Generalized Bayesian Quadrature with Spectral Kernels - Supplementary Material

Houston Warren[1] **Rafael Oliveira**[2] **Fabio Ramos**[1, 3]

[1]School of Computer Science, The University of Sydney, Australia
[2]Brain and Mind Centre, The University of Sydney, Australia
[3]NVIDIA, USA

A GENERALIZED BAYESIAN QUADRATURE DERIVATIONS

A.1 PRELIMINARIES

We start by deriving the general solution to indefinite integrals of the form:

$$\begin{aligned}\langle f\rangle &= \int \frac{1}{R}\sum_{r=1}^{R}\cos(\boldsymbol{\omega}_r^T(\boldsymbol{x}-\boldsymbol{X}))\,d\boldsymbol{x}\\ &\stackrel{\dagger}{=} \frac{1}{R}\sum_{r=1}^{R}\int \cos(\boldsymbol{\omega}_r^T(\boldsymbol{x}-\boldsymbol{X}))\,d\boldsymbol{x}\end{aligned} \tag{1}$$

where $\boldsymbol{\omega} \in \mathcal{R}^{R\times d}$, and $\dagger$ makes use of the fact that $\int x + x\,dx = \int x\,dx + \int x\,dx$. Equation 1 represents the integral of an RFF estimated kernel. Using u-substitution to integrate out a single x^j variable from the vector-valued $\boldsymbol{x}$ results in:

$$\langle f\rangle_{x^j} = \frac{1}{R}\sum_{r=1}^{R}\int_{\boldsymbol{x}^i\in\mathcal{R}^{d-1}} \frac{\sin(\boldsymbol{\omega}_r^T(\boldsymbol{x}-\boldsymbol{X}))}{\omega_r^j}\,d\boldsymbol{x}^i. \tag{2}$$

where

$$\boldsymbol{x}^i := \begin{bmatrix} x^1\\ \vdots \\ x^{j-1}\\ x^{j+1}\\ \vdots\\ x^d\end{bmatrix} \tag{3}$$

If we integrate (2) again over a new variable x^t in $\boldsymbol{x}^i$, we start to see a general pattern emerge:

$$\langle f\rangle_{x^j x^t} = \frac{1}{R}\sum_{r=1}^{R}\int_{\boldsymbol{x}^{i'}\in\mathcal{R}^{d-2}} \frac{-\cos(\boldsymbol{\omega}_r^T(\boldsymbol{x}-\boldsymbol{X}))}{\omega_r^j\omega_r^t}\,d\boldsymbol{x}^{i'}. \tag{4}$$

The repeated integral of the $\cos$ follows a repeating pattern through $h = [\sin, -\cos, -\sin, \cos, \sin, \dots]$, while the integral of u-substituted $\boldsymbol{\omega}^T(\boldsymbol{x}-\boldsymbol{X})$ simply results in the multiplication of the integrand denominator by w_r^j for each variable x^j in $\boldsymbol{x}$ we integrate over. Thus, integrating over the entirety all d dimensions of $\boldsymbol{x}$ will result in an indefinite integral of the form :

$$\langle f\rangle = \frac{1}{R}\sum_{r=1}^{R}\frac{h^d(\boldsymbol{\omega}_r^T(\boldsymbol{x}-\boldsymbol{X}))}{\prod_{j=1}^{d}\omega_r^j} \tag{5}$$

where h is defined as the repeating series above and h^d represents the d-th index of h. For an RFF kernel parametrized by $\boldsymbol{\omega}$, equation (5) represents the indefinite uniform expectation; in other words, the uniform kernel mean.

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

A.2 RFF KERNEL MEANS OVER GAUSSIAN MEASURES

Equipped with the knowledge of the anti-derivative pattern that arises in (5), we now turn our attention to indefinite integrals of the form:

$$
\begin{aligned}
\langle f \rangle &= \int \frac{1}{R} \sum_{r=1}^{R} \cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X})) p(\boldsymbol{x}) \, d\boldsymbol{x} \\
&= \frac{1}{R} \sum_{r=1}^{R} \int \cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X})) p(\boldsymbol{x}) \, d\boldsymbol{x} \,,
\end{aligned}
\tag{6}
$$

where $p(\boldsymbol{x})$ is a Gaussian. Equation 6 represents the RFF kernel approximation expectation over a Gaussian distribution, or the Gaussian kernel mean $\mu_{\boldsymbol{x}}(\boldsymbol{X})$.

As in the main paper, we parametrize $p(\boldsymbol{x})$ as an RFF approximation $q(\boldsymbol{x})$ to the multivariate Gaussian, and can rewrite (6) and substitute into the BQ integral mean formulation $\langle \bar{f} \rangle = \mu_{\boldsymbol{x}}(\boldsymbol{X})^T \boldsymbol{K}^{-1} \boldsymbol{y}$ resulting in:

$$
\begin{aligned}
\langle \bar{f} \rangle &= \boldsymbol{y}^T \boldsymbol{K}^{-1} \int \frac{1}{R} \sum_{r=1}^{R} \cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X})) \times \frac{1}{Z|(2\pi)^d \boldsymbol{\Sigma}|^{1/2}} \sum_{z=1}^{Z} \cos(\boldsymbol{\rho}_z^T (\boldsymbol{x} - \boldsymbol{\mu})) d\boldsymbol{x} \\
&= \frac{\boldsymbol{y}^T \boldsymbol{K}^{-1}}{RZ|(2\pi)^d \boldsymbol{\Sigma}|^{1/2}} \sum_{r=1}^{R} \sum_{z=1}^{Z} \int \cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X})) \cos(\boldsymbol{\rho}_z^T (\boldsymbol{x} - \boldsymbol{\mu})) d\boldsymbol{x} \,.
\end{aligned}
\tag{7}
$$

Looking at the integrand term in (7), we can apply the trigonometric identity $\cos(\alpha)\cos(\beta) = \cos(\alpha+\beta)/2 + \cos(\alpha-\beta)/2$ and rewrite the integrand as:

$$
\begin{aligned}
\cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X})) \cos(\boldsymbol{\rho}_z^T (\boldsymbol{x} - \boldsymbol{\mu})) d\boldsymbol{x} = &\frac{\cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X}) + \boldsymbol{\rho}_z^T (\boldsymbol{x} - \boldsymbol{\mu}))}{2} \\
&+ \frac{\cos(\boldsymbol{\omega}_r^T (\boldsymbol{x} - \boldsymbol{X}) - \boldsymbol{\rho}_z^T (\boldsymbol{x} - \boldsymbol{\mu}))}{2},
\end{aligned}
\tag{8}
$$

which we can reorganize, while also moving the division by two outside the integral in (7), as:

$$
\cos(\boldsymbol{x}^T (\boldsymbol{\omega}_r + \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} + \boldsymbol{\rho}_z^T \boldsymbol{\mu})) + \cos(\boldsymbol{x}^T (\boldsymbol{\omega}_r - \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} - \boldsymbol{\rho}_z^T \boldsymbol{\mu})).
\tag{9}
$$

Using the same method of u-substitution and properties of the anti-derivatives of $\cos$ and $\sin$ as in (2) and (5), applying the anti-derivative over the integrand term in (9) yields the indefinite form:

$$
\frac{h^d(\boldsymbol{x}^T (\boldsymbol{\omega}_r + \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} + \boldsymbol{\rho}_z^T \boldsymbol{\mu}))}{\prod_{j=1}^{d} (\omega_r^j + \rho_z^j)} + \frac{h^d(\boldsymbol{x}^T (\boldsymbol{\omega}_r - \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} - \boldsymbol{\rho}_z^T \boldsymbol{\mu}))}{\prod_{j=1}^{d} (\omega_r^j - \rho_z^j)} + c \,,
\tag{10}
$$

where c is a constant of integration. Substituting the above into the full GBQ formulation in (7) and applying over definite bounds $\boldsymbol{a} \le \boldsymbol{x} \le \boldsymbol{b}$ results in:

$$
\frac{\boldsymbol{y}^T \boldsymbol{K}^{-1}}{Q_{\boldsymbol{a}}^{\boldsymbol{b}} \times RZ|(2\pi)^d \boldsymbol{\Sigma}|^{1/2}} \left[\sum_{r=1}^{R} \sum_{z=1}^{Z} \frac{h^d(\boldsymbol{x}^T (\boldsymbol{\omega}_r + \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} + \boldsymbol{\rho}_z^T \boldsymbol{\mu}))}{\prod_{j=1}^{d} (\omega_r^j + \rho_z^j)} + \frac{h^d(\boldsymbol{x}^T (\boldsymbol{\omega}_r - \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} - \boldsymbol{\rho}_z^T \boldsymbol{\mu}))}{\prod_{j=1}^{d} (\omega_r^j - \rho_z^j)} \right]_{\boldsymbol{a}}^{\boldsymbol{b}},
\tag{11}
$$

which we can recognize as the same equation as definition 1 in the main paper: the GBQ mean $\langle \bar{f} \rangle$ of the approximation to the ingtegral of f over a Gaussian measure.

We note that in (11) we have also added $[Q_{\boldsymbol{a}}^{\boldsymbol{b}}]^{-1}$ to the formulation. This is due to the fact that integrating over RFF approximated Gaussian measure $q(\boldsymbol{x})$ on bounds $[\boldsymbol{a}, \boldsymbol{b}]$ will necessarily truncate $q(\boldsymbol{x})$ such that it is no longer a proper probability density function. To account for this, we can modify $q(\boldsymbol{x})$ to the truncated normal form $\frac{q(\boldsymbol{x})}{Q(\boldsymbol{b}) - Q(\boldsymbol{a})}$ in equation 7 where Q is the CDF of q. We can calculate Q analytically as described in section A.6.

A.2.1 Algorithm for Efficient Implementation over Definite Bounds

Before progressing further, we will briefly describe a method by which (11) can be applied over bounds $|_{\boldsymbol{a}}^{\boldsymbol{b}}$. If done naively, the application of the fundamental law of calculus to calculate the definite integral of (11) over multidimensional bounds will add a multiplicative factor of 2^d to the GBQ complexity, as all possible vector permutations of limits $\{\boldsymbol{a}, \boldsymbol{b}\}$ in each dimension would need to be evaluated. However, we will briefly describe here a method that allows for the application of GBQ to bounded integrals which shares the same complexity ($\mathcal{O}(dNRZ)$) of calculating the indefinite integral (11) at a single point.

First, we note that h^d in this case can be any of $[\cos, \sin, -\cos, -\sin]$. Our notation moving forward from hereon will assume $h^d = \cos$ but the transformation of of other cases to cosine form is trivial using the identity $\sin(x) = \cos(x - \frac{\pi}{2})$, the effect of which will not change the methodology.

Next, we will use the identity $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$ to further simplify the terms within the double summation in (11) to isolate those not involving $\boldsymbol{x}$ (including the $-\frac{\pi}{2}$ if the h^d transformation was necessary), resulting in:

$$\begin{aligned}&\Big[\prod_{j=1}^{d}(\omega_r^j + \rho_z^j)\Big]^{-1}\Big[\cos(\boldsymbol{\omega}_r^T\boldsymbol{X} + \boldsymbol{\rho}_z^T\boldsymbol{\mu})\cos(\boldsymbol{x}^T(\boldsymbol{\omega}_r + \boldsymbol{\rho}_z)) + \sin(\boldsymbol{\omega}_r^T\boldsymbol{X} + \boldsymbol{\rho}_z^T\boldsymbol{\mu})\sin(\boldsymbol{x}^T(\boldsymbol{\omega}_r + \boldsymbol{\rho}_z))\Big] \\ +&\Big[\prod_{j=1}^{d}(\omega_r^j - \rho_z^j)\Big]^{-1}\Big[\cos(\boldsymbol{\omega}_r^T\boldsymbol{X} - \boldsymbol{\rho}_z^T\boldsymbol{\mu})\cos(\boldsymbol{x}^T(\boldsymbol{\omega}_r - \boldsymbol{\rho}_z)) + \sin(\boldsymbol{\omega}_r^T\boldsymbol{X} - \boldsymbol{\rho}_z^T\boldsymbol{\mu})\sin(\boldsymbol{x}^T(\boldsymbol{\omega}_r - \boldsymbol{\rho}_z))\Big],\end{aligned} \tag{12}$$

Next, we will apply the harmonic addition theorem to further reduce terms.

Theorem 4 (Harmonic addition theorem)**.** *A linear combination of sinusoids which share a frequency x but have differing amplitudes $\{a, b\}$ can be reformulated as:*

$$a\cos(x) + b\sin(x) = \text{sign}(a)\sqrt{a^2 + b^2}\cos(x + \arctan(\frac{-b}{a}))\,. \tag{13}$$

Using the identities that $\cos^2(x) + \sin^2(x) = 1$ and $\frac{\sin(x)}{\cos(x)} = \tan(x)$, we can modify (12) to

$$\begin{aligned}&\Big[\prod_{j=1}^{d}(\omega_r^j + \rho_z^j)\Big]^{-1}\Big[\text{sign}\left(\cos\left(\boldsymbol{\omega}_r^T\boldsymbol{X} + \boldsymbol{\rho}_z^T\boldsymbol{\mu}\right)\right)\cos\left(\boldsymbol{x}^T(\boldsymbol{\omega}_r + \boldsymbol{\rho}_z) + \arctan\left(-\tan\left(\boldsymbol{\omega}_r^T\boldsymbol{X} + \boldsymbol{\rho}_z^T\boldsymbol{\mu}\right)\right)\right)\Big] \\ +&\Big[\prod_{j=1}^{d}(\omega_r^j - \rho_z^j)\Big]^{-1}\Big[\text{sign}\left(\cos\left(\boldsymbol{\omega}_r^T\boldsymbol{X} - \boldsymbol{\rho}_z^T\boldsymbol{\mu}\right)\right)\cos\left(\boldsymbol{x}^T(\boldsymbol{\omega}_r - \boldsymbol{\rho}_z) + \arctan\left(-\tan\left(\boldsymbol{\omega}_r^T\boldsymbol{X} - \boldsymbol{\rho}_z^T\boldsymbol{\mu}\right)\right)\right)\Big],\end{aligned} \tag{14}$$

which we will then re-parametrize with $[\alpha_{rz}, \beta_{rz}, \gamma_{rz}, \delta_{rz}]$ for terms that do not involve $\boldsymbol{x}$ and substitute back into (11) for

$$\frac{\boldsymbol{y}^T\boldsymbol{K}^{-1}}{Q_{\boldsymbol{a}}^{\boldsymbol{b}} \times RZ|(2\pi)^d\boldsymbol{\Sigma}|^{1/2}}\left[\sum_{r=1}^{R}\sum_{z=1}^{Z}\alpha_{rz} \times \cos\left(\boldsymbol{x}^T(\boldsymbol{\omega}_r + \boldsymbol{\rho}_z) + \beta_{rz}\right) + \gamma_{rz} \times \cos\left(\boldsymbol{x}^T(\boldsymbol{\omega}_r - \boldsymbol{\rho}_z) + \delta_{rz}\right)\right]_{\boldsymbol{a}}^{\boldsymbol{b}}. \tag{15}$$

Given that we can evaluate each dimension of the interval of integration independently ie.

$$f(\boldsymbol{x})\big|_{\boldsymbol{a}}^{\boldsymbol{b}} = f(x^1, \cdot)\big|_{a^1}^{b^1} \cdots f(x^d, \cdot)\big|_{a^d}^{b^d}\,, \tag{16}$$

we can apply the integration bounds iteratively to (15). By leveraging trigonometric identities, we can reduce the resulting terms at each step to provide computational advantages.

Consider the case of evaluating (15) over the bounds of a single dimension $d = 1$. Leveraging the identity $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$, we can the separate the terms within the double summation in (15) that involve $x^{d=1}$ from those that do not, resulting in

$$\begin{aligned}\alpha_{rz}\Big[&\cos([x^1]^T(\omega_r^1 + \rho_z^1) + \beta_{rz})\cos([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1} + \boldsymbol{\rho}_z^{d\neq1})) \\ &- \sin([x^1]^T(\omega_r^1 + \rho_z^1) + \beta_{rz})\sin([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1} + \boldsymbol{\rho}_z^{d\neq1}))\Big] \\ + \gamma_{rz}\Big[&\cos([x^1]^T(\omega_r^1 - \rho_z^1) + \delta_{rz})\cos([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1} - \boldsymbol{\rho}_z^{d\neq1})) \\ &- \sin([x^1]^T(\omega_r^1 - \rho_z^1) + \delta_{rz})\sin([\boldsymbol{x}_{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1} - \boldsymbol{\rho}_z^{d\neq1}))\Big]\,.\end{aligned} \tag{17}$$

If we now substitute a^1 and b^1 for x^1 as we apply over a single bound (ie. $f(x)|_a^b = f(b) - f(a)$), the result (within the double summation of (15)) is:

$$\begin{aligned}
&\alpha_{rz}\Big[\cos([b^1]^T(\omega_r^1+\rho_z^1)+\beta_{rz}) - \cos([a^1]^T(\omega_r^1+\rho_z^1)+\beta_{rz})\Big]\cos([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1}+\boldsymbol{\rho}_z^{d\neq1}))\\
&-\alpha_{rz}\Big[\sin([b^1]^T(\omega_r^1+\rho_z^1)+\beta_{rz}) - \sin([a^1]^T(\omega_r^1+\rho_z^1)+\beta_{rz})\Big]\sin([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1}+\boldsymbol{\rho}_z^{d\neq1}))\\
&+\gamma_{rz}\Big[\cos([b^1]^T(\omega_r^1-\rho_z^1)+\delta_{rz}) - \cos([a^1]^T(\omega_r^1-\rho_z^1)+\delta_{rz})\Big]\cos([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1}-\boldsymbol{\rho}_z^{d\neq1}))\\
&-\gamma_{rz}\Big[\sin([b^1]^T(\omega_r^1-\rho_z^1)+\delta_{rz}) - \sin([a^1]^T(\omega_r^1-\rho_z^1)+\delta_{rz})\Big]\sin([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1}-\boldsymbol{\rho}_z^{d\neq1}))\,.
\end{aligned} \tag{18}$$

We again have a form that can leverage the harmonic addition theorem 4 to collect the constant terms that do not involve $\boldsymbol{x}^{d\neq1}$. Applying the theorem and reparametrizing results in

$$\begin{aligned}
&\frac{\boldsymbol{y}^T\boldsymbol{K}^{-1}}{Q_{\boldsymbol{a}}^{\boldsymbol{b}}\times RZ|(2\pi)^d\boldsymbol{\Sigma}|^{1/2}}\\
&\quad\times\left[\sum_{r=1}^{R}\sum_{z=1}^{Z}\alpha_{rz}^*\times\cos\left([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1}+\boldsymbol{\rho}_z^{d\neq1})+\beta_{rz}^*\right)+\gamma_{rz}^*\times\cos\left([\boldsymbol{x}^{d\neq1}]^T(\boldsymbol{\omega}_r^{d\neq1}-\boldsymbol{\rho}_z^{d\neq1})+\delta_{rz}^*\right)\right]_{\boldsymbol{a}^{d\neq1}}^{\boldsymbol{b}^{d\neq1}},
\end{aligned} \tag{19}$$

where

$$\alpha_{rz}^* = \alpha_{rz}\times\operatorname{sign}\left(\cos([b^1]^T(\omega_r^1+\rho_z^1)+\beta_{rz}) - \cos([a^1]^T(\omega_r^1+\rho_z^1)+\beta_{rz})\right)\times\sqrt{2-2\cos((b^1-a^1)(\omega_r^1+\rho_z^1))} \tag{20}$$

$$\beta_{rz}^* = \arctan\left(-\cot\left(\frac{(b^1+a^1)(\omega_r^1+\rho_z^1)+2\beta_{rz}}{2}\right)\right) \tag{21}$$

$$\gamma_{rz}^* = \gamma_{rz}\times\operatorname{sign}\left(\cos([b^1]^T(\omega_r^1-\rho_z^1)+\delta_{rz}) - \cos([a^1]^T(\omega_r^1-\rho_z^1)+\delta_{rz})\right)\times\sqrt{2-2\cos((b^1-a^1)(\omega_r^1-\rho_z^1))} \tag{22}$$

$$\delta^* = \arctan\left(-\cot\left(\frac{(b^1+a^1)(\omega_r^1-\rho_z^1)+2\delta_{rz}}{2}\right)\right)\,. \tag{23}$$

We can see that (19) is of the same form as (15), but with updated parameters $[\alpha_{rz}^*, \beta_{rz}^*, \gamma_{rz}^*, \delta_{rz}^*]$ that rely only on the constant values $[a^1, b^1, \omega_r^1, \rho_z^1, \alpha_{rz}, \beta_{rz}, \gamma_{rz}, \delta_{rz}]$. Starting at (19) and repeating the steps in 15 through 19 for all $d-1$ remaining dimensions will result in the full evaluation of the GBQ integral estimate over bounds $[\boldsymbol{a}, \boldsymbol{b}]$. Complexity is further discussed in B.

A.3 GBQ VARIANCE OVER GAUSSIAN MEASURES

We begin with by stating the BQ formulation of the variance of the mean integration approximation $\langle\bar{f}\rangle$:

$$\mathbb{V}(\langle\bar{f}\rangle) = \int\int k(\boldsymbol{x}, X)\,p(\boldsymbol{x})p(\boldsymbol{X})\,d\boldsymbol{x}d\boldsymbol{X}\,. \tag{24}$$

To find the variance of $\langle\bar{f}\rangle$ over a Gaussian measure, we substitute the expectation of the Gaussian kernel mean, which we calculate in A.2, into (24) for $\int k(\boldsymbol{x}, X)\,p(\boldsymbol{x})\,d\boldsymbol{x}$. To do so, we first rewrite the Gaussian kernel mean approximation as

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{x}}(\boldsymbol{X}) &= [2RZ[(2\pi)^d|\boldsymbol{\Sigma}|]^{1/2}]^{-1}\sum_{r=1}^{R}\sum_{z=1}^{Z}\left[\frac{h^d(\boldsymbol{x}^T(\boldsymbol{\omega}_r+\boldsymbol{\rho}_z)-(\boldsymbol{\omega}_r^T\boldsymbol{X}+\boldsymbol{\rho}_z^T\boldsymbol{\mu}))}{\prod_{j=1}^d(\omega_r^j+\rho_z^j)}\right.\\
&\qquad\left.+\frac{h^d(\boldsymbol{x}^T(\boldsymbol{\omega}_r-\boldsymbol{\rho}_z)-(\boldsymbol{\omega}_r^T\boldsymbol{X}-\boldsymbol{\rho}_z^T\boldsymbol{\mu}))}{\prod_{j=1}^d(\omega_r^j-\rho_z^j)}\right]\\
&= \frac{1}{2RZ[(2\pi)^d|\boldsymbol{\Sigma}|]^{1/2}}\sum_{r=1}^{R}\sum_{z=1}^{Z}\left[\frac{h^d(\tau-\boldsymbol{\omega}_r^T\boldsymbol{X})}{\prod_{j=1}^d(\omega_r^j+\rho_z^j)}+\frac{h^d(\nu-\boldsymbol{\omega}_r^T\boldsymbol{X})}{\prod_{j=1}^d(\omega_r^j-\rho_z^j)}\right]
\end{aligned} \tag{25}$$

Where we have combined the terms inside the h^ds not involving $\boldsymbol{X}$ with τ and ν.

We next substitute (25) into (24), while also introducing the RFF estimate $q(\boldsymbol{X})$ to $p(\boldsymbol{X})$, to obtain the integral variance:

$$
\begin{aligned}
\mathbb{V}(\langle \bar{f} \rangle) &= \int \frac{\cos(\boldsymbol{\rho}_u^T(\boldsymbol{X}-\boldsymbol{\mu}))}{RZU[(2\pi)^d|\boldsymbol{\Sigma}|]} \sum_{r=1}^{R}\sum_{z=1}^{Z}\sum_{u=1}^{U}\left[\frac{h^d(\tau-\boldsymbol{\omega}_r^T\boldsymbol{X})}{2\prod_{j=1}^d(\omega_r^j+\rho_z^j)}+\frac{h^d(\nu-\boldsymbol{\omega}_r^T\boldsymbol{X})}{2\prod_{j=1}^d(\omega_r^j-\rho_z^j)}\right] d\boldsymbol{X} \\
&= L'\sum_{r=1}^{R}\sum_{z=1}^{Z}\sum_{u=1}^{U}\int \cos(\boldsymbol{\rho}_u^T(\boldsymbol{X}-\boldsymbol{\mu}))\left[\frac{h^d(\tau-\boldsymbol{\omega}_r^T\boldsymbol{X})}{2\prod_{j=1}^d(\omega_r^j+\rho_z^j)}+\frac{h^d(\nu-\boldsymbol{\omega}_r^T\boldsymbol{X})}{2\prod_{j=1}^d(\omega_r^j-\rho_z^j)}\right] d\boldsymbol{X}\,,
\end{aligned}
\tag{26}
$$

where have introduced another index of $\boldsymbol{\rho}$ in $u = 1, \ldots, U$, and substituted L' for $[RZU(2\pi)^d|\boldsymbol{\Sigma}|]^{-1}$

We note here that h^d could be any of $[\cos, \sin, -\cos, -\sin]$, so we cannot necessarily leverage the same identity we used previously on the products of cosines. However, the simple identity $sin(x) = cos(x - \frac{\pi}{2})$ can rectify this case. We will continue with the variance proof under the assumption that $h^d = \cos$, but it is straightforward to derive the variance alternative cases.

Using the identity $\cos(\alpha)\cos(\beta) = \cos(\alpha+\beta)/2 + \cos(\alpha-\beta)/2$, we simplify the integrand in (26) to:

$$
\begin{aligned}
&\frac{\cos(\boldsymbol{\rho}_u^T(\boldsymbol{X}-\boldsymbol{\mu})+\tau-\boldsymbol{\omega}_r^T\boldsymbol{X})}{4\prod_{j=1}^d(\omega_r^j+\rho_z^j)}+\frac{\cos(\boldsymbol{\rho}_u^T(\boldsymbol{X}-\boldsymbol{\mu})-\tau+\boldsymbol{\omega}_r^T\boldsymbol{X})}{4\prod_{j=1}^d(\omega_r^j+\rho_z^j)} \\
+&\frac{\cos(\boldsymbol{\rho}_u^T(\boldsymbol{X}-\boldsymbol{\mu})+\nu-\boldsymbol{\omega}_r^T\boldsymbol{X})}{4\prod_{j=1}^d(\omega_r^j-\rho_z^j)}+\frac{\cos(\boldsymbol{\rho}_u^T(\boldsymbol{X}-\boldsymbol{\mu})-\nu+\boldsymbol{\omega}_r^T\boldsymbol{X})}{4\prod_{j=1}^d(\omega_r^j-\rho_z^j)}
\end{aligned}
\tag{27}
$$

and simplify further to

$$
\begin{aligned}
&\frac{\cos(\boldsymbol{X}^T(\boldsymbol{\rho}_u-\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}+\tau)}{4\prod_{j=1}^d(\omega_r^j+\rho_z^j)}+\frac{\cos(\boldsymbol{X}^T(\boldsymbol{\rho}_u+\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}-\tau)}{4\prod_{j=1}^d(\omega_r^j+\rho_z^j)} \\
+&\frac{\cos(\boldsymbol{X}^T(\boldsymbol{\rho}_u-\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}+\nu)}{4\prod_{j=1}^d(\omega_r^j-\rho_z^j)}+\frac{\cos(\boldsymbol{X}^T(\boldsymbol{\rho}_u+\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}-\nu))}{4\prod_{j=1}^d(\omega_r^j-\rho_z^j)}\,.
\end{aligned}
\tag{28}
$$

Lastly, we use the anti-derivative methods described in A.1 to calculate the anti-derivative of this integrand, which represents the indefinite GBQ variance over a Gaussian measure:

$$
\begin{aligned}
\mathbb{V}(\langle \bar{f} \rangle) = L'\sum_{r=1}^{R}\sum_{z=1}^{Z}\sum_{u=1}^{U}\Bigg[&\frac{h^{2d}(\boldsymbol{X}^T(\boldsymbol{\rho}_u-\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}+\tau)}{4\prod_{j=1}^d(\omega_r^j+\rho_z^j)(\rho_u^j-\omega_r^j)}+\frac{h^{2d}(\boldsymbol{X}^T(\boldsymbol{\rho}_u+\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}-\tau)}{4\prod_{j=1}^d(\omega_r^j+\rho_z^j)(\rho_u^j+\omega_r^j)} \\
+&\frac{h^{2d}(\boldsymbol{X}^T(\boldsymbol{\rho}_u-\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}+\nu)}{4\prod_{j=1}^d(\omega_r^j-\rho_z^j)(\rho_u^j-\omega_r^j)}+\frac{h^{2d}(\boldsymbol{X}^T(\boldsymbol{\rho}_u+\boldsymbol{\omega}_r)-\boldsymbol{\rho}_u^T\boldsymbol{\mu}-\nu))}{4\prod_{j=1}^d(\omega_r^j-\rho_z^j)(\rho_u^j+\omega_r^j)}\Bigg]
\end{aligned}
\tag{29}
$$

where h^{2d} is defined as the $2d$-th index of function vector h as previously defined in A.1. As we again have an indefinite anti-derivative form that is a linear combination of trigonometric functions evaluated over the dot products $\boldsymbol{x}$ and $\boldsymbol{X}$ with their measure RFF frequencies $\boldsymbol{\rho}$, we can use a variant of the method outlined in A.2.1 to evaluate over definite bounds $[\boldsymbol{a}, \boldsymbol{b}]$ for both $\boldsymbol{x}$ and $\boldsymbol{X}$. We note that in this case it is again necessary to modify L' in (29) with the necessary truncation term $(Q_{\boldsymbol{a}}^{\boldsymbol{b}})^2$, as in (15) (where $Q_{\boldsymbol{a}}^{\boldsymbol{b}}$ is squared in this case due to integration over truncated forms of both $q(\boldsymbol{x})$ and $q(\boldsymbol{X})$).

A.4 GBQ INTEGRAL MEAN OVER UNIFORM MEASURES

We showed in A.1 that (5) represents the RFF kernel mean over a uniform measure. Given this, it is straightforward to derive GBQ over a uniform measure. We simply need to evaluate the definite form of (5) and then substitute it for $\mu_{\boldsymbol{x}}(\boldsymbol{X})$ in $\langle \bar{f} \rangle = \mu_{\boldsymbol{x}}(\boldsymbol{X})^T\boldsymbol{K}^{-1}\boldsymbol{y}$.

The GBQ integral mean posterior in full is

$$
\langle \bar{f} \rangle = \frac{\boldsymbol{y}^T\boldsymbol{K}^{-1}}{R}\sum_{r=1}^{R}\frac{h^d(\boldsymbol{\omega}_r^T(\boldsymbol{x}-\boldsymbol{X}))}{\prod_{j=1}^d\omega_r^j}\Bigg|_{\boldsymbol{a}}^{\boldsymbol{b}}\,.
\tag{30}
$$

We note that GBQ over the uniform measure is equivalent to direct analytical integration of the RFF-parametrized GP integrand $\bar{f}$.

A.5 GBQ INTEGRAL VARIANCE OVER UNIFORM MEASURES

To calculate the variance, substitute (30) for $\int k(\boldsymbol{x}, X)\, p(\boldsymbol{x})\, d\boldsymbol{x}$ in (24),

$$\begin{aligned} \mathbb{V}(\langle \bar{f} \rangle) &= \int \frac{1}{R} \sum_{r=1}^{R} \frac{h^d(\boldsymbol{\omega}_r^T(\boldsymbol{x} - \boldsymbol{X}))}{\prod_{j=1}^{d} \omega_r^j} p(\boldsymbol{X}) d\boldsymbol{X} \\ &= \frac{1}{R} \sum_{r=1}^{R} \int \frac{h^d(\boldsymbol{\omega}_r^T(\boldsymbol{x} - \boldsymbol{X}))}{\prod_{j=1}^{d} \omega_r^j} d\boldsymbol{X} \end{aligned} . \tag{31}$$

Using the same techniques as those in A.1 and A.4, we can easily arrive at the anti-derivative form of this variance estimate:

$$\mathbb{V}(\langle \bar{f} \rangle) = \frac{1}{R} \sum_{r=1}^{R} \frac{-1^d h^{2d}(\boldsymbol{\omega}_r^T(\boldsymbol{x} - \boldsymbol{X}))}{\prod_{j=1}^{d} \omega_r^j \omega_r^j} \tag{32}$$

where h^{2d} is the $2d$-th index of h as defined in A.1, and the term -1^d is introduced due to the fact that $\boldsymbol{X}$ is negative in the integrand.

For both the GBQ uniform mean and variance calculations, a simplified version of the algorithm described in section A.2.1 can be used for efficient implementation.

A.6 MULTIVARIATE CDF OF THE RFF FORMULATED GAUSSIAN

Using the established methods from A.1 and A.4 on the integrals of RFF-parametrized kernels and distributions, it is trivial to show through u-substitution and trigonometric anti-derivatives that the indefinite integral an RFF-parametrized Gaussian $q(\boldsymbol{x})$ is:

$$\begin{aligned} Q(\boldsymbol{x}) &= \int_{\boldsymbol{x} \in \mathcal{R}^d} \frac{1}{R[(2\pi)^d |\boldsymbol{\Sigma}|]^{1/2}} \sum_{r=1}^{R} \cos(\boldsymbol{\rho}^T(\boldsymbol{x} - \boldsymbol{\mu}))\, d\boldsymbol{x} \\ &= \frac{1}{R[(2\pi)^d |\boldsymbol{\Sigma}|]1/2} \sum_{r=1}^{R} \int_{\boldsymbol{x} \in \mathcal{R}^d} \cos(\boldsymbol{\rho}^T(\boldsymbol{x} - \boldsymbol{\mu}))\, d\boldsymbol{x} \\ &= \frac{h^d(\boldsymbol{\rho}^T(\boldsymbol{x} - \boldsymbol{\mu}))}{R[(2\pi)^d |\boldsymbol{\Sigma}|]^{1/2} \prod_{j=1}^{d} \rho_r^j} \end{aligned} \tag{33}$$

where h^d is defined as in A.1. A simplified form of the algorithm presented in A.2.1 can be used for the application of this indefinite integral over definite bounds $[\boldsymbol{a}, \boldsymbol{b}]$, which can then be used to estimate the CDF of a multivariate Gaussian approximation over a domain .

B COMPUTATIONAL COMPLEXITY

In this section we focus on deriving the complexity of the GBQ mean integral estimate $\langle \bar{f} \rangle$ and compare this complexity to that of traditional BQ.

B.1 GBQ OVER GAUSSIAN MEASURES

We can examine the equation of the GBQ integral estimate mean over a Gaussian measure, reproduced here, to derive the computational complexity of the mean estimation

$$\frac{\boldsymbol{y}^T \boldsymbol{K}^{-1}}{Q_{\boldsymbol{a}}^{\boldsymbol{b}} \times RZ|(2\pi)^d \boldsymbol{\Sigma}|^{1/2}} \left[\sum_{r=1}^{R} \sum_{z=1}^{Z} \frac{h^d(\boldsymbol{x}^T(\boldsymbol{\omega}_r + \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} + \boldsymbol{\rho}_z^T \boldsymbol{\mu}))}{\prod_{j=1}^{d} (\omega_r^j + \rho_z^j)} + \frac{h^d(\boldsymbol{x}^T(\boldsymbol{\omega}_r - \boldsymbol{\rho}_z) - (\boldsymbol{\omega}_r^T \boldsymbol{X} - \boldsymbol{\rho}_z^T \boldsymbol{\mu}))}{\prod_{j=1}^{d} (\omega_r^j - \rho_z^j)} \right]_{\boldsymbol{a}}^{\boldsymbol{b}} . \tag{11}$$

There are two potential terms which may dominate the complexity of GBQ: (1) the inversion of $\boldsymbol{K}^{-1}$, which is an operation shared by vanilla BQ and has a complexity of $\mathcal{O}(N^3)$; or (2) The kernel mean calculation inside the double summation and evaluation over bounds $[\boldsymbol{a}, \boldsymbol{b}]$.

We will focus on deriving the complexity of the bracketed double summation, which represents the RFF kernel mean $\mu_{\boldsymbol{x}}(\boldsymbol{X})$. We will first note that $\mu_{\boldsymbol{x}}(\boldsymbol{X}) = [\mu_{\boldsymbol{x}}(\boldsymbol{x}_1) \dots \mu_{\boldsymbol{x}}(\boldsymbol{x}_n)]$, which implies a baseline complexity of $\mathcal{O}(N)$ when we evaluate $\mu_{\boldsymbol{x}}(\boldsymbol{x}_i) \, \forall \, \boldsymbol{x}_i \in \boldsymbol{X}$.

Next, the double summation over R and Z implies additional multiplicative complexity of of RZ, for an aggregate complexity of $\mathcal{O}(NRZ)$. Finally, the operations within the double sum have at most complexity d, which results in a total complexity of $\mathcal{O}(dNRZ)$ in the indefinite form.

For application of the indefinite form over definite bounds as in (19), we can derive the complexity through the individual complexities of the single-dimension parameter update equations 20 through 23. A single iteration of the update equations are evaluated in $\mathcal{O}(1)$ time, but we must apply them d times for all dimensions of $\boldsymbol{x}$. In addition, the double summation over R and Z, and subsequent evaluation across all $\boldsymbol{x}_i \in \boldsymbol{X}$ results in a total complexity of $\mathcal{O}(dNRZ)$, which is the same as the evaluation of the indefinite form at a single point. Considering that a naive implementation of the indefinite integral over multidimensional bounds results in a multiplicative increase to the indefinite complexity of 2^d, the algorithm presented in A.2.1 represents a significant performance incentive.

B.2 GBQ OVER UNIFORM MEASURES

The complexity of GBQ over the uniform measure follows a very similar derivation to that of GBQ over the Gaussian measure. The complexity can be alternatively dominated by the inversion of the kernel matrix $\boldsymbol{K}$ or calculation of the kernel mean $\mu_{\boldsymbol{x}}(\boldsymbol{X})$.

We derive here the complexity of the kernel mean over a uniform measure. By the same reasoning through which we derive the indefinite Gaussian GBQ complexity as $\mathcal{O}(dNRZ)$, and the fact that (30) only contains a single summation over one set of Fourier features $\{\boldsymbol{\omega}_r\}_{r=1}^{R}$ rather than two, we can easily derive that the indefinite form of uniform GBQ has complexity $\mathcal{O}(dNR)$.

Similarly, we can use the implementation in A.2.1 when applying (30) over multidimensional bounds. As we have previously derived that the method in A.2.1 results in the same complexity as evaluation of the indefinite anti-derivative at a single point, we can similarly reason that GBQ over a uniform measure and multidimensional bounds has complexity $\mathcal{O}(dNR)$.

B.3 COMPARISON OF BQ AND GBQ COMPLEXITY

Traditional BQ scales in $\mathcal{O}(N^3)$ due to the necessary operation $\boldsymbol{K}^{-1}$, and from the previous sections we see that GBQ scales in $\mathcal{O}(N^3)$ or $\mathcal{O}(dNR)$ (uniform) / $\mathcal{O}(dNRZ)$ (Gaussian).

Eliminating common terms in $\mathcal{O}(N^3)$, $\mathcal{O}(dNR)$, and $\mathcal{O}(dNRZ)$ allows us to see that when $dR < N^2$, uniform GBQ shares the same complexity as traditional BQ in $\mathcal{O}(N^3)$. The same statement applies for Gaussian GBQ when $dRZ < N^2$.

As the number of RFF parameters R and Z are traditionally kept well below N in practice, and BQ is generally used in $d \leq 10$, these are very reasonable conditions under which, at medium-size N, BQ and GBQ share the same computational complexity for evaluation of the mean of the approximated integral, $\langle \bar{f} \rangle$.

C PROOFS FOR THE THEORETICAL RESULTS

C.1 BACKGROUND

We consider a standard GP posterior mean and variance, respectively, as:

$$\mu_n(\boldsymbol{x}) := \mathbf{k}_n(\boldsymbol{x})^\mathsf{T}(\mathbf{K}_n + \lambda\mathbf{I})^{-1}\mathbf{y}_n \tag{34}$$

$$\sigma_n^2(\boldsymbol{x}) := k(\boldsymbol{x}, \boldsymbol{x}') - \mathbf{k}_n(\boldsymbol{x})^\mathsf{T}(\mathbf{K}_n + \lambda\mathbf{I})^{-1}\mathbf{k}_n(\boldsymbol{x}) \tag{35}$$

where we use notation shortcuts for the vector $\mathbf{k}(\boldsymbol{x}) := [k(\boldsymbol{x}, \boldsymbol{x}_i)]_{i=1}^n \in \mathbb{R}^n$ and the kernel matrix $\mathbf{K} := [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1}^n \in \mathbb{R}^{n \times n}$. Correspondingly, the our method employs Fourier features to approximate a GP posterior mean as:

$$\hat{\mu}_n(\boldsymbol{x}) := \tilde{\mathbf{k}}(\boldsymbol{x})^\mathsf{T} \left(\mathbf{K}_n + \lambda\mathbf{I}\right)^{-1} \mathbf{y}\,, \tag{36}$$

where $\tilde{k} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is formally defined according to the next statement.

Definition 2. *Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ denote a translation-invariant positive-definite kernel on $\mathcal{X} \subset \mathbb{R}^d$, $d \in \mathbb{N}$. The random Fourier feature approximation is defined as:*

$$\tilde{k}(\boldsymbol{x}, \boldsymbol{x}') := \boldsymbol{\phi}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\phi}(\boldsymbol{x}'), \quad \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}\,, \tag{37}$$

where:

$$\boldsymbol{\phi}(\boldsymbol{x}) := \sqrt{\frac{1}{R}} \begin{bmatrix} \sin(\boldsymbol{\omega}_1^\mathsf{T}\boldsymbol{x}) \\ \cos(\boldsymbol{\omega}_1^\mathsf{T}\boldsymbol{x}) \\ \vdots \\ \sin(\boldsymbol{\omega}_R^\mathsf{T}\boldsymbol{x}) \\ \cos(\boldsymbol{\omega}_R^\mathsf{T}\boldsymbol{x}) \end{bmatrix}, \quad \boldsymbol{\omega}_i \overset{i.i.d.}{\sim} P_k\,, \quad \boldsymbol{x} \in \mathcal{X}\,, \tag{38}$$

with P_k denoting the probability distribution that corresponds to the Fourier transform of the kernel k. Equivalently, we can also write:

$$\tilde{k}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{R}\sum_{i=1}^{R} \cos(\boldsymbol{\omega}_i^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}')\,, \quad \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}\,. \tag{39}$$

C.2 AUXILIARY RESULTS

We will make use of guarantees for RFFs to bound the kernel approximation error. In particular, we consider the following result from Sutherland and Schneider (2015).

Lemma 3 (Sutherland and Schneider (2015, Proposition 1), full version)**.** *Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a continuous shift-invariant positive-definite kernel with $k(\boldsymbol{x}, \boldsymbol{x}) = 1$ and such that $\nabla^2 k(\boldsymbol{x}, \boldsymbol{x})$ exists, for all $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$. Suppose $\mathcal{X}$ is compact with diameter $\ell_\mathcal{X} < \infty$. Denote k's Fourier transform as P_k, which is a probability measure, and let $\sigma_k^2 := \mathbb{E}[\|\boldsymbol{\omega}\|_2^2]$ for $\omega \sim P_k$. Let $\tilde{k} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ denote k's RFF approximation with R frequencies according to Definition 2. For any $\xi > 0$, let:*

$$\alpha_\xi := \min\left(1, \sup_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} \frac{1}{2} + \frac{1}{2}k(2\boldsymbol{x}, 2\boldsymbol{x}') - k(\boldsymbol{x}, \boldsymbol{x}')^2 + \frac{1}{3}\xi\right)\,, \tag{40}$$

$$\beta_d := \left(\left(\frac{d}{2}\right)^{-\frac{d}{d+2}} + \left(\frac{d}{2}\right)^{\frac{2}{d+2}}\right) 2^{\frac{6d+2}{d+2}}\,. \tag{41}$$

Then the following holds for any $\xi > 0$:

$$\begin{aligned} \mathbb{P}\left[\sup_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}} |\tilde{k}(\boldsymbol{x}, \boldsymbol{x}') - k(\boldsymbol{x}, \boldsymbol{x}')| \geq \xi\right] &\leq \beta_d \left(\frac{\sigma_k \ell_\mathcal{X}}{\xi}\right)^{\frac{2}{1+\frac{2}{d}}} \exp\left(-\frac{R\xi^2}{4(d+2)\alpha_\xi}\right) \\ &\leq 66\left(\frac{\sigma_k \ell_\mathcal{X}}{\xi}\right)^2 \exp\left(-\frac{R\xi^2}{4(d+2)}\right)\,, \end{aligned} \tag{42}$$

where for the second statement we assume $\xi \leq \sigma_k \ell_\mathcal{X}$. Therefore, for any $\delta \in (0, 1)$, we can achieve pointwise approximation error less than ξ with probability at least $1 - \delta$ if:

$$R \geq \frac{4(d+2)\alpha_\xi}{\xi^2}\left(\frac{2}{1+\frac{2}{d}}\log\frac{\sigma_k \ell_\mathcal{X}}{\xi} + \log\frac{\beta_d}{\delta}\right)\,. \tag{43}$$

Compared to the original statement of the result in Sutherland and Schneider (2015), note that we use the number of Fourier frequencies R, instead of the dimensionality of the feature vector, i.e., $D := 2R$, so that some constants are changed. Considering the result above, as $\max_{d \in \mathbb{N}} \beta_d = 66$ (see Sutherland and Schneider, 2015) and $\alpha_\xi \leq 1$, we can also set the minimum number of features for a given error bound $\xi > 0$ and $\delta \in (0, 1)$ as:

$$R(\xi, \delta, \sigma_k) := \frac{4(d+2)}{\xi^2}\left(\frac{2}{1+\frac{2}{d}}\log\frac{\sigma_k \ell_\mathcal{X}}{\xi} + \log\frac{66}{\delta}\right)\,, \tag{44}$$

though a tighter bound is available via Equation 43. Therefore, the restatement of the result in the main paper as Lemma 2 is still valid.

The norm of the observations vector $\mathbf{y}$ in a Gaussian process can be bounded in terms of the integrand f's extremes and the number of data points, as in the following result.

Lemma 4. *Given* $\delta \in (0,1)$*, assuming i.i.d. Gaussian observation noise* $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$*, we have that:*

$$\mathbb{P}\left[\|\mathbf{y}\|_2 \leq \sqrt{n}\left(\|f\|_\infty + \sigma_\epsilon \sqrt{2 \log\left(\frac{n}{\delta}\right)}\right)\right] \geq 1 - \delta\,. \tag{45}$$

Proof. Starting from the definition of the 2-norm, we have:

$$\|\mathbf{y}\|_2^2 = \sum_{i=1}^n y_i^2 = \sum_{i=1}^n (f(\boldsymbol{x}_i) + \epsilon_i)^2 \leq n \max_{i \in \{1,\dots,n\}} (f(\boldsymbol{x}_i) + \epsilon_i)^2\,. \tag{46}$$

Assuming i.i.d. Gaussian observation noise $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, the following holds:

$$\forall \beta > 0, \quad \mathbb{P}\left[|\epsilon| \geq \beta\sigma_\epsilon\right] \leq \exp\left(-\beta^2/2\right)\,, \tag{47}$$

By applying a union bound, we have:

$$\begin{aligned}\mathbb{P}\left[\exists i \in \{1,\dots,n\} : y_i \geq f(\boldsymbol{x}_i) + \beta\sigma_\epsilon\right] &\leq \sum_{i=1}^n \mathbb{P}\left[\epsilon_i \geq \beta\sigma_\epsilon\right] \\ &\leq n\mathbb{P}\left[|\epsilon| \geq \beta\sigma_\epsilon\right] \\ &\leq n \exp\left(-\beta^2/2\right)\end{aligned} \tag{48}$$

Solving for $n \exp\left(-\beta^2/2\right) = \delta$ and taking the complement, we then obtain:

$$\mathbb{P}\left[\forall i \in \{1,\dots,n\}, \quad y_i \leq \|f\|_\infty + \sigma_\epsilon \sqrt{2 \log\left(\frac{n}{\delta}\right)}\right] \geq 1 - \delta\,. \tag{49}$$

The result then follows by applying the latter to Equation 46. □

C.3 THE PROBABILITY DISTRIBUTION APPROXIMATION VIA RFF

For the approximation of p by $\tilde{p}$, we use the following fact.

Theorem 5 (Bochner's theorem (Rudin, 1990))**.** *A function* $u : \mathcal{X} \to \mathbb{R}$*,* $\mathcal{X} \subset \mathbb{R}^d$ *is positive-definite if and only if it is the Fourier transform of a non-negative measure.*

By Bochner's theorem (Theorem 5), as previously applied to positive-definite kernels (Theorem 1, main paper), we can also trivially conclude that any *positive-definite* probability density function is by itself the Fourier transform of a probability measure, so that it admits a Fourier-feature representation of the form in Definition 2. A probability density function $p : \mathbb{R}^d \to [0, \infty)$ is positive-definite if, for all $n \in \mathbb{N}$, $\{\alpha_i\}_{i=1}^n \subset \mathbb{R}$ and all $\{\boldsymbol{x}_i\}_{i=1}^n \subset \mathbb{R}^d$ the following holds:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j p(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 0\,. \tag{50}$$

Not every probability density function is positive-definite, but examples include Gaussian and Student-T distributions (Rossberg, 1995). In particular, we can make a kernel k_p from a probability density function p on $\mathcal{X}$ by:

$$\begin{aligned} k_p : \mathcal{X} \times \mathcal{X} &\longrightarrow \mathbb{R} \\ \boldsymbol{x}, \boldsymbol{x}' &\longmapsto \begin{cases} p(\boldsymbol{x} - \boldsymbol{x}')\,, & \boldsymbol{x} - \boldsymbol{x}' \in \mathcal{X}, \\ 0\,, & \boldsymbol{x} - \boldsymbol{x}' \notin \mathcal{X}\,. \end{cases} \end{aligned} \tag{51}$$

It is easy to verify that a kernel defined as above is positive-definite if p is positive-definite. The kernel is also translation-invariant, since $k_p(\mathbf{v} + \boldsymbol{x}, \mathbf{v} + \boldsymbol{x}') = k_p(\boldsymbol{x}, \boldsymbol{x}')$, for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ and any $\mathbf{v} \in \mathbb{R}^d$. Similarly, we have the equivalence $p(\boldsymbol{x}) = k_p(\boldsymbol{x}, \mathbf{0})$ and a corresponding $\tilde{p}(\boldsymbol{x}) = \tilde{k}_p(\boldsymbol{x}, \mathbf{0})$, for $\boldsymbol{x} \in \mathcal{X}$, by applying Definition 2 to k_p. As a result, we can use Lemma 3 to k_p to bound the approximation error in $|p(\boldsymbol{x}) - \tilde{p}(\boldsymbol{x})|$.

Theorem 6 (Restatement of Theorem 2)**.** *Let $p : \mathcal{X} \to \mathbb{R}$ be a positive-definite probability density function defined on $\mathcal{X} \subset \mathbb{R}^d$ which is such that $\nabla^2 p(\mathbf{0})$ exists. Assume $\mathcal{X}$ is compact, and let $b_p > 0$ be any constant such that $b_p \geq \max_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x})$. Let $\tilde{k}_p$ denote an RFF approximation with $R_p \in \mathbb{N}$ frequencies to k_p as defined in Equation 51, and let $\tilde{p} : \boldsymbol{x} \mapsto \tilde{k}_p(\boldsymbol{x}, \mathbf{0})$, $\boldsymbol{x} \in \mathcal{X}$. Then, for any $\xi > 0$, the following holds:*

$$\begin{aligned}\mathbb{P}\left[\sup_{\boldsymbol{x}\in\mathcal{X}} |\tilde{p}(\boldsymbol{x}) - p(\boldsymbol{x})| \geq b_p \xi\right] &\leq \beta_d \left(\frac{\sigma_{k_p}\ell_{\mathcal{X}}}{\xi}\right)^{\frac{2}{1+\frac{2}{d}}} \exp\left(-\frac{R_p\xi^2}{4(d+2)\alpha_\xi}\right)\\ &\leq 66\left(\frac{\sigma_{k_p}\ell_{\mathcal{X}}}{\xi}\right)^2 \exp\left(-\frac{R_p\xi^2}{4(d+2)}\right)\end{aligned} \tag{52}$$

where for the second statement we assume $\xi \leq \sigma_{k_p}\ell_{\mathcal{X}}$, and σ_{k_p}, $\ell_{\mathcal{X}}$, α_ξ and β_ξ are the same as defined in Lemma 3 for $k := \frac{1}{b_p}k_p$.

Proof. The result follows by applying Lemma 3 to a normalised version $\bar{k}_p := \frac{1}{b_p}k_p$ of k_p (Equation 51), which is such that $\bar{k}_p(\boldsymbol{x}, \boldsymbol{x}') = 1$. Noticing that:

$$\begin{aligned}\sup_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} |\tilde{k}_p(\boldsymbol{x}, \boldsymbol{x}') - k_p(\boldsymbol{x}, \boldsymbol{x}')| &= \sup_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} |\tilde{k}_p(\boldsymbol{x} - \boldsymbol{x}', \mathbf{0}) - k_p(\boldsymbol{x} - \boldsymbol{x}', \mathbf{0})|\\ &= \sup_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}:\boldsymbol{x}-\boldsymbol{x}'\in\mathcal{X}} |\tilde{p}(\boldsymbol{x} - \boldsymbol{x}') - p(\boldsymbol{x} - \boldsymbol{x}')|\\ &\leq \sup_{\boldsymbol{x}\in\mathcal{X}} |\tilde{p}(\boldsymbol{x}) - p(\boldsymbol{x})|\,,\end{aligned} \tag{53}$$

so that $\sup_{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}} |\tilde{k}_p(\boldsymbol{x}, \boldsymbol{x}') - k_p(\boldsymbol{x}, \boldsymbol{x}')| \geq b_p\xi$ implies $\sup_{\boldsymbol{x}\in\mathcal{X}} |\tilde{p}(\boldsymbol{x}) - p(\boldsymbol{x})| \geq b_p\xi$, concludes the proof. □

Given $\xi_p > 0$ such that $\sup_{\boldsymbol{x}\in\mathcal{X}} |p(\boldsymbol{x}) - \tilde{p}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} \leq \xi_p$, the integration error is bounded by:

$$\int_{\mathcal{X}} |p(\boldsymbol{x}) - \tilde{p}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} \leq b_p\xi_p \int_{\mathcal{X}} \mathrm{d}\boldsymbol{x} \leq b_p\xi_p v_{\mathcal{X}}\,, \tag{54}$$

where $v_{\mathcal{X}} := \int_{\mathcal{X}} \mathrm{d}\boldsymbol{x}$ denotes the volume of the domain $\mathcal{X}$. The latter can be bounded by the volume of a hyper-sphere of diameter $\ell_{\mathcal{X}}$ in $\mathbb{R}^d$, i.e.:

$$v_{\mathcal{X}} \leq \frac{\pi^d \ell_{\mathcal{X}}^d}{2^d\Gamma\left(\frac{d}{2}+1\right)}\,, \tag{55}$$

where Γ denotes Euler's gamma function.

C.4 QUADRATURE APPROXIMATION ERROR

We now combine our results to bound the quadrature approximation error.

Theorem 7 (Restatement of Theorem 3)**.** *Let $f \in \mathcal{H}_k$, where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive-definite, translation-invariant kernel on $\mathcal{X} \subset \mathbb{R}^d$. Assume that:*

1. *$\mathcal{X}$ is compact with diameter $\ell_{\mathcal{X}} < \infty$ and volume $v_{\mathcal{X}} := \int_{\mathcal{X}} \mathrm{d}\boldsymbol{x} < \infty$;*
2. *$k(\mathbf{0}, \mathbf{0}) = 1$ and $\nabla^2 k(\mathbf{0}, \mathbf{0})$ exists;*
3. *and $p : \mathcal{X} \to [0, \infty)$ is a positive-definite probability density function.*

Then, given any $\delta \in (0, 1)$, the following holds with probability at least $1 - \delta$:

$$\begin{aligned}&\left|\int_{\mathcal{X}} f(\boldsymbol{x})p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{X}} \hat{\mu}_n(\boldsymbol{x})\tilde{p}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right|\\ &\leq \left(\frac{n}{\lambda}\beta_\epsilon\left(\frac{\delta}{4}\right)\xi_k + \beta_k\left(\frac{\delta}{4}\right)\max_{\boldsymbol{x}\in\mathcal{X}}\sigma_n(\boldsymbol{x})\right)(1 + b_p\xi_p v_{\mathcal{X}}) + \|f\|_\infty b_p\xi_p v_{\mathcal{X}}\,,\end{aligned} \tag{56}$$

for an RFF approximation to k *with* $R_k \geq R\left(\xi_k, \frac{\delta}{4}, \sigma_k\right)$ *frequencies and an RFF approximation to* p *with* $R_p \geq R\left(\xi_p, \frac{\delta}{4}, \sigma_{k_p}\right)$ *frequencies, given* $0 < \xi_k \leq \sigma_k \ell_{\mathcal{X}}$ *and* $0 < \xi_p \leq \sigma_{k_p} \ell_{\mathcal{X}}$*, where:*

$$\beta_\epsilon(\delta) := \|f\|_\infty + \sigma_\epsilon \sqrt{2 \log\left(\frac{n}{\delta}\right)} \tag{57}$$

$$\beta_k(\delta) := \|f\|_k + \sigma_\epsilon \sqrt{\frac{2}{\lambda} \log\left(\frac{\det(\mathbf{I} + \lambda^{-1}\mathbf{K}_n)^{1/2}}{\delta}\right)} \tag{58}$$

$$R(\xi, \delta, \sigma_k) := \frac{4(d+2)}{\xi^2} \left(\frac{2}{1+\frac{2}{d}} \log \frac{\sigma_k \ell_{\mathcal{X}}}{\xi} + \log \frac{66}{\delta}\right) . \tag{59}$$

Proof. In the spectral Bayesian quadrature formulation, we have the following approximation:

$$\int_{\mathcal{X}} f(\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \approx \mathbf{y}^\mathsf{T} (\mathbf{K}_n + \lambda \mathbf{I})^{-1} \int_{\mathcal{X}} \tilde{\mathbf{k}}_n(\boldsymbol{x}) \tilde{p}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\mathcal{X}} \hat{\mu}(\boldsymbol{x}) \tilde{p}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \,, \tag{60}$$

where $\hat{\mu}_n(\boldsymbol{x}) := \tilde{\mathbf{k}}_n(\boldsymbol{x})^\mathsf{T}(\mathbf{K}_n + \lambda \mathbf{I})^{-1}\mathbf{y}$ We will bound the approximation error by starting with the following decomposition:

$$\begin{aligned}
&\left| \int_{\mathcal{X}} f(\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \int_{\mathcal{X}} \hat{\mu}_n(\boldsymbol{x}) \tilde{p}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right| \\
&\leq \left| \int_{\mathcal{X}} f(\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \int_{\mathcal{X}} \hat{\mu}_n(\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right| + \left| \int_{\mathcal{X}} \hat{\mu}_n(\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \int_{\mathcal{X}} \hat{\mu}_n(\boldsymbol{x}) \tilde{p}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right| \\
&\leq \|f - \hat{\mu}_n\|_\infty + \|\hat{\mu}_n\|_\infty \int_{\mathcal{X}} |p(\boldsymbol{x}) - \tilde{p}(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x} \,.
\end{aligned} \tag{61}$$

We first observe that:

$$\forall \boldsymbol{x} \in \mathcal{X}, \quad |f(\boldsymbol{x}) - \hat{\mu}_n(\boldsymbol{x})| \leq |f(\boldsymbol{x}) - \mu(\boldsymbol{x})| + |\mu(\boldsymbol{x}) - \hat{\mu}(\boldsymbol{x})| \,. \tag{62}$$

Assuming $f \in \mathcal{H}_k$, given $\delta_\mu \in (0, 1)$, we can apply Lemma 1 (main paper) to bound the first term on the right-hand side as:

$$\mathbb{P}\left[\sup_{\boldsymbol{x} \in \mathcal{X}} |f(\boldsymbol{x}) - \mu_n(\boldsymbol{x})| \leq \sup_{\boldsymbol{x} \in \mathcal{X}} \beta_k(\delta_\mu) \sigma_n(\boldsymbol{x})\right] \geq 1 - \delta_\mu \,. \tag{63}$$

For the second-term on the right-hand side of Equation 62, we have that:

$$\begin{aligned}
|\mu_n(\boldsymbol{x}) - \hat{\mu}_n(\boldsymbol{x})| &\leq \|\mathbf{k}_n(\boldsymbol{x}) - \tilde{\mathbf{k}}_n(\boldsymbol{x})\|_2 \|(\mathbf{K}_n + \lambda \mathbf{I})^{-1}\mathbf{y}\|_2 \\
&\leq \|\mathbf{k}_n(\boldsymbol{x}) - \tilde{\mathbf{k}}_n(\boldsymbol{x})\|_2 \|(\mathbf{K}_n + \lambda \mathbf{I})^{-1}\|_2 \|\mathbf{y}\|_2 \,, \\
&\leq \frac{\|\mathbf{y}\|_2}{\lambda} \|\mathbf{k}_n(\boldsymbol{x}) - \tilde{\mathbf{k}}_n(\boldsymbol{x})\|_2 \,.
\end{aligned} \tag{64}$$

since $\|(\mathbf{K}_n + \lambda \mathbf{I})^{-1}\|_2 \leq \lambda^{-1}$. Applying Lemma 4, given $\delta_\epsilon \in (0, 1)$, yields:

$$\mathbb{P}\left[\|\mathbf{y}\|_2 \leq \sqrt{n} \beta_\epsilon(\delta_\epsilon)\right] \geq 1 - \delta_\epsilon \,. \tag{65}$$

where $\beta_\epsilon(\delta) := \|f\|_\infty + \sigma_\epsilon \sqrt{2 \log\left(\frac{n}{\delta}\right)}$. In addition, considering the kernel approximation guarantee in Lemma 3, for a given number of Fourier frequencies $R_k \geq R(\delta_k, \xi_k)$, leads us to:

$$\mathbb{P}\left[\sup_{\boldsymbol{x} \in \mathcal{X}} \|\mathbf{k}_n(\boldsymbol{x}) - \tilde{\mathbf{k}}_n(\boldsymbol{x})\|_2 \leq \sqrt{n} \xi_k\right] \geq 1 - \delta_k \,. \tag{66}$$

Therefore, we have:

$$\mathbb{P}\left[\|f - \hat{\mu}_n\|_\infty \leq \beta_k(\delta_\mu) \max_{\boldsymbol{x} \in \mathcal{X}} \sigma_n(\boldsymbol{x}) + \frac{1}{\lambda} n \xi_k \beta_\epsilon(\delta_\epsilon)\right] \geq 1 - \delta_\mu - \delta_\epsilon - \delta_k \,, \tag{67}$$

which follows by applying a union bound on the complementary events in the equations above. Lastly, note that, under the assumption that the event in Equation 67 holds, the following is also true:

$$\|\hat{\mu}_n\|_\infty \leq \|f\|_\infty + \frac{1}{\lambda} n\xi_k \beta_\epsilon(\delta_\epsilon) + \max_{\boldsymbol{x}\in\mathcal{X}} \beta_k(\delta_k)\sigma_n(\boldsymbol{x})\,. \tag{68}$$

Regarding the probability density approximation, let $v_{\mathcal{X}} := \int_{\mathcal{X}} \mathrm{d}\boldsymbol{x}$ represent the volume of $\mathcal{X}$. Assume $R_p \geq R(\delta_p, \xi_p)$ Fourier frequencies for $\tilde{p}$, for $\delta_p \in (0,1)$. Then Theorem 6 tells us that:

$$\mathbb{P}\left[\int_{\mathcal{X}} |p(\boldsymbol{x}) - \tilde{p}(\boldsymbol{x})| \,\mathrm{d}\boldsymbol{x} \leq b_p \xi_p v_{\mathcal{X}}\right] \geq 1 - \delta_p\,. \tag{69}$$

The final result follows by applying a union bound to combine the events in equations 67, 68 and 69 into Equation 61. □

D FULL EXPERIMENTAL RESULTS

D.1 5D CONTINUOUS EQUATION

Table 1: 5D Continuous Equation Integration Results (% Error).

N	MC	BQ	GBQ-U RBF	GBQ-G RBF
10	$\mathbf{9.67 \pm 8.43}$	20.39 ± 3.85	23.77 ± 4.33	20.35 ± 3.99
25	9.32 ± 7.7	3.21 ± 1.87	6.02 ± 2.46	$\mathbf{3.0 \pm 1.97}$
50	5.57 ± 4.14	$\mathbf{0.61 \pm 0.34}$	2.48 ± 0.51	0.88 ± 0.42
100	3.81 ± 2.1	2.05 ± 0.35	$\mathbf{0.89 \pm 0.44}$	2.25 ± 0.4
200	2.65 ± 2.23	2.29 ± 0.24	$\mathbf{0.52 \pm 0.26}$	2.45 ± 0.27
300	3.89 ± 2.44	2.28 ± 0.21	$\mathbf{0.4 \pm 0.17}$	2.44 ± 0.19
400	2.74 ± 1.7	2.28 ± 0.2	$\mathbf{0.33 \pm 0.14}$	2.44 ± 0.2
500	1.29 ± 1.06	2.27 ± 0.14	$\mathbf{0.22 \pm 0.12}$	2.43 ± 0.16
600	1.53 ± 0.97	2.28 ± 0.12	$\mathbf{0.21 \pm 0.09}$	2.44 ± 0.17
700	2.39 ± 2.43	2.29 ± 0.12	$\mathbf{0.16 \pm 0.1}$	2.43 ± 0.16
800	1.42 ± 1.06	2.24 ± 0.1	$\mathbf{0.17 \pm 0.12}$	2.38 ± 0.16
900	1.64 ± 1.21	2.24 ± 0.09	$\mathbf{0.15 \pm 0.09}$	2.38 ± 0.12
1000	1.79 ± 1.09	2.22 ± 0.08	$\mathbf{0.14 \pm 0.09}$	2.37 ± 0.13

D.2 5D DISJOINT EQUATION

Table 2: 5D Disjoint Equation Integration Results (% Error).

N	MC	BQ	GBQ-G RBF	GBQ-G M3/2
10	$\mathbf{23.94 \pm 13.0}$	33.32 ± 3.0	33.26 ± 3.12	38.11 ± 3.78
25	$\mathbf{16.84 \pm 20.99}$	18.26 ± 0.86	17.96 ± 1.08	22.17 ± 1.22
50	$\mathbf{7.58 \pm 5.92}$	15.15 ± 0.59	14.87 ± 0.6	16.69 ± 0.7
100	5.89 ± 3.64	$\mathbf{1.71 \pm 0.83}$	2.06 ± 1.28	5.53 ± 4.55
200	4.81 ± 4.15	2.24 ± 0.73	$\mathbf{2.17 \pm 1.44}$	5.58 ± 2.91
300	6.24 ± 4.39	$\mathbf{0.59 \pm 0.46}$	0.62 ± 0.44	2.07 ± 0.67
400	3.98 ± 2.28	1.11 ± 0.42	1.7 ± 0.55	$\mathbf{0.79 \pm 0.64}$
500	4.17 ± 4.17	$\mathbf{3.37 \pm 0.44}$	4.28 ± 0.7	3.55 ± 0.84
600	3.22 ± 2.83	2.63 ± 0.34	3.18 ± 0.68	$\mathbf{2.59 \pm 0.58}$
700	3.93 ± 2.51	1.03 ± 0.34	1.31 ± 0.61	$\mathbf{0.85 \pm 0.44}$
800	3.0 ± 1.8	$\mathbf{0.78 \pm 0.44}$	1.15 ± 0.6	0.89 ± 0.51
900	3.58 ± 2.21	$\mathbf{0.48 \pm 0.29}$	0.92 ± 0.54	$\mathbf{0.48 \pm 0.42}$
1000	3.24 ± 2.15	$\mathbf{0.38 \pm 0.24}$	0.89 ± 0.46	0.53 ± 0.5

REFERENCES

Rossberg, H.-J. (1995). “Positive definite probability densities and probability distributions.” In: *Journal of Mathematical Sciences* 76.1, pp. 2181–2197.

Rudin, Walter (1990). “The Basic Theorems of Fourier Analysis.” In: *Fourier Analysis on Groups*. 2. John Wiley & Sons, Ltd. Chap. 1, pp. 1–34.

Sutherland, Dougal J. and Jeff Schneider (2015). “On the error of random Fourier features.” In: *Uncertainty in Artificial Intelligence - Proceedings of the 31st Conference, UAI 2015*, pp. 862–871.