



Worst-case optimal approximation with increasingly flat Gaussian kernels

Toni Karvonen^{1,2} · Simo Särkkä¹

Received: 5 June 2019 / Accepted: 24 January 2020 /
Published online: 6 March 2020
© The Author(s) 2020

Abstract

We study worst-case optimal approximation of positive linear functionals in reproducing kernel Hilbert spaces induced by increasingly flat Gaussian kernels. This provides a new perspective and some generalisations to the problem of interpolation with increasingly flat radial basis functions. When the evaluation points are fixed and unisolvent, we show that the worst-case optimal method converges to a polynomial method. In an additional one-dimensional extension, we allow also the points to be selected optimally and show that in this case convergence is to the unique Gaussian quadrature-type method that achieves the maximal polynomial degree of exactness. The proofs are based on an explicit characterisation of the reproducing kernel Hilbert space of the Gaussian kernel in terms of exponentially damped polynomials.

Keywords Worst-case analysis · Reproducing kernel Hilbert spaces · Gaussian kernel · Gaussian quadrature

Mathematics Subject Classification (2010) 41A05 · 41A30 · 46E22 · 65D05 · 65D32

1 Introduction

Most popular kernels used in scattered data approximation [11, 37] and Gaussian process regression [27] are *isotropic* (i.e., radial basis functions), depending only on the Euclidean distance $\|\cdot\|_2$ between the points:

$$K_\ell(x, x') = \Phi\left(\frac{\|x - x'\|_2}{\ell}\right) \quad (1.1)$$

Communicated by: Robert Schaback

✉ Toni Karvonen
tkarvonen@turing.ac.uk

Simo Särkkä
simo.sarkka@aalto.fi

¹ Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland

² The Alan Turing Institute, London, United Kingdom

for a continuous positive-definite function $\Phi : [0, \infty) \rightarrow \mathbb{R}$ and a *length-scale* parameter $\ell > 0$. Given any function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ evaluated at distinct points $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$, such a kernel can be used to construct a unique *kernel interpolant* based on the translates $\{K_\ell(\cdot, x_n)\}_{n=1}^N$. The kernel interpolant is as follows:

$$s_{\ell, f, X}(x) = \sum_{n=1}^N f(x_n) u_{\ell, n}(x), \tag{1.2}$$

where u_n are the Lagrange cardinal functions that solve:

$$\begin{bmatrix} K_\ell(x_1, x_1) & \cdots & K_\ell(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K_\ell(x_N, x_1) & \cdots & K_\ell(x_N, x_N) \end{bmatrix} \begin{bmatrix} u_{\ell, 1}(x) \\ \vdots \\ u_{\ell, N}(x) \end{bmatrix} = \begin{bmatrix} K_\ell(x, x_1) \\ \vdots \\ K_\ell(x, x_N) \end{bmatrix} \tag{1.3}$$

and satisfy $u_{\ell, n}(x_m) = \delta_{nm}$. Uniqueness of the solution for each $x \in \mathbb{R}^d$ is guaranteed by positive-definiteness of the matrix on the left-hand side of this system.

When $\ell \rightarrow \infty$, the kernel K_ℓ becomes increasingly flat and the linear system (1.3) increasingly ill-conditioned.¹ Nevertheless, the corresponding kernel interpolant is typically well-behaved at this limit. Starting with the work of Driscoll and Fornberg [10], it has been shown that a certain unisolvency assumption on X implies that the kernel interpolant converges to (i) a polynomial interpolant if the kernel is infinitely smooth [10, 13, 19, 20, 31, 32] or (ii) a polyharmonic spline interpolant if the kernel is finitely smooth [21, 33]. Further generalisations appear in [22]. The former case covers kernels such as, Gaussians, multiquadrics, and inverse multiquadrics while the latter applies to, for example, Matérn kernels and Wendland’s functions. Among the most interesting of these results is the one by Schaback [31] who proved that the interpolant at the increasingly flat limit of the Gaussian kernel

$$K_\ell(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{2\ell^2}\right) \tag{1.4}$$

exists regardless of the geometry of X and coincides with the de Boor and Ron polynomial interpolant [6, 7]. Furthermore, numerical ill-conditioning for large ℓ , mentioned above, has necessitated the development of techniques for stable evaluation of the kernel interpolant [5, 12, 14, 38]. Increasingly flat kernels have been also discussed independently in the literature on the use of Gaussian processes for numerical integration [24, 26, 34], albeit accompanied only with non-rigorous arguments. Even though the intuition that the lowest degree terms in the Taylor expansion of the kernel dominate construction of the interpolant as $\ell \rightarrow \infty$ and that this ought to imply convergence to a polynomial interpolant is quite clear, this is not always translated into transparent proofs.

The purpose of this article is to generalise the aforementioned results on flat limits of kernel interpolants for worst-case optimal approximation of general positive linear functionals in the reproducing kernel Hilbert space (RKHS) of the Gaussian kernel

¹Note that most of the literature we cite parametrises the kernel in terms of the inverse length-scale $\varepsilon = 1/\ell$ and accordingly considers the case $\varepsilon \rightarrow 0$.

(1.4). That such generalisations are possible is not perhaps surprising; it is rather the simple proof technique made possible by the worst-case framework and an explicit characterisation [23] of the Gaussian RKHS that we find the most interesting aspect of the present work.

1.1 Worst-case optimal approximation

Let Ω be a subset of \mathbb{R}^d with a non-empty interior and $L : C(\Omega) \rightarrow \mathbb{R}$ a positive linear functional acting on continuous real-valued functions defined on Ω and satisfying $L[|p|] < \infty$ for every polynomial p on Ω . The functionals most often discussed in this article are the point evaluation and the integration functionals

$$L_x[f] = f(x) \quad \text{and} \quad L_\mu[f] = \int_\Omega f \, d\mu \quad \text{for a Borel measure } \mu \text{ on } \Omega, \quad (1.5)$$

respectively. Derivative evaluation functionals $L_x^{(n)}[f] = f^{(n)}(x)$ are also often considered. A cubature rule (quadrature if $d = 1$) $Q_X(w) : C(\Omega) \rightarrow \mathbb{R}$ with the distinct points $X = \{x_1, \dots, x_N\} \subset \Omega$ and weights $w = (w(1), \dots, w(N)) \in \mathbb{R}^N$ is a weighted approximation to L of the form

$$Q_X(w)[f] = \sum_{n=1}^N w(n) f(x_n) \approx L[f]. \quad (1.6)$$

When restricted on $\Omega \times \Omega$, the positive-definite kernel K_ℓ in (1.1) induces a unique reproducing kernel Hilbert space $\mathcal{H}(K_\ell) \subset C(\Omega)$ where the reproducing property $\langle f, K_\ell(\cdot, x) \rangle_{\mathcal{H}(K_\ell)} = f(x)$ holds for every $x \in \Omega$ and $f \in \mathcal{H}(K_\ell)$. With minor modifications everything in this section holds also when the kernel is not isotropic. Because the kernel is isotropic, $L[K_\ell(x, x)] \leq L[\Phi(0)] < \infty$ by the assumption that $L[p]$ is finite if p is a polynomial. This guarantees that $L[K_\ell(\cdot, x)] \in \mathcal{H}(K_\ell)$ for any $x \in \Omega$ and consequently that $L[f] < \infty$ for any $f \in \mathcal{H}(K_\ell)$.

The worst-case error $e_\ell(Q_X(w))$ of the cubature rule (1.6) in $\mathcal{H}(K_\ell)$ is as follows:

$$e_\ell(Q_X(w)) = \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} \left| L[f] - \sum_{n=1}^N w(n) f(x_n) \right|. \quad (1.7)$$

Given a fixed set of distinct points, we are interested in the kernel cubature rule $Q_X(w_\ell^*)$ whose weights are chosen so as to minimise the worst-case error:

$$w_\ell^* = \arg \min_{w \in \mathbb{R}^N} e_\ell(Q_X(w)) \quad \text{and} \quad e_\ell(Q_X(w_\ell^*)) = \inf_{w \in \mathbb{R}^N} e_\ell(Q_X(w)).$$

These weights are unique and available as the solution to the linear system [25, Section 3.2]

$$\begin{bmatrix} K_\ell(x_1, x_1) & \cdots & K_\ell(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K_\ell(x_N, x_1) & \cdots & K_\ell(x_N, x_N) \end{bmatrix} \begin{bmatrix} w_\ell^*(1) \\ \vdots \\ w_\ell^*(N) \end{bmatrix} = \begin{bmatrix} L[K_\ell(\cdot, x_1)] \\ \vdots \\ L[K_\ell(\cdot, x_N)] \end{bmatrix}. \quad (1.8)$$

Although our notation does not make this explicit, the weights obviously depend on the linear functional L and the evaluation points X . For each $x \in \mathbb{R}^d$, the kernel

interpolant $s_{\ell, f, X}(x)$ now arises as the kernel cubature rule for approximation of the point evaluation functional L_x in (1.5) and the Lagrange functions are $u_{\ell, n}(x) = w_{\ell}^*(n)$. In this case, the worst-case error coincides with the power function [30]. For an arbitrary L , the kernel cubature rule can be obtained by applying L to the kernel interpolant as follows:

$$Q_X(w_{\ell}^*) = L[s_{\ell, f, X}] = \sum_{n=1}^N f(x_n)L[u_{\ell, n}].$$

That is, the weights are $w_{\ell}^*(n) = L[u_{\ell, n}]$.

1.2 Contributions

Recall that we only consider the Gaussian kernel (1.4). This article contains two theoretical main contributions:

- In Section 2, we prove that if X is unisolvent with respect to a full polynomial space Π_m and $N = \dim \Pi_m$, then $Q_X(w_{\ell}^*)$ converges (as $\ell \rightarrow \infty$) to the unique cubature rule $Q_X(w_{\text{poi}})$ that satisfies $Q_X(w_{\text{poi}})[p] = L[p]$ for every polynomial p of degree at most m . This result, contained in Theorem 2.2 and Corollary 2.1, is a generalisation for arbitrary positive linear functionals of the interpolation results cited earlier. If Ω is bounded, the results hold for any positive linear functional satisfying the mild assumptions imposed earlier. However, boundedness of Ω is not necessary: at the end of Section 2, we supply an example involving integration over \mathbb{R}^d with respect to the Gaussian measure.
- In Section 3, we present a generalisation, based on a theorem of Barrow [2], for *optimal kernel quadrature rules* [25, Chapter 5] that have both their points and weights selected so as to minimise the worst-case error. The result, Theorem 3.2, states that such rules, if unique, converge to the N -point Gaussian quadrature rule for the functional L , which is the unique quadrature rule $Q_{X_G}(w_G)$ such that $Q_{X_G}(w_G)[p] = L[p]$ for every polynomial p of degree at most $2N - 1$. This partially settles a conjecture posed by O'Hagan [26, Section 3.3], and further discussed in [24, 34], on convergence of optimal kernel quadrature rules to Gaussian quadrature rules.

Some generalisations for other kernels and cubature rules of more general form than (1.6) are briefly discussed in Section 4.

2 Fixed points

The following theorem, which provides a characterisation of the RKHS of the Gaussian kernel (1.4), is the central tool of this article. This result is due to Steinwart [36] and Minh [23]; see also [35, Section 4.4] and [8, Example 3]. In this theorem (and the remainder of the article), \mathbb{N}_0^d stands for the collection of d -dimensional non-negative multi-indices: $\mathbb{N}_0^d = \{(\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d : \alpha_1, \dots, \alpha_d \in \mathbb{N}_0\}$. The absolute value and factorial of $\alpha \in \mathbb{N}_0^d$ are $|\alpha| = \alpha_1 + \dots + \alpha_d$ and $\alpha! = \alpha_1! \times \dots \times \alpha_d!$.

Theorem 2.1 (Steinwart 2006; Minh 2010) *Let Ω be a subset of \mathbb{R}^d with a non-empty interior. Then, the RKHS $\mathcal{H}(K_\ell)$ induced by the Gaussian kernel (1.4) with length-scale $\ell > 0$ consists of the functions*

$$f(x) = e^{-\|x\|_2^2/(2\ell^2)} \sum_{\alpha \in \mathbb{N}_0^d} f_\alpha x^\alpha \quad \text{such that} \quad \|f\|_{\mathcal{H}(K_\ell)}^2 = \sum_{\alpha \in \mathbb{N}_0^d} \ell^{2|\alpha|} \alpha! f_\alpha^2 < \infty, \tag{2.1}$$

where convergence is absolute. Its inner product is $\langle f, g \rangle_{\mathcal{H}(K_\ell)} = \sum_{\alpha \in \mathbb{N}_0^d} \ell^{2|\alpha|} \alpha! f_\alpha g_\alpha$. Furthermore, the collection

$$\left\{ \frac{1}{\ell^{|\alpha|} \sqrt{\alpha!}} e^{-\|x\|_2^2/(2\ell^2)} x^\alpha \right\}_{\alpha \in \mathbb{N}_0^d} \tag{2.2}$$

of functions forms an orthonormal basis of $\mathcal{H}(K_\ell)$.

Two crucial implications of this theorem are that $\mathcal{H}(K_\ell)$ consists of functions expressible as series of exponentially damped polynomials, the damping effect vanishing as $\ell \rightarrow \infty$, and that, due to the terms $\ell^{2|\alpha|}$ appearing in the RKHS norm, the high-degree terms contribute the most to the norm. Consequently, the worst-case error (1.7), taking into account only functions of at most unit norm, is dominated by low-degree terms when ℓ is large. The rest of this section formalises this intuition.

Let $\Pi_m \subset C(\Omega)$ stand for the space of d -variate polynomials of degree at most $m \in \mathbb{N}_0$:

$$\Pi_m = \text{span} \left\{ x^\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| \leq m \right\}.$$

In this section, we assume that the point set $X \subset \Omega \subset \mathbb{R}^d$ is Π_m -*unisolvant*. That is,

$$N = \#X = \dim \Pi_m = \binom{m+d}{d} = \frac{(m+d)!}{d!m!}$$

and the zero function is the only element of Π_m that vanishes on X . This is equivalent to non-singularity of the (generalised) Vandermonde matrix

$$P_\Pi = \begin{bmatrix} x_1^{\alpha_1} & \cdots & x_1^{\alpha_N} \\ \vdots & \ddots & \vdots \\ x_N^{\alpha_1} & \cdots & x_N^{\alpha_N} \end{bmatrix}, \tag{2.3}$$

where $\{\alpha_1, \dots, \alpha_N\} = \{\alpha \in \mathbb{N}_0^d : |\alpha| \leq m\} \subset \mathbb{N}_0^d$. It follows that there is a unique polynomial cubature rule $Q_X(w_{\text{pol}})$ such that $Q_X(w_{\text{pol}})[p] = L[p] < \infty$ for every $p \in \Pi_m$. Its weights solve the linear system $P_\Pi^T w_{\text{pol}} = L_\Pi$ of N equations, where the N -vector L_Π has the elements $[L_\Pi]_n = L[x^{\alpha_n}]$. In this section, we prove that the worst-case optimal weights w_ℓ^* for the Gaussian kernel (1.4) converge to w_{pol} as $\ell \rightarrow \infty$.

Define then

$$\phi_\alpha^\ell(x) = e^{-\|x\|_2^2/(2\ell^2)} x^\alpha, \tag{2.4}$$

so that functions in the Gaussian RKHS, characterised by Theorem 2.1, are of the form $f(x) = \sum_{\alpha \in \mathbb{N}_0^d} f_\alpha \phi_\alpha^\ell(x)$ for coefficients f_α decaying sufficiently fast. Since the exponential function has no real roots, the determinant of the matrix

$$P_{\phi, \ell} = \begin{bmatrix} \phi_{\alpha_1}^\ell(x_1) & \cdots & \phi_{\alpha_N}^\ell(x_1) \\ \vdots & \ddots & \vdots \\ \phi_{\alpha_1}^\ell(x_N) & \cdots & \phi_{\alpha_N}^\ell(x_N) \end{bmatrix} \tag{2.5}$$

satisfies $|P_{\phi, \ell}| = |P_\Pi| \exp\left(-\sum_{n=1}^N \|x_n\|_2^2 / (2\ell^2)\right) \neq 0$ and $P_{\phi, \ell}$ is hence non-singular. From non-singularity, it follows that there are unique weights $w_{\phi, \ell}$ such that $Q_X(w_{\phi, \ell})[\phi_\alpha^\ell] = L[\phi_\alpha^\ell]$ for every $\alpha \in \mathbb{N}_0^d$ satisfying $|\alpha| \leq m$. The weights solve $P_{\phi, \ell}^\top w_{\phi, \ell} = L_{\phi, \ell}$, where the N -vector $L_{\phi, \ell}$ has the elements $[L_{\phi, \ell}]_n = L[\phi_{\alpha_n}^\ell]$.² This auxiliary cubature rule plays an important role in our argument. To summarise, the following three weights (or sequences of weights) appear in the proofs below:

1. The weights w_ℓ^* , solved from (1.8), are the worst-case optimal weights for the Gaussian kernel (1.4). The results concern the behaviour of these weights as $\ell \rightarrow \infty$.
2. The weights w_{pol} are constructed such that the cubature rule defined by them is exact for all polynomials up to degree m : $Q_X(w_{\text{pol}})[p] = L[p]$ whenever $p \in \Pi_m$.
3. The auxiliary weights $w_{\phi, \ell}$ satisfy $Q_X(w_{\phi, \ell})[\phi_\alpha^\ell] = L[\phi_\alpha^\ell]$ for every $\ell > 0$ and $|\alpha| \leq m$.

Lemma 2.1 *Suppose that X is Π_m -unisolvent and $\lim_{\ell \rightarrow \infty} L[\phi_\alpha^\ell(x)] = L[x^\alpha]$ for every $|\alpha| \leq m$. Then, there is a constant $C_{\ell_0} \geq 0$ such that $\sup_{\ell \geq \ell_0} \sum_{n=1}^N |w_{\phi, \ell}(n)| \leq C_{\ell_0}$ for any $\ell_0 > 0$.*

Proof The assumption $\lim_{\ell \rightarrow \infty} L[\phi_\alpha^\ell(x)] = L[x^\alpha]$ and unisolvency of X imply that $\lim_{\ell \rightarrow \infty} w_{\phi, \ell} = w_{\text{pol}}$. Because $L[|p|] < \infty$ for any polynomial p , both the weights w_{pol} and $w_{\phi, \ell}$ are finite, which implies the claim. \square

Lemma 2.2 *Suppose that X is Π_m -unisolvent and $\lim_{\ell \rightarrow \infty} L[\phi_\alpha^\ell(x)] = L[x^\alpha]$ for every $|\alpha| \leq m$. If $(w_\ell)_{\ell > 0}$ is any sequence of weights such that*

$$\lim_{\ell \rightarrow \infty} \left| L[\phi_\alpha^\ell] - Q_X(w_\ell)[\phi_\alpha^\ell] \right| = 0 \quad \text{for every } |\alpha| \leq m,$$

then, $\lim_{\ell \rightarrow \infty} w_\ell = w_{\text{pol}}$.

²See [12] for an interpolation method based on a closely related basis derived from a Mercer eigendecomposition of the Gaussian kernel and [17] for an explicit construction of weights similar to $w_{\phi, \ell}$ in the case L is the Gaussian integral.

Proof We have $P_{\Pi}^{\top} w_{\Pi} = L_{\Pi}$ and

$$\|L_{\Pi} - P_{\Pi}^{\top} w_{\ell}\|_2 \leq \|L_{\Pi} - L_{\phi, \ell}\|_2 + \|L_{\phi, \ell} - P_{\phi, \ell}^{\top} w_{\ell}\|_2 + \|P_{\phi, \ell}^{\top} w_{\ell} - P_{\Pi}^{\top} w_{\ell}\|_2,$$

where each of the terms on the right-hand side vanishes as $\ell \rightarrow \infty$. Because $\|L_{\Pi} - P_{\Pi}^{\top} w_{\ell}\|_2 = \|P_{\Pi}^{\top}(w_{\text{pol}} - w_{\ell})\|_2$ and P_{Π} is non-singular, we conclude that $\lim_{\ell \rightarrow \infty} w_{\ell} = w_{\text{pol}}$. □

We are ready to prove the main result of the article for a fixed Π_m -unisolvent point set $X \subset \Omega$ consisting of N distinct points. First, by considering one of the basis functions (2.2), we show that $|L[\phi_{\alpha}^{\ell}] - Q_X(w_{\ell}^*)[\phi_{\alpha}^{\ell}]| \leq \sqrt{\alpha!} \ell^{|\alpha|} e_{\ell}(Q_X(w_{\ell}^*))$ for every $\alpha \in \mathbb{N}_0^d$. Second, the sub-optimal cubature rule $Q_X(w_{\phi, \ell})$ defined above can be used, in combination with (2.1), to establish the upper bound $e_{\ell}(Q_X(w_{\ell}^*)) \leq C \ell^{-(m+1)}$. These two bounds imply that $|L[\phi_{\alpha}^{\ell}] - Q_X(w_{\ell}^*)[\phi_{\alpha}^{\ell}]| \rightarrow 0$ for every $|\alpha| \leq m$. If $\lim_{\ell \rightarrow \infty} L[\phi_{\alpha}^{\ell}(x)] = L[x^{\alpha}]$, Lemma 2.2 then implies that $w_{\ell}^* \rightarrow w_{\text{pol}}$.

Theorem 2.2 *Let $N = \dim \Pi_m$ for some $m \in \mathbb{N}_0$ and X be Π_m -unisolvent. Suppose that $\lim_{\ell \rightarrow \infty} L[\phi_{\alpha}^{\ell}(x)] = L[x^{\alpha}]$ for every $\alpha \in \mathbb{N}_0^d$ such that $|\alpha| \leq m$ and that*

$$L \left[\sum_{|\alpha| \geq m+1} \frac{|a_{\alpha}|}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha!}} |x^{\alpha}| \right] \leq C_L < \infty \tag{2.6}$$

for some $\ell_0 > 1$ and any sequence $(a_{\alpha})_{\alpha \in \mathbb{N}_0^d}$ such that $\sum_{\alpha \in \mathbb{N}_0^d} a_{\alpha}^2 \leq 1$. Then,

$$\lim_{\ell \rightarrow \infty} w_{\ell}^* = w_{\text{pol}} \quad \text{and} \quad e_{\ell}(Q_X(w_{\ell}^*)) = \mathcal{O}(\ell^{-(m+1)}),$$

where w_{pol} are the weights of the unique polynomial cubature rule such that $Q_X(w_{\text{pol}})[p] = L[p]$ for every $p \in \Pi_m$.

Proof For every $\alpha \in \mathbb{N}_0^d$, select the function

$$g_{\alpha}(x) = \frac{1}{\ell^{|\alpha|} \sqrt{\alpha!}} e^{-\|x\|_2^2 / (2\ell^2)} x^{\alpha} = \frac{1}{\ell^{|\alpha|} \sqrt{\alpha!}} \phi_{\alpha}^{\ell}(x).$$

From Theorem 2.1, it follows that $\|g_{\alpha}\|_{\mathcal{H}(K_{\ell})}^2 = 1$ since g_{α} is one of the basis functions (2.2). Thus, by definition of the worst-case error,

$$\frac{1}{\ell^{|\alpha|} \sqrt{\alpha!}} |L[\phi_{\alpha}^{\ell}] - Q_X(w_{\ell}^*)[\phi_{\alpha}^{\ell}]| = |L[g_{\alpha}] - Q_X(w_{\ell}^*)[g_{\alpha}]| \leq e_{\ell}(Q_X(w_{\ell}^*)). \tag{2.7}$$

Next, we derive an appropriate upper bound on $e_{\ell}(Q_X(w_{\ell}^*))$ by considering the unique sub-optimal cubature rule $Q_X(w_{\phi, \ell})$ that is exact for every ϕ_{α}^{ℓ} with $|\alpha| \leq m$.

In the expansion (2.1) of a function in $\mathcal{H}(K_\ell)$, we have $L[\phi_\alpha^\ell] = Q_X(w_{\phi,\ell})[\phi_\alpha^\ell]$ for every term with $|\alpha| \leq m$. Consequently, the worst-case error is bounded as follows:

$$\begin{aligned} & e_\ell(Q_X(w_{\phi,\ell})) \\ &= \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} \left\| L \left[\sum_{|\alpha| \geq m+1} f_\alpha \phi_\alpha^\ell \right] - Q_X(w_{\phi,\ell}) \left[\sum_{|\alpha| \geq m+1} f_\alpha \phi_\alpha^\ell \right] \right\| \\ &\leq \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} L \left[\sum_{|\alpha| \geq m+1} |f_\alpha| |\phi_\alpha^\ell| \right] + \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} \left\| Q_X(w_{\phi,\ell}) \left[\sum_{|\alpha| \geq m+1} f_\alpha \phi_\alpha^\ell \right] \right\|, \end{aligned}$$

where f_α are the coefficients that define $f \in \mathcal{H}(K_\ell)$ in Theorem 2.1. A consequence of (2.1) is that $\|f\|_{\mathcal{H}(K_\ell)} \leq 1$ implies $|f_\alpha| \leq a_\alpha / (\ell^{|\alpha|} \sqrt{\alpha})$ for some real numbers $|a_\alpha| \leq 1$ such that $\sum_{\alpha \in \mathbb{N}_0^d} a_\alpha^2 \leq 1$. Therefore, for $\ell \geq \ell_0 > 1$,

$$\begin{aligned} \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} L \left[\sum_{|\alpha| \geq m+1} |f_\alpha| |\phi_\alpha^\ell| \right] &\leq L \left[\sum_{|\alpha| \geq m+1} \frac{|a_\alpha|}{\ell^{|\alpha|} \sqrt{\alpha}} |\phi_\alpha^\ell| \right] \\ &\leq \ell^{-(m+1)} L \left[\sum_{|\alpha| \geq m+1} \frac{|a_\alpha|}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha}} |\phi_\alpha^\ell| \right] \\ &\leq \ell^{-(m+1)} L \left[\sum_{|\alpha| \geq m+1} \frac{|a_\alpha|}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha}} |x^\alpha| \right] \\ &\leq C_L \ell^{-(m+1)} \end{aligned}$$

by assumption (2.6). Moreover, because

$$\max_{n=1, \dots, N} |\phi_\alpha^\ell(x_n)| \leq \max_{n=1, \dots, N} |x_n^\alpha| \leq C_X$$

for some $C_X > 0$ and every ℓ , we have the following:

$$\begin{aligned} & \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} \left\| Q_X(w_{\phi,\ell}) \left[\sum_{|\alpha| \geq m+1} f_\alpha \phi_\alpha^\ell \right] \right\| \\ &\leq \sup_{\|f\|_{\mathcal{H}(K_\ell)} \leq 1} \sum_{n=1}^N |w_{\phi,\ell}(n)| \sum_{|\alpha| \geq m+1} |f_\alpha| |\phi_\alpha^\ell(x_n)| \\ &\leq \ell^{-(m+1)} \sum_{n=1}^N |w_{\phi,\ell}(n)| \sum_{|\alpha| \geq m+1} \frac{|a_\alpha|}{\ell^{|\alpha|-(m+1)} \sqrt{\alpha}} |\phi_\alpha^\ell(x_n)| \end{aligned}$$

$$\begin{aligned} &\leq \ell^{-(m+1)} \sum_{n=1}^N |w_{\phi, \ell}(n)| \sum_{|\alpha| \geq m+1} \frac{C_X}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha!}} \\ &\leq \ell^{-(m+1)} \left(\sup_{\ell \geq \ell_0} \sum_{n=1}^N |w_{\phi, \ell}(n)| \right) \sum_{|\alpha| \geq m+1} \frac{C_X}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha!}} \\ &=: C_Q \ell^{-(m+1)} \end{aligned}$$

where $C_Q < \infty$ follows from convergence of the last term and Lemma 2.1. Thus,

$$e_\ell(Q_X(w_{\phi, \ell})) \leq (C_L + C_Q)\ell^{-(m+1)} =: C\ell^{-(m+1)} \tag{2.8}$$

when $\ell \geq \ell_0$. Since $Q_X(w_\ell^*)$ is worst-case optimal, we have thus established with (2.7) and (2.8) that, for sufficiently large ℓ ,

$$\frac{1}{\ell^{|\alpha|} \sqrt{\alpha!}} |L[\phi_\alpha^\ell] - Q_X(w_\ell^*)[\phi_\alpha^\ell]| \leq e_\ell(Q_X(w_\ell^*)) \leq e_\ell(Q_X(w_{\phi, \ell})) \leq C\ell^{-(m+1)}$$

for every $\alpha \in \mathbb{N}_0^d$ such that $|\alpha| \leq m$ and a constant C independent of ℓ . That is,

$$|L[\phi_\alpha^\ell] - Q_X(w_\ell^*)[\phi_\alpha^\ell]| \leq C\sqrt{\alpha!} \ell^{-(m+1)+|\alpha|} \leq C\sqrt{m!} \ell^{-1} \rightarrow 0 \quad \text{as } \ell \rightarrow \infty. \tag{2.9}$$

The claim then follows by setting $w_\ell = w_\ell^*$ in Lemma 2.2. □

Assumptions of Theorem 2.2 hold, for instance, if the domain Ω is bounded.

Corollary 2.1 *Let $N = \dim \Pi_m$ for some $m \in \mathbb{N}_0$ and X be Π_m -unisolvent. Suppose that Ω is bounded. Then,*

$$\lim_{\ell \rightarrow \infty} w_\ell^* = w_{\text{pol}} \quad \text{and} \quad e_\ell(Q_X(w_\ell^*)) = \mathcal{O}(\ell^{-(m+1)}),$$

where w_{pol} are the weights of the unique polynomial cubature rule such that $Q_X(w_{\text{pol}})[p] = L[p]$ for every $p \in \Pi_m$.

Proof On a bounded domain, the convergence $\phi_\alpha^\ell(x) \rightarrow x^\alpha$ as $\ell \rightarrow \infty$ is uniform. Thus,

$$|L[x^\alpha] - L[\phi_\alpha^\ell]| \leq L[1] \sup_{x \in \Omega} |x^\alpha - \phi_\alpha^\ell(x)| \rightarrow 0$$

as $\ell \rightarrow \infty$ for every $\alpha \in \mathbb{N}_0^d$. Assumption (2.6) is also satisfied as follows:

$$L \left[\sum_{|\alpha| \geq m+1} \frac{|a_\alpha|}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha!}} |x^\alpha| \right] \leq L \left[\sum_{|\alpha| \geq m+1} \frac{\beta^\alpha}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha!}} \right] < \infty,$$

where $\beta = (b, \dots, b) \in \mathbb{R}^d$ for $b = \sup_{z \in \Omega} \|z\|_2$ and finiteness follows from the assumption $L[1] < \infty$. □

However, boundedness of Ω is not necessary. Consider Gaussian integration as follows:

$$L[f] = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) e^{-\|x\|_2^2/2} dx = \prod_{i=1}^d \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-x_i^2/2} dx_i \right].$$

If $\alpha \in \mathbb{N}_0^d$ has an odd element, $L[\phi_\alpha^\ell] = L[x^\alpha] = 0$ for every $\ell > 0$ by symmetry. If $\alpha = 2\beta$ for some $\beta \in \mathbb{N}_0^d$, the convergence $L[\phi_\alpha^\ell(x)] \rightarrow L[x^\alpha]$ as $\ell \rightarrow \infty$ follows from the monotone convergence theorem. To verify (2.6), recall that the absolute moments of the standard Gaussian distribution are as follow:

$$\begin{aligned} L[|x^\alpha|] &= \pi^{-d/2} \prod_{i=1}^d 2^{\alpha_i/2} \Gamma\left(\frac{\alpha_i + 1}{2}\right) \\ &= \left[\prod_{\alpha_i \text{ odd}} \pi^{-1/2} 2^{\alpha_i/2} \left(\frac{\alpha_i - 1}{2}\right)! \right] \times \left[\prod_{\alpha_i \text{ even}} (\alpha_i - 1)!! \right], \end{aligned}$$

where $\Gamma(\cdot)$ is the Gamma function. Because $(n - 1)!! \leq \sqrt{n!}$ for any $n \in \mathbb{N}$ and

$$\pi^{-1/2} \frac{2^{n/2}}{\sqrt{n!}} \left(\frac{n - 1}{2}\right)! = \pi^{-1/2} \frac{2^{n/2}}{\sqrt{n!}} \times \frac{(n - 1)!!}{2^{(n-1)/2}} = \sqrt{\frac{2}{\pi}} \frac{(n - 1)!!}{\sqrt{n!}} \leq \sqrt{\frac{2}{\pi}} \leq 1$$

if n is odd, we have the following:

$$\frac{L[|x^\alpha|]}{\sqrt{\alpha!}} = \left[\prod_{\alpha_i \text{ odd}} \pi^{-1/2} \frac{2^{\alpha_i/2}}{\sqrt{\alpha_i!}} \left(\frac{\alpha_i - 1}{2}\right)! \right] \times \left[\prod_{\alpha_i \text{ even}} \frac{(\alpha_i - 1)!!}{\sqrt{\alpha_i!}} \right] \leq 1.$$

Thus,

$$L \left[\sum_{|\alpha| \geq m+1} \frac{|\alpha_\alpha|}{\ell_0^{|\alpha|-(m+1)} \sqrt{\alpha!}} |x^\alpha| \right] \leq \sum_{|\alpha| \geq m+1} \frac{1}{\ell_0^{|\alpha|-(m+1)}} < \infty$$

if $\ell_0 > 1$.

3 Optimal points in one dimension

Let $d = 1$ and $\Omega = [a, b]$ for $a < b$. In this section, we consider quadrature rules whose points are also selected so as to minimise the worst-case error. A kernel quadrature rule is *optimal* if its points and weights satisfy the following:

$$e_\ell \left(Q_{X_\ell^*}(w_\ell^*) \right) = \inf_{w \in \mathbb{R}^N, X \in \Omega^N} e_\ell (Q_X(w)).$$

In order to eliminate degrees of freedom in ordering the points, we require that the points are in ascending order (i.e., $x_n \leq x_{n+1}$). Even though optimal kernel quadrature rules have been studied since the 1970s [1, 3, 18, 28, 29] for the integration functional $L[f] = \int_a^b f(x)\omega(x)dx$, $\omega(x) > 0$, their theory is still not complete (the

main results have been recently collated by Oettershagen [25, Section 5.1]). Although uniqueness results have been proved only for totally positive isotropic kernels of the form (1.1) and integration when $\omega \equiv 1$ [4], there exists numerical evidence suggesting that the optimal rule is unique in more general settings [25, p. 97]. Note that the Gaussian kernel (1.4) we consider is totally positive.

In Theorem 3.2, we show that uniqueness of an optimal kernel quadrature rule for each $\ell > 0$ implies that its increasingly flat limit is $Q_G = Q_{X_G}(w_G)$, the N -point Gaussian quadrature rule for the linear functional L . This is the unique quadrature rule that is exact for every polynomial of degree at most $2N - 1$: $Q_G[x^n] = L[x^n]$ whenever $n \leq 2N - 1$. This degree of exactness is maximal; there are no N -point quadrature rules exact for all polynomials up to degree $2N$. The most familiar methods of this type are of course the classical Gaussian quadrature rules for numerical integration [15, Section 1.4]. For example, the Gauss–Legendre quadrature rule satisfies the following:

$$Q_G[p] = \int_{-1}^1 p(x)dx$$

for every polynomial p of degree at most $2N - 1$ and its points are the roots of the N th degree Legendre polynomial. Theorem 3.2 was conjectured by O’Hagan [26, Section 3.3] in 1991 in the form that the optimal kernel quadrature rule has the classical Gauss–Hermite quadrature rule as its increasingly flat limit if the kernel is Gaussian and L is the Gaussian integral. More discussion of this conjecture—but no rigorous proofs—can be found in [24, Section 4].

The proof of Theorem 3.2 is based on a general result by Barrow [2] on existence and uniqueness of *generalised Gaussian quadrature rules*. This result replaces the polynomials in a Gaussian quadrature rule with *generalised polynomials* formed out of functions that constitute an *extended Chebyshev system* [16, Chapter 1]. A collection $\{u_n\}_{n=0}^{m-1} \subset C^{m-1}([a, b])$ of functions is an extended Chebyshev system if any non-trivial linear combination of the functions has at most $m - 1$ zeroes, counting multiplicities. That is, if $u \in \text{span}\{u_n\}_{n=0}^{m-1}$ and $u^{(q_p)}(x_p) = 0$ for $x_p \in [a, b]$, $p = 1, \dots, P$, and $q_p = 0, \dots, Q_p - 1$, then $\sum_{p=1}^P Q_p \leq m - 1$. Any basis of the space of polynomials of degree at most $m - 1$ is an extended Chebyshev system. Importantly, the functions $\{\phi_n^\ell\}_{n=0}^{m-1}$ in (2.4) are an extended Chebyshev system for any $m \in \mathbb{N}$. To verify this, note that any $\phi \in \text{span}\{\phi_n^\ell\}_{n=0}^{m-1}$ can be written as $\phi(x) = e^{-x^2/(2\ell^2)} p(x)$ for some polynomial p of degree at most $m - 1$ and consequently

$$\phi^{(l)}(x) = e^{-x^2/(2\ell^2)} \left(\sum_{r=0}^{l-1} s_r(x) p^{(r)}(x) + p^{(l)}(x) \right)$$

for some polynomials s_r . From this expression, we see that $\phi^{(l)}(x) = 0$ for every $l = 0, \dots, q$ if and only if $p^{(l)}(x) = 0$ for every $l = 0, \dots, q$. Since p can have at most $m - 1$ zeroes, counting multiplicities, it follows that the same is true of ϕ .

Theorem 3.1 (Barrow 1978) *Let $\{u_n\}_{n=0}^{2N-1} \subset C^{2N-1}([a, b])$ be an extended Chebyshev system and L a positive linear functional on $\text{span}\{u_n\}_{n=0}^{2N-1}$. Then, there exist unique points $a < x_1 < \dots < x_N < b$ and positive weights $w \in \mathbb{R}^N$ such that*

$$Q_X(w)[u_n] = L[u_n] \quad \text{for every } n = 0, \dots, 2N - 1.$$

Lemma 3.1 *Let $\Omega \subset \mathbb{R}^d$ and suppose that a cubature rule $Q_X(w)$ with non-negative weights satisfies $Q_X(w)[u] = L[u]$ for some positive function $u: \Omega \rightarrow (0, \infty)$ such that $0 < c_l \leq u(x) \leq c_u$ for all $x \in \Omega$. Then,*

$$\max_{n=1, \dots, N} w(n) \leq \sum_{n=1}^N w(n) \leq L[1] \frac{c_u}{c_l}.$$

Proof The claim follows immediately from the inequalities

$$c_l \sum_{n=1}^N w(n) \leq \inf_{x \in \Omega} u(x) \sum_{n=1}^N w(n) \leq \sum_{n=1}^N w(n) u(x_n) = L[u] \leq L[1] c_u.$$

□

Lemma 3.2 *Let A be a metric space, $\ell_0 > 0$ a constant, and $g: [\ell_0, \infty) \times A \rightarrow [0, \infty)$ a function. If there is a continuous function $g_\infty: A \rightarrow [0, \infty)$ such that $g(\ell, \cdot) \rightarrow g_\infty$ uniformly as $\ell \rightarrow \infty$ and a unique minimiser x_∞^* for which $g_\infty(x_\infty^*) = 0$, then any function $z: [\ell_0, \infty) \rightarrow A$ such that $\lim_{\ell \rightarrow \infty} g(\ell, z(\ell)) = 0$ has $\lim_{\ell \rightarrow \infty} z(\ell) = x_\infty^*$.*

Proof The inequality $g_\infty(z(\ell)) \leq g(\ell, z(\ell)) + |g_\infty(z(\ell)) - g(\ell, z(\ell))|$ shows that $g_\infty(z(\ell)) \rightarrow 0$ since $g(\ell, z(\ell)) \rightarrow 0$ by assumption and $|g_\infty(z(\ell)) - g(\ell, z(\ell))| \rightarrow 0$ by uniformity of the convergence $g(\ell, \cdot) \rightarrow g_\infty$. Because g_∞ is continuous, non-negative, and has a unique minimiser x_∞^* , this implies that $z(\ell) \rightarrow x_\infty^*$. □

Theorem 3.2 *Suppose that $\Omega = [a, b]$ for $a < b$. If for every $\ell > 0$, there exists a unique optimal kernel quadrature rule $Q_\ell^* = Q_{X_\ell^*}(w_\ell^*)$, then its points and weights converge to those of the N -point Gaussian quadrature rule for L :*

$$\lim_{\ell \rightarrow \infty} X_\ell^* = X_G \quad \text{and} \quad \lim_{\ell \rightarrow \infty} w_\ell^* = w_G,$$

where X_G and w_G are the unique points and weights such that $Q_{X_G}(w_G)[x^n] = L[x^n]$ for every $0 \leq n \leq 2N - 1$. Moreover, $e_\ell(Q_\ell^*) = O(\ell^{-2N})$.

Proof In a manner identical to the proof of Theorem 2.2, we establish the lower bound

$$\frac{1}{\ell^n \sqrt{n!}} |L[\phi_n^\ell] - Q_\ell^*[\phi_n^\ell]| \leq e_\ell(Q_\ell^*)$$

that holds for every $n \geq 0$. Because $\{\phi_n^\ell\}_{n=0}^{2N-1}$ is an extended Chebyshev system, Theorem 3.1 guarantees the existence of a unique N -point quadrature rule

$Q_G^\ell = Q_{X_G^\ell}(w_G^\ell)$ such that $Q_G^\ell[\phi_n^\ell] = L[\phi_n^\ell]$ for every $n \leq 2N - 1$. The points $X_G^\ell = \{x_1^{G,\ell}, \dots, x_N^{G,\ell}\}$ of this rule are distinct and lie inside Ω and the weights w_G^ℓ are positive. We can then replicate the rest of the proof of Theorem 2.2 in one dimension but with $m = 2N - 1$ and Lemma 2.1 replaced with Lemma 3.1 (applied to the function $u = \phi_0^\ell$) to show that, for sufficiently large ℓ and a constant C independent of ℓ ,

$$\frac{1}{\ell^n \sqrt{n!}} |L[\phi_n^\ell] - Q_\ell^*[\phi_n^\ell]| \leq e_\ell(Q_\ell^*) \leq e_\ell(Q_G^\ell) \leq C\ell^{-2N}$$

for every $n \leq 2N - 1$. Consequently,

$$|L[\phi_n^\ell] - Q_\ell^*[\phi_n^\ell]| \leq C\sqrt{n!} \ell^{n-2N} \leq C\sqrt{(2N - 1)!} \ell^{-1} \rightarrow 0 \quad \text{as } \ell \rightarrow \infty \tag{3.1}$$

for every $n \leq 2N - 1$. We then fix $\ell_0 > 0$ and invoke Lemma 3.2 with the function

$$g(\ell, (X, w)) = \sum_{n=0}^{2N-1} |L[\phi_n^\ell] - Q_X(w)[\phi_n^\ell]|,$$

domain $A = (\Omega^N \times [0, \infty)^N)$, and $z(\ell) = (X_\ell^*, w_\ell^*)$. Because the domain $\Omega = [a, b]$ is bounded, $\lim_{\ell \rightarrow \infty} L[\phi_n^\ell] \rightarrow L[x^n]$ for every $n \in \mathbb{N}_0$. Thus,

$$g(\ell, (X, w)) \rightarrow g_\infty((X, w)) := \sum_{n=0}^{2N-1} |L[x^n] - Q_X(w)[x^n]| \quad \text{as } \ell \rightarrow \infty$$

uniformly on A . Since the unique minimiser of g_∞ is (X_G, w_G) , the claim follows from (3.1) and Lemma 3.2. □

4 Generalisations

This section discusses some straightforward generalisations of the results in Sections 2 and 3.

4.1 Damped power series kernels

Theorem 2.1 for the Gaussian kernel (1.4) is a consequence of the identity

$$\begin{aligned} K_\ell(x, x') &= e^{-\|x\|_2/(2\ell^2)} e^{-\|x'\|_2/(2\ell^2)} \sum_{\alpha \in \mathbb{N}_0^d} \frac{1}{\alpha! \ell^{2|\alpha|}} x^\alpha (x')^\alpha \\ &=: e^{-\|x\|_2/(2\ell^2)} e^{-\|x'\|_2/(2\ell^2)} K_\ell^{\text{pow}}(x, x'), \end{aligned}$$

where $K_\ell^{\text{pow}}(x, x')$ is a *power series kernel* [39]. Accordingly, the results in Sections 2 and 3 can be generalised for a class of kernels that we call *damped power series*

kernels. Let $G : \mathbb{R}^d \rightarrow \mathbb{R} \setminus \{0\}$ be a non-zero function and define $G_\ell(x) = G(\|x\|/\ell)$. Then, a damped power series kernel is defined as follows:

$$K_\ell(x, x') = G_\ell(x)G_\ell(x') \sum_{\alpha \in \mathbb{N}_0^d} \frac{\omega_\alpha}{(\alpha!)^2 \ell^{q|\alpha|}} x^\alpha (x')^\alpha \tag{4.1}$$

for $q > 0$ and weight parameters $\omega_\alpha > 0$ such that the series converges for any $\ell > 0$ and $x, x' \in \Omega$. Arguments identical to those used in [23, 39] establish that K_ℓ defined in (4.1) is a positive-definite kernel and that its RKHS $\mathcal{H}(K_\ell)$ consists of functions

$$f(x) = G_\ell(x) \sum_{\alpha \in \mathbb{N}_0^d} f_\alpha x^\alpha \quad \text{such that} \quad \|f\|_{\mathcal{H}(K_\ell)}^2 = \sum_{\alpha \in \mathbb{N}_0^d} \frac{(\alpha!)^2 \ell^{q|\alpha|}}{\omega_\alpha} f_\alpha^2 < \infty.$$

The Gaussian kernel is recovered by setting $G(x) = e^{-\|x\|_2^2/2}$, $q = 2$, and $\omega_\alpha = \alpha!$. Note that the Gaussian kernel is an exception; damped power series kernels are rarely stationary.

Denote $\psi_\alpha^\ell(x) = G_\ell(x)x^\alpha$. If we assume that (i) G is bounded, (ii) $\lim_{\ell \rightarrow \infty} L[\psi_\alpha^\ell(x)] = L[x^\alpha]$ for every $\alpha \in \mathbb{N}_0^d$, and (iii) a summability condition analogous to (2.6) holds, then a generalisation of Theorem 2.2 for damped power series kernels is readily obtained. To generalise Theorem 3.2, we also need to assume that $\{\psi_n\}_{n=0}^{2N-1}$ constitutes an extended Chebyshev system.

4.2 Taylor space kernels

Let $d = 1$. Taylor space kernels [9, 40] are obtained by selecting $G \equiv 1$ in (4.1). As $\ell \rightarrow \infty$, the corresponding kernel quadrature rules then converge to polynomial rules. Perhaps the two most interesting special cases are the exponential kernel are

$$K_\ell(x, x') = \exp\left(\frac{xx'}{\ell}\right) = \sum_{n=0}^{\infty} \frac{(xx')^n}{\ell^n n!}$$

and the Szegő kernel

$$K_\ell(x, x') = \frac{\ell^2}{\ell^2 - xx'} = \frac{1}{1 - \ell^{-2}xx'} = \sum_{n=0}^{\infty} \ell^{-2n} (xx')^n.$$

The Szegő kernel induces a Hardy space on a disk of radius ℓ . Interestingly, it has been pointed out already in the 1970s that approximation with the Szegő kernel yields polynomial methods as $\ell \rightarrow \infty$ [18, Section 3]. See also [24, Section 4]. An extensive numerical investigation has been recently published by Oettershagen [25, Section 6.2].

4.3 General information functionals

It would also be easy to replace the cubature rule (1.6) with a generalised version as follows:

$$Q[f] = \sum_{n=1}^N w(n)L_n[f],$$

where L_n are any bounded linear functionals. If L_n are such that the matrices

$$\begin{bmatrix} L_1[x^{\alpha_1}] & \cdots & L_1[x^{\alpha_N}] \\ \vdots & \ddots & \vdots \\ L_N[x^{\alpha_1}] & \cdots & L_N[x^{\alpha_N}] \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} L_1[\phi_{\alpha_1}^\ell(x)] & \cdots & L_1[\phi_{\alpha_N}^\ell(x)] \\ \vdots & \ddots & \vdots \\ L_N[\phi_{\alpha_1}^\ell(x)] & \cdots & L_N[\phi_{\alpha_N}^\ell(x)] \end{bmatrix},$$

which are generalisations of (2.3) and (2.5), are non-singular, then Theorem 2.2 and Corollary 2.1 can be generalised.

4.4 Non-unisolvent point sets

If the kernel is Gaussian but point set $X \subset \Omega$ is not unisolvent, Schaback [31] has proved that the kernel interpolant (1.2) converges the de Boor and Ron polynomial interpolant [6, 7], which is the unique interpolant to f at X in a point-dependent polynomial space Π_X having in a certain sense minimal degree. We expect that extensions for non-unisolvent points of the results in Section 2 are possible. The kernel cubature weights would presumably convergence to the weights w'_{pol} such that $Q_X(w'_{\text{pol}})[p] = L[p]$ for every $p \in \Pi_X$.

Acknowledgements We thank the reviewers for numerous comments that helped in improving the presentation.

Funding information Open access funding provided by Aalto University. This work was supported by the Aalto ELEC Doctoral School and the Academy of Finland.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Barrar, R.B., Loeb, H.L., Werner, H.: On the existence of optimal integration formulas for analytic functions. *Numer. Math.* **23**(2), 105–117 (1974)
2. Barrow, D.L.: On multiple node Gaussian quadrature formulae. *Math. Comput.* **32**(142), 431–439 (1978)

3. Bojanov, B.D.: On the existence of optimal quadrature formulae for smooth functions. *Calcolo* **16**(1), 61–70 (1979)
4. Braess, D., Dyn, N.: On the uniqueness of monosplines and perfect splines of least L_1 - and L_2 -norm. *J. d'Analyse Math.* **41**(1), 217–233 (1982)
5. Cavoretto, R., Fasshauer, G.E., McCourt, M.: An introduction to the Hilbert-Schmidt SVD using iterated Brownian bridge kernels. *Num. Algorithms* **68**(2), 393–422 (2015)
6. de Boor, C.: Polynomial interpolation in several variables. In: Rice, J., DeMillo, R.A. (eds.) *Studies in Computer Science*, pp. 87–109 (1994)
7. de Boor, C., Ron, A.: The least solution for the polynomial interpolation problem. *Math. Z.* **210**(1), 347–378 (1992)
8. De Marchi, S., Schaback, R.: Nonstandard kernels and their applications. *Dolomites Research Notes on Approximation* **2**(1), 16–43 (2009)
9. Dick, J.: A Taylor space for multivariate integration. *Monte Carlo Methods Appl.* **12**(2), 99–112 (2006)
10. Driscoll, T.A., Fornberg, B.: Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.* **43**(3–5), 413–422 (2002)
11. Fasshauer, G., McCourt, M.: *Kernel-based Approximation Methods Using MATLAB*. Number 19 in *Interdisciplinary Mathematical Sciences*. World Scientific Publishing (2015)
12. Fasshauer, G.E., McCourt, M.J.: Stable evaluation of Gaussian radial basis function interpolants. *SIAM J. Sci. Comput.* **34**(2), A737–A762 (2012)
13. Fornberg, B., Wright, G., Larsson, E.: Some observations regarding interpolants in the limit of flat radial basis functions. *Comput. Math. Appl.* **47**(1), 37–55 (Jan 2004)
14. Fornberg, B., Lehto, E., Powell, C.: Stable calculation of Gaussian-based RBF-FD stencils. *Comput. Math. Appl.* **65**(4), 627–637 (2013)
15. Gautschi, W.: *Orthogonal Polynomials: Computation and Approximation*. Numerical mathematics and scientific computation. Oxford University Press, Oxford (2004)
16. Karlin, S.: *Techebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, New York (1966)
17. Karvonen, T., Särkkä, S.: Gaussian kernel quadrature at scaled Gauss–Hermite nodes. *BIT Numer. Math.* **59**(4), 877–902 (2019)
18. Larkin, F.M.: Optimal approximation in Hilbert spaces with reproducing kernel functions. *Math. Comput.* **24**(112), 911–921 (1970)
19. Larsson, E., Fornberg, B.: Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.* **49**(1), 103–130 (2005)
20. Lee, Y.J., Yoon, G.J., Yoon, J.: Convergence of increasingly flat radial basis interpolants to polynomial interpolants. *SIAM J. Math. Anal.* **39**(2), 537–553 (2007)
21. Lee, Y.J., Micchelli, C.A., Yoon, J.: On convergence of flat multivariate interpolation by translation kernels with finite smoothness. *Constr. Approx.* **40**(1), 37–60 (2014)
22. Lee, Y.J., Micchelli, C.A., Yoon, J.: A study on multivariate interpolation by increasingly flat kernel functions. *J. Math. Anal. Appl.* **427**(1), 74–87 (2015)
23. Minh, H.Q.: Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constr. Approx.* **32**(2), 307–338 (2010)
24. Minka, T.: *Deriving quadrature rules from Gaussian processes*. Technical report, Statistics Department, Carnegie Mellon University (2000)
25. Oettershagen, J.: *Construction of Optimal Cubature Algorithms with Applications to Econometrics and Uncertainty Quantification*. PhD thesis. Institut für Numerische Simulation, Universität Bonn (2017)
26. O'Hagan, A.: Bayes–Hermite quadrature. *J. Stat. Plan. Inference* **29**(3), 245–260 (1991)
27. Rasmussen, C.E., Williams, C.K.I.: *Gaussian Processes for Machine Learning Adaptive Computation and machine learning*. MIT Press, Cambridge (2006)
28. Richter, N.: Properties of minimal integration rules. *SIAM J. Numer. Anal.* **7**(1), 67–79 (1970)
29. Richter-Dyn, N.: Properties of minimal integration rules. II. *SIAM J. Numer. Anal.* **8**(3), 497–508 (1971)
30. Schaback, R.: Comparison of radial basis function interpolants. In: *Multivariate approximation: From CAGD to wavelets*, pp. 293–305. World Scientific (1993)
31. Schaback, R.: Multivariate interpolation by polynomials and radial basis functions. *Constr. Approx.* **21**(3), 293–317 (2005)

32. Schaback, R.: Limit problems for interpolation by analytical radial basis functions. *J. Comput. Appl. Math.* **212**(2), 127–149 (2008)
33. Song, G., Riddle, J., Fasshauer, G.E., Hickernell, F.J.: Multivariate interpolation with increasingly flat radial basis functions of finite smoothness. *Adv. Comput. Math.* **36**(3), 485–501 (2012)
34. Särkkä, S., Hartikainen, J., Svensson, L., Sandblom, F.: On the relation between Gaussian process quadratures and sigma-point methods. *J. Adv. Inform. Fus.* **11**(1), 31–46 (2016)
35. Steinwart, I., Christmann, A.: *Support Vector Machines. Information science and statistics.* Springer, Berlin (2008)
36. Steinwart, I., Hush, D., Scovel, C.: An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Trans. Inf. Theory* **52**(10), 4635–4643 (2006)
37. Wendland, H.: *Scattered Data Dpproximation.* Number 17 in Cambridge monographs on applied and computational mathematics. Cambridge University Press, Cambridge (2005)
38. Wright, G.B., Fornberg, B.: Stable computations with flat radial basis functions using vector-valued rational approximations. *J. Comput. Phys.* **331**, 137–156 (2017)
39. Zwicknagl, B.: Power series kernels. *Constr. Approx.* **29**(1), 61–84 (2009)
40. Zwicknagl, B., Schaback, R.: Interpolation and approximation in Taylor spaces. *J. Approx. Theory* **171**, 65–83 (2013)

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