Maximum mean discrepancies of Farey sequences

Toni Karvonen[1] — Anatoly Zhigljavsky[2]

[1] *Lappeenranta–Lahti University of Technology LUT, Finland.*
[2] *Cardiff University, United Kingdom*

October 6, 2025

Abstract

We identify a large class of positive-semidefinite kernels for which a certain polynomial rate of convergence of maximum mean discrepancies of Farey sequences is equivalent to the Riemann hypothesis. This class includes all Matérn kernels of order at least one-half.

1 Introduction

A kernel $K\colon \Omega \times \Omega \to \mathbb{R}$ on a domain Ω is positive-semidefinite if the kernel Gram matrix with elements $K(x_i, x_j)$ for $i, j = 1, \ldots, N$ is positive-semidefinite for all $N \in \mathbb{N}$ and $X = (x_1, \ldots, x_N) \subseteq \Omega$. Each positive-semidefinite kernel induces a unique reproducing kernel Hilbert space H (RKHS) that is a Hilbert space of certain real-valued functions defined on Ω equipped with a norm $\|\cdot\|_H$ [3, 18]. Properties such as differentiability and boundedness of the kernel determine which functions are elements of H. Under mild measurability assumptions one can then define the *maximum mean discrepancy* (MMD)

$$\mathrm{MMD}(P, X) = \sup_{\|f\|_H \leq 1} \left| \int_\Omega f(x) P(\mathrm{d}x) - \frac{1}{N} \sum_{i=1}^N f(x_i) \right|$$

between a probability measure P on Ω and the empirical measure $\xi = {}^1\!/\!{}_N \sum_{i=1}^N \delta_{x_i}$ of the point set X. The MMD measures how well the empirical measure approximates P or, in other words, how well the points are P-distributed. Among other things, the MMD is routinely used as a test statistic and for hypothesis testing in non-parametric statistics and machine learning [9, 19].

The Riemann hypothesis (RH) states that the non-trivial roots of the Riemann zeta function $\zeta(s) = \sum_{k=1}^\infty k^{-s}$ on the complex plane all have real part $1/2$. The purpose of this note is to show that the RH is equivalent to a certain polynomial rate of convergence of the MMD of the uniform measure on $[0, 1]$ and the empirical measure of the *Farey sequence.* The equivalence holds for a large class of commonly used kernels, including Matérns. The nth Farey sequence $F_n = (a_{i,n})_{i=1}^N$ is the increasing sequence of reduced

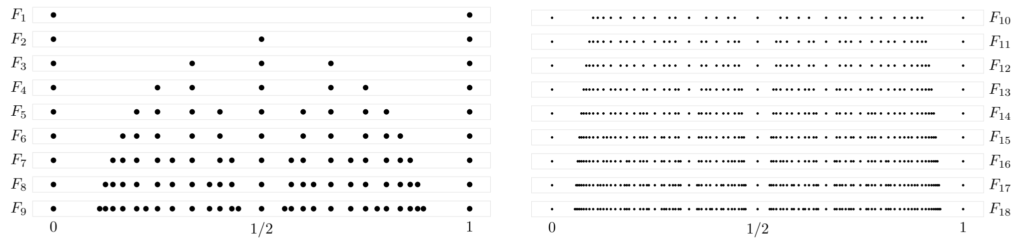

Figure 1: The first 18 Farey sequences $F_1, \ldots, F_{18}$.

fractions on $[0,1]$ whose denominators do not exceed n. The first six Farey sequences are

$$\begin{aligned}
F_1 &= \left(\frac{\mathbf{0}}{\mathbf{1}}, \frac{\mathbf{1}}{\mathbf{1}}\right) = (a_{1,1},\ a_{2,1}),\\
F_2 &= \left(\frac{0}{1}, \frac{\mathbf{1}}{\mathbf{2}}, \frac{1}{1}\right) = (a_{1,2},\ a_{2,2},\ a_{3,2}),\\
F_3 &= \left(\frac{0}{1}, \frac{\mathbf{1}}{\mathbf{3}}, \frac{1}{2}, \frac{\mathbf{2}}{\mathbf{3}}, \frac{1}{1}\right) = (a_{1,3},\ a_{2,3},\ a_{3,3},\ a_{4,3},\ a_{5,3}),\\
F_4 &= \left(\frac{0}{1}, \frac{\mathbf{1}}{\mathbf{4}}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{\mathbf{3}}{\mathbf{4}}, \frac{1}{1}\right) = (a_{1,4},\ a_{2,4},\ a_{3,4},\ a_{4,4},\ a_{5,4},\ a_{6,4},\ a_{7,4}),\\
F_5 &= \left(\frac{0}{1}, \frac{\mathbf{1}}{\mathbf{5}}, \frac{1}{4}, \frac{1}{3}, \frac{\mathbf{2}}{\mathbf{5}}, \frac{1}{2}, \frac{\mathbf{3}}{\mathbf{5}}, \frac{2}{3}, \frac{3}{4}, \frac{\mathbf{4}}{\mathbf{5}}, \frac{1}{1}\right) = (a_{1,5},\ a_{2,5}, \ldots,\ a_{10,5},\ a_{11,5}),\\
F_6 &= \left(\frac{0}{1}, \frac{\mathbf{1}}{\mathbf{6}}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{\mathbf{5}}{\mathbf{6}}, \frac{1}{1}\right) = (a_{1,6},\ a_{2,6}, \ldots,\ a_{12,6},\ a_{13,6}),
\end{aligned}$$

where we have bolded the points in F_n that do not appear in F_{n-1}. The first 18 Farey sequences are shown in Figure 1. Note that n refers to the index of a Farey sequence, not to the number of points in the nth Farey sequence, which is $N = |F_n| = \Phi(n) + 1$, where Φ is the summatory totient function.

Farey sequences have long history and many appearances in different branches of mathematics [4]. The most intriguing appearance of Farey sequences is related to their connections with the RH, which originate with a classical result by Franel [8], who proved that

$$\sum_{i=1}^{N} \left(\frac{i}{N} - a_{i,n}\right)^2 = O(n^{-1+\varepsilon}) \qquad \text{as} \qquad n \to \infty \tag{1.1}$$

for every $\varepsilon > 0$ is equivalent to the RH. Equation (1.1) is a statement about the uniformity of the distribution of Farey sequences. We refer to [14] for a comprehensive review. In this note we exploit these connections and recast Franel's result using the statistical concept of MMD, hence providing an equivalent formulation of the RH with statistical flavour.

Equation (1.1) can be understood as a statement about the discretised L^2-discrepancy; see [6, 16] for results on the star-discrepancy. Let $X = (x_1, \ldots, x_N) \subset [0,1]$ be an increasing sequence. The local discrepancy of X on $[0,1]$ is given by

$$D_{\mathrm{loc}}(A, X) = \left|\frac{|A \cap X|}{N} - \mathrm{meas}(A)\right| \qquad \text{for measurable} \qquad A \subseteq [0,1].$$

The L^2-discrepancy measures the uniformity of X and is defined as

$$D^2(X) = \sqrt{\int_0^1 D_{\mathrm{loc}}([0,\alpha], X)^2 \, \mathrm{d}\alpha}.$$

An approximation $\tilde{D}^2(X)$ to $D^2(X)$ can be obtained by discretising at $x_1, \ldots, x_N$:

$$\tilde{D}^2(X) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} D_{\text{loc}}([0,x_i],X)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{i}{N} - x_i\right)^2} \approx D^2(X),$$

where the second equality follows from the facts that $\text{meas}([0,x_i]) = x_i$ and, since X is an increasing sequence, $|[0,x_i] \cap X| = i$. A theorem of Mertens [10, § 18.5] gives the asymptotic equivalence

$$N = |F_n| \sim \frac{n^2}{2\zeta(2)} = \frac{3n^2}{\pi^2}, \tag{1.2}$$

where $\zeta(2) = \pi^2/6$ is the famous value of the Riemann zeta function at $s = 2$ originally computed by Euler. Franel's result in Equation (1.1) and the asymptotics in Equation (1.2) then show that the RH is equivalent to

$$\tilde{D}^2(F_n) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{i}{N} - a_{i,n}\right)^2} = O(n^{-3/2+\varepsilon}) = O(N^{-3/4+\varepsilon})$$

for every $\varepsilon > 0$. That is, the RH is equivalent to a statement about the discretised L^2-discrepancies of Farey sequences. It should now come as no suprise that the RH can also be formulated in terms of the MMD.

2 Results and remarks

For the MMD between the uniform measure on $[0,1]$ and the empirical measure for a sequence points $X = (x_1, \ldots, x_N) \subset [0,1]$ we use the simplified notation

$$\text{MMD}(X) = \sup_{\|f\|_H \leq 1} \left| \int_0^1 f(x)\,\mathrm{d}x - \frac{1}{N}\sum_{i=1}^{N} f(x_i) \right|. \tag{2.1}$$

The following theorem connects the rate of convergence of MMDs of Farey sequences to the RH. We give the proof in Section 4.

Theorem 2.1. *Let K be a positive-semidefinite kernel on $[0,1]$ and H its RKHS. Suppose that*

(a) H is a subset of $W^{1,2}([0,1])$, the Sobolev space of order one on $[0,1]$, and

(b) H contains the function $x \mapsto a + bx + x^\beta$ for some $a, b \in \mathbb{R}$ and $\beta \in [2,\gamma] \cup \{4,5\}$, where $\gamma = 1 + {}^6\!/\!\sqrt{3D} \approx 3.405$ and $D = {}^{\pi^2}\!/6 + {}^2\!/3 \log 2 - {}^2\!/3$.

Then the RH is equivalent to

$$\text{MMD}(F_n) = O(n^{-3/2+\varepsilon}) = O(N^{-3/4+\varepsilon}) \qquad \textit{for every} \qquad \varepsilon > 0. \tag{2.2}$$

Note that assumption (b) is satisfied if H contains the monomial $x \mapsto x^\beta$ for some $\beta \in \{2,3,4,5\}$. Figure 2 shows how the MMD behaves. There are three commonly used classes of kernels that are covered by Theorem 2.1:

1. Let $\lambda > 0$ be a correlation length parameter. The Matérn kernel of order $\nu > 0$ is given by

$$K_\nu(x,y) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}|x-y|}{\lambda}\right)^\nu \mathcal{K}_\nu\left(\frac{\sqrt{2\nu}|x-y|}{\lambda}\right), \tag{2.3}$$

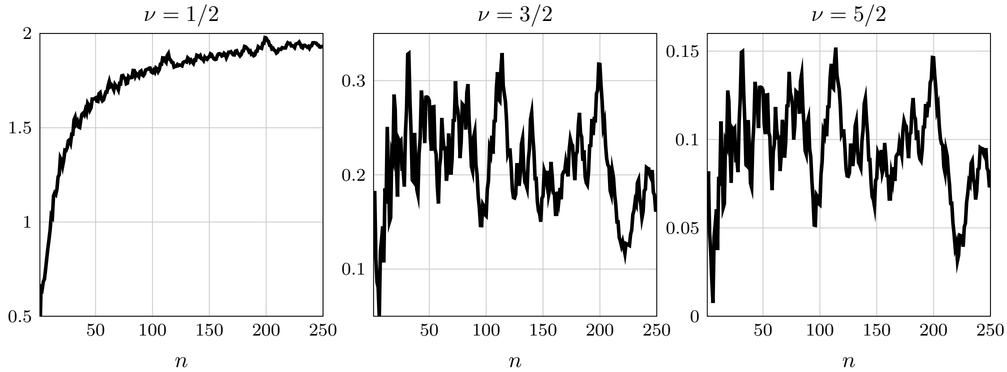

Figure 2: The plots show $\mathrm{MMD}(F_n) \cdot n^{3/2}$, the normalised MMDs of Farey sequences, up to $n = 250$ for Matérn kernels with (i) $\nu = 1/2$ and $\lambda = 1$, (ii) $\nu = 3/2$ and $\lambda = 3^{1/2}$, and (iii) $\nu = 5/2$ and $\lambda = 5^{1/2}$. Note that n is the index of the Farey sequence, not the number of points. For $n = 250$ we have $N = |F_n| = 19{,}025$.

where Γ is the gamma function and $\mathcal{K}_\nu$ the modified Bessel function of the second kind of order ν. Matérn kernels are widely used to define Gaussian random field models in spatial statistics [20]. It is well known that the RKHS of a Matérn kernel of order ν on $[0, 1]$ is norm-equivalent to the (possibly fractional) Sobolev space $W^{\nu+1/2,2}([0, 1])$; see [23, Thm. 6.13 and Cor. 10.48]. Assumptions (a) and (b) are thus satisfied if $\nu \geq 1/2$ because Sobolev spaces are nested and contain all polynomials.

2. By integrating $\min\{x, y\}$, the covariance kernel of Brownian motion, m times and adding a polynomial part that removes boundary conditions at the origin one obtains the released m-fold integrated Brownian motion kernel

$$K_m(x, y) = \sum_{k=0}^{m} \frac{(xy)^k}{(k!)^2} + \int_0^1 \frac{(x - t)_+^m (y - t)_+^m}{(m!)^2} \,\mathrm{d}t, \tag{2.4}$$

where $(x)_+ = \max\{0, x\}$. Because its RKHS is norm-equivalent to $W^{m+1,2}([0, 1])$, this kernel satisfies the assumptions of Theorem 2.1 [3, p. 322].

3. The energy-distance kernel

$$K_\alpha(x, y) = |x|^\alpha + |y|^\alpha - |x - y|^\alpha \tag{2.5}$$

is positive-semidefinite for $\alpha \in (0, 2)$. The energy-distance kernel is, up to scaling, the covariance kernel of the fractional Brownian motion with Hurst index $\beta/2$; for $\alpha = 1$ it reduces to the Brownian motion kernel. In Section 4.2 we use a characterisation by Barton and Poor [2] to verify that the RKHS of an energy-distance kernel satisfies the assumptions of Theorem 2.1 for every $\alpha \in [1, 2)$.

We summarise these observations in the following proposition.

Proposition 2.2. *Suppose that K is (i) any Matérn kernel of order $\nu \geq 1/2$ in Equation* (2.3)*, (ii) a released m-fold integrated Brownian motion kernel in Equation* (2.4) *with $m \geq 0$, or (iii) an energy-distance kernel in Equation* (2.5) *with $\alpha \in [1, 2)$. Then the RH is equivalent to Equation* (2.2).

Remark 2.3. Matérn kernels, integrated Brownian motion kernels, and energy-distance kernel are only finitely differentiable. Theorem 2.1 applies also to some infinitely differentiable kernels, such as $K(x, y) = \exp(xy)$. The RKHS of this particular kernel is contained in every Sobolev space and includes all polynomials by the virtue of the Taylor expansion $K(x, y) = \sum_{k=0}^{\infty} (xy)^k/k!$; see [18, Sec. 2.1] and [25]. The theorem does not cover the popular Gaussian kernel $K(x, y) = \exp(-(x-y)^2/(2\lambda^2))$ because its RKHS does not contain any non-trivial polynomials [15].

Remark 2.4. There are a number of results which state that the RH is equivalent to

$$\left| \int_0^1 f(x) \,\mathrm{d}x - \frac{1}{N} \sum_{i=1}^{N} f(a_{i,n}) \right| = O(n^{-3/2+\varepsilon})$$

for every $\varepsilon > 0$ if f is a suitable fixed function. The proof of Theorem 2.1 uses such results from [12, 24] that apply to certain monomials. Other options from [12, 13, 24] include functions whose derivatives satisfy a particular inequality and certain functions with singularity at the origin, such as $f(x) = \log(x)$. Kanemitsu and Yoshimoto [11] provide a collection of suitable functions expressed as Fourier cosine series.

3 Implications for energy-distance kernels

Theorem 2.1 has some curious implications when applied to the energy-distance kernel K_α in Equation (2.5). By direct calculation, we obtain the following expression for the squared MMD of the nth Farey sequence:

$$\mathrm{MMD}^2(F_n) = \frac{4}{(\alpha+1)N} \sum_{i=1}^{N} a_{i,n}^{\alpha+1} - \frac{1}{N^2} \sum_{i,j=1}^{N} |a_{i,n} - a_{j,n}|^\alpha - \frac{2}{(\alpha+1)(\alpha+2)}. \tag{3.1}$$

For $\alpha = 1$, the energy-distance kernel reduces to the Brownian motion kernel $K_1(x, y) = \min\{x, y\}$ that we use in the proof of Theorem 2.1 (in particular, see Lemma 4.1). Here we consider $\alpha \in (1, 2)$. Denote

$$S_{\alpha,N} = \frac{1}{N} \sum_{i=1}^{N} a_{i,n}^{\alpha+1} \quad \text{and} \quad T_{\alpha,N} = \frac{1}{N^2} \sum_{i,j=1}^{N} |a_{i,n} - a_{j,n}|^\alpha,$$

so that

$$\mathrm{MMD}^2(F_n) = \left(\frac{4}{\alpha+1} S_{\alpha,N} - \frac{4}{(\alpha+1)(\alpha+2)} \right) + \left(\frac{2}{(\alpha+1)(\alpha+2)} - T_{\alpha,N} \right) =: \delta_{1,N} + \delta_{2,N}.$$

By the asymptotic uniformity of F_n [see Equation (4.3)],

$$S_{\alpha,N} \to \int_0^1 x^{\alpha+1} \,\mathrm{d}x = \frac{1}{\alpha+2}$$

with the rate $O(n^{-3/2+\varepsilon})$ for every $\varepsilon > 0$ as $N \to \infty$. That is, $\delta_{1,N} = O(n^{-3/2+\varepsilon})$. However, Theorem 2.1 and Proposition 2.2 state that the RH is equivalent to $\mathrm{MMD}^2(F_n) = O(n^{-3+\varepsilon})$ for any $\alpha \in [1, 2)$. This has two consequences. First,

$$\delta_{2,N} = \frac{2}{(\alpha+1)(\alpha+2)} - T_{\alpha,N} = \int_0^1 \int_0^1 |x-y|^\alpha \,\mathrm{d}x \,\mathrm{d}y - \frac{1}{N^2} \sum_{i,j=1}^{N} |a_{i,n} - a_{j,n}|^\alpha$$
$$= O(n^{-3/2+\varepsilon})$$

if the RH holds. Second, the RH requires there to be substantial cancellations between the terms $\delta_{1,N}$ and $\delta_{2,N}$ which individually tend to zero as $O(n^{-3/2+\varepsilon})$ but whose sum must have the much faster rate $O(n^{-3+\varepsilon})$ under the RH.

4 Proofs

This section contains proofs for the results in Section 2.

4.1 Proof of Theorem 2.1

The proof of Theorem 2.1 uses the following lemma, which in fact states that the MMD for a piecewise linear kernel equals the L^2-discrepancy [5, 22].

Lemma 4.1. *Let $N \in \mathbb{N}$ be odd and $K(x, y) = 1 + \min\{x, y\}$. If an increasing sequence $X = (x_1, \ldots, x_N)$ contains $1/2$ and is symmetric on $[0, 1]$, in that $1 - x \in X$ if $x \in X$, then*

$$\mathrm{MMD}(X)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{i}{N} - x_i\right)^2 - \frac{1}{6N^2}.$$

Proof. From the reproducing property of the kernel in H it follows that the squared MMD admits the closed form expression

$$\mathrm{MMD}(X)^2 = \int_0^1\int_0^1 K(x, y)\,\mathrm{d}x\,\mathrm{d}y - \frac{2}{N}\sum_{i=1}^{N}\int_0^1 K(x, x_i)\,\mathrm{d}x + \frac{1}{N^2}\sum_{i,j=1}^{N} K(x_i, x_j) \quad (4.1)$$

in terms of the kernel and its integrals; see, for example, [9, Lem. 6] or [17, Sec. 10.2]. For the kernel $K(x, y) = 1 + \min\{x, y\}$ it is straightforward to compute that

$$\int_0^1 K(x, y)\,\mathrm{d}x = 1 + \frac{1}{2}(2 - y)y \quad \text{and} \quad \int_0^1\int_0^1 K(x, y)\,\mathrm{d}x\,\mathrm{d}y = \frac{4}{3}.$$

Therefore Equation (4.1) gives

$$\begin{aligned}\mathrm{MMD}(X)^2 &= \frac{4}{3} - \frac{2}{N}\sum_{i=1}^{N}\left(1 + \frac{1}{2}(2 - x_i)x_i\right) + \frac{1}{N^2}\sum_{i,j=1}^{N}(1 + \min\{x_i, x_j\}) \\ &= \frac{1}{3} - \frac{1}{N}\sum_{i=1}^{N}(2 - x_i)x_i + \frac{1}{N^2}\sum_{i=1}^{N}(2N - 2i + 1)x_i \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{i}{N} - x_i\right)^2 + \frac{1}{N^2}\sum_{i=1}^{N} x_i - \frac{3N + 1}{6N^2}.\end{aligned}$$

Because the sequence X contains $1/2$ and is symmetric,

$$\frac{1}{N^2}\sum_{i=1}^{N} x_i - \frac{3N + 1}{6N^2} = \frac{1}{N^2}\left(\frac{1}{2} + \frac{N - 1}{2}\right) - \frac{3N + 1}{6N^2} = -\frac{1}{6N^2}.$$

This concludes the proof. □

Proof of Theorem 2.1. Consider first the kernel $K_0(x, y) = 1+\min\{x, y\}$ from Lemma 4.1. Because they contain $1/2$ for $n \geq 2$ and are symmetric, we may apply the lemma to the Farey sequences F_n and so obtain

$$\mathrm{MMD}_0(F_n)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{i}{N} - a_{i,n}\right)^2 - \frac{1}{6N^2} \quad (4.2)$$

for the K_0-MMD of F_n when $n \geq 2$. The result by Franel in Equation (1.1) and the asymptotics $N = |F_n| \sim 3n^2/\pi^2$ from Equation (1.2) imply that the RH is equivalent to

$$\sum_{i=1}^{N} \left(\frac{i}{N} - a_{i,n} \right)^2 = O(n^{-1+\varepsilon}) = O(N^{-1/2+\varepsilon})$$

for every $\varepsilon > 0$. Inserting this in Equation (4.2) and observing that the second term is negligible shows that the RH is equivalent to $\mathrm{MMD}_0(F_n) = O(N^{-3/4+\varepsilon})$, from which Equation (2.2) follows for $K = K_0$ by using Equation (1.2) again.

It is well known (e.g., [1, Sec. 3.1] or [21, Sec. 10]) that the RKHS H_0 of $K_0(x, y) = 1 + \min\{x, y\}$ is norm-equivalent to the Sobolev space $W^{1,2}([0, 1])$. By an RKHS inclusion theorem of Aronszajn [18, Thm. 5.1], the norm of any RKHS $H \subseteq H_0$ satisfies $\|f\|_{H_0} \leq c\|f\|_H$ for a positive c and all $f \in H$. From this it follows that the unit ball of H is contained in the c-ball of H_0. Because the MMD in Equation (2.1) is a supremum over the unit ball of the RKHS, the MMD of F_n for any kernel K that satisfies assumption (a) is bounded from above by a constant multiple of the MMD of F_n for K_0. Consequently,

$$\mathrm{MMD}(F_n) = O(\,\mathrm{MMD}_0(F_n)\,) = O(n^{-3/2+\varepsilon}) = O(N^{-3/4+\varepsilon})$$

for every $\varepsilon > 0$ if the RH is true. To show that this rate implies the RH we use assumption (b). Let $f_\beta(x) = a + bx + x^\beta$ for $a, b \in \mathbb{R}$ and $\beta \in [2, \gamma] \cup \{4, 5\}$, where

$$\gamma = 1 + \frac{6}{\sqrt{3D}} \approx 3.405 \quad \text{and} \quad D = \frac{\pi^2}{6} + \frac{2}{3}\log 2 - \frac{2}{3}.$$

Because Farey sequences are symmetric about $1/2$,

$$\left| \int_0^1 f_\beta(x)\,\mathrm{d}x - \frac{1}{N}\sum_{i=1}^{N} f(a_{i,n}) \right| = \left| \int_0^1 x^\beta\,\mathrm{d}x - \frac{1}{N}\sum_{i=1}^{N} a_{i,n}^\beta \right|.$$

Results by Mikolás [12, Thm. 5] and Yoshimoto [24, pp. 302–3] state, in combination with the asymptotics in Equation (1.2), that the RH is equivalent to

$$\left| \int_0^1 x^\beta\,\mathrm{d}x - \frac{1}{N}\sum_{i=1}^{N} a_{i,n}^\beta \right| = O(n^{-3/2+\varepsilon}) = O(N^{-3/4+\varepsilon}) \tag{4.3}$$

for every $\varepsilon > 0$. Under assumption (b) the unit ball of H contains a constant multiple of f_β, so that the rate $\mathrm{MMD}(F_n) = O(n^{-3/2+\varepsilon})$ implies Equation (4.3) and consequently also the RH. This concludes the proof. □

4.2 Proof of Proposition 2.2 for energy-distance kernels

We first prove that RKHSs of energy-distance kernels contain suitable polynomials.

Lemma 4.2. *Let m be a positive integer. The RKHS of the energy-distance kernel K_α in* (2.5) *on* $[0, 1]$ *contains a polynomial of degree m for every $\alpha \in (0, 2)$.*

Proof. Let $H_\alpha(\mathbb{R})$ denote the RKHS of the energy-distance kernel (2.5) on $\mathbb{R}$. Theorem 4.1 in [2] states that $f \in H_\alpha(\mathbb{R})$ if and only if there is $g\colon \mathbb{R} \to \mathbb{C}$ satisfying $\int_\mathbb{R} |g(\omega)|^2 |\omega|^{1-\alpha}\,\mathrm{d}\omega < \infty$ such that

$$f(x) = \int_\mathbb{R} \frac{e^{i\omega x} - 1}{i\omega} \overline{g(\omega)} |\omega|^{1-\alpha}\,\mathrm{d}\omega \quad \text{for all} \quad x \in \mathbb{R}. \tag{4.4}$$

We select

$$g(\omega) = -i\omega \,\mathrm{sinc}^{m+1}\left(\frac{\omega}{2\pi}\right)\frac{1}{|\omega|^{1-\alpha}}.$$

For this function the integral

$$\int_{\mathbb{R}} |g(\omega)|^2 |\omega|^{1-\alpha} \,\mathrm{d}\omega = \int_{\mathbb{R}} |\omega|^{1+\alpha} \mathrm{sinc}^{2(m+1)}\left(\frac{\omega}{2\pi}\right) \mathrm{d}\omega = \int_{\mathbb{R}} \frac{\sin^{2(m+1)}(\omega/2\pi)}{|\omega|^{2m+1-\alpha}} \,\mathrm{d}\omega$$

is finite if $2m > \alpha$. Therefore the function f defined in Equation (4.4) is in $H_\alpha(\mathbb{R})$ for any $\alpha \in (0, 2)$. This function is

$$\begin{aligned} f(x) &= \int_{\mathbb{R}} \left[e^{i\omega x} \mathrm{sinc}^{m+1}\left(\frac{\omega}{2\pi}\right) - \mathrm{sinc}^{m+1}\left(\frac{\omega}{2\pi}\right)\right] \mathrm{d}\omega \\ &= \int_{\mathbb{R}} [\cos(x\omega) + i \sin(x\omega)] \,\mathrm{sinc}^{m+1}\left(\frac{\omega}{2\pi}\right) \mathrm{d}\omega - \text{constant} \\ &= \int_{\mathbb{R}} \cos(x\omega) \,\mathrm{sinc}^{m+1}\left(\frac{\omega}{2\pi}\right) \mathrm{d}\omega - \text{constant}, \end{aligned}$$

where the integral is the cosine transform of the $(m+1)$th power of sinc. By Equation (16) in Section 1.6 of [7], this cosine transform equals a polynomial of degree m on some non-empty interval $[0, \delta]$. By applying a suitable scaling we obtain a function in $H_\alpha(\mathbb{R})$ that is a polynomial of degree m on $[0, 1]$ with a leading coefficient one. This proves the claim because the RKHS of K_α on $[0, 1]$ consists of restrictions onto $[0, 1]$ of functions in $H_\alpha(\mathbb{R})$ [3, p. 25]. □

Proof of Proposition 2.2 for energy-distance kernels with $\alpha \in [1, 2)$. Assumption (b) holds by setting $m = 2$ in Lemma 4.2. For $\alpha = 1$ and $x, y \geq 0$ the energy-distance kernel becomes

$$K_1(x, y) = |x| + |y| - |x - y| = \min\{x, y\}.$$

In the proof of Theorem 2.1 we noted that the RKHS of $K(x, y) = 1 + \min\{x, y\}$ on $[0, 1]$ is norm-equivalent to $W^{1,2}([0, 1])$. The RKHS of K consists of sums of constant functions with elements of the RKHS of K_1 [3, p. 24]. From the characterisation of $H_\alpha(\mathbb{R})$ in [2] that we used in the proof of Lemma 4.2 it also follows that $H_\alpha(\mathbb{R}) \subseteq H_\gamma(\mathbb{R})$ if $\alpha \geq \gamma$. This inclusion is inherited by RKHSs on $[0, 1]$. Therefore K_α satisfies assumption (a) for $\alpha \in [1, 2)$. □

Acknowledgements

TK was supported by the Research Council of Finland projects 338567 ("Scalable, adaptive and reliable probabilistic integration"), 359183 ("Flagship of Advanced Mathematics for Sensing, Imaging and Modelling"), and 368086 ("Inference and approximation under misspecification").

References

[1] R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer Monographs in Mathematics. Springer, 2007.

[2] R. J. Barton and H. V. Poor. Signal detection in fractional Gaussian noise. *IEEE Trans. Autom. Control*, 34(5):943–959, 1988.

[3] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics.* Springer, 2004.

[4] C. Cobeli and A. Zaharescu. The Haros-Farey sequence at two hundred years. *Acta Univ. Apulensis. Math. - Inf.*, 5:1–38, 2003.

[5] J. Dick and F. Pillichshammer. Discrepancy theory and quasi-Monte Carlo integration. In *A Panorama of Discrepancy Theory*, pages 539–619. Springer Cham, 2014.

[6] F. Dress. Discrépance des suites de Farey. *J. Théor. Nr. Bordx*, 11(2):345–367, 1999.

[7] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi. *Tables of Integral Transforms*, volume 1. McGraw-Hill, 1954.

[8] J. Franel. Les suites de Farey et le problème des nombres premiers. *Nachr. Ges. Wiss. Göttingen*, pages 198–201, 1924.

[9] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, 2012.

[10] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Oxford University Press, 4th edition, 1960.

[11] S. Kanemitsu and M. Yoshimoto. Euler products, Farey series and the Riemann hypothesis. *Publ. Math. Debrecen*, 56(3–4):431–449, 2000.

[12] M. Mikolás. Farey series and their connection with the prime number problem. I. *Acta Sci. Math. (Szeged)*, 13:93–117, 1949.

[13] M. Mikolás. Farey series and their connection with the prime number problem. II. *Acta Sci. Math. (Szeged)*, 14:5–21, 1951.

[14] M. Mikolás and K. Sato. On the asymptotic behaviour of Franel's sum and the Riemann hypothesis. *Results Math.*, 21:368–378, 1992.

[15] H. Q. Minh. Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constr. Approx.*, 32(2):307–338, 2010.

[16] H. Niederreiter. The distribution of Farey points. *Math. Ann.*, 201:341–345, 1973.

[17] E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems. Volume II: Standard Information for Functionals.* Number 12 in EMS Tracts in Mathematics. European Mathematical Society, 2010.

[18] V. I. Paulsen and M. Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces.* Number 152 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016.

[19] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Stat.*, 41(5):2263–2291, 2013.

[20] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer Series in Statistics. Springer, 1999.

[21] A. W. van der Vaart and J. H. van Zanten. Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, pages 200–222. Institute of Mathematical Statistics, 2008.

[22] T. T. Warnock. Computational investigations of low-discrepancy point sets. In *Applications of Number Theory to Numerical Analysis*, pages 319–343. Elsevier, 1972.

[23] H. Wendland. *Scattered Data Approximation.* Number 17 in Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005.

[24] M. Yoshimoto. Farey series and the Riemann hypothesis. II. *Acta Math. Hungar.*, 78(4):287–304, 1998.

[25] B. Zwicknagl and R. Schaback. Interpolation and approximation in Taylor spaces. *J. Approx. Theory*, 171:65–83, 2013.