal covering radius, i.e.,  $\rho_n \asymp n^{-1/d}$ , then Theorem 1 yields the optimal rate for the worst case integration error

$$\text{wce}(\{(x_j, \widehat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \lesssim n^{-s/d},$$

cf. (5).

Several point sets on  $\mathbb{S}^2$  with asymptotically optimal covering radius are discussed in Hardin et al. (2016), see quasi-uniform point sequences therein, and see Breger et al. (2018) for general  $\mathcal{M}$ . The covering radius of random points is studied in Brauchart et al. (2018), Oates et al. (2018) and Reznikov and Saff (2015), which leads to almost optimal bounds on the worst case error in the subsequent section. Although we shall consider independent random points, it is noteworthy that it is verified in Oates et al. (2018) that the required estimates on the covering radius still hold for random points arising from a Markov chain instead of being independent. Note also that results related to Theorem 1 are derived in Mhaskar (2018) for more general spaces  $\mathcal{M}$ .

### 4 Consequences for random points

For random points  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  and any type of weights  $\{w_j\}_{j=1}^n \subset \mathbb{R}$ , no matter if random or not, (5) implies, for all  $r > 0$ ,

$$n^{-s/d} \lesssim \left( \mathbb{E}[\text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M}))^r] \right)^{1/r}, \tag{9}$$

where the constant may depend on  $s, p$ , and  $\mathcal{M}$ . Note that if  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  are random points, then the weights  $\{\widehat{w}_j^{\geq 0; p}\}_{j=1}^n$  are random as well. We shall deduce that Theorem 1 implies that the optimal worst case error rate is (almost) matched in these cases:

**Corollary 1** *Let  $\{x_j\}_{j=1}^n \subset \mathcal{M}$  be random points, independently distributed according to  $\mu$ . Suppose  $1 \leq p \leq \infty$  and  $s > d/p$ , then, for each  $r \geq 1/s$ , it holds*

$$\begin{aligned} & \left( \mathbb{E}[\text{wce}(\{(x_j, \widehat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M}))^r] \right)^{1/r} \\ & \lesssim n^{-s/d} \log(n)^{s/d}. \end{aligned} \tag{10}$$

Note that Corollary 1 yields the optimal rate up to the logarithmic factor  $\log(n)^{s/d}$ , cf. (9), and that the constant in (10) may depend on  $s, \mathcal{M}, p$ , and  $r$ .

**Proof** From Reznikov and Saff (2015), Theorem 3.2, Corollary 3.3, we deduce that, for each  $r \geq 1$ ,

$$(\mathbb{E} \rho_n^r)^{1/r} \asymp n^{-1/d} \log(n)^{1/d},$$

where the constant may depend on  $\mathcal{M}$  and  $r$ . Thus, Theorem 1 implies

$$\mathbb{E}[\text{wce}(\{(x_j, \widehat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M}))^r] \lesssim \left( \frac{\log(n)}{n} \right)^{sr/d},$$

for each  $r \geq 1/s$ . □

**Remark 4** Let  $\nu$  be a probability measure on  $\mathcal{M}$  that is absolutely continuous with respect to  $\mu$  and its density is bounded away from zero, i.e.,  $\nu = f\mu$  with  $f(x) \geq c > 0$ , for all  $x \in \mathcal{M}$ . Corollary 1 still holds for independent samples from  $\nu$ , where the constant in (10) then also depends on  $c$ . This is due to  $\approx \frac{2}{c}n$  independent samples from  $\nu$  covering  $\mathcal{M}$  at least as well as  $n$  independent samples from  $\mu$ .

Corollary 1 yields bounds on the moments of the worst case integration error. The results in Reznikov and Saff (2015) also enable us to derive probability estimates:

**Corollary 2** *Under the assumptions of Corollary 1, there are positive constants  $c_1, \dots, c_4$  depending on  $\mathcal{M}$  and where  $c_2$  may additionally depend on  $s$  and  $p$ , such that,*

$$\begin{aligned} & \mathbb{P}\left( \text{wce}(\{(x_j, \widehat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \geq c_2 \left( \frac{r \log(n)}{n} \right)^{s/d} \right) \\ & \leq c_3 \left( \frac{1}{n} \right)^{c_4 r - 1}, \end{aligned}$$

for all  $r \geq c_1$ .

**Proof** By applying Theorem 1, we deduce that there is a constant  $c > 0$ , which may depend on  $\mathcal{M}, s$ , and  $p$ , such that

$$\text{wce}(\{(x_j, \widehat{w}_j^{\geq 0; p})\}_{j=1}^n, H_p^s(\mathcal{M})) \leq c \rho_n^s.$$

According to Reznikov and Saff (2015), Theorem 2.1, there are constants  $c_1, \tilde{c}_2, c_3, c_4 > 0$ , which may depend on  $\mathcal{M}$ , such that, for all  $r \geq c_1$ ,

$$\mathbb{P}\left( \rho_n \geq \tilde{c}_2 \left( \frac{r \log(n)}{n} \right)^{1/d} \right) \leq c_3 \left( \frac{1}{n} \right)^{c_4 r - 1}.$$

Raising the left inequality to the power  $s$  and multiplying by  $c$  yields the desired result with  $c_2 := c \tilde{c}_2^s$ . □

Our bounds address functions in  $H_p^s(\mathcal{M})$  exclusively. For bounds beyond Sobolev functions in misspecified settings, we refer to Kanagawa et al. (2019) and references therein.

**Remark 5** Corollaries 1 and 2 also hold for weights  $\{\tilde{w}_j^n\}_{j=1}^n$  that minimize (7) up to a constant factor as discussed in Remark 2, and, in particular, for the unconstrained minimizer

$$\{\hat{w}_j^n\}_{j=1}^n := \arg \min_{\{w_j\}_{j=1}^n \subset \mathbb{R}} \text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M})). \quad (11)$$

Nonnegative weights are often more desirable for numerical applications of cubatures points, but solving the constrained minimization problem (7) is usually more involved than dealing with the unconstrained problem (11).

### 5 Implications for Bayesian Monte Carlo

Our results have consequences for *Bayesian cubature*, cf. Larkin (1972), an integration method whose output is not a scalar but a distribution. Bayesian cubature enables a statistical quantification of integration error, useful in the context of a wider computational work-flow to measure the impact of integration error on subsequent output, cf. Briol et al. (2018) and Cockayne et al. (2017).

Consider a linear topological space  $\mathcal{L}$  of continuous functions on  $\mathcal{M}$  such as a reproducing kernel Hilbert space on  $\mathcal{M}$ . The integrand  $f$  in Bayesian cubature is treated as a Gaussian random process; that is,  $f : \mathcal{M} \times \Omega \rightarrow \mathbb{R}$ , where  $f(\cdot, \omega) \in \mathcal{L}$  for each  $\omega \in \Omega$ , and the random variables  $\omega \mapsto Lf(\cdot, \omega) \in \mathcal{L}$  are (univariate) Gaussian for all continuous linear functionals  $L$  on  $\mathcal{L}$ , such as integration ( $\mathcal{I}f = \int_{\mathcal{M}} f(x) d\mu(x)$ ) and point evaluation ( $\delta_x f = f(x)$ ) operators, cf. Bogachev (1998). The Bayesian approach is then taken, wherein the process  $f$  is constrained to interpolate the values  $\{(x_j, f(x_j))\}_{j=1}^n$ . Formally, this is achieved by conditioning the process on the data provided through the point evaluation operators  $\delta_{x_j}(f) = f(x_j)$ , for  $\{x_j\}_{j=1}^n \subset \mathcal{M}$ . The conditioned process, denoted  $f_n$ , is again Gaussian, cf. Bogachev (1998), and as such the linear functional  $\mathcal{I}f_n$  is a (univariate) Gaussian; this is the output of the Bayesian cubature method. This distribution, defined on the real line, provides statistical uncertainty quantification for the (unknown) true value of the integral.

Concretely, let  $K(x, y) = \text{cov}(f(x), f(y))$  denote the covariance function that characterizes the Gaussian probability model. The output of Bayesian cubature is the univariate Gaussian distribution with mean

$$\sum_{j=1}^n \hat{w}_j f(x_j),$$

which takes the form of a weighted integration method with weights  $\hat{w} = (\hat{w}_1, \dots, \hat{w}_n)^\top$  implicitly defined by  $\mathcal{K}\hat{w} = b$  where

$$\mathcal{K} := \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_n) \\ \vdots & & \vdots \\ K(x_n, x_1) & \dots & K(x_n, x_n) \end{bmatrix} \quad (12)$$

$$b := \begin{bmatrix} \int_{\mathcal{M}} K(x_1, y) d\mu(y) \\ \vdots \\ \int_{\mathcal{M}} K(x_n, y) d\mu(y) \end{bmatrix}, \quad (13)$$

cf. Briol et al. (2018). The integrals appearing in (13) have an elegant interpretation in terms of the *kernel mean embedding* of the distribution  $\mu$  (Muandet et al. 2017).

Any symmetric positive definite covariance function  $K$  can be viewed as a reproducing kernel. In particular, the Bessel kernel

$$K_B^{(s)}(x, y) = \sum_{\ell=0}^{\infty} (1 + \lambda_\ell)^{-s} \varphi_\ell(x) \varphi_\ell(y), \quad x, y \in \mathcal{M}, \quad (14)$$

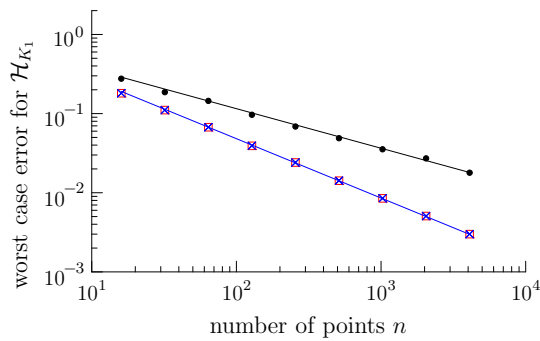
reproduces  $H^s(\mathcal{M}) := H_2^s(\mathcal{M})$ , for  $s > d/2$ . Observe that the weights  $\hat{w}$  just defined solve the unconstrained minimization problem (11) for  $p = 2$ . The latter follows from the quadratic minimization form in (3) as well as from the posterior mean being an  $L_2$ -optimal estimator (Kimeldorf and Wahba 1970).

The variance of the Gaussian measure can be shown to be formally equal to (3) when these weights are substituted, see Briol et al. (2018). The special case where the points  $\{x_j\}_{j=1}^n$  are random was termed *Bayesian Monte Carlo* in Rasmussen and Ghahramani (2003). Therefore, our results in Sect. 4 have direct consequences for Bayesian Monte Carlo. Due to Remark 5 within this Bayesian setting, Corollaries 1 and 2 generalize earlier work of Briol et al. (2018) to a general smooth, connected, closed Riemannian manifold.

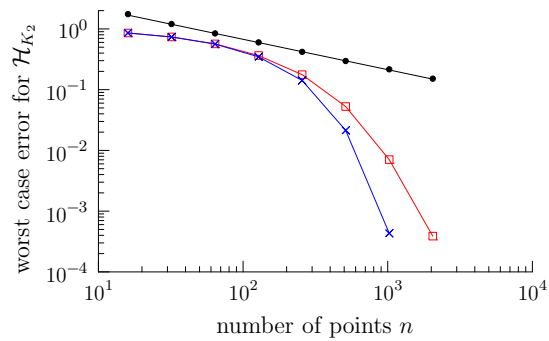
### 6 Numerical experiments for the sphere and the Grassmannian

The numerical computation of the worst case error  $\text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(\mathcal{M}))$  is difficult in general but, for  $p = 2$ , it is expressed in terms of the reproducing kernel in (3). Therefore, our numerical experiments are designed for  $p = 2$ . However, the kernel  $K_B^{(s)}$  itself, see (14), may still be difficult to evaluate numerically, so that we would like to allow for other kernels in numerical experiments. If  $K$  is any positive definite kernel on  $\mathcal{M}$  that reproduces  $H^s(\mathcal{M})$  with equivalent norms, then

$$\text{wce}(\{(x_j, w_j)\}_{j=1}^n, H^s(\mathcal{M})) \asymp \text{wce}(\{(x_j, w_j)\}_{j=1}^n, \mathcal{H}_K).$$



(a) Optimized weights for  $K_1$  yield decay  $-3/4$



(b) Fixed weights  $1/n$  for  $K_2$  are stuck with decay  $-1/2$

**Fig. 1** The worst case integration error (wce) for  $\mathcal{H}_{K_1}$  and  $\mathcal{H}_{K_2}$  averaged over 20 instances of random points in logarithmic scalings. The lines with exact slope  $-1/2$  and  $-3/4$  are included as a visual aid

(—●— average wce with equal weights; —□— average wce with nonnegative weights; —×— average wce with optimal weights)

Therefore, the asymptotic results in Corollaries 1 and 2 are the same when replacing  $\{\widehat{w}_j^{\geq 0; 2}\}_{j=1}^n$  with the minimizer

$$\{\widehat{w}_j^{\geq 0; K}\}_{j=1}^n := \arg \min_{w_1, \dots, w_n \geq 0} \text{wce}(\{(x_j, w_j)\}_{j=1}^n, \mathcal{H}_K). \quad (15)$$

Dropping the nonnegativity constraints yields  $\widehat{w}^K$ , which is given by  $\widehat{w} = \mathcal{K}^{-1}b$ , where  $\mathcal{K}$  and  $b$  are as in (12) and (13). To provide numerical experiments for Sobolev spaces on the sphere  $\mathbb{S}^2 \subset \mathbb{R}^3$  and on the Grassmannian  $\mathcal{G}_{2,4}$ , we shall specify suitable kernels in the following. We shall consider two kernels  $K_1, K_2$  on the sphere  $\mathbb{S}^2$  and two kernels  $K_3, K_4$  on the Grassmannian  $\mathcal{G}_{2,4}$ .

The numerical results are produced by taking sequences of random points  $\{x_j\}_{j=1}^n$  with increasing cardinality  $n$ . We compute each of the three worst case errors  $\text{wce}(\{(x_j, \frac{1}{n})\}_{j=1}^n, \mathcal{H}_{K_i})$ ,  $\text{wce}(\{(x_j, \widehat{w}_j^{K_i})\}_{j=1}^n, \mathcal{H}_{K_i})$ , and  $\text{wce}(\{(x_j, \widehat{w}_j^{\geq 0; K_i})\}_{j=1}^n, \mathcal{H}_{K_i})$ , for  $i = 1, \dots, 4$ , and averaged these results over 20 instantiations of the random points. The constrained minimization problem for the latter two quantities is solved by using the Python CVXOPT library. It should be mentioned that numerical experiments on the sphere for the unconstrained optimizer  $\widehat{w}^{K_1}$  are also contained in Briol et al. (2018).

The kernel

$$K_1(x, y) := 2 - \|x - y\|, \quad x, y \in \mathbb{S}^2,$$

reproduces the Sobolev space  $H^{3/2}(\mathbb{S}^2)$  with an equivalent norm, cf. Gräf (2013), Section 6.4.1. To compute (3), it is sufficient to notice

$$\int_{\mathbb{S}^2} K_1(x, y) d\mu(y) = \frac{2}{3}, \quad \text{for all } x \in \mathbb{S}^2.$$

By plotting the worst case error versus the number of points in a logarithmic scale, we are supposed to observe

lines whose slopes coincide with the decay rate  $-s/d$  for the optimized weights and slope  $-1/2$  for the weights  $1/n$ . Indeed, we see in Fig. 1a that  $\text{wce}(\{(x_j, \frac{1}{n})\}_{j=1}^n, \mathcal{H}_{K_1})$  for random points matches the error rate  $-1/2$  predicted by (4) with  $d = 2$ . When optimizing the weights, we observe the decay rate  $-3/4$  for both optimizations,  $\widehat{w}^{\geq 0; K}$  in (15) and the unconstrained minimizer  $\widehat{w}^K$ . Hence, the numerical results match the rate predicted by the theoretical findings in (9), (10) with  $p = 2$  and  $r = 1$ . The logarithmic factor in (10) is not visible.

The smooth kernel

$$K_2(x, y) := 48 \exp(-12\|x - y\|^2), \quad x, y \in \mathbb{S}^2,$$

generates a space  $\mathcal{H}_{K_2}$  of smooth functions contained in  $H^s(\mathbb{S}^2)$ , for all  $s > 0$ , and satisfies

$$\int_{\mathbb{S}^2} K_2(x, y) d\mu(y) = 1 - \exp(-48), \quad \text{for all } x \in \mathbb{S}^2.$$

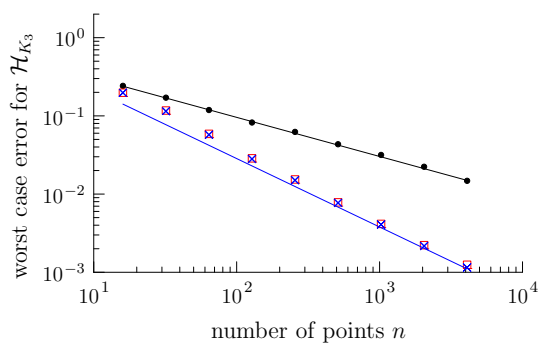
Our numerical experiments in Fig. 1b suggest that the decay rate for the optimized weights is indeed faster than  $-1/2$ . Note that the equal weight case is stuck at the  $-1/2$  rate, although we are now dealing with arbitrarily smooth functions.

The dimension of the Grassmannian  $\mathcal{G}_{2,4}$  is  $d = 4$ , and we consider the two reproducing kernels

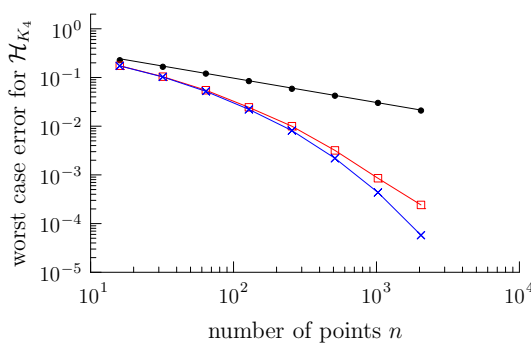
$$K_3(x, y) := \sqrt{(2 - \text{trace}(xy))^3 + 2 \text{trace}(xy)},$$

$$K_4(x, y) := \frac{3}{2} \exp(\text{trace}(xy) - 2).$$

Note that  $K_3$  reproduces  $H^{7/2}(\mathcal{G}_{2,4})$  with an equivalent norm, and  $\mathcal{H}_{K_4}$  is contained in  $H^s(\mathcal{G}_{2,4})$ , for all  $s > 0$ . The worst case integration error (3) is computable from



(a) Optimized weights for  $K_3$  yield decay  $-7/8$



(b) Fixed weights  $1/n$  for  $K_4$  are stuck with decay  $-1/2$

**Fig. 2** The worst case integration error for  $\mathcal{H}_{K_3}$  and  $\mathcal{H}_{K_4}$  averaged over 20 instances with random points in logarithmic scalings. The lines with exact slope  $-1/2$  and  $-7/8$  are included as a visual aid (—●—)

$$\int_{\mathcal{G}_{2,4}} K_3(x, y) d\mu(y) = 2 + \frac{74}{75} \sqrt{2} - \frac{2}{5} \log(1 + \sqrt{2}),$$

$$\int_{\mathcal{G}_{2,4}} K_4(x, y) d\mu(y) = \frac{3}{2} \exp(-1) \int_0^1 \frac{\sinh(t)}{t} dt.$$

for all  $x \in \mathcal{G}_{2,4}$ .

For  $K_3$ , we observe in Fig. 2a that the random points with equal weights yield decay rate  $-1/2$  and optimizing weights leads to  $-7/8$  matching the optimal rate in (9), (10) with  $d = 4$ . In Fig. 2b, it seems that the worst case error for  $\mathcal{H}_{K_4}$  decays faster than the  $-1/2$  rate when optimizing the weights for random points on the Grassmannian  $\mathcal{G}_{2,4}$  outperforming the case where weights are equal.

### 7 Beyond closed manifolds

We shall make use of the push-forward to transfer our results on the worst case integration error from closed manifolds to more general sets. Suppose  $S$  is a topological space and  $h : \mathcal{M} \rightarrow S$  is Borel measurable and surjective. We endow  $S$  with the push-forward measure  $h_*\mu$  defined by  $(h_*\mu)(A) = \mu(h^{-1}A)$  for any Borel measurable subset  $A \subset S$ . By abusing notation, let  $\text{dist}_{\mathcal{M}}(A, B) := \inf_{a \in A; b \in B} \text{dist}_{\mathcal{M}}(a, b)$  for  $A, B \subset \mathcal{M}$ , and we put

$$\text{dist}_{S,h}(x, y) := \text{dist}_{\mathcal{M}}(h^{-1}x, h^{-1}y), \quad x, y \in S. \quad (16)$$

For  $s > d/p$ , we define

$$H_p^s(S)_h := \{f : S \rightarrow \mathbb{R} : h^*f \in H_p^s(\mathcal{M})\}$$

with  $\|f\|_{H_p^s(S)_h} := \|h^*f\|_{H_p^s(\mathcal{M})}$ , where  $h^*f$  denotes the pullback  $f \circ h$ . This enables us to formulate the analogue of Theorem 1:

average wce with equal weights; —□— average wce with nonnegative weights; —×— average wce with optimal weights)

**Theorem 2** Given  $1 \leq p \leq \infty$  with  $s > d/p$  and  $\{x_j\}_{j=1}^n \subset S$ , suppose that the following two conditions are satisfied,

- (a)  $\#h^{-1}x$  is finite, for all  $x \in S$ ,
- (b)  $\text{dist}_{\mathcal{M}}(\{a\}, h^{-1}x) \asymp \text{dist}_{\mathcal{M}}(h^{-1}h(a), h^{-1}x)$ , for all  $a \in \mathcal{M}, x \in S$ .

Then there are nonnegative weights  $\{w_j\}_{j=1}^n$  such that

$$\text{wce}(\{(x_j, w_j)\}_{j=1}^n, H_p^s(S)_h) \lesssim \rho_n^s,$$

where  $\rho_n$  denotes the covering radius of  $\{x_j\}_{j=1}^n$  taken with respect to (16).

Note that (16) is a quasi-metric on  $S$  if the assumptions in Theorem 2 are satisfied, i.e., the conditions of a metric are satisfied except for the triangular inequality that still holds up to a constant factor.

Conventional integration bounds on standard Euclidean domains  $S$ , see Kanagawa et al. (2019), Proposition 4, for instance, usually require bounded densities. Theorem 2 also applies to standard Euclidean domains that now inherit the measure and the distance from the closed manifold. The measure can now have unbounded density because there is a potential compensation by the induced distance. This shall be observed for the unit ball, the cube, and the simplex in the present section.

**Proof of Theorem 2** Denote  $\{z_{j,i}\}_{i=1}^{n_j} = h^{-1}x_j$ , for  $j = 1 \dots, n$ . According to Theorem 1, there exist nonnegative weights  $\{w_{j,i}\}_{i=1}^{n_j}$ , for  $j = 1 \dots, n$ , such that, for all  $f \in H_p^s(S)_h$ ,

$$\left| \int_{\mathcal{M}} (h^*f)(z) d\mu(z) - \sum_{j=1}^n \sum_{i=1}^{n_j} w_{j,i} (h^*f)(z_{j,i}) \right|$$

$$\lesssim \rho_{\mathcal{M}} \left( \bigcup_{j=1}^n h^{-1}x_j \right)^s \|h^*f\|_{H_p^s(\mathcal{M})}, \tag{17}$$

where  $\rho_{\mathcal{M}}(\bigcup_{j=1}^n h^{-1}x_j)$  denotes the covering radius of  $\bigcup_{j=1}^n h^{-1}x_j \subset \mathcal{M}$ . The assumptions imply that  $\rho_n \asymp \rho_{\mathcal{M}}(\bigcup_{j=1}^n h^{-1}x_j)$ , so that  $w_j := \sum_{i=1}^{n_j} w_{j,i}$ ,  $j = 1, \dots, n$ , and (17) lead to

$$\left| \int_S f(x) d(h_*\mu)(x) - \sum_{j=1}^n w_j f(x_j) \right| \lesssim \rho_n^s \|f\|_{H_p^s(S)_h},$$

which concludes the proof. □

**Remark 6** Since independent random points  $\{x_j\}_{j=1}^n$  distributed according to  $h_*\mu$  on  $S$  with covering radius  $\rho_n$  are generated by independent random points  $\{z_j\}_{j=1}^n$  with respect to  $\mu$  on  $\mathcal{M}$  with  $x_j = h(z_j)$ , for  $j = 1, \dots, n$ , the observation

$$\rho_n \lesssim \rho_{\mathcal{M}}(\{z_j\}_{j=1}^n)$$

implies that also Corollaries 1, and 2 hold for  $H_p^s(S)_h$  and  $h_*\mu$ .

The impact of Theorem 2 depends on whether or not the choices of  $h$  yield reasonable function spaces  $H_p^s(S)_h$ , distances  $\text{dist}_{S,h}$ , and measures  $h_*\mu$ . For instance, if  $h$  is also injective with measurable  $h^{-1}$ , then  $H^s(S)_h$  is the reproducing kernel Hilbert space with kernel

$$\sum_{\ell=0}^{\infty} (1 + \lambda_{\ell})^{-s} \psi_{\ell}(x)\psi_{\ell}(y),$$

where  $\psi_{\ell} := \varphi_{\ell} \circ h^{-1}$ , so that  $\{\psi_{\ell}\}_{\ell=0}^{\infty}$  is an orthonormal basis for the square integrable functions with respect to  $h_*\mu$ . In the following, we shall discuss a few special cases, in which  $h$  is not injective. By using the results in Xu (1998, 2001), we shall determine  $H^s(S)_h$  for  $S$  being the unit ball  $\mathbb{B}^d := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$ , the cube  $[-1, 1]^d$ , and the simplex  $\Sigma^d := \{x \in \mathbb{R}^d : x_1, \dots, x_d \geq 0; \sum_{i=1}^d x_i \leq 1\}$ .

Let  $h : \mathbb{S}^d \rightarrow \mathbb{B}^d$  be the projection onto the first  $d$  coordinates, i.e.,  $h(x) = (x_1, \dots, x_d) \in \mathbb{B}^d$ . The push-forward measure  $h_*\mu_{\mathbb{S}^d}$  on  $\mathbb{B}^d$  is given by

$$\frac{\Gamma(d/2 + 1/2)}{\pi^{d/2+1/2}} \frac{dx}{\sqrt{1 - \|x\|^2}}, \tag{18}$$

and the assumptions in Theorem 2 are satisfied. Let  $\{T_{k,\ell} : \ell = 0, 1, 2, \dots; k = 1, \dots, r_{\ell}^d\}$  with  $r_{\ell}^d := \binom{\ell+d-1}{\ell}$  be orthonormal polynomials with respect to the measure (18), and each  $T_{k,\ell}$  has total degree  $\ell$ . For  $d = 1$ , this corresponds

to Chebyshev polynomials. The case  $d = 2$  relates to generalized Zernike polynomials, cf. Wünsche (2005).

**Proposition 1** (The unit ball) *For  $s > d/2$  and  $h : \mathbb{S}^d \rightarrow \mathbb{B}^d$  as above, the space  $H^s(\mathbb{B}^d)_h$  is reproduced by the kernel*

$$K_{\mathbb{B}^d}^s(x, y) := \sum_{\ell=0}^{\infty} (1 + \ell(\ell + d - 1))^{-s} \sum_{k=1}^{r_{\ell}^d} T_{k,\ell}(x)T_{k,\ell}(y) \tag{19}$$

for  $x, y \in \mathbb{B}^d$ .

For related results on approximation on  $\mathbb{B}^d$ , we refer to Petrushev and Xu (2008) and references therein.

**Proof** For  $\ell = 0, 1, 2, \dots$ , the eigenfunctions of the Laplace-Beltrami operator on the sphere associated to the eigenvalue  $-\lambda_{\ell} = -\ell(\ell + d - 1)$  are the spherical harmonics of order  $\ell$ , given by the homogeneous harmonic polynomials in  $d + 1$  variables of exact total degree  $\ell$  restricted to  $\mathbb{S}^d$ . Each eigenspace  $E_{\ell}$  associated to  $\lambda_{\ell}$  splits orthogonally into  $E_{\ell} = E_{\ell}^{(1)} \oplus E_{\ell}^{(2)}$ , where

$$E_{\ell}^{(1)} = \{f \in E_{\ell} : f(x) = f(h(x), -x_{d+1}), \forall x \in \mathbb{S}^d\},$$

$$E_{\ell}^{(2)} = \{f \in E_{\ell} : f(x) = -f(h(x), -x_{d+1}), \forall x \in \mathbb{S}^d\} \tag{20}$$

for  $\ell = 0, 1, 2, \dots$ . We deduce from Xu (1998), Theorem 3.3, Example 3.4 that the functions

$$Z_{k,\ell}^{(1)}(z) := \|z\|^{\ell} (h^*T_{k,\ell}) \left( \frac{z}{\|z\|} \right), \quad z \in \mathbb{R}^{d+1}$$

are homogeneous polynomials of total degree  $\ell$ , and their restrictions  $Y_{k,\ell}^{(1)} := Z_{k,\ell}^{(1)}|_{\mathbb{S}^d}$ , for  $k = 1, \dots, r_{\ell}^d$ , are an orthonormal basis for  $E_{\ell}^{(1)}$ . Note that  $f \in H^s(\mathbb{B}^d)_h$  if and only if  $f \circ h$  is contained in

$$H^s(\mathbb{S}^d)^{\text{sym}} := \{f \in H^s(\mathbb{S}^d) : f|_{h^{-1}x} \text{ is constant } \forall x \in \mathbb{B}^d\}$$

$$= \{f \in H^s(\mathbb{S}^d) : f(x) = f(h(x), -x_{d+1}), x \in \mathbb{S}^d\}.$$

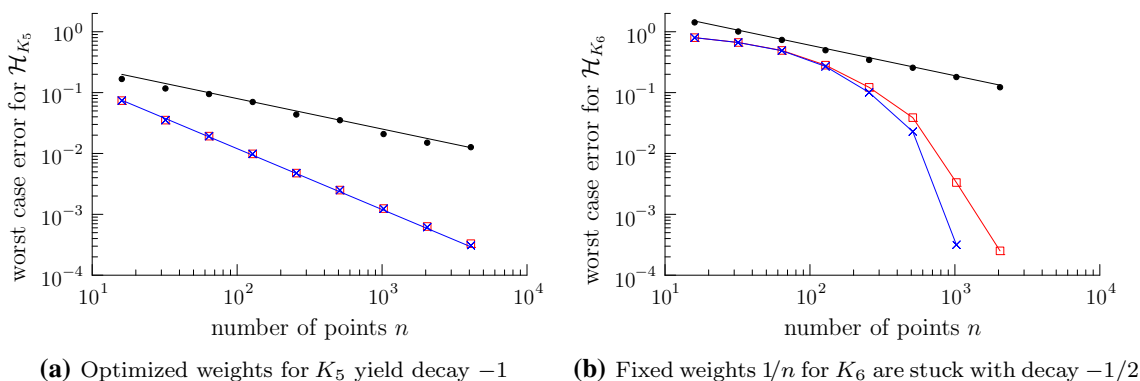
According to (14) and due to the decomposition induced by (20) and using  $Y_{k,\ell}^{(1)} = h^*T_{k,\ell}$ , the reproducing kernel of  $H^s(\mathbb{S}^d)^{\text{sym}}$  is

$$K_{\mathbb{S}^d, \text{sym}}^s(x, y) = \sum_{\ell=0}^{\infty} (1 + \lambda_{\ell})^{-s} \sum_{k=1}^{r_{\ell}^d} Y_{k,\ell}^{(1)}(x)Y_{k,\ell}^{(1)}(y)$$

$$= \sum_{\ell=0}^{\infty} (1 + \lambda_{\ell})^{-s} \sum_{k=1}^{r_{\ell}^d} (h^*T_{k,\ell})(x)(h^*T_{k,\ell})(y)$$

for  $x, y \in \mathbb{S}^d$ . Thus,  $H^s(\mathbb{B}^d)_h$  is indeed reproduced by (19). □





**Fig. 3** Worst case integration error for  $\mathcal{H}_{K_5}$  and  $\mathcal{H}_{K_6}$  with  $r = 0.97$ , averaged over random points in logarithmic scalings (—●— average wce with equal weights; —□— average wce with nonnegative weights; —×— average wce with optimal weights)

**Example 1** ( $\mathbb{B}^1$ ) For  $d = 1$ , the even and odd spherical harmonics,

$$\begin{aligned} Y_\ell^{(1)}(\cos(\alpha), \sin(\alpha)) &= \sqrt{2} \cos(\ell\alpha), \\ Y_\ell^{(2)}(\cos(\alpha), \sin(\alpha)) &= \sqrt{2} \sin(\ell\alpha), \quad \alpha \in [0, 2\pi], \end{aligned} \tag{21}$$

with  $\ell = 1, 2, 3, \dots$ , and  $Y_0^{(1)} = 1$ , form orthonormal bases for the respective spaces  $E_\ell^{(1)}$  and  $E_\ell^{(2)}$  in (20). We observe  $\text{dist}_{[-1,1],h}(x, y) := |\arccos(x) - \arccos(y)|$ ,  $x, y \in [-1, 1]$ , and recognize the Chebyshev measure in (18). The Chebyshev polynomials  $T_\ell$  of the first kind, scaled by the factor  $\sqrt{2}$  for  $\ell = 1, 2, 3, \dots$ , indeed satisfy the characteristic identities  $T_\ell(\cos(\alpha)) = \sqrt{2} \cos(\ell\alpha)$  for  $\alpha \in [0, 2\pi]$ ,  $\ell = 1, 2, 3, \dots$ , and  $T_0 = 1$ .

To simplify numerical experiments, we observe that the kernel

$$\begin{aligned} K_5(x, y) &:= 2 - \sqrt{1 - xy + |x - y|} \\ &= 2 + \frac{4}{\pi} \sum_{\ell=0}^{\infty} \frac{1}{4\ell^2 - 1} T_\ell(x)T_\ell(y) \end{aligned}$$

reproduces  $H^1(\mathbb{B}^1)_h$  with an equivalent norm, and, for fixed  $0 < r < 1$ , the smooth kernel

$$\begin{aligned} K_6(r; x, y) &:= \frac{(1 - r^2)(1 - 2rxy + r^2)}{1 + r^4 - 4xy(r + r^3) + r^2(4x^2 + 4y^2 - 2)} \\ &= \frac{1}{2} + \frac{1}{2} \sum_{\ell=0}^{\infty} r^\ell T_\ell(x)T_\ell(y) \end{aligned}$$

reproduces a function space that is continuously embedded into  $H^s(\mathbb{B}^1)_h$  for all  $s > 1/2$ . As in our previous examples, our numerical experiments in Fig. 3 are in accordance with the theoretical results.

**Example 2** ( $\mathbb{B}^2$ ) For  $r \geq 1$ , we define the family of kernels

$$\begin{aligned} L_r(x, y) &:= 3 - \frac{3}{2\sqrt{2r + 2}} \\ &\quad \sqrt{r - \langle x, y \rangle + \sqrt{(r - \langle x, y \rangle)^2 - (1 - \|x\|^2)(1 - \|y\|^2)}}. \end{aligned}$$

These kernels are positive definite on  $\mathbb{B}^2$  and satisfy

$$\int_{\mathbb{B}^2} L_r(x, y) \frac{dx}{2\pi\sqrt{1 - \|x\|^2}} = \frac{3 + r + 2\sqrt{r^2 - 1}}{1 + r + \sqrt{r^2 - 1}}.$$

Note that  $L_1$  reproduces  $H^{3/2}(\mathbb{B}^2)$  with an equivalent norm and  $L_r$  for  $r > 1$  reproduces a space that is continuously embedded into each  $H^s(\mathbb{B}^2)$  for  $s > 1$ . In our numerical experiments, we set

$$K_7(x, y) := L_1(x, y), \quad K_8(x, y) := L_{51/50}(x, y),$$

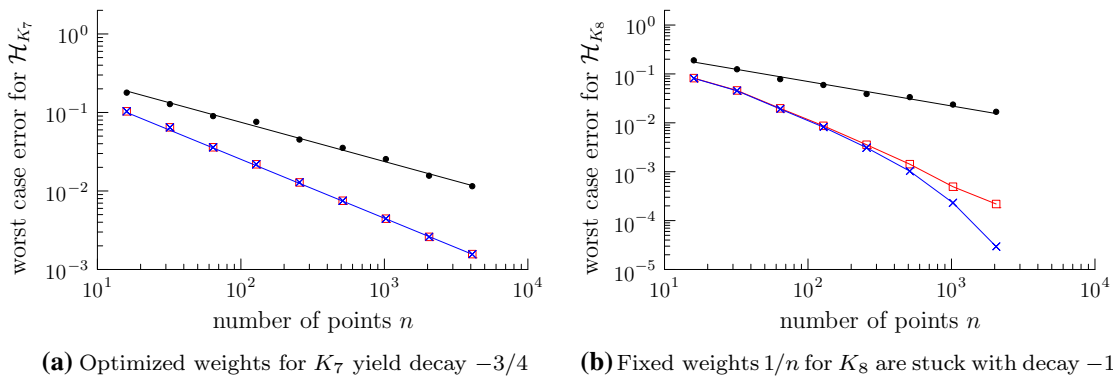
and Fig. 4 supports our theoretical results. There, however, the worst case error for nonnegative weights does not show faster decay for smooth functions in Fig. 4b, but we speculate that this is due to a numerical artifact of the very last data point.

The  $d$ -dimensional torus  $\mathbb{T}^d := \mathbb{S}^1 \times \dots \times \mathbb{S}^1$  leads to  $h : \mathbb{T}^d \rightarrow [-1, 1]^d$  defined by

$$h(x_1, \dots, x_d) = (x_{1,1}, \dots, x_{d,1}),$$

where  $x_i = (x_{i,1}, x_{i,2})^\top \in \mathbb{S}^1$ . The push-forward of the Riemannian measure on  $\mathbb{T}^d$  under  $h$  is

$$\frac{dx_1 \cdots dx_d}{\pi^d \sqrt{(1 - x_1^2) \cdots (1 - x_d^2)}}. \tag{22}$$



**Fig. 4** Worst case integration error for  $\mathcal{H}_{K_7}$  and  $\mathcal{H}_{K_8}$ , averaged over random points in logarithmic scalings (—●— average wce with equal weights; —□— average wce with nonnegative weights; —×— average wce with optimal weights)

A suitable basis of orthonormal polynomials characterizes  $H^s([-1, 1]^d)_h$ :

**Proposition 2** (The cube) For  $s > d/2$  and  $h : \mathbb{T}^d \rightarrow [-1, 1]^d$  as above, the space  $H^s([-1, 1]^d)_h$  is reproduced by

$$K_{[-1,1]^d}^s(x, y) := \sum_{\ell \in \mathbb{N}^d} (1 + \|\ell\|^2)^{-s} T_\ell(x) T_\ell(y),$$

$x, y \in [-1, 1]^d$ , where  $T_\ell(x) := T_{\ell_1}(x_1) \cdots T_{\ell_d}(x_d)$ ,  $x \in [-1, 1]^d$ ,  $\ell \in \mathbb{N}^d$ .

**Proof** The polynomials  $\{T_\ell : \ell \in \mathbb{N}^d\}$  are orthonormal with respect to (22). By using the eigenspace decomposition (20) for  $\mathbb{S}^1$ , we deduce that the space

$$H^s(\mathbb{T}^d)^{\text{sym}} := \{f \in H^s(\mathbb{T}^d) : f|_{h^{-1}x} \text{ is constant } \forall x \in [-1, 1]^d\}$$

is reproduced by the kernel

$$K_{\mathbb{T}^d, \text{sym}}^s(x, y) = \sum_{\ell \in \mathbb{N}^d} (1 + \|\ell\|^2)^{-s} Y_\ell^{(1)}(x) Y_\ell^{(1)}(y),$$

$x, y \in \mathbb{T}^d$ , with  $Y_\ell^{(1)}(x) := Y_{\ell_1}^{(1)}(x_1) \cdots Y_{\ell_d}^{(1)}(x_d)$  and  $Y_{\ell_i}^{(1)}$  are as in (21). Observing  $Y_\ell^{(1)} = h^* T_\ell$  concludes the proof.  $\square$

By following Xu (2001), we derive an analogous construction for the simplex. Define  $h : \mathbb{S}^d \rightarrow \Sigma^d$  by  $h(x) := (x_1^2, \dots, x_d^2)$  and observe that the assumptions in Theorem 2 are satisfied. The push-forward measure  $h_* \mu_{\mathbb{S}^d}$  on  $\Sigma^d$  is given by

$$\frac{\Gamma(d/2 + 1/2)}{\pi^{d/2+1/2}} \frac{du}{\sqrt{u_1 \cdots u_d (1 - \sum_{i=1}^d u_i)}}. \tag{23}$$

Let  $\{R_{k,\ell} : \ell = 0, 1, 2, \dots; k = 1, \dots, r_\ell^d\}$  be a system of orthonormal polynomials with respect to (23) on  $\Sigma^d$ , so that each  $R_{k,\ell}$  has total degree  $\ell$ .

**Proposition 3** (The simplex) For  $s > d/2$  and  $h : \mathbb{T}^d \rightarrow \Sigma^d$  as above, the space  $H^s(\Sigma^d)_h$  is reproduced by

$$K_{\Sigma^d}^s(u, v) := \sum_{\ell=0}^{\infty} (1 + 2\ell(2\ell + d - 1))^{-s} \sum_{k=1}^{r_\ell^d} R_{k,\ell}(u) R_{k,\ell}(v),$$

for  $u, v \in \Sigma^d$ .

**Proof** Let us define

$$Z_{k,2\ell}(z) := \|z\|^{2\ell} R_{k,\ell} \left( \frac{z_1^2}{\|z\|^2}, \dots, \frac{z_d^2}{\|z\|^2} \right), \quad z \in \mathbb{R}^{d+1}.$$

Note that the restrictions  $Y_{k,2\ell} := Z_{k,2\ell}|_{\mathbb{S}^d}$  satisfy  $Y_{k,2\ell} = h^* R_{k,\ell}$ . We deduce from Xu (2001) that the collection  $\{Y_{k,2\ell} : k = 1, \dots, r_\ell^d\}$  is an orthonormal system of spherical harmonics of order  $2\ell$  and that the space

$$H^s(\mathbb{S}^d)^{\text{sym}} := \{f \in H^s(\mathbb{S}^d) : f|_{h^{-1}x} \text{ is constant } \forall x \in \Sigma^d\}$$

is reproduced by the kernel

$$\begin{aligned} K(x, y) &:= \sum_{\ell=0}^{\infty} (1 + 2\ell(2\ell + d - 1))^{-s} \sum_{k=1}^{r_\ell^d} Y_{k,2\ell}(x) Y_{k,2\ell}(y) \\ &= \sum_{\ell=0}^{\infty} (1 + 2\ell(2\ell + d - 1))^{-s} \sum_{k=1}^{r_\ell^d} (h^* R_{k,\ell})(x) (h^* R_{k,\ell})(y), \end{aligned}$$

which concludes the proof.  $\square$

**Remark 7** Our Theorem 2 is an elementary way to transfer results from closed manifolds to more general settings. Our treatment of the unit ball, the cube, and the simplex were based on this transfer. The proof of the underlying Theorem 1 is based on results in Filbir and Mhaskar (2010), and we restricted attention to closed manifolds although the setting in Filbir and Mhaskar (2010) is more general. Alternatively, we could have stated our Theorem 1 in more generality and then attempted to check that the technical requirements in Filbir and Mhaskar (2010) hold. For instance, technical requirements for  $[-1, 1]$  were checked in Coulhon et al. (2012), and the recent work (Kerkycharian et al. 2019) covers technical details for the unit ball and the simplex.

## 8 Perspectives

Re-weighting techniques for statistical and numerical integration have attracted attention in different disciplines. Partially complementing findings in Briol et al. (2018) and Oettershagen (2017), we have here established that re-weighting random points can yield almost optimal approximation rates of the worst case integration error for isotropic Sobolev spaces on closed Riemannian manifolds. Our results suggest several directions for future work, for instance, allowing for more general spaces  $\mathcal{M}$ , considering other smoothness classes than  $H_p^s(\mathcal{M})$ , considering other types of point processes such as determinantal point processes (Bardenet and Hardy 2016), and replacing the expected worst case error by alternative error functionals such as the average error, cf. Novak and Wozniakowski (2010) and Ritter (2000).

Our results have direct consequences for the Bayesian Monte Carlo method, as indicated in Sect. 5. Indeed, there has been recent interest in exploiting Bayesian cubature in applications including global illumination in computer vision (Marques et al. 2015), signal processing (Prüher and Šimandl 2016), uncertainty quantification (Oettershagen 2017) and Bayesian computation (Briol et al. 2018). The results in this paper justify the use of a *random* point set in these applications, in situations where a deterministic point set would otherwise need to be explicitly constructed.

**Acknowledgements** Open access funding provided by University of Vienna. ME and MG were funded by the Vienna Science and Technology Fund (WWTF) through project VRG12-009. CJO was supported by the Lloyd’s Register Foundation program on data-centric engineering at the Alan Turing Institute, UK. This research was supported by the National Science Foundation, USA, under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect

the views of the above-named funding bodies and research institutions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bakhvalov, N.S.: On the approximate calculation of multiple integrals (in Russian). *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem.* **4**, 3–18 (1959)
- Bardenet, R., Hardy, A.: Monte Carlo with determinantal point processes. *Ann. Appl. Probab.* (**in press**)
- Bogachev, V.I.: *Gaussian Measures*. AMS, London (1998)
- Brandolini, L., Choirat, C., Colzani, L., Gigante, G., Seri, R., Travaglini, G.: Quadrature rules and distribution of points on manifolds. *Ann. Sc. Norm. Super. Pisa Classe di Sci. XII* **I**(4), 889–923 (2014)
- Brauchart, J.S., Saff, E.B., Sloan, I.H., Womersley, R.S.: QMC designs: optimal order quasi Monte Carlo integration schemes on the sphere. *Math. Comput.* **83**, 2821–2851 (2014)
- Brauchart, J.S., Reznikov, A.B., Saff, E.B., Sloan, I.H., Wang, Y.G., Womersley, R.S.: Random point sets on the sphere: hole radii, covering, and separation. *Exp. Math.* **27**(1), 62–81 (2018)
- 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*, vol. 1, Birkhauser/Springer (2017)
- Breger, A., Ehler, M., Gräf, M.: Points on manifolds with asymptotically optimal covering radius. *J. Complex.* **48**, 1–14 (2018)
- Briol, F.X., Oates, C.J., Girolami, M., Osborne, M.A., Sejdinovic, D.: Probabilistic integration: a role in statistical computation? *Stat. Sci.* (to appear) (2018)
- Chavel, I.: *Eigenvalues in Riemannian Geometry*. Academic Press Inc., Cambridge (1984)
- Cockayne, J., Oates, C.J., Sullivan, T., Girolami, M.: Bayesian Probabilistic Numerical Methods. [arXiv:1702.03673](https://arxiv.org/abs/1702.03673) (2017)
- Coulhon, T., Kerkycharian, G., Petrushev, P.: Heat kernel generated frames in the setting of Dirichlet spaces. *J. Fourier Anal. Appl.* **18**(5), 995–1066 (2012)
- Ehler, M., Gräf, M.: Numerically optimizing weights for Monte Carlo integration on smooth compact manifolds. In: *19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, pp. 393–396 (2017)
- Filbir, F., Mhaskar, H.N.: A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel. *J. Fourier Anal. Appl.* **16**(5), 629–657 (2010)
- Gräf, M.: *Efficient Algorithms for the Computation of Optimal Quadrature Points on Riemannian Manifolds*. Universitätsverlag Chemnitz (2013)
- Hardin, D., Michaels, T., Saff, E.: A comparison of popular point configurations on  $\mathbb{S}^2$ . *Dolomit. Res. Notes Approx.* **9**, 16–49 (2016)
- Hellekalek, P., Kritzer, P., Pillichshammer, F.: Open type quasi-Monte Carlo integration based on Halton sequences in weighted Sobolev spaces. *J. Complex.* **33**, 169–189 (2016)
- Hinrichs, A.: Optimal importance sampling for the approximation of integrals. *J. Complex.* **26**(2), 125–134 (2010)
- Hinrichs, A., Markhasin, L., Oettershagen, J., Ullrich, T.: Optimal quasi-Monte Carlo rules on order 2 digital nets for the numerical integration of multivariate periodic functions. *Numer. Math.* **134**(1), 163–196 (2016)

- Hsu, E.P.: Estimates of derivatives of the heat kernel on a compact Riemannian manifold. *Proc. Am. Math. Soc.* **127**(12), 3739–3744 (1999)
- Kanagawa, M., Sriperumbudur, B.K., Fukumizu, K.: Convergence analysis of deterministic Kernel-based quadrature rules in misspecified settings. *Found. Comput. Math.* (2019). <https://doi.org/10.1007/s10208-018-09407-7>
- Karvonen, T., Särkkä, S.: Gaussian Kernel quadrature at scaled Gauss-hermite nodes. *BIT Numer. Math.* (2019). <https://doi.org/10.1007/s10543-019-00758-3>
- Kerkyacharian, G., Petrushev, P., Xu, Y.: Gaussian bounds for the weighted heat Kernels on the interval, ball and simplex. *Constr. Approx.* (2019). <https://doi.org/10.1007/s00365-019-09458-1>
- Kimeldorf, G.S., Wahba, G.: A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat.* **41**(2), 495–502 (1970)
- Larkin, F.M.: Gaussian measure in Hilbert space and applications in numerical analysis. *Rocky Mt. J. Math.* **2**(3), 379–421 (1972)
- Liu, Q., Lee, J.D.: Black-Box Importance Sampling. [arXiv:1610.05247](https://arxiv.org/abs/1610.05247) (2016)
- Marques, R., Bouville, C., Santos, L.P., Bouatouch, K.: Efficient quadrature rules for illumination integrals: from quasi Monte Carlo to Bayesian Monte Carlo. *Synth. Lect. Comput. Gr. Anim.* **7**(2), 1–92 (2015)
- Mhaskar, H.N.: Approximate quadrature measures on data-defined spaces. In: Dick, J., Kuo, F., Wozniakowski, H. (eds.) *Contemporary Computational Mathematics: A Celebration of the 80th Birthday of Lan Sloan*. Springer, Berlin (2018)
- Mhaskar, H.N., Narcowich, F.J., Ward, J.D.: Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature. *Math. Comput.* **70**, 1113–1130 (2001)
- Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B., et al.: Kernel mean embedding of distributions: a review and beyond. *Found. Trends Mach. Learn.* **10**(1–2), 1–141 (2017)
- Niederreiter, H.: Some current issues in quasi-Monte Carlo methods. *J. Complex.* **19**(3), 428–433 (2003)
- Novak, E., Wozniakowski, H.: *Tractability of Multivariate Problems*. EMS Tracts in Mathematics, vol. 12. EMS Publishing House, Zürich (2010)
- Oates, C.J., Girolami, M., Chopin, N.: Control functionals for Monte Carlo integration. *J. R. Stat. Soc. B* **79**(3), 659–718 (2017)
- Oates, C.J., Cockayne, J., Briol, F., Girolami, M.: *Convergence Rates for a Class of Estimators Based on Stein's Method*. Bernoulli, Basel (2018)
- Oettershagen, J.: *Construction of optimal cubature algorithms with applications to econometrics and uncertainty quantification*. Ph.D. Thesis, University of Bonn (2017)
- Petrushev, P., Xu, Y.: Localized polynomial frames on the ball. *Constr. Approx.* **27**, 121–148 (2008)
- Plaskota, L., Wasilkowski, G.W., Zhao, Y.: New averaging technique for approximating weighted integrals. *J. Complex.* **25**, 268–291 (2009)
- Prüher, J., Šimandl, M.: Bayesian quadrature variance in sigma-point filtering. In: *Informatics in Control, Automation and Robotics 12th International Conference, ICINCO 2015 Colmar, France, July 21–23, 2015 Revised Selected Papers*, Springer, pp. 355–370 (2016)
- Rasmussen, C.E., Ghahramani, Z.: Bayesian Monte Carlo. *Adv. Neural Inf. Process. Syst. (NIPS)* **15**, 489–496 (2003)
- Reznikov, A., Saff, E.B.: The covering radius of randomly distributed points on a manifold. *Int. Math. Res. Not.* **2016**, 6065–6094 (2015)
- Ritter, K.: *Average-Case Analysis of Numerical Problems*. Lecture Notes in Mathematics. Springer, Berlin (2000)
- Robert, C., Casella, G.: *Monte Carlo Statistical Methods*. Springer, Berlin (2013)
- Sommariva, A., Vianello, M.: Numerical cubature on scattered data by radial basis functions. *Computing* **76**(3–4), 295–310 (2006)
- Trefethen, L.N.: Cubature, approximation, and isotropy in the hypercube. *SIAM Rev.* **59**(3), 469–491 (2017a)
- Trefethen, L.N.: Multivariate polynomial approximation in the hypercube. *Proc. Am. Math. Soc.* **145**(11), 4837–4844 (2017b)
- Triebel, H.: *Theory of Function Spaces II*. Birkhäuser, Basel (1992)
- Ullrich, M.: A Monte Carlo method for integration of multivariate smooth functions. *SIAM J. Numer. Anal.* **55**(3), 1188–1200 (2017)
- Wünsche, A.: Generalized Zernike or disc polynomials. *J. Comput. Appl. Math.* **174**, 135–163 (2005)
- Xu, Y.: Orthogonal polynomials and cubature formulae on spheres and on balls. *SIAM J. Math. Anal.* **29**(3), 779–793 (1998)
- Xu, Y.: Orthogonal polynomials and cubature formulae on the balls, simplices, and spheres. *J. Comput. Appl. Math.* **127**, 349–368 (2001)

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