Spectral decomposition of $H^1(\mu)$ and Poincaré inequality on a compact interval - Application to kernel quadrature

O. Roustant, N. Lüthen, F. Gamboa

Data Sheet

Spectral decomposition of $H^1(\mu)$ and Poincaré inequality on a compact interval – Application to kernel quadrature.

Olivier Roustant[1], Nora Lüthen[2], and Fabrice Gamboa[3]

[1] *UMR CNRS 5219, Institut de Mathématiques de Toulouse, INSA, Université de Toulouse, France*
[2] *Chair of Risk, Safety, and Uncertainty Quantification, ETH Zürich, 8093 Zürich, Switzerland*
[3] *UMR CNRS 5219, Institut de Mathématiques de Toulouse, Université Paul Sabatier, France*

November 29, 2022

Abstract

Motivated by uncertainty quantification of complex systems, we aim at finding quadrature formulas of the form $\int_a^b f(x)d\mu(x) = \sum_{i=1}^n w_i f(x_i)$ where f belongs to $H^1(\mu)$. Here, μ belongs to a class of continuous probability distributions on $[a, b] \subset \mathbb{R}$ and $\sum_{i=1}^n w_i \delta_{x_i}$ is a discrete probability distribution on $[a, b]$. We show that $H^1(\mu)$ is a reproducing kernel Hilbert space with a continuous kernel K, which allows to reformulate the quadrature question as a kernel (or Bayesian) quadrature problem. Although K has not an easy closed form in general, we establish a correspondence between its spectral decomposition and the one associated to Poincaré inequalities, whose common eigenfunctions form a T-system (Karlin and Studden, 1966). The quadrature problem can then be solved in the finite-dimensional proxy space spanned by the first eigenfunctions. The solution is given by a generalized Gaussian quadrature, which we call Poincaré quadrature.

We derive several results for the Poincaré quadrature weights and the associated worst-case error. When μ is the uniform distribution, the results are explicit: the Poincaré quadrature is equivalent to the midpoint (rectangle) quadrature rule. Its nodes coincide with the zeros of an eigenfunction and the worst-case error scales as $\frac{b-a}{2\sqrt{3}} n^{-1}$ for large n. By comparison with known results for $H^1(0, 1)$, this shows that the Poincaré quadrature is asymptotically optimal. For a general μ, we provide an efficient numerical procedure, based on finite elements and linear programming. Numerical experiments provide useful insights: nodes are nearly evenly spaced, weights are close to the probability density at nodes, and the worst-case error is approximately $O(n^{-1})$ for large n.

Keywords Sobolev space, RKHS, Mercer representation, Poincaré inequality, Sturm-Liouville theory, Tchebytchev system (T-system), Bayesian quadrature, kernel quadrature, Gaussian quadrature.

Contents

1 Introduction

Motivation. This research is motivated by uncertainty quantification of complex systems, where a typical task is to compute integrals $I = \int G(x)dP(x)$. Here, G is a multivariate function representing a quantity of interest of the system, x is a vector of $\mathbb{R}^d$ representing the input variables and P is a probability distribution representing the uncertainty on x. In this context, the evaluation of G is often time-consuming, and cubature formula may be preferred to sampling techniques to compute the integral I. Assuming that the input variables are independent, cubature formulas then boil down to 1-dimensional quadrature formulas, by tensorization or using sparse grids (Smolyak, 1963; Garcke and Griebel, 2012).

Problem considered. For a given finite interval $[a, b]$ of $\mathbb{R}$, we aim at finding accurate approximations of integrals $\int_a^b f(x)d\mu(x)$ replacing μ by a discrete probability distribution. We thus deal with quadrature formulas

$$\int_a^b f(x)d\mu(x) = \sum_{i=1}^n w_i f(x_i), \tag{1.1}$$

where, for $i = 1, \cdots, n$, the quadrature nodes x_i lie in $[a, b]$. The quadrature weights (w_i) are non-negative and sum to 1. We will denote by X (resp. w) the sequence of nodes (x_i) (resp. of weights (w_i)). Considering minimal regularity conditions, we assume that f belongs to the Sobolev space $H^1(\mu) = \{f \in L^2(\mu), s.t. f' \in L^2(\mu)\}$, where the derivatives are defined in a weak sense. More generally, for an integer $p \geq 2$, we define the Sobolev space $H^p(\mu)$ as the subset of functions f of $H^{p-1}(\mu)$ such that $f^{(p)} \in L^2(\mu)$. In the whole paper, we consider the usual norm of $H^1(\mu)$ defined by $\|f\|^2_{H^1(\mu)} = \|f\|^2 + \|f'\|^2$, where $\|f\|^2 = \int_a^b f^2(x)d\mu(x)$ is the norm of $L^2(\mu)$. For technical reasons, we assume that μ is a *bounded perturbation* of the uniform distribution on $[a, b]$, meaning that it admits a positive continuous probabity density function ρ on $[a, b]$. This includes a wide range of probability distributions used in practice, such as the truncated normal one, obtained by conditioning a Gaussian variable to vary in a finite domain. This assumption implies that the sets $L^2(\mu), H^1(\mu)$ contain the same equivalence classes of functions than $L^2(a, b), H^1(a, b)$, associated to the uniform distribution on $[a, b]$, with equivalent norms.

Kernel quadrature formulation. When μ is the uniform probability distribution, it is well known that $H^1(a, b)$ is a reproducing kernel Hilbert space (RKHS) (Duc-Jacquet, 1973; Thomas-Agnan, 1996). Under the previous assumption on μ, we will show that $H^1(\mu)$ is also a RKHS (Section 3). In that case, a suitable criterion to evaluate the accuracy of a quadrature (X, w) is the worst-case error, defined by

$$\mathrm{wce}(X, w, \mathcal{H}) = \sup_{h \in \mathcal{H}, \|h\|_{\mathcal{H}} \leq 1} \left| \int h(x)d\mu(x) - \sum_{i=1}^n w_i h(x^i) \right|.$$

Here, $\mathcal{H}$ is some particular given functional space. Indeed, when $\mathcal{H}$ is a RKHS, $\mathrm{wce}(X, w, \mathcal{H})$ can be explicitly computed as a function of the kernel K (see Section 2.3 and e.g. Novak and Woźniakowski, 2008, Chapter 10). An interesting quadrature problem is so the following minimization problem:

$$(P): \quad \min_{X,w} \mathrm{wce}(X, w, \mathcal{H}),$$

where one wish to identify the minimizing quadrature. Such problem is often called *kernel quadrature*, or *Bayesian quadrature* in a probabilistic framework. There, the prior information is that the integrand is drawn from a Gaussian process with kernel K, and the worst-case error is equal to the posterior variance of the integral. We refer to O'Hagan (1991); Karvonen et al. (2019) for a presentation of this latter point of view.

Originality of the problem. We remark that, apart from the case of the uniform distribution, the problem does not reduce to the more standard quadrature problem with a weight function

$$\int_a^b f(x)\rho(x)dx = \sum_{i=1}^n w_i f(x_i) \tag{1.2}$$

where f belongs to $H^1(a, b)$. Indeed, the unit balls $\{h \in H^1(\mu), \|h\|_{H^1(\mu)} \leq 1\}$ and $\{h \in H^1(a, b), \|h\|_{H^1(a,b)} \leq 1\}$ are different if ρ is not a constant function. Thus $\mathrm{wce}(X, w, H^1(\mu)) \neq \mathrm{wce}(X, w, H^1(a, b))$ and the weighted quadrature problem, formulated as a worst-case error minimization problem, will in general not give the same solutions as (P).

Spectral decomposition of $H^1(\mu)$, Poincaré inequality and T-systems. The resolution of problem (P) is in general intractable, as K has not an explicit form. A key result of this paper is that there is a correspondence between the spectral decomposition (Mercer representation) of K and the one associated to Poincaré inequalities. Furthermore, the common eigenfunctions form a Tchebytchev system (T-system, see (Karlin and Studden, 1966)). Interestingly, even if K is not know explicitly in general, one can compute numerically its Mercer representation with a finite element technique (Roustant et al., 2017).

Problem resolution in a finite-dimensional proxy space. Thanks to the spectral decomposition of $H^1(\mu)$, we can replace problem (P) by the tractable proxy problem

$$(P_M): \quad \min_{X,w} \mathrm{wce}(X, w, \mathcal{H}_M)$$

where $H^1(\mu)$ has been replaced by its projection $\mathcal{H}_M$ onto the space spanned by the first M eigenfunctions. Indeed, similarly to polynomials, T-systems admit a Gaussian quadrature and for a given number of nodes n, there exists a unique quadrature (X, w) with positive weights for which $\mathrm{wce}(X, w, K_M) = 0$, where $M = 2n - 1$ is maximal. We call this optimal quadrature *Poincaré quadrature*. For a general probability distribution μ, the Poincaré quadrature can be computed easily and efficiently by linear programming.

Properties of the Poincaré quadrature. We derive several results for the connection between the kernel associated to $H^1(\mu)$, the Poincaré quadrature nodes and weights, and the associated worst-case error. When μ is the uniform distribution, the results are explicit (Section 5): the Poincaré quadrature is equal to the midpoint (rectangle) quadrature rule, its nodes coincide with the zeros of an eigenfunction, as for the Gaussian quadrature of polynomials, and the worst-case error scales as $\frac{b-a}{2\sqrt{3}} n^{-1}$ for large n. Furthermore, in the case of $H^1(0, 1)$, the kernel is given explicitly, and it is possible in this frame to compute the optimal kernel quadrature without relying on a finite-dimensional approximation. The results obtained by Duc-Jacquet (1973) show that the optimal kernel quadrature has evenly space nodes and weights asymptotically equal to $\frac{1}{n}$, which shows that the Poincaré quadrature is asymptotically optimal.

In the general case, numerical experiments provide empirical insights (Section 6): nodes are nearly evenly spaced, weights are close to the probability density at nodes, and the worst-case error is approximately proportional to n^{-1} for large n.

Links with literature. The Mercer representation of Sobolev spaces seems unexplored for non-uniform probability distributions. For the uniform case, by connecting RKHS to Green's functions, Fasshauer (2012) gives Mercer representations of various kernels associated to Sobolev spaces, such as the anchored Sobolev space $H^1(0,1) = \{f \in H^1(0,1), f(0) = f(1) = 0\}$ with the usual norm. Dick et al. (2014) provides the Mercer representation of $H^1(0,1)$ for the norm given by $\|f\|^2 = (\int_0^1 f(x)dx)^2 + \int_0^1 f'(x)^2 dx$. While close to these examples, the case of $H^1(0,1)$ with its usual norm seems not reported in the literature.

Similarly, considering Sobolev spaces with a non-uniform probability distribution seems unexplored for quadrature problems. As mentioned above, this does not boil down to a quadrature with weights for the uniform distribution, as the unit balls are different. The case of $H^1(0,1)$ (uniform case) has been studied by several authors, with various choices of norms and weight functions (Equation 1.2). For instance, Zhang and Novak (2019) provide expressions of the radius of information (worst-case error for the optimal quadrature) in function of the nodes, for the semi-norm $\int_0^1 f'(x)^2 dx$ and centered weight functions. For a constant weight function, and the usual norm of $H^1(0,1)$ considered here, Duc-Jacquet (1973) obtains the optimal kernel quadrature. The link between T-systems and kernel quadrature has been also exploited in Oettershagen (2017). There, the kernel is assumed to have an explicit form, and the T-system is obtained by considering the kernel function at nodes $K(x_i, .)$, which is different than our approach based on the spectral decomposition of K. Therein, the case of $H^1(0,1)$ is considered on numerical experiments, but with a different norm associated to the Bernoulli polynomials, given by $\|f\|^2 = (\int_0^1 f(x)dx)^2 + \int_0^1 f'(x)^2 dx$.

Paper organization. Section 2 gives the prerequisites on Poincaré inequalities, T-systems, RKHS and kernel quadrature. The RKHS structure of $H^1(\mu)$ is discussed in Section 3, where a connection is established between Poincaré inequalities and the kernel of $H^1(\mu)$. Section 4 gives general formulas for the optimal quadrature weights and the associated worst-case error as a function of the kernel. Section 5 focuses on the case of the uniform distribution. Section 6 presents numerical experiments in the general case. Section 7 concludes and gives an outlook for future research.

2 Background

In the whole paper, we consider a bounded interval of the real line $[a, b]$, with $-\infty < a < b < \infty$. We consider a probability distribution μ supported on $[a, b]$ which is a bounded perturbation of the uniform distribution, in the following sense.

Definition 1 (Bounded perturbation of the uniform distribution). *Let μ be a continuous probability distribution on $[a, b]$, with density ρ. We say that μ is a* bounded perturbation of the uniform distribution *if ρ is a positive continuous and piecewise C^1 function on $[a, b]$. We denote by $\mathcal{B}$ the set of bounded perturbations of the uniform distribution on $[a, b]$.*
We also denote by $V = -\log(\rho)$ the so-called potential *associated to μ. Equivalently, $\rho(t) = e^{-V(t)}$.*

Remark 1. *(i) Obviously, if μ fulfils the previous definition, then ρ is bounded from below and above by positive constants. That is, there exist m, M in $\mathbb{R}$ such that*

$$\forall t \in [a, b], \qquad 0 < m \leq \rho(t) \leq M < +\infty.$$

(ii) When $\mu \in \mathcal{B}$, it is straightforward that the sets $L^2(\mu), H^1(\mu)$ contain the same equivalence classes of functions than $L^2(a, b), H^1(a, b)$, associated to the uniform distribution on $[a, b]$, with equivalent norms.

2.1 Poincaré inequalities and basis

This section is based on Roustant et al. (2017) (see also Bakry et al., 2014). Let μ be a probability distribution on $[a, b]$. For $f, g \in L^2(\mu)$, let $\|f\| = \left(\int f^2 d\mu\right)^{1/2}$ be the usual norm, and $\langle f, g \rangle = \int fg d\mu$ the usual scalar product. Denote by $\mathbb{V}\mathrm{ar}_\mu(f)$ the variance of f:

$$\mathbb{V}\mathrm{ar}_\mu(f) := \left\| f - \int f d\mu \right\|^2.$$

We first recall the notion of Poincaré inequality.

Definition 2 (Poincaré inequality). *We say that μ satisfies a Poincaré inequality if there exists a positive constant C such that for all $f \in H^1(\mu)$:*

$$\mathbb{V}ar_\mu(f) \leq C\|f'\|^2.$$

In this case, the smallest possible constant C above is denoted $C_P(\mu)$, and called the Poincaré constant of μ.

When it exists, the Poincaré constant is obtained by minimizing the so-called Rayleigh ratio $J(f) = \frac{\|f'\|^2}{\|f\|^2}$ over all centered functions of $H^1(\mu)$. An important result is that a bounded perturbation of the uniform distribution satisfies a Poincaré inequality, which is related to a spectral decomposition as described in the following theorem (proved e.g. in Roustant et al., 2017):

Theorem 1 (Spectral theorem). *Let μ be a probability distribution in $\mathcal{B}$. Consider the three following problems:*

(P1) Find $f \in H^1(\mu)$ with $\int f d\mu = 0$ s.t. $\quad J(f) = \frac{\|f'\|^2}{\|f\|^2} \quad$ is minimum.

(P2) Find $f \in H^1(\mu)$ and $\lambda > 0$ s.t. $\quad \langle f', g' \rangle = \lambda \langle f, g \rangle \quad \forall g \in H^1(\mu)$.

(P3) Find $f \in H^2(\mu)$ and $\lambda > 0$ s.t. $\quad f'' - V'f' = -\lambda f \quad$ and $\quad f'(a) = f'(b) = 0$.

Then the eigenvalue problems (P2) and (P3) are equivalent. The eigenvalues form an increasing sequence $(\lambda_m)_{m \geq 0}$ of non-negative real numbers that tends to infinity. They are all simple, and $\lambda_0 = 0$. The eigenvectors $(\varphi_m)_{m \geq 0}$ form a Hilbert basis of $L^2(\mu)$, and φ_0 is a constant function.
Furthermore when $\lambda = \lambda_1$, the first positive eigenvalue, (P2) and (P3) are equivalent to (P1) and the minimum of (P1) is attained for $f = \varphi_1$. Thus $C_P(\mu) = 1/\lambda_1$.

In this paper, our interest is on the whole spectral decomposition. In particular, we define the Poincaré basis of $L^2(\mu)$ as follows.

Definition 3 (Poincaré basis). *Let μ be a probability distribution in $\mathcal{B}$. The Poincaré basis is the orthonormal basis formed by the eigenfunctions $(\varphi_m)_{m \geq 0}$ defined in the spectral theorem (Theorem 1).*

Remark 2. *As all eigenvalues are simple, the Poincaré basis is unique up to a sign change for each eigenfunction. We set $\varphi_0 = 1$.*

We conclude this section by linking the previous material with the Sturm-Liouville theory of second-order differential equations.

Proposition 1. *Let μ be a probability distribution in $\mathcal{B}$. Then the Poincaré basis consists of the eigenfunctions of the Sturm-Liouville eigenproblem*

$$L(f)(x) = \beta r(x) f(x) \tag{2.1}$$

with Neumann conditions $f'(a) = f'(b) = 0$, where $L(f)(x) = -(p(x)f'(x))' + q(x)f(x)$ and $q = r = p = e^{-V}$. Furthermore, the Sturm-Liouville problem is regular, in the sense that all its eigenvalues are positive.

Proof. Recall that $\langle f, g \rangle_{H^1(\mu)} = \langle f, g \rangle + \langle f', g' \rangle$. From the proof of Theorem 2 in Roustant et al. (2017), the eigenfunctions of the Poincaré operator are solutions of the spectral problem: to find $f \in H^1(\mu)$ and β such that for all $g \in H^1(\mu)$,

$$\langle f, g \rangle_{H^1(\mu)} = \beta \langle f, g \rangle. \tag{2.2}$$

The corresponding eigenvalues are $\beta_m = 1 + \lambda_m$. In particular, $\beta_m > 0$, as $\lambda_m \geq 0$. Moreover, Problem (2.2) is equivalent to the second order differential equation

$$f''(x) - V'(x)f'(x) - f(x) = -\beta f(x)$$

with Neumann conditions $f'(a) = f'(b) = 0$ (see also Roustant et al., 2017, proof of Theorem 2). Multiplying by $e^{-V(x)}$, which by definition of $\mathcal{B}$ does not vanish, we obtain the equivalent Sturm-Liouville form (2.1). $\quad\square$

2.2 Quadrature with T-systems

This section is based on Karlin and Studden (1966).

Definition 4 (T-systems, generalized polynomials). *Let $(u_n)_{n\in\mathbb{N}}$ be a family of real-valued continuous functions defined on a compact interval $[a,b] \subset \mathbb{R}$. We say that (u_n) is a* complete Tchebytchev system, *or simply* T-system, *if for all integer $n \geq 1$ and for all sequence of distinct points $a \leq t_1 < \cdots < t_n \leq b$, the determinant of the* generalized Vandermonde matrix

$$V(u_0,\ldots,u_{n-1};t_1,\ldots,t_n) := \begin{pmatrix} u_0(t_1) & \ldots & u_0(t_n) \\ \vdots & \ddots & \vdots \\ u_{n-1}(t_1) & \ldots & u_{n-1}(t_n) \end{pmatrix}$$

is positive. The finite linear combinations of the u_n's are called generalized polynomials *or u-polynomials.*

A prototype of T-system on any interval of the real line is given by the polynomial functions $u_i(t) = t^i$, and the determinant of V is nothing more than the Vandermonde determinant $\det V(1,t,\ldots,t^{n-1},t_1,\ldots,t_n) = \prod_{1\leq i<j\leq n}(t_j - t_i)$ (see e.g. Karlin and Studden, 1966, page 1). An equivalent definition of T-systems (up to a sign change) is that any generalized polynomial, i.e., any linear combination of $u_0, u_1, \ldots, u_n$, has at most n zeros (Karlin and Studden, 1966, Theorem 4.1.). This extends the property that (ordinary) polynomials of degree n have at most n zeros. In that sense, T-systems can be viewed as a generalization of polynomials, which justifies the name *generalized polynomials*.

In the context of quadrature problems, the definition of T-systems guarantees that for any set of distinct quadrature nodes in $[a,b]$, there exists a unique set of quadrature weights such that the quadrature formula (1.1) is exact at order $n-1$, i.e. for all functions in $\mathrm{span}(u_0,\ldots,u_{n-1})$. Indeed, up to reordering, the equations above define a linear system whose matrix is invertible and equal, up to a sign change, to $V(u_0,\ldots,u_{n-1};t_1,\ldots,t_n)$. This quadrature formula thus generalizes the Newton-Cotes quadrature of polynomials, and suffers in general from the same drawback: the weights can be negative and the resulting quadrature formula can be instable (Oettershagen, 2017; Brass and Petras, 2011).

Interestingly, extending the Gaussian quadrature of ordinary polynomials to more general functions, T-systems admit a unique quadrature that has positive weights and is exact at order $2n-1$. Contrarily to the polynomial case, however, the nodes of this quadrature do in general not coincide with the zeros of a (generalized) polynomial. The computation of the nodes and weights uses a different approach, relying on geometry. More precisely, consider the *moment space*:

$$\mathcal{M}_{n+1} = \left\{ c \in \mathbb{R}^{n+1}, \text{ with } c_i = \int_a^b u_i(t)d\sigma(t), \text{ for all } i = 0,\ldots,n, \right.$$

$$\left. \text{where } \sigma \text{ is a finite measure with support } [a,b] \right\}. \quad (2.3)$$

$\mathcal{M}_{n+1}$ is the closed convex hull of the curve $(u(t))_{t\in[a,b]}$ where $u(t) = (u_0(t),\ldots,u_n(t)) \in \mathbb{R}^{n+2}$ ($t \in [a,b]$). Then, we may enounce precisely the previous quadrature results.

Proposition 2. *Let $u = (u_n)_{n\in\mathbb{N}}$ be a T-system, and μ be a probability distribution in $\mathcal{B}$. Then,*

1. *For all $n \in \mathbb{N}$, the vector $\bar{c} := \left(\int_a^b u_i(t)d\mu(t)\right)_{0\leq i\leq 2n-1}$ is an interior point of $\mathcal{M}_{2n}$.*

2. *There exists a unique quadrature (1.1) with positive weights which is exact at order $2n-1$ (i.e. exact on the vector space spanned by $u_0,\ldots,u_{2n-1}$). It uses a minimal number of nodes, equal to n. The nodes are all in the open interval (a,b) and the weights sum to 1.*

3. *This quadrature is obtained by solving the minimization problem*

$$\min_{\sigma\in V_{2n-1}(c)} \int_a^b u_{2n}(t)d\sigma(t) \quad (2.4)$$

over the set $V_{2n-1}(c)$ of all probability distributions subject to moment conditions $\int_a^b u_i(t)d\sigma(t) = \bar{c}_i$, for $i = 0, 1, \ldots, 2n - 1$.

Proof of Proposition 2. The proof is based on different results of Karlin and Studden (1966), that we will refer to. Let us denote $N = 2n - 1$ the quadrature order.

First, the separation Lemma 9.2., page 65, states that c is an interior point of $\mathcal{M}_N$ if and only if for all non negative non trivial u-polynomial $g = \sum_{i=0}^N a_i u_i$, we have $m_g(c) = \sum_{i=0}^N a_i c_i > 0$. Taking $c = \bar{c}$, we have $m_g(\bar{c}) = \int_a^b g d\mu$ and thus $m_g(\bar{c}) \geq 0$. As the density function of μ is continuous and does not vanish on $[a, b]$, we can conclude that $m_g(\bar{c}) > 0$, which shows that c is an interior point of $\mathcal{M}_N$.

Now, following Karlin and Studden (1966), §3, case (ii), page 46, let define the notion of index of a sequence of nodes, as the number of nodes in $[a, b]$, with a half weight for the nodes equal to the endpoints a, b (if any). Then, as N is odd, there are exactly two quadratures with positive weights and the smallest possible index, equal to $(N + 1)/2 = n$. These quadrature are called *principal representations* in this context. For one of them, called *upper principal representation*, the nodes include the endpoints, and thus the quadrature involves $n + 1$ nodes formed by the endpoints a, b and $n - 1$ nodes in (a, b). The other one, called *lower principal representation*, involves n nodes in the open interval (a, b). Thus, it is the only quadrature with positive weights and containing the smallest number of nodes, equal to n.

Furthermore, they show in Theorem 1.1. page 80, that, if u is a T-system, the solution of (2.4) is unique and equal to the lower representation of u. In particular, the weights are positive and sum to one.

Finally, if $u_i(t) = t^i$, we have equality with the Gaussian quadrature by uniqueness of the quadrature, since the Gaussian quadrature has n distinct nodes in (a, b), positive weights summing to one, and is exact for $u_0, \ldots, u_{2n-1}$ (Karlin and Studden, 1966, Chapter IV). $\qquad\square$

Remark 3. *Replacing the minimization problem in (2.4) by maximization, we obtain another valid quadrature of oder $2n - 1$, called* upper principal representation. *However, it involves one more node, i.e. $n + 1$ nodes, including the endpoints a, b. Equivalently, for a fixed number n of nodes, this quadrature has order $2n - 3$, compared to $2n - 1$ for the Gaussian quadrature. It generalizes the Lobatto quadrature for polynomials (Brass and Petras, 2011).*

Definition 5 (Gaussian quadrature for T-systems)**.** *The unique quadrature of Prop. 2 is called* lower principal representation *of u. By analogy with polynomials, and following (Brass and Petras, 2011; Oettershagen, 2017), we will call it the generalized Gaussian quadrature, or simply the* Gaussian quadrature *of the T-system $(u_n)_{n\in\mathbb{N}}$.*

We now show that a Poincaré basis is a T-system. This is an immediate consequence of the example of eigenfunctions of Sturm-Liouville problems (see e.g. Example 7 in Karlin and Studden, 1966). A proof can be found in Gantmakher and Krejn (2002). However, the proof is not easy to read as it is split in several parts. We recall below the main steps and give a roadmap for the interested reader.

Proposition 3. *Consider the notations and the assumption of Theorem 1. Then the Poincaré basis $(\varphi_m)_{m\in\mathbb{N}}$ is a T-system.*

Proof. By Prop. 1, the eigenfunctions of the Poincaré operator are eigenfunctions of the regular Sturm-Liouville problem (2.1). Then the result is a particular case of a more general result stating that the eigenfunctions of a regular Sturm-Liouville operator form a T-system. A proof can be found in Gantmakher and Krejn (2002). We give here the three main steps, and pointers to the corresponding sections.

Firstly, it is proved in IV.10.4, (pages $236 - 238$), that the eigenfunctions of a Sturm-Liouville operator under boundary constraints verify an integral equation of the form

$$\varphi(x) = \lambda \int_a^b K(x, s)\varphi(s)d\sigma(s) \tag{2.5}$$

where $\lambda > 0$, σ is a probability distribution, and K is the so-called Green function of L, defined by:

$$K(x, s) = \psi(\min(x, s))\chi(\max(x, s)),$$

where ψ, χ are particular solutions of the homogeneous equations $L(f) = 0$, such that $\frac{\psi}{\chi}$ is a non-decreasing function.

Secondly, it is proved that K is an *oscillatory kernel*, in the sense of[1] Definition 1 page 178. Roughly speaking,

---

[1]see also Definition 1' page 179, and definitions 4 and 6, pages 74 and 76.

it means that every matrix extracted from K *in ascending order*, i.e. $(K(x_i, s_j))_{1 \leq i,j \leq n}$ with $x_1 < \cdots < x_n$ and $s_1 < \cdots < s_n$ is positive semidefinite. The proof starts at page 78 (Example 5, 'Single-pair' matrices), continues at page 103 (Theorem 12), page 220 (Criterion A) and ends at page 238 (Theorem 16).

Thirdly, if K is an oscillatory kernel, then the solutions of the integral equation (2.5) form a T-sytem, which is proved in Theorem 1, page 181. $\qquad\square$

2.3 Kernel quadrature

RKHS. We first recall some facts on *reproducing kernel Hilbert spaces* (RKHS), refering to Berlinet and Thomas-Agnan (2011) for more details. For a given set T, let $\mathcal{H}$ be a Hilbert space of functions $T \to \mathbb{R}$, with norm $\|.\|$ and scalar product $\langle .,. \rangle$. We say that $\mathcal{H}$ is a *RKHS* if for all $x \in T$, the evaluation functions $h \in \mathcal{H} \mapsto h(x)$ are continuous. It can be shown that a RKHS is in bijection with a semi-definite positive function, also called *kernel*. If K is a kernel associated to $\mathcal{H}$, we write $\mathcal{H} = \mathcal{H}_K$. The RKHS $\mathcal{H}_K$ is characterized by the so-called *reproducing property*

$$\forall x \in T, \forall h \in \mathcal{H}_K, \qquad \langle K(x,.), h \rangle = h(x).$$

In particular, choosing $h = K(y,.)$, we get

$$\forall x, y \in T, \qquad \langle K(x,.), K(y,.) \rangle = K(x,y).$$

Worst-case error in RKHS. In RKHS, worst-case quantities of linear functionals can be computed explicitly. Indeed, for instance, the Cauchy-Schwartz inequality gives that for all $h \in \mathcal{H}_K$ and all $x \in T$:

$$|h(x)| = |\langle K(x,.), h \rangle| \leq \|K(x,.)\| \|h\|,$$

from which it is obvious that $\sup_{h \in \mathcal{H}_K, \|h\| \leq 1} |h(x)| = \|K(x,.)\|$. Furthermore, by the reproducing property, we have $\|K(x,.)\|^2 = K(x,x)$, and finally, for all $x \in T$,

$$\sup_{h \in \mathcal{H}_K, \|h\| \leq 1} |h(x)| = \sqrt{K(x,x)}.$$

A similar computation can be done for linear functionals defined by quadrature formulas. Let us first define the worst-case error of a general quadrature.

Definition 6 (worst-case error of a quadrature). *Let (X, w) be a quadrature composed of a set of nodes $X = (x_1, \ldots, x_n) \in [a,b]^n$ and a set of weights $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$. Let $\mathcal{H}$ be a set of functions on $[a,b] \to \mathbb{R}$. The worst-case error of (X, w) on $\mathcal{H}$ is defined by*

$$\mathrm{wce}(X, w, \mathcal{H}) = \sup_{h \in \mathcal{H}, \|h\| \leq 1} \left| \int h(x) d\mu(x) - \sum_{i=1}^{n} w_i h(x_i) \right|.$$

If $\mathcal{H}$ is a RKHS $\mathcal{H}_K$, we simply denote $\mathrm{wce}(X, w, K) = \mathrm{wce}(X, w, \mathcal{H}_K)$.

By a direct extension of the computation above with $T = [a,b]$, we have

$$\mathrm{wce}(X, w, K) = \left\| \int K(x,.) d\mu(x) - \sum_{i=1}^{n} w_i K(x_i,.) \right\|_{\mathcal{H}_K}.$$

Using the reproducing property, one obtains the well-known analytical expression (see e.g. Novak and Woźniakowski, 2008, Chapter 10):

$$\mathrm{wce}(X, w, K)^2 = \iint K(x,x') d\mu(x) d\mu(x') - 2 \sum_{i=1}^{n} w_i \int K(x_i, x) d\mu(x) + \sum_{i,j} w_i w_j K(x_i, x_j)$$

which can be rewritten in the matricial form

$$\mathrm{wce}(X, w, K)^2 = w^\top K(X, X) w - 2\ell_K(X)^\top w + c_K \tag{2.6}$$

where $K(X, X) = (K(x_i, x_j))_{1 \leq i,j \leq n}$ is the Gram matrix, $\ell_K(X) = (\int K(x_i, x) d\mu(x))_{1 \leq i \leq n}$ is the column vector formed by the primitive function of the kernel at x_i and $c_K = \iint K(x, x') d\mu(x) d\mu(x')$ is a constant.

Kernel quadrature. A kernel quadrature is obtained by minimizing the worst-case error. We will need the following assumption:

Assumption 1. *If for all $i \neq j, x_i \neq x_j$ then the Gram matrix $K(X, X)$ is invertible.*

Under Assumption 1, for a given set of nodes X formed by different nodes, then (2.6) defines a strictly convex function. Thus, it has a unique minimum, denoted by $w^\star(X, K)$. By solving the first order conditions, we directly get the exact expression of the vector of optimal weights:

$$w^\star(X, K) = K(X, X)^{-1} \ell_K(X). \tag{2.7}$$

After some algebra, we obtain the corresponding minimal value for the worst case error:

$$wce(X, w^\star, K)^2 = c_K - \ell_K(X)^\top K(X, X)^{-1} \ell_K(X). \tag{2.8}$$

Kernel quadrature and optimal kernel quadratures can then be defined as follows.

Definition 7 (Kernel quadrature, optimal kernel quadrature). *Let X be a set of nodes, and assume that Assumption 1 is verified. 1) The* kernel quadrature *associated to X on $\mathcal{H}_K$ is the quadrature $(X, w^\star(X, K))$ that minimizes the worst-case error $wce(X, w, K)$ over all sets of weights in $\mathbb{R}^n$.*
2) An optimal kernel quadrature *is a quadrature $(X^\star, w^\star)$ that minimizes the worst-case error $wce(X, w, K)$ among all quadratures (X, w), or equivalently, that minimizes $wce(X, w^\star(X, w), K)$ over all sets of nodes X.*

Remark 4. *Notice that the weights of a kernel quadrature are constrained neither to be positive, nor to sum to 1.*

3 Spectral decomposition of $H^1(\mu)$ with the Poincaré basis

We show our main result: when μ is a bounded perturbation of the uniform distribution, then $H^1(\mu)$ is a RKHS whose kernel eigenfunctions coincide with the Poincaré basis. This comes from two connections: the first one between Poincaré basis and Sturm-Liouville problems, the second one between Sturm-Liouville problems and RKHS via Green's functions. We illustrate the result on two examples where explicit computations can be made.

3.1 Main result

Proposition 4 (Mercer's representation of $H^1(\mu)$ with the Poincaré basis). *Assume that μ is a probability distribution in $\mathcal{B}$. Let denote by $(\lambda_m, \varphi_m)_{m \in \mathbb{N}}$ the eigenvalues and (normalized) eigenfunctions of the Poincaré operator. Let, for $m \in \mathbb{N}$, $\alpha_m = (1 + \lambda_m)^{-1}$. Then,*

1. *$H^1(\mu)$, with its usual Hilbert norm $\|f\|_{H^1(\mu)}^2 = \|f\|^2 + \|f'\|^2$, is a RKHS. Its kernel K is continuous on $[a, b]^2$ and verifies $\int_a^b K(x, y) d\mu(y) = 1$ for all $x \in [a, b]$.*

2. *The Mercer decomposition of K is written*

$$K(x, y) = \sum_{m=0}^\infty \alpha_m \varphi_m(x) \varphi_m(y) \qquad (x, y \in [a, b]) \tag{3.1}$$

   where the convergence is uniform on $[a, b]^2$.

3. *K is a* single-pair kernel *(Gantmakher and Krejn, 2002), meaning that it has the form*

$$K(x, y) = \frac{1}{C} \psi(\min(x, y)) \chi(\max(x, y)). \qquad (x, y \in [a, b]) \tag{3.2}$$

   Furthermore, ψ, χ are two linearly independent solutions of the homogeneous equation $f'' - f'V' - f = 0$ such that $\psi'(a) = 0$ and $\chi'(b) = 0$, and $C = \chi(b) \int_a^b \psi(x) d\mu(x) = \psi(a) \int_a^b \chi(y) d\mu(y)$ is a normalization constant.

Proof. Under the assumption on μ, for all $f \in L^2[a, b]$,

$$m(b-a)\int_a^b f^2 d\lambda \leq \int_a^b f^2 d\mu \leq M(b-a)\int_a^b f^2 d\lambda$$

where m, M are defined in Remark 1 and λ is the uniform distribution on $[a, b]$. Thus $L^2(\mu) = L^2(a, b)$ (with an equivalent norm). Similarly $H^1(\mu) = H^1(a, b)$ (with an equivalent norm). Now, it is well known that $H^1(a, b)$ is a RKHS, with an explicit kernel (see e.g. Duc-Jacquet, 1973; Thomas-Agnan, 1996). Thus, for all $x \in [a, b]$, the evaluation $f \in H^1(a, b) \mapsto f(x)$ is continuous. By equivalence of the norms, $f \in H^1(\mu) \to f \in H^1(a, b)$ is continuous. Hence by composition, $f \in H^1(\mu) \to f(x)$ is continuous. This shows that $H^1(\mu)$ is a RKHS. Let us denote by K its kernel.

The link between K and the Poincaré inequality can be understood through the bilinear form $a(f, g) = \langle f, g \rangle_{H^1(\mu)}$. Consider the spectral problem: find $f \in H^1(\mu)$ and β such that for all $g \in H^1(\mu)$,

$$a(f, g) = \beta \langle f, g \rangle \tag{3.3}$$

From (Roustant et al., 2017, Theorem 2 and its proof), under the assumption on μ, there exists a countable sequence of solutions, which is given by $\beta_m = 1 + \lambda_m$ and φ_m ($m \in \mathbb{N}$). Notice that $\beta_m > 0$ for all $m \in \mathbb{N}$. Furthermore β_m, φ_m are defined in an unique way (up to a change sign of the eigenfunctions) because the eigenvalues are simple and the eigenfunctions have norm 1.

Now, since $H^1(\mu)$ is a RKHS, the functions $K(x, .)$ are dense in $H^1(\mu)$ ($x \in [a, b]$). Thus, Problem (3.3) is equivalent to:

$$\forall x \in [a, b], \quad a(f, K(x, .)) = \beta \langle f, K(x, .) \rangle.$$

By the reproducing property, $a(f, K(x, .)) = \langle f, K(x, .) \rangle_{H^1(\mu)} = f(x)$. Hence, (3.3) is equivalent to: find $f \in H^1(\mu)$ and β such that

$$\forall x \in [a, b], \quad f(x) = \beta \int_a^b K(x, y) f(y) d\mu(y), \tag{3.4}$$

which is equivalent to the spectral decomposition of the Hilbert-Schmidt operator associated to K (recall that $\beta > 0$).

Moreover, by Proposition 1 and its proof, Problem (3.3) is equivalent to the regular Sturm-Liouville problem (2.1). Thus, from (Gantmakher and Krejn, 2002, Section 10, pages 234-238) or (Fasshauer, 2012), we also obtain that the solution of (2.1) is equivalent to the solution of (3.4). In this context, K is called *Green function*. But this point of view gives sharp information and leads to

$$K(x, y) = \frac{1}{C} \psi(\min(x, y)) \chi(\max(x, y)), \qquad (x, y \in [a, b])$$

where ψ, χ are two linearly independent solutions of the homogeneous equation $L(f) = 0$ such that $\psi'(a) = 0$ and $\chi'(b) = 0$ (Gantmakher and Krejn, 2002, section 7). The constant C is determined such that $\int_a^b K(x, y) d\mu(y) = 1$ for all $x \in [a, b]$. Indeed, as the constant function 1 belongs to $H^1(\mu)$, the RKHS reproducing property gives, for all $x \in [a, b]$:

$$1 = 1(x) = \langle 1, K(x, .) \rangle_{H^1(\mu)} = \int_a^b K(x, y) d\mu(y).$$

For instance, choosing $y = b$ or $x = a$, we obtain $C = \chi(b) \int_a^b \psi(x) d\mu(x) = \psi(a) \int_a^b \chi(y) d\mu(y)$.

Now ψ and χ are continuous, as elements of $H^1(\mu)$ (whose functions are equal to those of $H^1(a, b)$). As $\min, \max$ are continuous functions, we obtain, by composition, that K is continuous on $[a, b]^2$. Hence by Mercer theorem (see e.g. Berlinet and Thomas-Agnan, 2011), K is written in terms of the solutions of (3.4) as

$$K(x, y) = \sum_{m \in \mathbb{N}} \alpha_m \varphi_m(x) \varphi_m(y), \qquad (x, y, \in [a, b]) \tag{3.5}$$

with $\alpha_m = \frac{1}{\beta_m} = \frac{1}{1+\lambda_m}$, and the convergence is uniform on $[a, b]^2$. $\qquad\square$

Remark 5. *We mention another way to obtain the Mercer's representation of K. A property of the Poincaré basis is that it is an orthogonal basis of $H^1(\mu)$ with $\|\varphi_m\|^2_{H^1(\mu)} = 1 + \lambda_m = 1/\alpha_m$ (Lüthen et al., 2021). Thus, the functions $e_m = \sqrt{\alpha_m}\varphi_m$ ($m \geq 0$) define an orthonormal basis of $H^1(\mu)$. Then, the representation (3.5) is obtained with the usual representation of a kernel in a separable RKHS (Berlinet and Thomas-Agnan, 2011):*

$$K(x,y) = \sum_{m\in\mathbb{N}} \alpha_m \varphi_m(x)\varphi_m(y) = \sum_{m\in\mathbb{N}} e_m(x)e_m(y).$$

However, following this way leads only to pointwise convergence. So that, it is not clear whether K is continuous on $[a,b]^2$. Thus, another argument has been used here, coming from the Green function point of view, to prove the kernel continuity.

Remark 6. *At first look, it may be surprising that the Neumann conditions $f'(a) = f'(b) = 0$ do not appear in the RKHS, whereas all basis functions φ_m satisfy it (while being dense in $L^2(\mu)$). Actually, it is not difficult to see that any function f of $H^1(\mu)$ can be approximated by a function $f_\epsilon \in H^1(\mu)$ that verifies the Neumann condition, simply by truncating f on $[a+\epsilon, b-\epsilon]$ and extending it continuously by a constant on $[a, a+\epsilon]$ and $[b-\epsilon, b]$. As functions of $H^1(\mu)$ are continuous on the compact interval $[a,b]$ (still under our assumption on μ), the approximation error can be made arbitrary small.*

3.2 Examples

Example 1 (Case of the uniform distribution). *Let μ be the uniform distribution on $[a,b]$. Then L is the Laplacian operator, and the spectral problem becomes*

$$\varphi''(x) = -\lambda\varphi(x) \qquad \forall x \in [a,b],$$

with Neumann conditions $\varphi'(a) = \varphi'(b) = 0$. The solutions are given by $\lambda_0 = 0, \varphi_0 = 1$ and for $m \geq 1$,

$$\lambda_m = m^2\omega^2, \qquad \varphi_m(x) = \sqrt{2}\cos(m\omega(x-a)),$$

with $\omega = \pi/(b-a)$. The kernel of $H^1(\mu)$ (with its usual norm) has been obtained by Duc-Jacquet (1973) in the 70's. English-written proofs can be found in Atteia (1992), Example 1.4, or Thomas-Agnan (1996)). The kernel may be written as

$$K(x,y) = \frac{b-a}{\sinh(b-a)} \cosh[\min(x,y)-a] \cosh[b-\max(x,y)]$$

where $\cosh, \sinh$ denote the hyperbolic functions: $\cosh(x) = \frac{e^x+e^{-x}}{2}$, $\sinh(x) = \frac{e^x-e^{-x}}{2}$. Applying Proposition 4, we deduce that for all $(x,y) \in [a,b]^2$ such that $x \leq y$,

$$K(x,y) = \frac{\pi/\omega}{\sinh(\pi/\omega)} \cosh(x-a)\cosh(b-y) = 1 + 2\sum_{m=1}^{+\infty} \frac{1}{1+m^2\omega^2}\cos[m\omega(x-a)]\cos[m\omega(y-a)].$$

As a by-product, we can derive the value of some 'shifted' Riemann series. For instance, from $x = y = a$ and $x = a, y = b$ and using $r = 1/\omega$, we get the formulas (with $\tanh(x) = \sinh(x)/\cosh(x)$), valid for all $r > 0$:

$$\sum_{n=1}^{+\infty} \frac{1}{n^2+r^2} = \frac{1}{2r^2}\left(\frac{\pi r}{\tanh(\pi r)} - 1\right), \qquad \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n^2+r^2} = \frac{1}{2r^2}\left(1 - \frac{\pi r}{\sinh(\pi r)}\right).$$

It can be shown that these formulas are also valid when r tends to zero. The limit case gives the well-known expression at $s = 2$ of the Riemann and Dirichlet functions: $\zeta(2) = \pi^2/6$ and $\eta(2) = \pi^2/12$.

As for the classical space $H^1(a,b)$, the kernel of $H^1(\mu)$ and its Mercer representation are completely tractable for the truncated exponential distribution.

Example 2 (Truncated exponential distribution). *Consider the exponential distribution, truncated on $[a, b] \subseteq \mathbb{R}^+$: $\frac{d\mu}{dx} = \exp(-V(x)) = \frac{e^{-x}}{e^{-a} - e^{-b}} \mathbb{1}_{[a,b]}(x)$. Noting that $V'(x) = 1$ on $[a, b]$, leads to linear differential equations with constant coefficients. Following Roustant et al. (2017, Proof of Proposition 5), the spectral problem*

$$\varphi'' - \varphi' = -\lambda\varphi$$

with Neumann conditions $\varphi'(a) = \varphi'(b) = 0$, admits the solutions, $\lambda_0 = 0, \varphi_0 \equiv 1$ and for $m \geq 1$,

$$\lambda_m = \frac{1}{4} + (m\omega)^2, \qquad \varphi_m(x) = c_m e^{x/2}(2m\omega\cos(m\omega(x - a)) - \sin(m\omega(x - a))) \tag{3.6}$$

where $\omega = \pi/(b - a)$ and c_m is a normalizing constant ensuring that φ_m has $L^2(\mu)$ norm equal to 1. More precisely, we have

$$c_m = \left(\frac{e^{-a} - e^{-b}}{b - a} \frac{1}{2\lambda_m} \right)^{1/2}.$$

From Proposition 4 and following Gantmakher and Krejn (2002), the Green function associated to this spectral problem can be computed by considering the linear homogeneous equation

$$\varphi'' - \varphi' - \varphi = 0.$$

The solutions are spanned by $e^{r_0 x}, e^{r_1 x}$, where $r_0 = \frac{1 - \sqrt{5}}{2}$ and $r_1 = \frac{1 + \sqrt{5}}{2}$ (the gold number). Solutions ψ, χ satisfying one-sided Neumann conditions $\psi'(a) = 0$ and $\chi'(b) = 0$ are given by

$$\psi(x) = (r_1 e^{r_1 a})e^{r_0 x} - (r_0 e^{r_0 a})e^{r_1 x}, \qquad \chi(x) = (r_1 e^{r_1 b})e^{r_0 x} - (r_0 e^{r_0 b})e^{r_1 x}. \tag{3.7}$$

Finally, the normalization constant is given by $C = \int_a^b \psi(x)\chi(b)d\mu(x)$. After some algebra, we obtain

$$C = \frac{(e^{r_0 a + r_1 b} - e^{r_0 b + r_1 a})(r_1 - r_0)}{e^{-a} - e^{-b}}.$$

Finally the kernel of $H^1(\mu)$ is given explicitly by its single-pair form (3.2) or by its Mercer representation (3.1), where $\alpha_m = (1 + \lambda_m)^{-1}$ and λ_m, φ_m are given by (3.6).

4 Poincaré quadrature and optimal kernel quadrature in $H^1(\mu)$

Proposition 4 shows that the spectral decomposition associated to Poincaré inequalities is in correspondence with the spectral decomposition of the kernel of $H^1(\mu)$, viewed as a RKHS. However, the kernel of $H^1(\mu)$ is in general not directly available, which makes optimal kernel quadrature in $H^1(\mu)$ intractable. This motivates us to focus on the quadratures defined from the Poincaré basis, as eigenfunctions of $H^1(\mu)$.

4.1 Definitions and notations

Definition 8 (Poincaré quadrature). *We call* Poincaré quadrature *the Gaussian quadrature of the T-system of the Poincaré basis of μ, as defined in Definition 5. We denote it by (X_P, w_P).*

We now establish a connection between the Poincaré quadrature and the kernel quadrature spanned by the Poincaré basis functions. To reach this goal, we first check that these kernel quadratures are properly defined, by showing that Assumption 1 is verified both for K and its finite-dimensional approximation K_M, defined below.

Definition 9 (Finite-dimensional kernel for $H^1(\mu)$). *Let M a non-zero integer, and consider the notations of Proposition 4 and its assumptions. We set*

$$K_M(x, y) = \sum_{m=0}^{M} \alpha_m \varphi_m(x)\varphi_m(y), \qquad (x, y \in [a, b]) \tag{4.1}$$

the truncated Mercer representation of the kernel of $H^1(\mu)$.

Proposition 5 (Invertibility of the Gram matrix for truncated Mercer representation of $H^1(\mu)$). *Assumption 1 is verified for K_M when X contains at most $M + 1$ distinct points.*

Proof. Let $x_0, \ldots, x_n$ a set of distinct points with $n \leq M$. For $i = 0, \ldots, n$, denote $e_i = K_M(x_i, .) \in H^1(\mu)$. Then $K_M(x_i, x_j) = \langle e_i, e_j \rangle_{H^1(\mu)}$. Thus $K_M(X, X)$ is the Gram matrix of the vectors e_i for the scalar product in $H^1(\mu)$. By a classical result, it is invertible if and only if $e_0, e_1, \ldots, e_n$ are linearly independent. It is enough to prove that $e_0, e_1, \ldots, e_M$ are linearly independent, since $n \leq M$. Now, by definition, we have

$$e_i = K_M(x_i, .) = \sum_{m=0}^{M} A_{i,m} \varphi_m(.)$$

with $A_{i,m} = \alpha_m \varphi_m(x_i)$. Hence, if $e = (e_1, \ldots, e_M)^\top$ and $\varphi = (\varphi_0, \ldots, \varphi_M)^\top$ are column vectors whose elements are in $H^1(\mu)$, then we have $e = A\varphi$. Remember that the φ_m's are linearly independent ($m \geq 0$), as they form an Hilbertian basis of $L^2(\mu)$. Thus the coordinates of e are linearly independent if and only if A is invertible. Remarking that the mth column of A is proportional to the vector $(\varphi_m(x_i))_{0 \leq i \leq M}$ with a non-zero multiplicative coefficient α_m. Hence, the rank of A is equal to the rank of the matrix $(\varphi_m(x_i))_{0 \leq m, i \leq M}$. As the φ_m's form a T-system, this matrix is invertible, which completes the proof. $\qquad\square$

Proposition 6 (Invertibility and form of the Gram matrix for $H^1(\mu)$). *Let K be the kernel of $H^1(\mu)$, where μ is a probability distribution in $\mathcal{B}$. Then, Assumption 1 is verified for all set X composed of distinct knots. Furthermore, in that case, if the elements of X are sorted in ascending (or decreasing) order, the precision matrix $K(X, X)^{-1}$ is a one-band matrix (or Jacobi matrix) of the form:*

$$\begin{pmatrix} a_1 & b_1 & 0 & 0 & \ldots & 0 \\ b_1 & a_2 & b_2 & 0 & & 0 \\ 0 & b_2 & a_3 & b_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & 0 & b_{n-2} & a_{n-1} & b_{n-1} \\ 0 & 0 & \ldots & 0 & b_{n-1} & a_n \end{pmatrix}$$

Proof. By Proposition 4 and Theorem 16 (page 238) of Gantmakher and Krejn (2002), the single-pair kernel K is oscillatory in the sense of definition 1 page 178 of the same reference, implying in particular that for all X composed of distinct knots, $K(X, X)$ is invertible. The form of the precision matrix is derived in section II.3., example 6, pages 79-82 of the same reference. $\qquad\square$

Remark 7 (Link with Gaussian graphical models). *The one-band form of the precision matrix of Proposition 6 shows that a Gaussian process associated to the kernel of $H^1(\mu)$ is a graphical Gaussian model (Maathuis et al., 2018). Indeed, consider the graph whose vertices are the quadrature nodes sorted in ascending order $x_1 < \cdots < x_n$ and edges are formed by successive nodes (x_i, x_{i+1}) with $i = 1, \ldots, n-1$. An associated Gaussian process Z verifies that, for $|i - j| > 1$, $Z(x_i)$ and $Z(x_j)$ are independent conditionaly on $\{Z(x_\ell) \text{ s.t. } \ell \notin \{i, j\}\}$. This is equivalent to the one-band form of the precision matrix of Z computed at $x_1, \ldots, x_n$.*

4.2 Equivalence of Poincaré and optimal kernel quadratures in $H^1(\mu)$

In the previous section, we proved that kernel quadrature is well defined both for the kernel K of $H^1(\mu)$ and its finite-dimensional approximation K_M. Now, we show that the Poincaré quadrature can be viewed as an optimal kernel quadrature with positive weights for the finite dimensional approximation of the kernel of $H^1(\mu)$ in its Mercer's representation.

Proposition 7 (Equivalence of Poincaré quadrature and optimal kernel quadrature in $H^1(\mu)$). *Let (X_P, w_P) be the Poincaré quadrature of $H^1(\mu)$ with n nodes and order $M = 2n - 1$. Then, $\mathrm{wce}(X_P, w_P, K_M) = 0$ and (X_P, w_P) is an optimal kernel quadrature for $\mathcal{H}_{K_M}$, with positive weights.*
Conversely, if (X, w) defines a kernel quadrature for $\mathcal{H}_{K_M}$ such that $\mathrm{wce}(X, w, K_M) = 0$ and the weights are

positive, then $X = X_P$ and $w = w_P$.

Furthermore, w_P minimizes over all (possibly negative) weights the worst-case error given X_P:

$$w_P = w^\star(X_P, K_M) = \operatorname{argmin}_{w \in \mathbb{R}^n} \operatorname{wce}(X_P, w, K_M).$$

Remark 8. *As a consequence of Proposition 7, the optimal weights $w^\star(X_P, K_M)$ are equal to w_P, and are thus positive and sum to 1, which was not obvious a priori as they are defined by an optimization problem on $\mathbb{R}^n$.*

Proof. Let us first consider the Poincaré quadrature. We know that

$$\int \varphi_m(x) d\mu(x) = \sum_{j=1}^n w_j \varphi_m(x^j), \qquad m = 0, \dots, M. \tag{4.2}$$

Let $x' \in [a, b]$. By linearity, we deduce from (4.2) that

$$\sum_{m=0}^M \alpha_m \varphi_m(x') \int \varphi_m(x) d\mu(x) = \sum_{m=0}^M \alpha_m \varphi_m(x') \sum_{j=1}^n w_j \varphi_m(x^j),$$

or equivalently

$$\int \sum_{m=0}^M \alpha_m \varphi_m(x') \varphi_m(x) d\mu(x) = \sum_{j=1}^n w_j \left(\sum_{m=0}^M \alpha_m \varphi_m(x') \varphi_m(x^j) \right),$$

i.e.

$$\int K_M(x, x') d\mu(x) = \sum_{i=1}^n w_i K(x^i, x').$$

This proves that $\operatorname{wce}(X_P, w_P, K_M) = 0$.

Conversely, let (X, w) be a kernel quadrature for $\mathcal{H}_{K_M}$ such that $\operatorname{wce}(X_P, w_P, K_M) = 0$ and with positive weights. By definition of the worst-case error, this implies that for all f in the unit ball of $\mathcal{H}_{K_M}$, $\int f(x) d\mu(x) = \sum_{i=1}^n w_i f(x_i)$. By considering $f / \|f\|_{\mathcal{H}_{K_M}}$, this identity is true for all $f \in \mathcal{H}_{K_M}$. In particular for $f = \varphi_m$ with $m \le M$, we deduce that the quadrature defined by (X, w) is exact for all φ_m such that $m \le M$. Thus, by uniqueness of the Gaussian quadrature of the T-system $(\varphi_m)_{m \in \mathbb{N}}$ (Proposition 2), we deduce that $X = X_P$ and $w = w_P$.

Now, denote $X_P = \{x_1, \dots, x_n\}$. Recall that the nodes x_i are all different by a property of the (generalized) Gaussian quadrature. By Proposition 5, the Gram matrix $(K_M(x^i, x^j))_{0 \le i, j \le n}$ is then invertible. This implies that the minimization problem

$$\min_w \operatorname{wce}(X_P, w, K_M)$$

has a unique solution $w^\star(X_P, K_M)$. Since $\operatorname{wce}(X_P, w_P, K_M) = 0$ and $\operatorname{wce}(X_P, w, K_M) \ge 0$ for all w, we obtain $w^\star(X_P, K_M) = w_P$, which concludes the proof. □

4.3 Formulas for optimal weights

Exploiting the equivalence of Poincaré quadrature and kernel quadrature, we obtain several formulas for the optimal weights.

Proposition 8 (Expression of the optimal weights and associated worst-case error for $H^1(\mu)$)**.** *Let K be the kernel of $H^1(\mu)$. Then, for all set X composed of n distinct quadrature knots, we have*

$$w^\star(X, K) = K(X, X)^{-1} \mathbb{1}, \tag{4.3}$$

where $\mathbb{1}$ is the vector of ones of length n, and

$$(\operatorname{wce}(X, w^\star(X, K), K))^2 = 1 - \mathbb{1}^\top K(X, X)^{-1} \mathbb{1} = 1 - \sum_{i=1}^n w_i^\star(X, K). \tag{4.4}$$

14

Similarly, when X is formed by at most $M + 1$ distinct points, Equations (4.3) and (4.4) are true when replacing K by K_M. In particular, let (X_P, w_P) be the Poincaré quadrature of $H^1(\mu)$ with n nodes and order $M = 2n - 1$. Then we have:

$$w_P = w^\star(X_P, K_M) = K_M(X_P, X_P)^{-1}\mathbb{1}. \tag{4.5}$$

.

Proof. First recall that Assumption 1 is verified both for K and K_M (Prop. 6 and 5), in the latter case when X has at most $M + 1$ points. Let us first consider the case of the kernel of $H^1(\mu)$. By Proposition 4, 1., we have, for all $y \in [a, b]$:

$$\ell_K(y) = \int_a^b K(x, y)d\mu(x) = 1.$$

From (2.7), we then deduce (4.3). Finally, from (2.8), we deduce (4.4), using that $c_K = \int_a^b \ell_K(y)d\mu(y) = 1$. The same proof applies when $K = K_M$, remarking that, by orthonormality of the Poincaré basis,

$$\ell_{K_M}(y) = \int_a^b K_M(x, y)d\mu(x) = \sum_{m=0}^{M} \alpha_m \varphi_m(y) \int_a^b \varphi_m(x)d\mu(x) = 1.$$

This implies that formulas (4.3) and (4.4) hold when K is replaced by K_M.

Now, let us set $X = X_P$. From Proposition 7, we have $w_P = w^\star(X_P, K_M)$. This implies (4.5). Notice that, as the the weights w_P sum to one, we recover one statement of Proposition 7 that $\mathrm{wce}(X_P, w_P, K_M) = 0$. $\square$

4.4   Quadrature error

The quadrature error can be quantified using the *radius of information*, which is defined as the smallest worst-case error of the optimal kernel quadrature with n nodes:

$$r(n) = \inf_{X,w} \mathrm{wce}(X, w, H^1(\mu)). \tag{4.6}$$

In our case, the radius of information can hardly be computed. However, we can consider the Poincaré quadrature, which is the optimal kernel quadrature with positive weights of the finite-dimensional approximation of $H^1(\mu)$, and compute the corresponding worst-case error in $H^1(\mu)$. More precisely, if (X_P, w_P) denotes the Poincaré quadrature with n nodes and order $M = 2n - 1$, we set:

$$\mathrm{wce}(n) = \mathrm{wce}(X_P, w_P, H^1(\mu)). \tag{4.7}$$

By definition we have $\mathrm{wce}(n) \geq r(n)$. Furthemore, when n is large, K_M tends to K and we can hope that $\mathrm{wce}(n)$ is a good approximation of $r(n)$. We now provide formulas for $\mathrm{wce}(n)$.

Proposition 9 (Quadrature error). *The worst-case error of the Poincaré quadrature with n nodes and order $M = 2n - 1$ can be expressed with the Mercer representation of $H^1(\mu)$, by:*

$$\mathrm{wce}(n)^2 = \sum_{m \geq M+1} \alpha_m \left(\sum_{i=1}^{n} w_i \varphi_m(x_i) \right)^2, \tag{4.8}$$

or with formulas involving the kernel of $H^1(\mu)$:

$$
\begin{aligned}
\mathrm{wce}(n)^2 &= w_P^\top (K(X_P, X_P) - K_M(X_P, X_P))w_P &(4.9)\\
&= \mathbb{1}^\top K_M(X_P, X_P)^{-1}(K(X_P, X_P) - K_M(X_P, X_P))K_M(X_P, X_P)^{-1}\mathbb{1} &(4.10)
\end{aligned}
$$

Furthermore, we have, for all $n \in \mathbb{N}$,

$$\mathrm{wce}(n) \leq \sqrt{\|K - K_{2n-1}\|_\infty},$$

which goes to zero when n tends to infinity.

Proof of Proposition 9. Denote by L the linear form on $\mathcal{H}_K$ defined by

$$L(f; X, w) = \int f(x)d\mu(x) - \sum_{i=1}^n w_i f(x_i)$$

and

$$R_M(x, x') = \sum_{m \geq M+1} \alpha_m \varphi_m(x) \varphi_m(x'). \tag{4.11}$$

Notice that $K = K_M + R_M$. For the Poincaré quadrature, we have:

$$L(K_M(., x'); X_P, w_P) = \sum_{m=0}^M \alpha_m \varphi_m(x') \left(\int \varphi_m(x)d\mu(x) - \sum_{i=1}^n w_i \varphi_m(x_i) \right) = 0$$

as the quadrature is exact for the eigenfunctions up to order M. Hence, by linearity, it holds for all x':

$$L(K(., x'); X_P, w_P) = L(R_M(., x'); X_P, w_P).$$

Now, remark that for all quadrature formula (X, w),

$$\text{wce}(X, w, K) = \|x' \mapsto L(K(., x'))\|_{\mathcal{H}_K}.$$

Therefore, we have $\text{wce}(X_P, w_P, K_M) = 0$ and $\text{wce}(X_P, w_P, K) = \text{wce}(X_P, w_P, R_M)$. Thus, the quantity of interest reduces to:

$$\text{wce}(n) := |\text{wce}(X_P, w_P, K_M) - \text{wce}(X_P, w_P, K)| = \text{wce}(X_P, w_P, R_M).$$

Now, when $m \geq M + 1$, we have $\int \varphi_m d\mu = 0$. Thus,

$$\begin{aligned}
\text{wce}(n)^2 &= \|x' \mapsto L(R_M(., x'); X_P, w_P)\|_{\mathcal{H}_K}^2 \\
&= \left\| \sum_{m \geq M+1} \alpha_m \left(\sum_{i=1}^n w_i \varphi_m(x_i) \right) \varphi_m \right\|_{\mathcal{H}_K}^2
\end{aligned}$$

As (φ_m) is an orthogonal basis of $H^1(\mu)$ with $\|\varphi_m\|_{\mathcal{H}_K}^2 = 1 + \lambda_m = \alpha_m^{-1}$, we immediately obtain

$$\text{wce}(n)^2 = \sum_{m \geq M+1} \alpha_m \left(\sum_{i=1}^n w_i \varphi_m(x_i) \right)^2.$$

Remarking that $R_M(x_i, .) = \sum_{m \geq M+1} \alpha_m \varphi_m(x_i) \varphi_m(.)$, we get

$$\text{wce}(n)^2 = \left\| \sum_{i=1}^n w_i R_M(x_i, .) \right\|_{\mathcal{H}_K}^2$$

which gives, using the reproducing property (as $R_M(x_i, .) \in \mathcal{H}_K$),

$$\text{wce}(n)^2 = \sum_{1 \leq i, j \leq n} w_i w_j R_M(x_i, x_j)$$

which gives (4.9). Using (4.5), we deduce (4.10). By Mercer theorem, as K is continuous, the series $\sum_m \alpha_m \varphi_m(x) \varphi_m(y)$ converges to $K(x, y)$ uniformly on $[a, b] \times [a, b]$. Thus R_M goes to zero uniformly on $[a, b]^2$, and, using the positivity of the weights,

$$\text{wce}(n)^2 \leq \left(\sum_{1 \leq i, j \leq n} w_i w_j \right) \|R_M\|_\infty.$$

The results follows by remarking that $\sum_{1 \leq i, j \leq n} w_i w_j = \left(\sum_{i=1}^n w_i \right)^2 = 1$. □

5 The case of $H^1(a, b)$

5.1 Nodes coincide with zeros of a basis function

For polynomials, the nodes of the Gaussian quadrature coincide with the zeros of orthogonal polynomials (Szegö, 1959). The main result of this section can be viewed as a variant of this property on the circle. Indeed, for the uniform distibution, the Poincaré basis consists of trigonometric functions that verify a stability-by-product property, similarly to power functions. The key argument is that the quadrature is not only exact for the functions of the T-system (up to some order) but also for their products. This guarantees that the quadrature nodes coincide with the zeros of an element of the T-system.

Proposition 10. *The nodes of the Poincaré quadrature of $H^1(a, b)$ with n nodes are equal to the zeros of the Poincaré basis function φ_n.*

Proof. We first prove that the Poincaré quadrature with n nodes is exact for the product $\varphi_j \varphi_k$ of eigenfunctions with $j, k \in \mathbb{N}$ such that $j < n$ and $k \leq n$.

Recall that when μ is the uniform distribution on $[a, b]$, we have $\varphi_m(x) = \sqrt{2}\cos(m\omega(x - a))$ with $\omega = \pi/(b - a)$. Using the trigonometric identity $\cos(u)\cos(v) = \frac{1}{2}\left(\cos(u + v) + \cos(u - v)\right)$, it follows that

$$\varphi_n(x)\varphi_m(x) = \frac{1}{2}\left(\varphi_{n+m}(x) + \varphi_{n-m}(x)\right).$$

By Definition 5, the Gaussian quadrature of the T-system $(\varphi_m)_{m \in \mathbb{N}}$ with n nodes is exact for φ_m for $0 \leq m \leq 2n - 1$. Let $j, k \in \mathbb{N}$ such that $j < n$, $k \leq n$. Without loss of generality, assume $k \geq j$. This implies that both $k + j$ and $k - j$ are non-negative and lower or equal than $2n - 1$. Thus, the quadrature is exact for φ_{k+j} and φ_{k-j} and we have:

$$
\begin{aligned}
\delta_{kj} = \int \varphi_k(x)\varphi_j(x)d\mu(x) &= \frac{1}{2}\int \varphi_{k+j}(x)d\mu(x) + \frac{1}{2}\int \varphi_{k-j}(x)d\mu(x) \\
&= \frac{1}{2}\sum_{i=1}^{n} w_i\varphi_{k+j}(x_i) + \frac{1}{2}\sum_{i=1}^{n} w_i\varphi_{k-j}(x_i) \\
&= \sum_{i=1}^{n} w_i\varphi_k(x_i)\varphi_j(x_i).
\end{aligned}
$$

This proves that the Poincaré quadrature with n nodes is exact for $\varphi_j \varphi_k$, with $j < n$ and $k \leq n$. We now deduce that the quadrature nodes coincides with the zeros of φ_n.

Following (Karlin and Studden, 1966, p. 20), as $(\varphi_m)_{m \in \mathbb{N}}$ is a T-system, there exists a non-zero generalized polynomial $p_n = \alpha_0\varphi_0 + \ldots + \alpha_n\varphi_n$ that vanishes at $x_1, \ldots, x_n$. It can be defined from the generalized Vandermonde matrix by:

$$p_n(x) = \det V(\varphi_0, \ldots, \varphi_{n-1}, \varphi_n; x_1, \ldots, x_n, x).$$

Notice that $\alpha_n \neq 0$. Indeed, by expanding the determinant, we have $\alpha_n = \det V(\varphi_0, \ldots, \varphi_{n-1}; x_1, \ldots, x_n)$, which is non-zero by the T-system property. Now, let $j \in \mathbb{N}$ with $j < n$. By the first part above, for all $k \in \mathbb{N}$ such that $k \leq n$, the Poincaré quadrature is exact for $\varphi_j \varphi_k$. This implies that p_n is orthogonal to φ_j:

$$
\begin{aligned}
\int p_n(x)\varphi_j(x)d\mu(x) &= \sum_{k=0}^{n} \alpha_k \int \varphi_k(x)\varphi_j(x)d\mu(x) \\
&= \sum_{k=0}^{n} \alpha_k \sum_{i=1}^{n} w_i\varphi_k(x_i)\varphi_j(x_i) \\
&= \sum_{i=1}^{n} w_i \underbrace{\sum_{k=0}^{n} \alpha_k\varphi_k(x_i)}_{=p_n(x_i)=0} \varphi_j(x_i) = 0.
\end{aligned}
$$

Thus p_n is orthogonal to $\varphi_0, \ldots, \varphi_{n-1}$. Since $(\varphi_m)_{m \in \mathbb{N}}$ is a system of orthogonal functions, we deduce that $p_n = \alpha_n\varphi_n$. As $\alpha_n \neq 0$, we have $\varphi_n = \alpha_n^{-1}p_n$. From the definition of p_n, this implies that φ_n is zero at the quadrature nodes, which was to prove. $\square$

5.2 Explicit quadrature formulas and quadrature error

We first give a standard result about trigonometric sums.

Lemma 1. *For all $m \in \mathbb{Z}$ and all $n \in \mathbb{N}^\star$,*

$$\sum_{i=1}^{n} \cos\left(\left(i - \frac{1}{2}\right)\frac{m\pi}{n}\right) = \begin{cases} 0 & \text{if } m \text{ is not a multiple of } 2n \\ n(-1)^p & \text{if } m = (2n)p, \text{ for all } p \in \mathbb{Z}. \end{cases}$$

Proof. Let $x = (m\pi)/n$. If m is a multiple of $2n$, then $x = 2p\pi$ for some $p \in \mathbb{Z}$. Then

$$\sum_{i=1}^{n} \cos\left(\left(i - \frac{1}{2}\right)x\right) = \sum_{i=1}^{n} \cos(-p\pi) = \sum_{i=1}^{n}(-1)^p = n(-1)^p.$$

Now, assume that m is not a multiple of $2n$. Then $x/2$ is not a multiple of π and $\sin(x/2) \neq 0$. Using the trigonometric identity $2\cos a \sin b = \sin(a+b) - \sin(a-b)$, we have

$$2\cos\left(\left(i - \frac{1}{2}\right)x\right)\sin\left(\frac{x}{2}\right) = \sin(ix) - \sin((i-1)x).$$

Summing with respect to i from 1 to n then gives the announced result. $\square$

We can now deduce the two main propositions of the section.

Proposition 11. *The Poincaré quadrature of $H^1(a,b)$ with n nodes corresponds to the midpoint (or rectangle) quadrature rule*

$$\int_a^b f(x)dx = \frac{b-a}{n}\sum_{i=1}^{n} f\left(a + \left(i - \frac{1}{2}\right)\frac{b-a}{n}\right).$$

Thus, the optimal weights are equal to $1/n$, and the optimal nodes are evenly spaced on $[a,b]$ and located at the middle of each interval $\left[a + (i-1)\dfrac{b-a}{n}, a + i\dfrac{b-a}{n}\right]$, $i = 1, \ldots, n$. This quadrature has order $2n-1$ with respect to the generalized polynomials: it is exact for all $\varphi_m \propto \cos\left(m\pi\frac{x-a}{b-a}\right)$ with $m \leq 2n - 1$. Furthermore it is also exact for all φ_m such that m is not a multiple of $2n$, and for polynomials of order 1.

Proof. From the previous section, the nodes of the Poincaré quadrature of $H^1(a,b)$ coincide to the zeros of $\varphi_n = \sqrt{2}\cos\left(n\pi\frac{x-a}{b-a}\right)$ on $[a,b]$. Hence they are equal to $x_i = a + \left(i - \frac{1}{2}\right)\frac{b-a}{n}$, for $i = 1, \ldots, n$.
Now, $\varphi_m(x_i) = \sqrt{2}\cos(m\omega(x_i - a)) = \sqrt{2}\cos\left(\left(i - \frac{1}{2}\right)\frac{m\pi}{n}\right)$. Thus, by Lemma 1, $\sum_{i=1}^{n}\varphi_m(x_i) = 0$ if m is not a multiple of $2n$. Hence, if we set $w_i = \frac{1}{n}$ and if m is not a multiple of $2n$, then

$$\int_a^b \varphi_m(x)\frac{dx}{b-a} = \delta_{m,0} = \sum_{i=1}^{n} w_i\varphi_m(x_i),$$

where the case $m = 0$ is equivalent to $w_1 + \cdots + w_n = 1$. Recall that the quadrature weights are uniquely determined by the first n equations above, corresponding to $m = 0, \ldots, n-1$. Indeed, the matrix of the linear system is $(\varphi_m(x_i))_{0 \leq m, i \leq n-1}$, which is invertible by the T-system property. This proves that the optimal weights are equal to $1/n$. Furthermore, the same equations show that the quadrature rule is exact for all φ_m such that m is not a multiple of $2n$. Finally, the quadrature is interpreted as the midpoint quadrature rule, which is exact for all polynomials of order 1. $\square$

Proposition 12 (Quadrature error)**.** *Consider the quadrature error defined as the worst-case error of the Poincaré quadrature of $H^1(a,b)$ with n nodes (see Section 4.4). We have:*

$$\mathrm{wce}(n) = \left(\frac{\frac{b-a}{2n}}{\tanh\left(\frac{b-a}{2n}\right)} - 1\right)^{1/2} \tag{5.1}$$

18

and goes to zero at a linear speed when n tends to infinity:

$$\text{wce}(n) \sim \frac{b-a}{2\sqrt{3}} \frac{1}{n}.$$

Proof. Recall that the optimal weights are $w_i = 1/n$, and the optimal nodes are $x_i = a + \left(i - \frac{1}{2}\right)\frac{b-a}{n}$. Then, using (4.8), we have:

$$\text{wce}(n)^2 = \sum_{m=2n}^{\infty} \alpha_m \frac{1}{n^2} \left(\sum_{i=1}^{n} \varphi_m(x_i) \right)^2.$$

Now, $\varphi_m(x_i) = \sqrt{2}\cos(m\omega(x_i - a)) = \sqrt{2}\cos\left(\left(i - \frac{1}{2}\right)\frac{m\pi}{n}\right)$, with $\omega = \frac{\pi}{b-a}$. By Lemma 1,

$$\sum_{i=1}^{n} \varphi_m(x_i) = \begin{cases} 0 & \text{if } m \text{ is not a multiple of } 2n \\ \sqrt{2}n(-1)^p & \text{if } m = (2n)p, \text{ for all positive integer } p. \end{cases}$$

Thus in this sum above, the non-zeros terms are such that $m = 2pn$. Observe that $m \geq 2n$ is then equivalent to $p \geq 1$. Hence, reparameterizing by p, we obtain

$$\text{wce}(n)^2 = 2\sum_{p=1}^{\infty} \alpha_{2np} = 2\sum_{p=1}^{\infty} \frac{1}{1 + p^2/r^2}.$$

with $r = (2n\omega)^{-1}$. Following the computations of Example 1, we have

$$\text{wce}(n)^2 = 2\sum_{p=1}^{\infty} \frac{1}{1 + p^2/r^2} = 2r^2 \sum_{p=1}^{\infty} \frac{1}{p^2 + r^2} = \frac{\pi r}{\tanh(\pi r)} - 1.$$

This gives the explicit form of $\text{wce}(n)$. To obtain the speed of convergence, notice that by an immediate application of Lebesgue theorem, $\sum_{p=1}^{\infty} \frac{1}{p^2 + r^2}$ tends to $\zeta(2) = \frac{\pi^2}{6}$ when r tends to zero. Hence, when n tends to infinity,

$$\text{wce}(n)^2 \sim \frac{1}{2n^2}\frac{1}{\omega^2}\frac{\pi^2}{6} = \frac{(b-a)^2}{12}\frac{1}{n^2}.$$

$\square$

5.3 Asymptotical optimality of the Poincaré quadrature

In the particular case of the uniform distribution, the kernel of $H^1(a,b)$ is given explicitly (see Section 3.2). Thus it is possible to derive the optimal kernel quadrature for $H^1(a,b)$, which can be found in (Duc-Jacquet, 1973, Theorem IV.5). For $a = 0, b = 1$, it is proved that the optimal nodes are $x_i^\star = \frac{2i-1}{2n}$, thus corresponding to the nodes of the rectangle quadrature, and the optimal weights are $w_i^\star = 2\tanh\left(\frac{1}{2n}\right)$. For large n, $w_i^\star \sim \frac{1}{n}$. Thus, the Poincaré quadrature, here equal to the rectangle quadrature, is asymptotically equivalent to the optimal kernel quadrature. Furthermore, the radius of information (worst-case error for the optimal quadrature, see Section 4.4) verifies

$$r(n)^2 = 1 - 2n\tanh\left(\frac{1}{2n}\right) \sim \frac{1}{12n^2}. \tag{5.2}$$

This is the same convergence speed as the worst-case error of the Poincaré quadrature, which we derived in Prop. 12:

$$\frac{\text{wce}(n)}{r(n)} \underset{n \to \infty}{\to} 1.$$

Therefore, we can conclude that the Poincaré quadrature is asymptotically optimal for $H^1(0,1)$. This result is intuitive as the finite-dimensional kernel K_M, for which the Poincaré quadrature is optimal, tends to the kernel of $H^1(\mu)$.

6 Numerical experiments

6.1 Numerical computation of the Poincaré quadrature

Computation of the spectral decomposition. The first step to compute numerically the Poincaré quadrature is to obtain the spectral decomposition of Theorem 1. This was investigated by Roustant et al. (2017, Section 4.2.), who proposed a finite element technique. It consists of solving Problem (P2) in the finite-dimensional space spanned by N_{mesh} piecewise linear functions whose knots are evenly spaced, which then boils down to a matrix diagonalisation problem. The theory of finite elements quantifies the speed of convergence when N_{mesh} tends to infinity, depending on the regularity of the probability density function ρ of $\mu \in \mathcal{B}$. If ρ is of class C^{ℓ}, the Poincaré basis functions are of class $C^{\ell+1}$, and the order of convergence is $O(N_{\text{mesh}}^{-2\ell})$ for the eigenvalues and $O(N_{\text{mesh}}^{-\ell})$ for the eigenfunctions. Notice that the value of N_{mesh}, controlling the mesh size, should be of course must larger than the order of the eigenvalue (or eigenfunction) to estimate. In practice we choose $N_{\text{mesh}} = 1\,000$. Figure 1 illustrates the result for the uniform distribution on $(0,1)$ and the exponential distribution truncated on $(0,3)$, for which the spectral decomposition is known in closed-form (as detailed in Section 3.2). We can see that the numerical approximation is accurate.
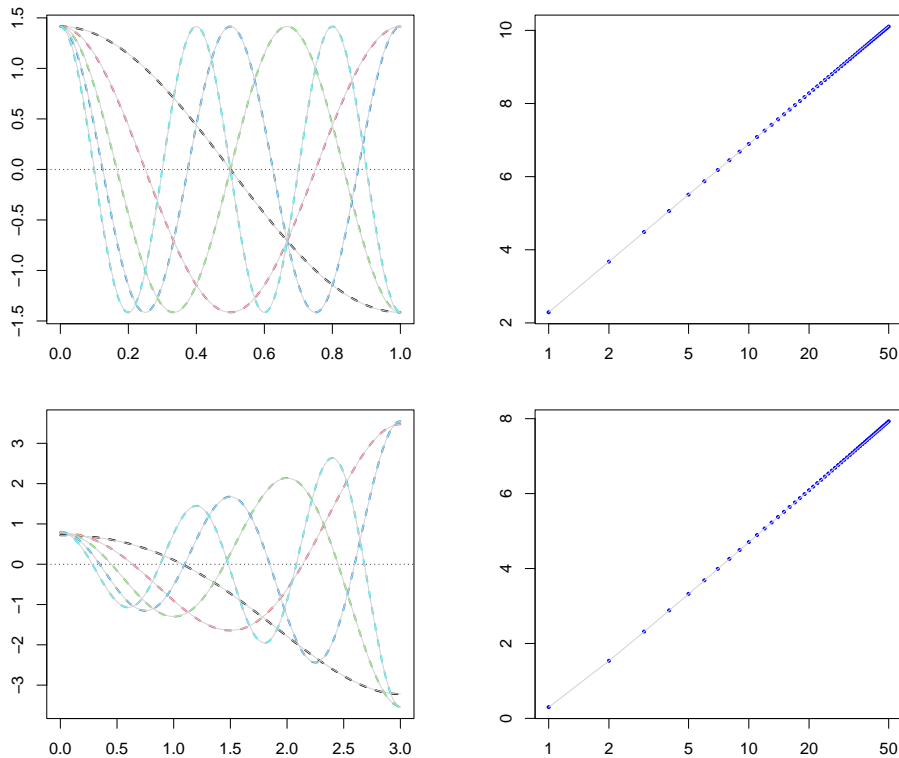

Figure 1: Poincaré spectral decomposition for the uniform distribution on $(0,1)$ (top) and the exponential distribution $d\mu(x) = e^{-x}1_{\mathbb{R}^+}(x)$, truncated on $(0,3)$ (bottom). The left panel represent the first five eigenfunctions (Poincaré basis) and the right panel the first 50 eigenvalues in log scale. The colored dotted lines are the approximations computed by finite elements, and the (superimposed) grey solid lines are the true expressions.

Computation of the Poincaré quadrature. We now assume that the Poincaré basis has been computed numerically, as explained in the previous paragraph. We aim at computing the Poincaré quadrature with n nodes. By definition, the Poincaré quadrature is the (generalized) Gaussian quadrature of the Poincaré

basis. Thus, it can be obtained by solving the minimization problem (2.4) of Proposition 2 over probability distributions σ subject to moment conditions. More precisely, inspired by the work of Ryu and Boyd (2015) for the usual Gaussian quadrature, we search σ as a discrete mesure supported on a fine uniform grid. We thus choose a large integer $N \gg n$ and consider the grid formed by evenly spaced points $z_j = a + jh$ ($j = 0, \ldots, N-1$) where $h = \frac{b-a}{N-1}$ is the grid size. Searching for σ of the form $\sigma = \sum_{j=0}^{N-1} w_j \delta_{z_j}$, (2.4) is then rewritten as the following linear programming (LP) problem:

$$w^* = \arg \min_{w \in [0,1]^N} \sum_{j=0}^{N-1} w_j \varphi_{2n}(z_j) \tag{6.1}$$

$$\text{subject to} \quad \sum_{j=0}^{N-1} w_j = 1$$

$$\text{and} \quad \sum_{j=0}^{N-1} w_j \varphi_i(z_j) = \int_a^b \varphi_i d\mu = \delta_{i,0} \qquad (i = 0, \ldots, 2n-1).$$

The problem can be solved numerically by standard LP solvers. However, as the grid points may not contain exactly the unknown nodes, the solution $\sigma^\star = \sum_{j=0}^{N-1} w_j^\star \delta_{z_j}$ is generally supported on more than n points. Thus, as a postprocessing step, we follow Ryu and Boyd (2015) and apply a clustering technique with n clusters (typically the k-means algorithm) to approximate the support points of the distribution. In each cluster $\mathcal{C}_i$ ($i = 1, \ldots, n$), a node x_i is defined as a convex combination of its elements z_j with weights proportional to $w_j^\star$ ($j \in \mathcal{C}_i$). Finally, the weight w_i associated to this node is defined as the sum of the weights in $\mathcal{C}_i$.

Due to the finite grid and the heuristic clustering and averaging technique, the moment conditions are in general not fulfilled exactly. To improve the accuracy of the obtained solution (x, w) further, we include a second optimization step, again following Ryu and Boyd (2015), in which we minimize the squared norm of the misfit for the moments

$$(X^P, w^P) = \arg \min_{x,w} \sum_{i=0}^{2n-1} \left(\delta_{i,0} - \sum_{j=1}^n w_j \varphi_i(x_j) \right)^2 \tag{6.2}$$

$$\text{subject to} \quad a \leq x_j \leq b \qquad (j = 1, \ldots, n)$$

$$\text{and} \quad 0 \leq w_j \leq 1 \qquad (j = 1, \ldots, n)$$

over both x and w using the interior-point algorithm (since we compute lower principal representations, which are always in (a, b)). The nodes and weights obtained from postprocessing the solution to (6.1) are used as the starting point for (6.2). The objective function value of the solution (X^P, w^P) of (6.2) is usually in the order of 10^{-7}. For the uniform distribution, it is in the order of 10^{-16}.

To evaluate the accuracy of the numerical quadrature, we compute the Poincaré quadrature of the uniform distribution where the analytical result is known (see Section 5). We have used the exact expression of the Poincaré basis. Figure 6.1 shows the results for $n = 3$ and $n = 5$ nodes. We can see that the nodes found coincide with the zeros of φ_n, as expected by the theory. Furthermore, the weights are equal to $1/n$ (up to machine precision), which is also expected.

Although the numerical computation of the Poincaré basis has been found to be accurate (see above in this section), we also investigate its influence by replacing the exact expression of the Poincaré basis in the previous experiments by its numerical approximation. The results are almost the same (the difference is in the order of 10^{-12}), showing that the whole procedure gives accurate results for the uniform distribution.

6.2 Further properties of Poincaré quadratures

Empirically, we find that the nodes and weights of a n-point Poincaré quadrature have the following properties, independently of the density:
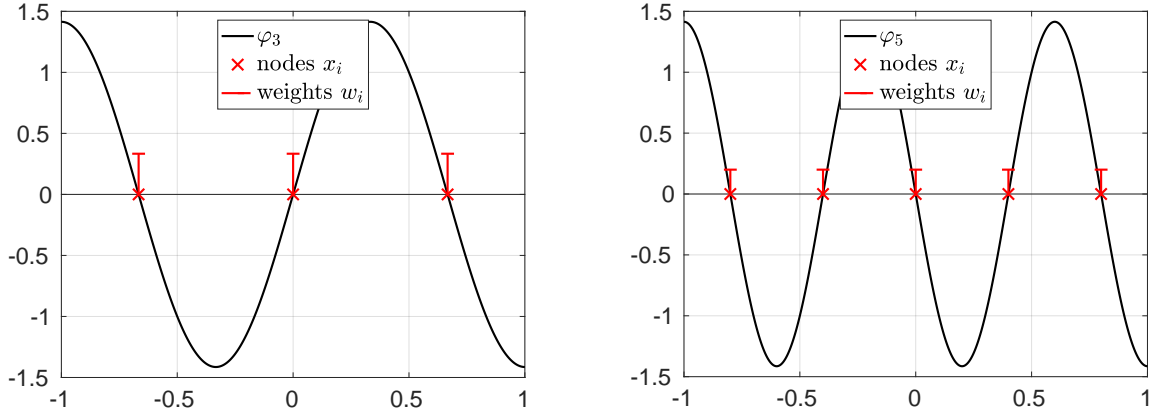
Figure 2: Poincaré quadrature for the uniform distribution on $(0, 1)$, with $n = 3$ nodes (left) and $n = 5$ nodes (right). The curve represents the Poincaré basis function with n roots, and the red crosses and lines the quadrature nodes and weights obtained by the numerical procedure.

1. The nodes are almost – but in general not exactly – equal to the zeros of φ_n. The difference is not due to numerical error. It is present even if the basis functions are known analytically, such as in the case of the truncated exponential.

2. The nodes are almost evenly spaced, but slightly skewed towards the concentration of probability mass.

3. The weights mimic the shape of the probability density function.

These three observations are illustrated in Fig. 3 for the truncated exponential distribution and for a nonparametric density.

(a) Truncated exponential ($n = 8$)

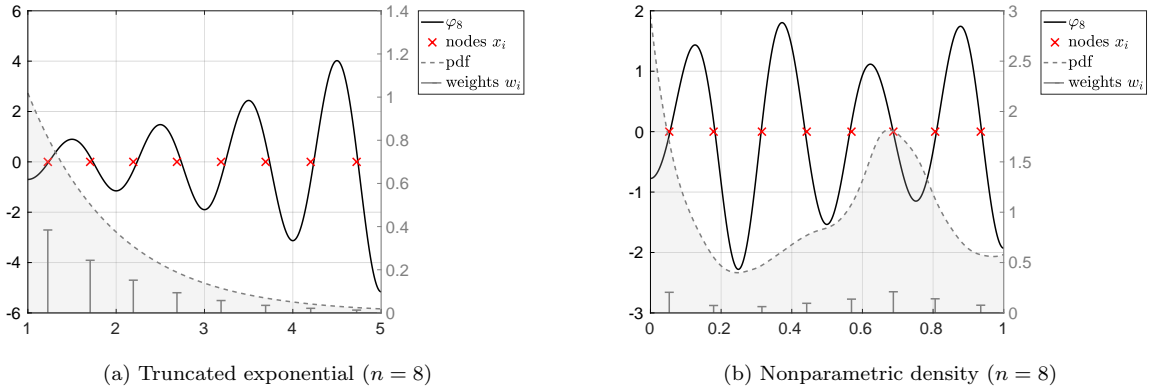(b) Nonparametric density ($n = 8$)

Figure 3: Poincaré nodes and weights for the truncated exponential distribution (interval $[1, 5]$, left) and a nonparametric density (right). Note that each plot has two y-axes with different scaling: the left one (in black) is for the basis function φ_n and for the nodes x_i, while the right one (in gray) is used for the probability density and the weights w_i.

Experiments with random densities. To better understand the properties of the Poincaré quadrature, we investigate it for a number of randomly generated continuous probability distributions. Their probability density functions (pdfs) are generated as follows. On the interval $[0, 1]$, we sample independent realizations of a Gaussian process with mean zero and a Matern-$\frac{5}{2}$ covariance kernel with parameter $\theta = 0.3$. Denote

22

one such realization with $g(x)$. Then, we define a pdf by $\exp(g(x))$ (up to a normalization constant). In case where the minimal value on $[0, 1]$ is smaller than 0.05, we reject this pdf, in order to avoid numerical issues (recall that μ must be a bounded perturbation of the uniform distribution, and thus its pdf does not vanish on the support interval). One example together with its Poincaré quadrature for $n = 8$ is shown in Fig. 3b. A set of 100 such pdfs is visualized in Figure 4a. As expected, the configurations are various, and often provide multimodal pdfs.

Ratio of pdf and weights. To further investigate the second and third observations mentioned above, namely, that the weights of the Poincaré quadrature mimic the associated density, we analyse the nodes and weights associated to 100 random densities.

In Fig. 4b, we display boxplots of the locations of the $n = 5$ nodes. We see that the nodes are nearly evenly spaced, which would correspond to the locations $(0.1, 0.3, 0.5, 0.7, 0.9)$.

Furthermore, in Fig. 4c we display boxplots of the ratio $\frac{nw_i}{\rho(x_i)}$ for the $n = 5$ quadrature nodes, where the weights are scaled by n for convenience. As already guessed from Fig. 3, we see that this ratio is quite close to 1. This suggests that the Poincaré quadrature might be a good quantization for the density ρ, which we investigate in the following.

(a) 100 randomly generated densities

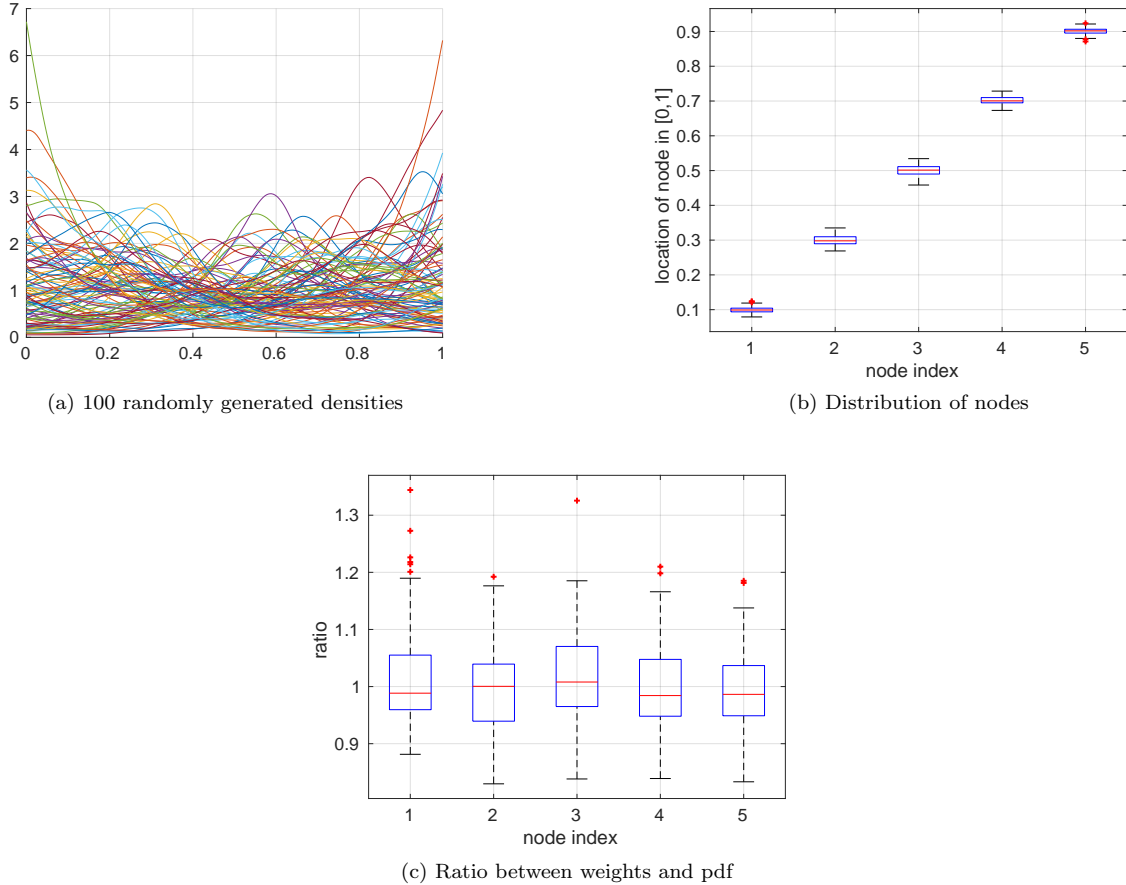(b) Distribution of nodes

(c) Ratio between weights and pdf

Figure 4: Left panel: 100 random pdfs generated by the procedure described in Section 6.2. Right panel: Location of nodes for the Poincaré quadratures associated to the same densities, with $n = 4$. Bottom panel: Ratio $\frac{w_i n}{\rho(x_i)}$, where ρ is the pdf associated to μ.

Wasserstein-optimal quantization.   It is interesting to compare the Poincaré quadrature to other standard quantization procedures, where quantization means an approximation of a continuous probability distribution by a discrete one. Here, we will focus on the optimal quantization associated to the Wasserstein distance, called Wasserstein-optimal quantization. The Wasserstein distance between two cumulative density functions (cdf) F, G is defined by

$$W(F, G) = \left(\int_0^1 (F^{-1}(p) - G^{-1}(p))^2 dp \right)^{1/2}. \tag{6.3}$$

Then, the corresponding optimal quantization with n points is the discrete probability distribution that has the smallest Wasserstein distance to the density associated to the measure μ. It can be computed efficiently using Lloyd's algorithm (Graf and Luschgy, 2007).

For each random pdf, for a fixed number of nodes $n = 5$, we compute the Poincaré quadrature, the standard Gaussian quadrature (associated to ordinary polynomials), and the Wasserstein-optimal quantization. We report the location of the nodes, as well as the zeros of φ_n, in Figure 5a. We observe that the Poincaré nodes are quite evenly spaced with small variability, and close (but not equal) to the zeros of the Poincaré basis function φ_n (denoted by red crosses). Furthermore, the Poincaré nodes are more evenly spaced than the support points of the Wasserstein-optimal quantization (denoted by yellow triangles). Finally, we observe that the Gaussian nodes are more spread out than the Poincaré nodes: the outermost Gaussian nodes are closer to the boundary than the outermost Poincaré nodes.

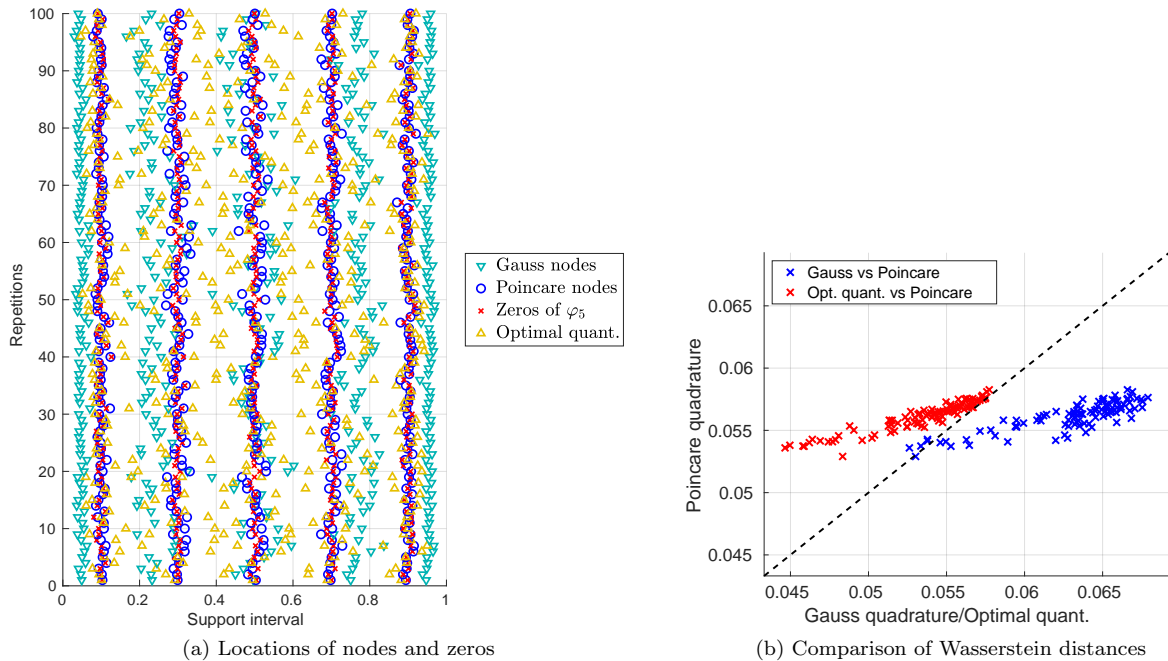(a) Locations of nodes and zeros   (b) Comparison of Wasserstein distances

Figure 5: Analysis of the properties of Poincaré quadrature rules with $n = 5$ nodes for the random probability distributions displayed in Fig. 4a. The left panel shows the corresponding nodes for the (usual) Gaussian quadrature and the Poincaré quadrature, as well as the support points for the Wasserstein-optimal quantization. The red crosses indicate the zeros of the Poincaré basis function φ_n. Right panel: Wasserstein distances between the continuous probability distribution μ and three different quadrature rules: Gaussian, Poincaré, and Wasserstein-optimal. Each blue point corresponds to a random density and shows the Wasserstein distances of Gaussian vs Poincaré quadratures. Similarly, each red point corresponds to a random density and shows the Wasserstein distances of Wasserstein-optimal vs Poincaré quadratures. In this way, each random density corresponds to two points in the plot. The black dashed line visualizes the identity $x = y$.

To further quantify the comparison, we measure the Wasserstein distance of the standard Gaussian quadrature and the Poincaré quadrature to μ. Results with $n = 5$ are presented in Figure 5b We see that in

most cases, the Poincaré quadrature has a smaller Wasserstein distance than the Gaussian one. The optimal quantization is by construction better than both, but often actually not much better than the Poincaré quadrature.

6.3   Worst-case error

Finally, we investigate the behaviour of the worst-case error $\mathrm{wce}(n)$ as a function of n. To compute $\mathrm{wce}(n)$ numerically, we use Equation (4.9), based on $K - K_M$ with $M = 2n - 1$. As in general, K is not known analytically, we approximate it by K_T (Equation 4.1), obtained by truncating the Mercer representation of K for a large value of $T \gg M$. Thus, $\mathrm{wce}(n) = \mathrm{wce}(X_P, w_P, K)$ is approximated by $\mathrm{wce}(X_P, w_P, K_T)$.

We first investigate the quality of this approximation for the uniform distribution, where both K and the Poincaré quadrature (X_P, w_P) have a closed-form expression. Here we have two options: a fully numerical computation, where we compute numerically the Poincaré quadrature, or a semi-analytical computation where we use its closed-form expressions. The resulting estimates of $\mathrm{wce}(n)$ differ by less than 10^{-4}, which is negligible compared to the actual value of $\mathrm{wce}(n)$. This reinforces our previous statement that the numerical computation of the Poincaré quadrature yields accurate results (Section 6.1). Therefore, in the next experiment, we only show the results for the fully numerical computation. Figure 6 shows the proxy worst-case error $\mathrm{wce}(X_P, w_P, K_T)$ as a function of the number of nodes n for different values of T. We compare it to the exact value of $\mathrm{wce}(n)$, given here by Equation (5.1), as well as to $\frac{1}{\sqrt{12}} n^{-1}$, which is asymptotically equivalent. We observe that as we include more terms in the truncated kernel K_T, the estimates approach $\mathrm{wce}(n)$. Because of the truncation, $\mathrm{wce}(X_P, w_P, K_T)$ underestimates the true error (see also Equation 4.8).
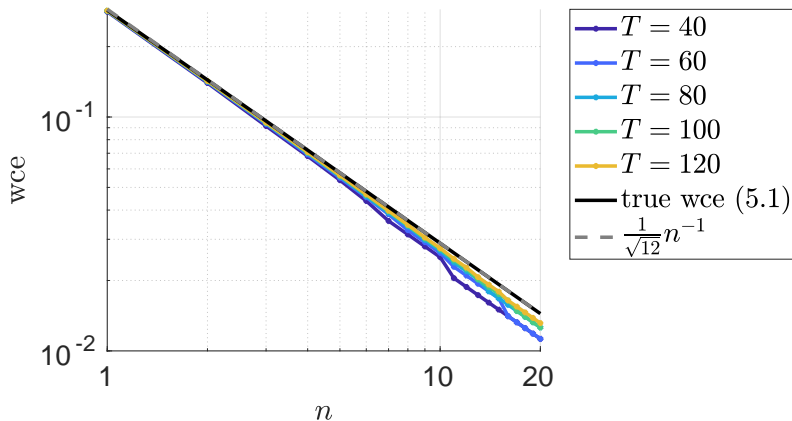

Figure 6: Approximation of $\mathrm{wce}(n)$ by $\mathrm{wce}(X_p, w_P, K_T)$ for the uniform distribution, and comparison with the closed-form expression of $\mathrm{wce}(n)$ and its asymptotical speed of convergence.

Secondly, we repeat the experiment for 5 random densities generated as described in Section 6.2. The worst-case error is computed with the fully numerical procedure presented above, with $T = 100$. The results are shown in Figure 7. Since we use a truncated kernel, the estimate likely underestimates the true $\mathrm{wce}(n)$. We observe that the worst-case error for the random densities appears to be close to $\frac{1}{\sqrt{12}} n^{-1}$, the asymptotic speed of convergence of $\mathrm{wce}(n)$ for $H^1(0, 1)$, suggesting that this speed of convergence might be universal.
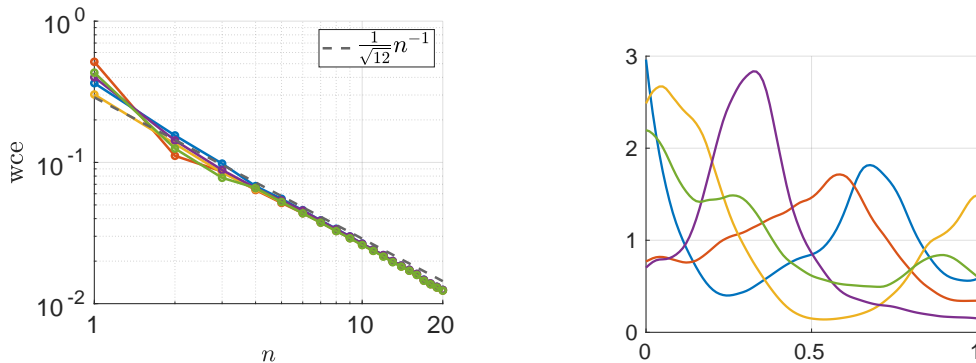
Figure 7: Worst-case error (left) for 5 random densities supported on $[0, 1]$ (right). The worst-case error is computed using Equation (4.9) as $\text{wce}(X_P, w_P, K_T)$ for $T = 100$.

7   Conclusion and outlook

In this paper, we have considered the wide class of probability distributions μ with continuous and non-vanishing pdf on a finite interval $[a, b]$. In this frame, we have brought three main contributions.

Firstly, remarking that the Sobolev space $H^1(\mu)$ is a RKHS, we have made a connection between the spectral decomposition of its kernel, Poincaré inequality on a compact interval, and T-systems. The form of the kernel is 'single-pair', and is linked to graphical Gaussian models.

Secondly, we have exploited this connection to achieve optimal kernel quadrature in $H^1(\mu)$. The T-system property guarantees the existence of a generalized Gaussian quadrature, extending the Gaussian quadrature of polynomials. We have investigated some particular cases where the results are explicit, in particular the standard (unweighted) Sobolev space with its usual Hilbert norm.

Thirdly, we have proposed an efficient numerical procedure. Using densities with and without analytically known kernel, we have explored the distribution of Poincaré quadrature nodes, the magnitude of the weights, and the speed of convergence for the worst-case error. We have gathered evidence for the hypothesis that asymptotically, the Poincaré quadrature nodes might be evenly spaced, and that the worst-case error of Poincaré quadrature scales as $\frac{1}{\sqrt{12}} n^{-1}$, as for the uniform case.

A direction for future research may be to theoretically investigate the asymptotical properties with respect to the number of quadrature nodes, that have been observed in the numerical experiments. This may involve specific techniques for studying large covariance matrices.

Acknowledgement

References

Atteia, M. (1992). *Hilbertian Kernels and Spline Functions*. Studies in Computational Mathematics, 4.

Bakry, D., I. Gentil, and M. Ledoux (2014). *Analysis and geometry of Markov diffusion operators, volume 348 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham.

Berlinet, A. and C. Thomas-Agnan (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.

Brass, H. and K. Petras (2011). *Quadrature theory: the theory of numerical integration on a compact interval.* Number 178. American Mathematical Soc.

Dick, J., D. Nuyens, and F. Pillichshammer (2014). Lattice rules for nonperiodic smooth integrands. *Numerische Mathematik 126*(2), 259–291.

Duc-Jacquet, M. (1973). *Approximation des fonctionnelles linéaires sur les espaces hilbertiens autoreproduisants.* Ph. D. thesis, Université Joseph-Fourier-Grenoble I.

Fasshauer, G. E. (2012). Green's functions: Taking another look at kernel approximation, radial basis functions, and splines. In *Approximation Theory XIII: San Antonio 2010*, pp. 37–63. Springer.

Gantmakher, F. R. and M. G. Krejn (2002). *Oscillation matrices and kernels and small vibrations of mechanical systems.* Number 345. American Mathematical Soc.

Garcke, J. and M. Griebel (2012). *Sparse grids and applications*, Volume 88. Springer Science & Business Media.

Graf, S. and H. Luschgy (2007). *Foundations of quantization for probability distributions.* Springer.

Karlin, S. and W. Studden (1966). T-systems: with applications in analysis and statistics. *Pure Appl. Math., Interscience Publishers, New York, London, Sidney.*

Karvonen, T., M. Kanagawa, and S. Särkkä (2019). On the positivity and magnitudes of bayesian quadrature weights. *Statistics and Computing 29*(6), 1317–1333.

Lüthen, N., O. Roustant, F. Gamboa, B. Iooss, S. Marelli, and B. Sudret (2021). Global sensitivity analysis using derivative-based sparse Poincaré chaos expansions.

Maathuis, M., M. Drton, S. Lauritzen, and M. Wainwright (2018). *Handbook of graphical models.* CRC Press.

Novak, E. and H. Woźniakowski (2008). *Tractability of Multivariate Problems: Standard information for functionals*, Volume 2. European Mathematical Society.

Oettershagen, J. (2017). *Construction of optimal cubature algorithms with applications to econometrics and uncertainty quantification.* Verlag Dr. Hut.

O'Hagan, A. (1991). Bayes–hermite quadrature. *Journal of statistical planning and inference 29*(3), 245–260.

Roustant, O., F. Barthe, and B. Iooss (2017). Poincaré inequalities on intervals – application to sensitivity analysis. *Electronic Journal of Statistics 11*(2), 3081 – 3119.

Ryu, E. K. and S. P. Boyd (2015). Extensions of Gauss quadrature via linear programming. *Foundations of Computational Mathematics 15*(4), 953–971.

Smolyak, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Doklady Akademii Nauk*, Volume 148, pp. 1042–1045. Russian Academy of Sciences.

Szegö, G. (1959). Orthogonal polynomials. In *Amer. Math. Soc. Colloquium, 1959.*

Thomas-Agnan, C. (1996). Computing a family of reproducing kernels for statistical applications. *Numerical Algorithms 13*(1), 21–32.

Zhang, S. and E. Novak (2019). Optimal quadrature formulas for the Sobolev space H^1. *Journal of Scientific Computing 78*(1), 274–289.