Random Fourier Features for Kernel Ridge Regression: Approximation Bounds and Statistical Guarantees Appendix: Proofs

A. Preliminaries

Our upper and lower bound analysis relies predominantly on Fourier analysis, so we now introduce some additional notation and state some useful facts about these.

A.1. Properties of Fourier Transforms

Definition 16 (Fourier Transform)**.** The *Fourier transform* of a continuous function $f : \mathbb{R}^d \to \mathbb{C}$ in $L_1(\mathbb{R}^n)$ is defined to be the function $\mathcal{F}f : \mathbb{R}^d \to \mathbb{C}$ as follows:

$$(\mathcal{F}f)(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} f(\mathbf{t}) e^{-2\pi i \mathbf{t}^T \boldsymbol{\xi}} \, d\mathbf{t}.$$

We also sometimes use the notation $\hat{f}$ for the Fourier transform of f. We often informally refer to f as representing the function in *time domain* and $\hat{f}$ as representing the function in *frequency domain*.

The original function f can also be obtained from $\hat{f}$ by the *inverse Fourier transform*:

$$f(\mathbf{t}) = \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\xi}) e^{2\pi i \boldsymbol{\xi}^T \mathbf{t}} \, d\boldsymbol{\xi}$$

Definition 17 (Convolution)**.** The *convolution* of two functions $f : \mathbb{R}^d \to \mathbb{C}$ and $g : \mathbb{R}^d \to \mathbb{C}$ is defined to be the function $(f * g) : \mathbb{R}^d \to \mathbb{C}$ given by

$$(f * g)(\boldsymbol{\eta}) = \int_{\mathbb{R}^d} f(\mathbf{t}) g(\boldsymbol{\eta} - \mathbf{t}) \, d\mathbf{t}.$$

The convolution theorem shows that the Fourier transform of the convolution of two functions is simply the product of the individual Fourier transforms:

Claim 18 (Convolution Theorem)**.** *Given functions $f : \mathbb{R}^d \to \mathbb{C}$ and $g : \mathbb{R}^d \to \mathbb{C}$ whose convolution is $h = f * g$, we have*

$$\hat{h}(\boldsymbol{\xi}) = \hat{f}(\boldsymbol{\xi}) \cdot \hat{g}(\boldsymbol{\xi})$$

for all $\boldsymbol{\xi} \in \mathbb{R}^d$.

Now, suppose $d = 1$, i.e., the functions we consider take inputs in $\mathbb{R}$. We define the *rectangle function* and *normalized sinc function*, which we use extensively in our analysis.

Definition 19 (Rectangle Function)**.** We define the *rectangle function* $\mathrm{rect}_a : \mathbb{R} \to \mathbb{C}$ as

$$\mathrm{rect}_a(x) = \begin{cases} 0 & \text{if } |x| > a/2 \\ \frac{1}{2} & \text{if } |x| = a/2 \\ 1 & \text{if } |x| < a/2 \end{cases}.$$

If $a = 1$, then we often omit the subscript and simply write rect.

Definition 20 (Normalized Sinc Function)**.** We define the *normalized sinc function* $\mathrm{sinc} : \mathbb{R} \to \mathbb{C}$ as

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

It is well known that the Fourier transform of the rectangle function (with $a = 1$) is the normalized sinc function:

$$\mathcal{F}(\text{rect}) = \text{sinc}$$

We use δ to denote the *Dirac delta function*. Recall that the Dirac delta function satisfies the following useful property for any function f:

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx = f(a),$$

i.e. the integral of a function multiplied by a shifted Dirac delta functions picks out the value of the function at a particular point. Thus, it is not hard to see that the Fourier transform of a δ is the constant function which is 1 everywhere:

$$(\mathcal{F}\delta)(\xi) = \int_{-\infty}^{\infty} e^{-2\pi i t\xi} \cdot \delta(t)\,dt = e^{-2\pi i \cdot 0 \cdot \xi} = 1$$

for all ξ. Similarly, the Fourier transform of a shifted delta function is as follows:

$$(\mathcal{F}\delta(\cdot - a))(\xi) = \int_{-\infty}^{\infty} e^{-2\pi i t\xi} \cdot \delta(t-a)\,dt = e^{-2\pi i a\xi}.$$

Moreover, it is not hard to see that convolving a function by a shifted delta funciton results in a shift of the original function:

$$(f * \delta(\cdot - a))(x) = f(x-a).$$

Thus, by the convolution theorem, we obtain the following identity:

Claim 21. *Given a function $f : \mathbb{R} \to \mathbb{C}$, we have*

$$(\mathcal{F}f(\cdot - a))(\xi) = (\mathcal{F}(f * \delta(\cdot - a)))(\xi) = \hat{f}(\xi) \cdot e^{-2\pi i a\xi}.$$

Similarly,

Claim 22. *Given a function $f : \mathbb{R} \to \mathbb{C}$, we have*

$$(\mathcal{F}(f(x) \cdot e^{2\pi i a x}))(\xi) = \hat{f}(\xi - a).$$

Finally, we introduce a useful function known as the *Dirac comb function*:

Definition 23. The *Dirac comb function* with period T is defined as f satisfying

$$f(x) = \sum_{j=-\infty}^{\infty} \delta(x - jT).$$

It is a standard fact that the Fourier transform of a Dirac comb function is another Dirac comb function which is scaled and has the inverse period:

Claim 24. *Let*

$$f(x) = \sum_{j=-\infty}^{\infty} \delta(x - jT)$$

be the Dirac comb function with period T. Then,

$$(\mathcal{F}f)(\xi) = \frac{1}{T} \sum_{j=-\infty}^{\infty} \delta\left(\xi - \frac{j}{T}\right).$$

We use the Dirac comb function in our lower bound constructions.

A.2. Properties of Gaussian Distributions

We also need several useful facts about Gaussian distributions. The following is a standard fact about the cumulative distribution function of the standard Gaussian distribution:

Claim 25 ((Feller, 1968)). *For any $x > 0$, we have*

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \le \frac{e^{-x^2/2}}{x\sqrt{2\pi}}.$$

Moreover, as a direct consequence, for any $\sigma, x > 0$, we have that

$$\frac{1}{\sqrt{2\pi}\sigma} \int_x^\infty e^{-t^2/2\sigma^2}\, dt \le \frac{\sigma e^{-x^2/2\sigma^2}}{x\sqrt{2\pi}}.$$

Also, if $x \ge 1$, then

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \le \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2}\, dt.$$

We also need the following property about Gaussian samples.

Claim 26. *Let $t \ge 10$, and $a_1, a_2, \ldots, a_t$ be sampled according to the Gaussian distribution given by probability density function $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Also, let $a^* = \max_{1\le j\le t} |a_j|$. Then,*

$$\Pr\left[\frac{1}{\sqrt{2\pi}} e^{-a^{*2}/2} \le \frac{8\sqrt{\log t}}{t}\right] \ge \frac{1}{2}.$$

Proof. Choose q_1 such that

$$\int_{q_1}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{t}. \tag{14}$$

Note that by Claim 25, we have

$$\int_{2\sqrt{\log t}}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \le \frac{1}{2\sqrt{2\pi} t^2 \sqrt{\log t}} \le \frac{1}{t}.$$

Thus, $q_1 \le 2\sqrt{\log t}$.

Also, since $\frac{1}{t} \le \frac{1}{4}$, we have that $q_1 \ge \frac{6}{5}$. Thus, by another application of Claim 25,

$$\frac{1}{t} = \int_{q_1}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \ge \left(\frac{1}{q_1} - \frac{1}{q_1^3}\right) \frac{1}{\sqrt{2\pi}} e^{-q_1^2/2} \ge \frac{1}{4q_1} \cdot \frac{1}{\sqrt{2\pi}} e^{-q_1^2/2},$$

and so,

$$\frac{1}{\sqrt{2\pi}} e^{-q_1^2/2} \le \frac{4q_1}{t} \le \frac{8\sqrt{\log t}}{t}.$$

Therefore,

$$\begin{aligned}
\Pr\left[\frac{1}{\sqrt{2\pi}} e^{-a^{*2}/2} \le \frac{8\sqrt{\log t}}{t}\right] &\ge \Pr[a^* \ge q_1] \\
&= 1 - \left(1 - \frac{1}{t}\right)^t \\
&\ge 1 - \frac{1}{e} \\
&\ge \frac{1}{2},
\end{aligned}$$

as desired. $\square$

B. Proof of Lemma 2

Note that $\mathbf{A} \preceq \mathbf{B}$ implies that $\mathbf{B}^{-1} \preceq \mathbf{A}^{-1}$ so for the bias term we have:

$$\mathbf{f}^{\mathrm{T}}(\tilde{\mathbf{K}} + \lambda \mathbf{I}_n)^{-1}\mathbf{f} \leq (1-\Delta)^{-1}\mathbf{f}^{\mathrm{T}}(\mathbf{K} + \lambda \mathbf{I}_n)^{-1}\mathbf{f}. \quad (15)$$

We now consider the variance term. Denote $s = \mathbf{rank}(\tilde{\mathbf{K}})$, and let $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ denote the eigenvalues of a matrix $\mathbf{A}$. We have:

$$\begin{aligned}
s_\lambda(\tilde{\mathbf{K}}) = \mathbf{Tr}\left((\tilde{\mathbf{K}} + \lambda \mathbf{I}_n)^{-1}\tilde{\mathbf{K}}\right) &= \sum_{i=1}^{s} \frac{\lambda_i(\tilde{\mathbf{K}})}{\lambda_i(\tilde{\mathbf{K}}) + \lambda} \\
&= s - \sum_{i=1}^{s} \frac{\lambda}{\lambda_i(\tilde{\mathbf{K}}) + \lambda} \\
&\leq s - (1+\Delta)^{-1} \sum_{i=1}^{s} \frac{\lambda}{\lambda_i(\mathbf{K}) + \lambda} \\
&= s - \sum_{i=1}^{s} \frac{\lambda}{\lambda_i(\mathbf{K}) + \lambda} + \frac{\Delta}{1+\Delta} \sum_{i=1}^{s} \frac{\lambda}{\lambda_i(\mathbf{K}) + \lambda} \\
&\leq n - \sum_{i=1}^{n} \frac{\lambda}{\lambda_i(\mathbf{K}) + \lambda} + \frac{\Delta \cdot s}{1+\Delta} \\
&= s_\lambda(\mathbf{K}) + \frac{\Delta \cdot s}{1+\Delta} \\
&\leq (1-\Delta)^{-1} s_\lambda(\mathbf{K}) + \frac{\Delta \cdot s}{1+\Delta}
\end{aligned}$$

where we use the fact that $\mathbf{A} \preceq \mathbf{B}$ implies that $\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{B})$ (this is a simple consequence of the Courant-Fischer minimax theorem).

Combining the above variance bound with the bias bound in (15) yields:

$$\widehat{\mathcal{R}}_{\tilde{\mathbf{K}}}(\mathbf{f}) \leq (1-\Delta)^{-1}\widehat{\mathcal{R}}_{\mathbf{K}}(\mathbf{f}) + \frac{\Delta}{(1+\Delta)} \cdot \frac{\mathbf{rank}(\tilde{\mathbf{K}})}{n} \cdot \sigma_\nu^2$$

and the bound $\mathcal{R}(\tilde{f}) \leq \widehat{\mathcal{R}}_{\tilde{\mathbf{K}}}(\mathbf{f})$ completes the proof.

C. Proof of Proposition 4

Since k is positive definite and $k(0) = 1$, $|k(\mathbf{x}, \mathbf{z})| \leq 1$ for all $\mathbf{x}$ and $\mathbf{z}$. This implies that the maximum eigenvalue of $\mathbf{K}$ is bounded by n, and the lower bound follows immediately. The upper bound on $\tau_\lambda(\boldsymbol{\eta})$ follows from the fact that $\|\mathbf{z}(\boldsymbol{\eta})\|_2^2 = n$ and all eigenvalues of $\mathbf{K} + \lambda \mathbf{I}_n$ are bounded from below by λ. The bound also establishes that the integral converges. We now have,

$$\begin{aligned}
\int_{\mathbb{R}^d} \tau_\lambda(\boldsymbol{\eta}) d\boldsymbol{\eta} &= \int_{\mathbb{R}^d} p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda \mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) d\boldsymbol{\eta} \\
&= \int_{\mathbb{R}^d} \mathbf{Tr}\left(p(\boldsymbol{\eta})(\mathbf{K} + \lambda \mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*\right) d\boldsymbol{\eta} \\
&= \mathbf{Tr}\left(\int_{\mathbb{R}^d} p(\boldsymbol{\eta})(\mathbf{K} + \lambda \mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^* d\boldsymbol{\eta}\right) \\
&= \mathbf{Tr}\left((\mathbf{K} + \lambda \mathbf{I}_n)^{-1} \int_{\mathbb{R}^d} p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^* d\boldsymbol{\eta}\right) \\
&= \mathbf{Tr}\left((\mathbf{K} + \lambda \mathbf{I}_n)^{-1}\mathbf{K}\right) = s_\lambda(\mathbf{K}) .
\end{aligned}$$

The second equality is due to the fact that $\mathbf{z}(\eta)$ is a rank one matrix, and third equality is due to linearity of the trace operation and the fact that all diagonal entries are positive.

D. Proof of Lemma 6 and Theorem 7

To prove Lemma 6 we need the following lemma which is essentially a restatement of Corollary 7.3.3 from (Tropp, 2015). However, the minimum t in the following statement is much lower than the bound that appears in (Tropp, 2015) which is unnecessarily loose (possibly, a typo in (Tropp, 2015)). For completeness, we include a proof.

Lemma 27. *Let $\mathbf{B}$ be a fixed $d_1 \times d_2$ matrix. Construct a $d_1 \times d_2$ random matrix $\mathbf{R}$ that satisfies*

$$\mathbb{E}[\mathbf{R}] = \mathbf{B} \quad \textit{and} \quad \|\mathbf{R}\|_2 \leq L.$$

Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be semidefinite upper bounds for the expected squares:

$$\mathbb{E}[\mathbf{R}\mathbf{R}^*] \preceq \mathbf{M}_1 \quad \textit{and} \quad \mathbb{E}[\mathbf{R}^*\mathbf{R}] \preceq \mathbf{M}_2.$$

Define the quantities

$$m = \max(\|\mathbf{M}_1\|_2, \|\mathbf{M}_2\|_2) \quad \textit{and} \quad d = (\mathbf{Tr}(\mathbf{M}_1) + \mathbf{Tr}(\mathbf{M}_2))/m.$$

Form the matrix sampling estimator

$$\bar{\mathbf{R}}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{R}_k$$

where each $\mathbf{R}_k$ is an independent copy of $\mathbf{R}$. Then, for all $t \geq \sqrt{m/n} + 2L/3n$,

$$\Pr(\|\bar{\mathbf{R}}_n - \mathbf{B}\|_2 \geq t) \leq 4d \exp\left(\frac{-nt^2/2}{m + 2Lt/3}\right). \tag{16}$$

Proof. The proof mirrors the proof of Corollary 6.2.1 in (Tropp, 2015), using Theorem 7.3.1 instead of Theorem 6.1.1 (both from (Tropp, 2015)).

Since $\mathbb{E}[\mathbf{R}] = \mathbf{B}$, we can write

$$\mathbf{Z} \equiv \bar{\mathbf{R}}_n - \mathbf{B} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{R}_k - \mathbb{E}[\mathbf{R}]) = \sum_{k=1}^{n} \mathbf{S}_k,$$

where we have define $\mathbf{S}_k \equiv n^{-1}(\mathbf{R}_k - \mathbb{E}[\mathbf{R}])$. These random matrices are i.i.d and each has zero mean.

Now, we can bound each of the summands:

$$\|\mathbf{S}_k\|_2 \leq \frac{1}{n}(\|\mathbf{R}_k\|_2 + \|\mathbb{E}[\mathbf{R}]\|_2) \leq \frac{1}{n}(\|\mathbf{R}_k\|_2 + \mathbb{E}[\|\mathbf{R}\|_2]) \leq \frac{2L}{n},$$

where the first inequality is the triangle inequality and the second is Jensen's inequality.

To find semidefinite upper bounds $\mathbf{V}_1$ and $\mathbf{V}_2$ on the matrix-valued variances we note that

$$\begin{aligned} \mathbb{E}[\mathbf{S}_1\mathbf{S}_1^*] &= n^{-2}\mathbb{E}[(\mathbf{R} - \mathbb{E}[\mathbf{R}])(\mathbf{R} - \mathbb{E}[\mathbf{R}])^*] \\ &= n^{-2}\left(\mathbb{E}[\mathbf{R}\mathbf{R}^*] - \mathbb{E}[\mathbf{R}]\,\mathbb{E}[\mathbf{R}]^*\right) \\ &\preceq n^{-2}\mathbb{E}[\mathbf{R}\mathbf{R}^*]. \end{aligned}$$

Likewise, $\mathbb{E}[\mathbf{S}_1^*\mathbf{S}_1] \preceq n^{-2}\mathbb{E}[\mathbf{R}^*\mathbf{R}]$. Since the summands are i.i.d, if we define $\mathbf{V}_1 \equiv n^{-1}\mathbf{M}_1$ and $\mathbf{V}_2 \equiv n^{-1}\mathbf{M}_2$, we have $\mathbb{E}[\mathbf{Z}\mathbf{Z}^*] \preceq \mathbf{V}_1$ and $\mathbb{E}[\mathbf{Z}^*\mathbf{Z}] \preceq \mathbf{V}_2$.

We now calculate,

$$\nu \equiv \max(\|\mathbf{V}_1\|_2, \|\mathbf{V}_2\|_2) = \frac{m}{n}$$

and

$$\frac{\mathbf{Tr}(\mathbf{V}_1) + \mathbf{Tr}(\mathbf{V}_2)}{\max(\|\mathbf{V}_1\|_2, \|\mathbf{V}_2\|_2)} = d.$$

Noticing, that the condition $t \geq \sqrt{m/n} + 2L/3n$ meets the required lower bound in Theorem 7.3.1 in (Tropp, 2015) we can now apply this theorem, which along with the above calculations translates to (16). □

We can now prove Lemma 6.

Proof of Lemma 6. Let $\mathbf{K}+\lambda\mathbf{I}_n = \mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^2\mathbf{V}$ be an eigendecomposition of $\mathbf{K}+\lambda\mathbf{I}_n$. Note that the Δ-spectral approximation guarantee (2) is equivalent to

$$\mathbf{K} - \Delta(\mathbf{K} + \lambda\mathbf{I}_n) \preceq \mathbf{Z}\mathbf{Z}^* \preceq \mathbf{K} + \Delta(\mathbf{K} + \lambda\mathbf{I}_n)\,,$$

so by multiplying by $\mathbf{\Sigma}^{-1}\mathbf{V}$ on the left and $\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}$ on the right we find that it suffices to show that

$$\|\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{K}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\|_2 \leq \Delta \tag{17}$$

holds with probability of at least $1-\rho$. Let

$$\mathbf{Y}_l = \frac{p(\boldsymbol{\eta}_l)}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\,.$$

Note that $\mathbb{E}\left[\mathbf{Y}_l\right] = \mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{K}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}$ and $\frac{1}{s}\sum_{l=1}^{s}\mathbf{Y}_l = \mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{Z}\mathbf{Z}^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}$. Thus, we can use the matrix concentration result above to prove (17).

To apply this bound we need to bound the norm of $\mathbf{Y}_l$ and the stable rank $\mathbb{E}\left[\mathbf{Y}_l^2\right]$. Since $\mathbf{Y}_l$ is always a rank one matrix we have

$$\begin{aligned}
\|\mathbf{Y}_l\|_2 &= \frac{p(\boldsymbol{\eta}_l)}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)}\mathbf{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\right)\\
&= \frac{p(\boldsymbol{\eta}_l)}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)}\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\\
&= \frac{p(\boldsymbol{\eta}_l)}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)}\mathbf{z}(\boldsymbol{\eta}_l)^*(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}_l)\\
&= \frac{s_{\tilde{\tau}}\cdot\tau(\boldsymbol{\eta}_l)}{\tilde{\tau}(\boldsymbol{\eta}_l)} \leq s_{\tilde{\tau}}
\end{aligned}$$

since $\tilde{\tau}(\boldsymbol{\eta}_l) \geq \tau(\boldsymbol{\eta}_l)$. We also have

$$\begin{aligned}
\mathbf{Y}_l^2 &= \frac{p(\boldsymbol{\eta}_l)^2}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)^2}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\\
&= \frac{p(\boldsymbol{\eta}_l)^2}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)^2}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\mathbf{z}(\boldsymbol{\eta}_l)^*(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\\
&= \frac{p(\boldsymbol{\eta}_l)\tau(\boldsymbol{\eta}_l)}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)^2}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{z}(\boldsymbol{\eta}_l)\mathbf{z}(\boldsymbol{\eta}_l)^*\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\\
&= \frac{\tau(\boldsymbol{\eta}_l)}{p_{\tilde{\tau}}(\boldsymbol{\eta}_l)}\mathbf{Y}_l\\
&= \frac{s_{\tilde{\tau}}\tau(\boldsymbol{\eta}_l)}{\tilde{\tau}(\boldsymbol{\eta}_l)}\mathbf{Y}_l \preceq s_{\tilde{\tau}}\mathbf{Y}_l
\end{aligned}$$

Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\mathbf{K}$. We have

$$\begin{aligned}
\mathbb{E}\left[s_{\tilde{\tau}}\mathbf{Y}_l\right] &= s_{\tilde{\tau}}\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{K}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\\
&= s_{\tilde{\tau}}\left(\mathbf{I}_n - \lambda\mathbf{\Sigma}^{-2}\right)\\
&= s_{\tilde{\tau}}\cdot\mathbf{diag}\left(\lambda_1/(\lambda_1+\lambda),\ldots,\lambda_n/(\lambda_n+\lambda)\right) := \mathbf{D}\,.
\end{aligned}$$

So,

$$
\begin{aligned}
\Pr\left(\left\|\frac{1}{s}\sum_{l=1}^{s}\mathbf{Y}_l-\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{K}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\right\|_2\geq\Delta\right) &\leq \frac{8\mathbf{Tr}\left(\mathbf{D}\right)}{\|\mathbf{D}\|_2}\exp\left(\frac{-s\Delta^2/2}{\|\mathbf{D}\|_2+2s_{\tilde{\tau}}\Delta/3}\right)\\
&\leq 8\frac{s_{\tilde{\tau}}\cdot s_{\lambda}(\mathbf{K})}{s_{\tilde{\tau}}\cdot\lambda_1/(\lambda_1+\lambda)}\exp\left(\frac{-s\Delta^2}{2s_{\tilde{\tau}}(1+2\Delta/3)}\right)\\
&\leq 16 s_{\lambda}(\mathbf{K})\exp\left(\frac{-s\Delta^2}{2s_{\tilde{\tau}}(1+2\Delta/3)}\right)\\
&\leq 16 s_{\lambda}(\mathbf{K})\exp\left(\frac{-3s\Delta^2}{8s_{\tilde{\tau}}}\right)\leq\rho
\end{aligned}
$$

where the third inequality is due to the assumption that $\lambda_1=\|\mathbf{K}\|_2\geq\lambda$ and the last inequality is due to the bound on s. □

Proof of Theorem 7. Define $\tilde{\tau}(\boldsymbol{\eta})=p(\boldsymbol{\eta})\cdot n_\lambda$ and note that $\tilde{\tau}(\boldsymbol{\eta})\geq\tau_\lambda(\boldsymbol{\eta})$ by Proposition 4 and that $s_{\tilde{\tau}}=n_\lambda$. Finally, note that $p_{\tilde{\tau}}(\boldsymbol{\eta})=p(\boldsymbol{\eta})$, the classic Fourier features sampling probability. □

E. Proof of Lemmas 11 and 12

Let $R(\mathbf{\Phi})\subseteq\mathbb{C}^n$ denote the range of $\mathbf{\Phi}$. Here we first prove that the operator $\mathbf{\Phi}$ is defined on all $L_2(d\mu)$ and is a bounded linear operator. Indeed, for $y\in L_2(d\mu)$ we have:

$$
\begin{aligned}
\|\mathbf{\Phi}y\|_2^2 &= \left\|\int_{\mathbb{R}^d}\mathbf{z}(\boldsymbol{\xi})y(\boldsymbol{\xi})d\mu(\boldsymbol{\xi})\right\|_2^2\\
\text{(by Jensen's inequality)}\quad &\leq \int_{\mathbb{R}^d}\|\mathbf{z}(\boldsymbol{\xi})y(\boldsymbol{\xi})\|_2^2 d\mu(\boldsymbol{\xi})\\
&= \int_{\mathbb{R}^d}\|y(\xi)\|_2^2\cdot\|\mathbf{z}(\boldsymbol{\xi})\|_2^2 d\mu(\boldsymbol{\xi})\\
&= n\cdot\|y\|_{L_2(d\mu)}^2\,.
\end{aligned}
$$

Thus, $R(\mathbf{\Phi})$ is a linear subspace of $\mathbb{C}^n$. Therefore, there is a unique adjoint operator $\mathbf{\Phi}^*:R(\mathbf{\Phi})\rightarrow L_2(d\mu)$, such that $\langle\mathbf{\Phi}y,\mathbf{x}\rangle_{\mathbb{C}^n}=\langle y,\mathbf{\Phi}^*\mathbf{x}\rangle_{L_2(d\mu)}$ for every $y\in L_2(d\mu)$ and $\mathbf{x}\in R(\mathbf{\Phi})$. It is easy to verify that $(\mathbf{\Phi}^*\mathbf{x})(\boldsymbol{\eta})=\mathbf{z}(\boldsymbol{\eta})^*\mathbf{x}$. We now have the following:

Proposition 28.

$$\mathbf{\Phi}\mathbf{\Phi}^*=\mathbf{K}$$

Proof. We have that for every $\mathbf{x}\in\mathbb{C}^n$,

$$
\begin{aligned}
\mathbf{\Phi}\mathbf{\Phi}^*\mathbf{x} &= \int_{\mathbb{R}^d}\mathbf{z}(\boldsymbol{\xi})(\mathbf{\Phi}^*\mathbf{x})(\boldsymbol{\xi})d\mu(\boldsymbol{\xi})\\
&= \int_{\mathbb{R}^d}\mathbf{z}(\boldsymbol{\xi})\mathbf{z}(\boldsymbol{\xi})^*\mathbf{x}d\mu(\boldsymbol{\xi})\\
&= \left(\int_{\mathbb{R}^d}\mathbf{z}(\boldsymbol{\xi})\mathbf{z}(\boldsymbol{\xi})^*d\mu(\boldsymbol{\xi})\right)\mathbf{x}=\mathbf{K}\mathbf{x}
\end{aligned}
$$

so $\mathbf{\Phi}\mathbf{\Phi}^*=\mathbf{K}$. □

We are now ready to prove the two lemmas.

Proof of Lemma 11. The minimizer of the right-hand side of (11) can be obtained from the usual normal equations, and simplified using the matrix inversion lemma for operators (Ogawa, 1988):

$$\begin{aligned} y^\star &= \sqrt{p(\boldsymbol{\eta})}(\boldsymbol{\Phi}^*\boldsymbol{\Phi} + \lambda \mathbf{I}_{L_2(d\mu)})^{-1}\boldsymbol{\Phi}^*\mathbf{z}(\boldsymbol{\eta}) \\ &= \sqrt{p(\boldsymbol{\eta})}\boldsymbol{\Phi}^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^* + \lambda \mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) \\ &= \sqrt{p(\boldsymbol{\eta})}\boldsymbol{\Phi}^*(\mathbf{K} + \lambda \mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})\,. \end{aligned}$$

So, $y^\star(\xi) = \sqrt{p(\boldsymbol{\eta})}\mathbf{z}(\xi)^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})$. We now have

$$\begin{aligned} \|y^\star\|^2_{L_2(d\mu)} &= p(\boldsymbol{\eta})\int_{\mathbb{R}^d} |\mathbf{z}(\boldsymbol{\xi})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})|^2 d\mu(\boldsymbol{\xi}) \\ &= p(\boldsymbol{\eta})\int_{\mathbb{R}^d} \mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\xi)\mathbf{z}(\boldsymbol{\xi})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) d\mu(\boldsymbol{\xi}) \\ &= p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\left(\int_{\mathbb{R}^d}\mathbf{z}(\boldsymbol{\xi})\mathbf{z}(\boldsymbol{\xi})^* d\mu(\boldsymbol{\xi})\right)(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) \\ &= p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{K}(\mathbf{K} + \lambda I_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) \\ &= p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}(\mathbf{K} + \lambda I_n - \lambda I_n)(\mathbf{K} + \lambda I_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) \\ &= p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) - \lambda p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-2}\mathbf{z}(\boldsymbol{\eta}) \end{aligned}$$

and

$$\begin{aligned} \|\boldsymbol{\Phi}y^\star - \sqrt{p(\boldsymbol{\eta})}\mathbf{z}(\boldsymbol{\eta})\|_2^2 &= p(\boldsymbol{\eta})\|\boldsymbol{\Phi}\boldsymbol{\Phi}^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) - \mathbf{z}(\boldsymbol{\eta})\|_2^2 \\ &= p(\boldsymbol{\eta})\|(\mathbf{K}(\mathbf{K} + \lambda\mathbf{I}_n)^{-1} - \mathbf{I}_n)\mathbf{z}(\boldsymbol{\eta})\|_2^2 \\ &= p(\boldsymbol{\eta})\|\big((\mathbf{K} + \lambda I_n - \lambda I_n)(\mathbf{K} + \lambda\mathbf{I}_n)^{-1} - \mathbf{I}_n\big)\mathbf{z}(\boldsymbol{\eta})\|_2^2 \\ &= p(\boldsymbol{\eta})\|\big(\lambda(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\big)\mathbf{z}(\boldsymbol{\eta})\|_2^2 \\ &= \lambda^2 p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-2}\mathbf{z}(\boldsymbol{\eta})\,. \end{aligned}$$

Now plugging these into (11) gives:

$$\begin{aligned} &\|y^\star\|^2_{L_2(d\mu)} + \lambda^{-1}\|\boldsymbol{\Phi}y^\star - \sqrt{p(\boldsymbol{\eta})}\mathbf{z}(\boldsymbol{\eta})\|_2^2 \\ &\quad = p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) - \lambda p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-2}\mathbf{z}(\boldsymbol{\eta}) \\ &\quad + \lambda p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-2}\mathbf{z}(\boldsymbol{\eta}) \\ &\quad = p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K} + \lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta}) \\ &\quad = \tau_\lambda(\boldsymbol{\eta}) \end{aligned}$$

□

Proof of Lemma 12. The optimization problem (11) can equivalently be reformulated as the following problem:

$$\begin{aligned} \tau(\boldsymbol{\eta}) = \text{minimum} \quad & \|y\|^2_{L_2(d\mu)} + \|\mathbf{u}\|_2^2 \\ y \in L_2(d\mu); \quad & \mathbf{u} \in \mathbb{C}^n \\ \text{subject to:} \quad & \boldsymbol{\Phi}y + \sqrt{\lambda}\mathbf{u} = \sqrt{p(\boldsymbol{\eta})}\mathbf{z}(\boldsymbol{\eta}) \end{aligned}$$

First we show that for any $\boldsymbol{\alpha} \in \mathbb{C}^n$, the argument of the minimization problem in (12) is no bigger than $\tau_\lambda(\boldsymbol{\eta})$. That is because for the optimal solution to above optimization, namely $\bar{\mathbf{u}}$ and $\bar{y}$, we have:

$$\boldsymbol{\Phi}\bar{y} + \sqrt{\lambda}\bar{\mathbf{u}} = \sqrt{p(\boldsymbol{\eta})}\mathbf{z}(\boldsymbol{\eta})$$

hence,

$$\begin{aligned}|\sqrt{p(\boldsymbol{\eta})}\boldsymbol{\alpha}^*\mathbf{z}(\boldsymbol{\eta})| &= |\boldsymbol{\alpha}^*(\boldsymbol{\Phi}\bar{y}+\sqrt{\lambda}\bar{\mathbf{u}})| \\ &= |\boldsymbol{\alpha}^*\boldsymbol{\Phi}\bar{y}+\boldsymbol{\alpha}^*\sqrt{\lambda}\bar{\mathbf{u}}| \\ &\le |\boldsymbol{\alpha}^*\boldsymbol{\Phi}\bar{y}| + |\boldsymbol{\alpha}^*\sqrt{\lambda}\bar{\mathbf{u}}| \\ &= |\langle\boldsymbol{\alpha},\boldsymbol{\Phi}\bar{y}\rangle_{\mathbb{C}^n}| + |\boldsymbol{\alpha}^*\sqrt{\lambda}\bar{\mathbf{u}}| \\ &= |\langle\boldsymbol{\Phi}^*\boldsymbol{\alpha},\bar{y}\rangle_{L_2(d\mu)}| + |\boldsymbol{\alpha}^*\sqrt{\lambda}\bar{\mathbf{u}}| \\ &\le \|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}\cdot\|\bar{y}\|_{L_2(d\mu)} + \sqrt{\lambda}\|\boldsymbol{\alpha}^*\|_2\cdot\|\bar{\mathbf{u}}\|_2\end{aligned}$$

where the last inequality follows from Cauchy-Schwarz inequality ($|\boldsymbol{\alpha}^*\boldsymbol{\Phi}\bar{y}| = |(\boldsymbol{\alpha}^*\boldsymbol{\Phi}\bar{y})^*| = |(\boldsymbol{\Phi}\bar{y})^*\boldsymbol{\alpha}| = |\langle\bar{y},\boldsymbol{\Phi}^*\boldsymbol{\alpha}\rangle_{L_2(d\mu)}| \le \|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}\cdot\|\bar{y}\|_{L_2(d\mu)}$). By another use of Cauchy-Schwarz we have:

$$\begin{aligned}p(\boldsymbol{\eta})|\boldsymbol{\alpha}^*\mathbf{z}(\boldsymbol{\eta})|^2 &\le \left(\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}\|\bar{y}\|_{L_2(d\mu)} + \sqrt{\lambda}\|\boldsymbol{\alpha}^*\|_2\cdot\|\bar{\mathbf{u}}\|_2\right)^2 \\ &\le \left(\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)} + \lambda\|\boldsymbol{\alpha}^*\|_2^2\right)\cdot\left(\|\bar{y}\|^2_{L_2(d\mu)} + \|\bar{\mathbf{u}}\|_2^2\right)\end{aligned}$$

therefore, for every $\boldsymbol{\alpha}\in\mathbb{C}^n$,

$$\frac{p(\boldsymbol{\eta})|\boldsymbol{\alpha}^*\mathbf{z}(\boldsymbol{\eta})|^2}{\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)} + \lambda\|\boldsymbol{\alpha}\|_2^2} \le \|\bar{y}\|^2_{L_2(d\mu)} + \|\bar{\mathbf{u}}\|_2^2 = \tau_\lambda(\boldsymbol{\eta}) \tag{18}$$

Now it is enough to show that at the optimal $\boldsymbol{\alpha}$ the dual problem gives the leverage scores. We show that $\bar{\boldsymbol{\alpha}} = \sqrt{p(\boldsymbol{\eta})}(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})$ matches the leverage scores. First note that for any $\boldsymbol{\alpha}\in\mathbb{C}^n$ we have

$$\begin{aligned}\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)} + \lambda\|\boldsymbol{\alpha}\|_2^2 &= \langle\boldsymbol{\Phi}^*\boldsymbol{\alpha},\boldsymbol{\Phi}^*\boldsymbol{\alpha}\rangle_{L_2(d\mu)} + \lambda\boldsymbol{\alpha}^*\boldsymbol{\alpha} \\ &= \langle\boldsymbol{\Phi}\boldsymbol{\Phi}^*\boldsymbol{\alpha},\boldsymbol{\alpha}\rangle_{\mathbb{C}^n} + \lambda\boldsymbol{\alpha}^*\boldsymbol{\alpha} \\ &= \langle\mathbf{K}\boldsymbol{\alpha},\boldsymbol{\alpha}\rangle_{\mathbb{C}^n} + \lambda\boldsymbol{\alpha}^*\boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^*(\mathbf{K}+\lambda\mathbf{I}_n)\boldsymbol{\alpha}\end{aligned}$$

Now by substituting $\bar{\boldsymbol{\alpha}} = \sqrt{p(\boldsymbol{\eta})}(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})$ we have:

$$\begin{aligned}\frac{p(\boldsymbol{\eta})|\bar{\boldsymbol{\alpha}}^*\mathbf{z}(\boldsymbol{\eta})|^2}{\|\boldsymbol{\Phi}^*\bar{\boldsymbol{\alpha}}\|^2_{L_2(d\mu)} + \lambda\|\bar{\boldsymbol{\alpha}}\|_2^2} &= \frac{p(\boldsymbol{\eta})^2|\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})|^2}{p(\boldsymbol{\eta})\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}(\mathbf{K}+\lambda\mathbf{I}_n)(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})} \\ &= p(\boldsymbol{\eta})|\mathbf{z}(\boldsymbol{\eta})^*(\mathbf{K}+\lambda\mathbf{I}_n)^{-1}\mathbf{z}(\boldsymbol{\eta})| \\ &= \tau_\lambda(\boldsymbol{\eta}) \end{aligned} \tag{19}$$

□

F. Proof of Theorem 13

Recall from Lemma 11 that

$$\tau_\lambda(\eta) = \min_{y\in L_2(d\mu)} \lambda^{-1}\|\boldsymbol{\Phi}y - \sqrt{p(\eta)}\mathbf{z}(\eta)\|_2^2 + \|y\|^2_{L_2(d\mu)} \tag{20}$$

To upper bound $\tau_\lambda(\eta)$ for any η, we will exhibit some test function, $y_\eta(\cdot)$, and compute the quantity under the minimum. $y_\eta(\cdot)$ will be a 'softened spike function' given by:

Definition 29 (Softened spike function)**.** For any η, and any u define:

$$y_{\eta,u}(t) = \frac{\sqrt{p(\eta)}}{p(t)}\cdot e^{-(t-\eta)^2u^2/4}\cdot v\cdot\mathbf{sinc}\left(v(t-\eta)\right) \tag{21}$$

where $v = 2(R+u\sqrt{2\log n_\lambda})$.

The reweighted function $g_{\eta,u}(t) = p(t) \cdot y_{\eta,u}(t)$ is just a Gaussian with standard deviation $\Theta(1/u)$ multiplied by a sinc function with width $\tilde{O}(1/(u+R))$, both centered at η. Taking the Fourier transform of this function yields a Gaussian with standard deviation $\Theta(u)$ convolved with a box of width $\tilde{O}(u) + R$. This width is wide enough such that when centered between $[-R, R]$ the box covers nearly all the mass of the Gaussian, and so the Fourier transform is nearly identically 1 on the range $[-R, R]$. Shifting by η, means that it is very close to a pure cosine wave with frequency η on this range, and hence makes the first term of (20) small. We make this argument formal below.

F.1. Bounding $\lambda^{-1}\|\mathbf{\Phi} y_{\eta,u} - \sqrt{p(\eta)}\mathbf{z}(\eta)\|_2^2$

Lemma 30 (Test Function Fourier Transform Bound)**.** *For any integer n, every parameter $0 < \lambda \le n$ and every η, u, and any kernel density function $p(\eta)$ if $x_j \in [-R, +R]$ for all $j \in [n]$ for any radius $R > 0$, then:*

$$\lambda^{-1}\|\mathbf{\Phi} y - \sqrt{p(\eta)}\mathbf{z}(\eta)\|_2^2 = \frac{1}{\lambda}\sum_{j=1}^{n}\left|\hat{g}_{\eta,u}(x_j) - \sqrt{p(\eta)} \cdot z(\eta)_j\right|^2 \le p(\eta). \tag{22}$$

where $g_{\eta,u}(t) \equiv p(t) y_{\eta,u}(t)$.

Proof. We have $g_{\eta,u}(t) = p(t)y_{\eta,u}(t) = p(\eta)e^{-(t-\eta)^2u^2/4} \cdot v \cdot \mathbf{sinc}\,(v(t-\eta))$ and $\hat{g}_{\eta,u}(x_j) = (\mathbf{\Phi} y)_j$. We thus have:

$$\begin{aligned}\hat{g}_{\eta,u}(x_j) &= \sqrt{p(\eta)}\int_{\mathbb{R}} e^{-2\pi i t x_j} e^{-(t-\eta)^2u^2/4} \cdot v \cdot \mathbf{sinc}\,(v(t-\eta))\,dt \\ &= \sqrt{p(\eta)}e^{-2\pi i x_j \eta}\int_{\mathbb{R}} e^{-2\pi i t x_j} e^{-t^2u^2/4} \cdot v \cdot \mathbf{sinc}\,(vt)\,dt \\ &= \sqrt{p(\eta)} \cdot z(\eta)_j \cdot h(x_j)\end{aligned} \tag{23}$$

where $h(x) = \frac{2\sqrt{\pi}}{u}e^{-4\pi^2x^2/u^2} * \mathrm{rect}_v(x)$ by the fact that multiplication in time domain becomes convolution in the Fourier domain (Claim 18), $\mathcal{F}(e^{-t^2u^2/4}) = \frac{2\sqrt{\pi}}{u}e^{-4\pi^2x^2/u^2}$, and $\mathcal{F}(v \cdot \mathbf{sinc}\,(vt)) = \mathrm{rect}_v(x)$.

For any x, we have $h(x) \le \int_{\mathbb{R}} \frac{2\sqrt{\pi}}{u}e^{-4\pi^2x^2/u^2} = 1$. Additionally, for any $x \in [-R, R]$ we have by Claim 25 and the fact that $v = 2R + 2u\sqrt{2\log n_\lambda}$:

$$\begin{aligned}h(x) &= \int_{x-\frac{v}{2}}^{x+\frac{v}{2}} \frac{2\sqrt{\pi}}{u}e^{-4\pi^2x^2/u^2}\,dx \\ &\ge 1 - 2\int_{v/2-R}^{\infty} \frac{2\sqrt{\pi}}{u}e^{-4\pi^2x^2/u^2}\,dx \\ &\ge 1 - \frac{1}{4\sqrt{\pi}} \cdot \frac{u}{v/2-R}e^{-4\pi^2(v/2-R)^2/u^2} && \text{(by second part of Claim 25)} \\ &\ge 1 - \frac{1}{4\sqrt{\pi}\sqrt{2\log n_\lambda}} \cdot \frac{1}{\sqrt{n_\lambda}} && \text{(since } v = 2R + 2u\sqrt{2\log n_\lambda}\text{)} \\ &\ge 1 - \frac{1}{\sqrt{n_\lambda}} && \text{(by assumption } n_\lambda \ge 2\text{).}\end{aligned}$$

Plugging into (23) gives

$$\begin{aligned}\left|\hat{g}_{\eta,u}(x_j) - \sqrt{p(\eta)} \cdot z(\eta)_j\right|^2 &= p(\eta)\,|h(x_j) - 1|^2 \\ &\le \frac{p(\eta)}{n_\lambda},\end{aligned}$$

and so,

$$\frac{1}{\lambda}\sum_{j=1}^{n}\left[\hat{g}(x_j) - \sqrt{p(\eta)} \cdot z(\eta)_j\right]^2 \le n_\lambda \cdot p(\eta) \cdot \frac{\lambda}{n} < p(\eta)$$

proving the claim. ∎

F.2. Bounding $\|y_{\eta,u}\|_{L_2(d\mu)}^2$

Having established Lemma 30, we note that showing that the weighted Fourier transform of $y_{\eta,u}$ is close to $\sqrt{p(\eta)}\mathbf{z}(\eta)$ reduces to bounding the norm of the test function. To that effect, we show the following:

Lemma 31 (Test Function ℓ_2 Norm Bound)**.** *For any integer n, any parameter $0 < \lambda \le \frac{n}{2}$, every $|\eta| \le 10\sqrt{\log n_\lambda}$, and every $2000 \log n_\lambda \le u \le 500 \log^{1.5} n_\lambda$, if $y_{\eta,u}(t)$ is defined as in (20), as per Definition 29, then we have*

$$\|y\|_{L_2(d\mu)}^2 \le 12\Big(R + u\sqrt{2\log n_\lambda}\Big) \tag{24}$$

Before proving Lemma 31, we first prove a claim:

Claim 32. *Suppose $|\eta| \le 100\sqrt{\log n_\lambda}$, and*

$$\eta - \frac{c\sqrt{\log n_\lambda}}{b} \le t \le \eta + \frac{c\sqrt{\log n_\lambda}}{b}$$

for some absolute constant $c > 0$. If $b \ge 100c \cdot \log n_\lambda$ then,

$$e^{-\frac{t^2}{2} + \frac{\eta^2}{2}} \le 3.$$

Proof. Let $\Delta = t - \eta$. Then, note that $|\Delta| \le c\sqrt{\log n_\lambda}/b$, and so,

$$\begin{aligned} e^{-\frac{t^2}{2} + \frac{\eta^2}{2}} &= e^{-\frac{(\Delta+\eta)^2}{2} + \frac{\eta^2}{2}} \\ &= e^{-\Delta\eta - \frac{\Delta^2}{2}} \\ &\le e^{|\Delta\eta| - \frac{\Delta^2}{2}} \\ &\le e^{|\Delta|\cdot|\eta|} \\ &\le e^{(c\sqrt{\log n_\lambda}/b)(100\sqrt{\log n_\lambda})} \\ &\le e \le 3, \end{aligned}$$

since $b \ge 100c \cdot \log n_\lambda$. □

Now, we are ready to prove Lemma 31:

Proof of Lemma 31. Recall that for the Gaussian kernel, we have $p(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2}$. We calculate:

$$\begin{aligned} \int_{\mathbb{R}} |y_{\eta,u}(t)|^2 d\mu(t) = p(\eta) \int_{\mathbb{R}} \sqrt{2\pi} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} \cdot v^2 \left(\mathbf{sinc}\left(v(t-\eta)\right)\right)^2 dt \\ = \sqrt{2\pi} p(\eta) \cdot v^2 \int_{\eta - \frac{20\sqrt{\log n_\lambda}}{u}}^{\eta + \frac{20\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} \left(\mathbf{sinc}\left(v(t-\eta)\right)\right)^2 dt \\ + \sqrt{2\pi} p(\eta) \cdot v^2 \int_{|t-\eta| \ge \frac{20\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} \left(\mathbf{sinc}\left(v(t-\eta)\right)\right)^2 dt \end{aligned} \tag{25}$$

For the integral over $|t - \eta| \ge 20\frac{\sqrt{\log n_\lambda}}{u}$ we have:

$$\begin{aligned} \int_{|t-\eta| \ge 20\frac{\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} \left(\mathbf{sinc}\left(v(t-\eta)\right)\right)^2 dt &\le \frac{1}{\left(v \cdot 20\frac{\sqrt{\log n_\lambda}}{u}\right)^2} \int_{|t-\eta| \ge 20\frac{\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} dt \\ &\le \frac{1}{v} \int_{|t-\eta| \ge 20\frac{\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} dt \end{aligned} \tag{26}$$

The first inequality above is because by definition of $\mathbf{sinc}\left(\cdot\right)$ we have the following for all $|t-\eta| \geq \frac{20\sqrt{\log n_\lambda}}{u}$:

$$|\mathbf{sinc}\left(v(t-\eta)\right)|^2 = \frac{\sin^2(\pi v(t-\eta))}{(\pi v(t-\eta))^2} \leq \frac{1}{(v(t-\eta))^2} \leq \frac{1}{\left(v \cdot \frac{20\sqrt{\log n_\lambda}}{u}\right)^2}$$

The last inequality in (26) is because of the following reason:

$$\begin{aligned}
\frac{1}{\left(v \cdot \frac{20\sqrt{\log n_\lambda}}{u}\right)^2} &= \frac{1}{v} \cdot \frac{1}{v \cdot \left(\frac{20\sqrt{\log n_\lambda}}{u}\right)^2} \\
&\leq \frac{1}{v} \cdot \frac{1}{800\left(\frac{\log^{1.5} n_\lambda}{u}\right)} \quad \text{(since } v = 2(R + u\sqrt{2\log n_\lambda}) \geq 2u\sqrt{2\log n_\lambda}\text{, see Definition 29)} \\
&\leq \frac{1}{v} \quad \text{(since } u \leq 500 \log^{1.5} n_\lambda\text{)}
\end{aligned}$$

Now note that $t^2 \leq 2(t-\eta)^2 + 2\eta^2$. We have the following for all $|t-\eta| \geq \frac{20\sqrt{\log n_\lambda}}{u}$:

$$\begin{aligned}
t^2 &\leq 2(t-\eta)^2 + 2\eta^2 \\
&\leq 2(t-\eta)^2 + 200\log n_\lambda \quad \text{(by the assumption } |\eta| \leq 10\sqrt{\log n_\lambda}\text{)} \\
&\leq 2(t-\eta)^2 + (t-\eta)^2u^2/2 \quad \text{(by the assumption } |t-\eta| \geq \frac{20\sqrt{\log n_\lambda}}{u}\text{)} \\
&\leq \frac{2}{3}(t-\eta)^2u^2
\end{aligned}$$

where the last inequality follows from $u \geq 2000 \log n_\lambda \geq 600$ (because $n_\lambda \geq 1/2$). Hence,

$$\begin{aligned}
\frac{1}{v}\int_{|t-\eta|\geq 20\frac{\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2}dt &\leq \frac{1}{v}\int_{|t-\eta|\geq 20\frac{\sqrt{\log n_\lambda}}{u}} e^{-(t-\eta)^2u^2/3}dt \\
&\leq \frac{1}{v} \cdot n_\lambda^{100}
\end{aligned} \tag{27}$$

Now, the first integral in (25):

$$\begin{aligned}
\int_{\eta-20\frac{\sqrt{\log n_\lambda}}{u}}^{\eta+20\frac{\sqrt{\log n_\lambda}}{u}} e^{t^2/2} \cdot e^{-(t-\eta)^2u^2/2} \left(\mathbf{sinc}\left(v(t-\eta)\right)\right)^2 dt &\leq 3e^{\frac{\eta^2}{2}} \int_{\mathbb{R}} \left(\mathbf{sinc}\left(v(t-\eta)\right)\right)^2 dt \\
&= \frac{3e^{\frac{\eta^2}{2}}}{v}.
\end{aligned} \tag{28}$$

where the inequality follows from Claim 32 with $c = 20$ and $b = u$, since, by assumption, $u \geq 2000\log n_\lambda$ and $|t| \leq |\eta| + |t-\eta| \leq 10\sqrt{\log n_\lambda} + \frac{20}{u}\sqrt{\log n_\lambda} \leq 100\sqrt{\log n_\lambda}$ whenever $t \in \left[\eta - \frac{20\log n_\lambda}{u}, \eta + \frac{20\log n_\lambda}{u}\right]$.

By incorporating (27) and (28) into (25) we have:

$$\int_{\mathbb{R}} |y_{\eta,u}(t)|^2 dt \leq \sqrt{2\pi}p(\eta) \cdot v^2\left(\frac{1}{v} \cdot n_\lambda^{-100} + \frac{3e^{\frac{\eta^2}{2}}}{v}\right) \leq 6v \tag{29}$$

where the last inequality uses that $\sqrt{2\pi}p(\eta) = \frac{\sqrt{2\pi}}{\sqrt{2\pi}}e^{-\eta^2/2} \leq 1$. □

Proof of Theorem 13. By the assumptions of the theorem n is an integer, parameter $0 < \lambda \le n/2$, and $R > 0$, and all $x_1, ..., x_n \in [-R, R]$ and $p(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}$, therefore all the preconditions of Lemmas 31, and 30 are satisfied and hence the lemmas go through and the upper bounds in (22) and (24) hold true. The theorem follows by setting $u = 2000 \log n_\lambda$ and then plugging upper bounds (22) and (24) into (20). ☐

G. Proof of Theorem 14

With the choice of the Gaussian kernel with $\sigma = (2\pi)^{-1}$ we have $p(\eta) = (2\pi)^{-1/2} \exp(-\eta^2/2)$. Recall from Lemma 12 that

$$\tau_\lambda(\eta) = \max_{\boldsymbol{\alpha} \in \mathbb{C}^n} \frac{p(\eta) \cdot |\boldsymbol{\alpha}^* \mathbf{z}(\eta)|^2}{\|\boldsymbol{\Phi}^* \boldsymbol{\alpha}\|_{L_2(d\mu)}^2 + \lambda \|\boldsymbol{\alpha}\|_2^2}. \tag{30}$$

In particular, this gives us a method of bounding the leverage scores from below, namely, by exhibiting some $\boldsymbol{\alpha}$ and computing the quantity under the maximum.

The rest of this section is organized as follows. In Section G.1, we construct our candidate set of data points $x_1, x_2, \ldots, x_n$ along with the vector $\boldsymbol{\alpha}$. In particular, $\boldsymbol{\alpha}$ will be chosen to be a vector of samples of a function $f_{\Delta,b,v}$ at each of the data points. Section G.2 then describes basic Fourier properties of the function $f_{\Delta,b,v}$ and $\boldsymbol{\alpha}$ that we will require later. The remaining sections then bound each of the relevant quantities that appear in (30) for our specific choice of $x_1, x_2, \ldots, x_n$ and $\boldsymbol{\alpha}$. In particular, Section G.3 shows a lower bound for $\boldsymbol{\alpha}^* \mathbf{z}(\eta)$, while Section G.4 shows an upper bound for $\|\boldsymbol{\alpha}\|_2^2$ and Section G.5 shows an upper bound for $\|\boldsymbol{\Phi}^* \boldsymbol{\alpha}\|_{L_2(d\mu)}^2$.

G.1. Construction of Data Point Set and the Vector of Coefficients $\boldsymbol{\alpha}$

Definition 33. For parameters $\Delta, b > 0$ and $v > 0$, let the function $f_{\Delta,b,v}$ be defined as follows:

$$\begin{aligned} f_{\Delta,b,v}(x) &= 2\cos(2\pi\Delta x) \left(\frac{1}{\sqrt{2\pi} b} e^{-(\cdot)^2/2b^2} * \mathrm{rect}_v \right)(x) \\ &= 2\cos(2\pi\Delta x) \int_{x-\frac{v}{2}}^{x+\frac{v}{2}} \frac{1}{\sqrt{2\pi} b} e^{-t^2/2b^2}\, dt \end{aligned}$$

Lemma 34. *For any $\Delta > 0$, $v > 0$, and $b > 0$, if we define the function $f_{\Delta,b,v}$ as in Definition 33, then*

$$\mathcal{F}\left(f_{\Delta,b,v}\right)(z) = e^{-2\pi^2 b^2 (z-\Delta)^2} (v \cdot \mathbf{sinc}\left(v(z-\Delta)\right) + e^{-2\pi^2 b^2 (z+\Delta)^2} (v \cdot \mathbf{sinc}\left(v(z+\Delta)\right)).$$

Proof. Note that

$$\mathcal{F}\left(\frac{1}{\sqrt{2\pi} b} e^{-(\cdot)^2/2b^2} \right)(z) = e^{-2\pi^2 b^2 z^2}.$$

Thus, by the convolution theorem (see Claim 18),

$$\mathcal{F}\left(\frac{1}{\sqrt{2\pi} b} e^{-(\cdot)^2/2b^2} * \mathrm{rect}_v \right)(z) = e^{-2\pi^2 b^2 z^2} \cdot v \cdot \mathbf{sinc}\left(v(z)\right).$$

Now by the duality of phase shift in time domain and frequency shift in the Fourier domain,

$$\begin{aligned} \mathcal{F}(f_{\Delta,b,v})(z) &= \mathcal{F}\left((e^{2\pi i \Delta(\cdot)} + e^{-2\pi i \Delta(\cdot)}) \cdot \left(\frac{1}{\sqrt{2\pi} b} e^{-(\cdot)^2/2b^2} * \mathrm{rect}_v \right) \right)(z) \\ &= \mathcal{F}\left(\frac{1}{\sqrt{2\pi} b} e^{-(\cdot)^2/2b^2} * \mathrm{rect}_v \right)(z-\Delta) + \mathcal{F}\left(\frac{1}{\sqrt{2\pi} b} e^{-(\cdot)^2/2b^2} * \mathrm{rect}_v \right)(z+\Delta) \\ &= e^{-2\pi^2 b^2 (z-\Delta)^2} \cdot v \cdot \mathbf{sinc}\left(v(z-\Delta)\right) + e^{-2\pi^2 b^2 (z+\Delta)^2} \cdot v \cdot \mathbf{sinc}\left(v(z+\Delta)\right). \end{aligned}$$

☐

Intuition for Theorem 14 If, instead of a discrete set of data points, we had a continuum of points, $\boldsymbol{\alpha}$ would be a function (or, alternatively, an infinite-dimensional vector coresponding to the evaluation of the function on the continuum of points). The intuition is that in this case, we would essentially like to choose $\boldsymbol{\alpha}$ to be the function $f_{\Delta,b,v}$ for some suitable choice of parameters Δ, b, v. In this case, the computation of bounds for the various quantities appearing in (30) would be relatively simple and involve bounding integrals. However, since our data points are actually discrete and $\boldsymbol{\alpha}$ is finite-dimensional, we must instead choose $\boldsymbol{\alpha}$ to be the vector of samples of $f_{\Delta,b,v}$ on the data points, and the bounds we deduce require computing Fourier transforms of $f_{\Delta,b,v}$ multiplied by suitable Dirac combs (see Lemma 36). Computation of the necessary bounds is further complicated by the fact that the data points are bounded in $[-R, R]$, which requires us to truncate the aforementioned Dirac combs and have appropriate Fourier tail bounds (see Lemma 37).

Let us provide some intuition about the quantities $|\boldsymbol{\alpha}^*\mathbf{z}(\eta)|^2$, $\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)}$ and $\|\boldsymbol{\alpha}\|_2^2$ that arise in (30) along these lines. If we have $\approx 2R$ equally spaced data points between $-R$ and R, then note that the points are separated by distance ≈ 1. This approximately corresponds to dealing with the continuous case in which $\boldsymbol{\alpha}$ is a function $f_{\Delta,b,v}$ and, therefore, sums in the discrete case can be approximated by corresponding integrals over continuous functions. Suppose $\Delta = \eta$ and $v = R$.

Note that the quantity $\boldsymbol{\alpha}^*\mathbf{z}(\xi)$ corresponds to

$$\begin{aligned}\boldsymbol{\alpha}^*\mathbf{z}(\xi) &\approx \int_{-\infty}^{\infty} f_{\eta,b,R}(x)e^{-2\pi i\xi x}\,dx \\ &\approx \mathcal{F}(f_{\eta,b,R})(\xi) \\ &= e^{-2\pi^2b^2(\xi-\eta)^2}\cdot R\cdot \mathbf{sinc}\left(R(\xi-\eta)\right) + e^{-2\pi^2b^2(\xi+\eta)^2}\cdot R\cdot \mathbf{sinc}\left(R(\xi+\eta)\right). \end{aligned} \tag{31}$$

Thus, $\boldsymbol{\alpha}^*\mathbf{z}(\xi)$ (which we bound rigorously in Section G.3) can be approximated as follows:

$$\boldsymbol{\alpha}^*\mathbf{z}(\xi) \approx R(1 + e^{-8\pi^2b^2\eta^2}\mathbf{sinc}\left(2R\eta\right)) \approx \Omega(R), \tag{32}$$

where the last transition uses the fact that $\mathbf{sinc}\left(\cdot\right) \geq -1/4$. Next, note that the quantity $\|\boldsymbol{\alpha}\|_2^2$ (which we bound rigorously in Section G.4) is roughly

$$\begin{aligned}\|\boldsymbol{\alpha}\|_2^2 \approx \int_{-\infty}^{\infty} f_{\eta,b,R}(x)^2\,dx &= \int_{-\infty}^{\infty} 4\cos^2(2\pi\eta x)\left(\int_{x-\frac{R}{2}}^{x+\frac{R}{2}} \frac{1}{\sqrt{2\pi}b}e^{-t^2/2b^2}\,dt\right)^2 dx \\ &\approx 4\int_{-\frac{3R}{2}}^{\frac{3R}{2}}\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}b}e^{-t^2/2b^2}\,dt\right)^2 dx \\ &\approx O(R). \end{aligned} \tag{33}$$

Finally, note that $\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)}$ (which we bound rigorously in Section G.5) is roughly

$$\begin{aligned}\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)} &\approx \int_{-\infty}^{\infty} |\boldsymbol{\alpha}^*\mathbf{z}(\xi)|^2\cdot\frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}\,d\xi \\ &\approx \int_{-\infty}^{\infty}\left(e^{-2\pi^2b^2(\xi-\eta)^2}\cdot R\cdot\mathbf{sinc}\left(R(\xi-\eta)\right) + e^{-2\pi^2b^2(\xi+\eta)^2}\cdot R\cdot\mathbf{sinc}\left(R(\xi+\eta)\right)\right)^2\cdot\frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}\,d\xi \\ &\approx \frac{1}{\sqrt{2\pi}}e^{-\eta^2/2}R^2\int_{-\infty}^{\infty}\mathbf{sinc}\left(R(\xi-\eta)\right)^2\,d\xi \\ &\approx O(p(\eta)R), \end{aligned} \tag{34}$$

using (31).

Now, going back to the discrete case, consider what happens if we scale up the number of points from $2R$ to n, keeping the points evenly spaced in the interval $[-R, R]$. In this case, the spacing between points decreases by a factor of $\gamma \approx n/2R$. Thus, this corresponds to the measure of integration over $\mathbb{R}$ scaling up by a factor of γ. Hence, $|\boldsymbol{\alpha}^*\mathbf{z}(\eta)|$ and $\|\boldsymbol{\alpha}\|_2^2$ can be expected to scale up by a factor of γ, while $\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)}$ would scale up by a factor of γ^2. Thus, along with (32), (33), and (34), we get that

$$\frac{p(\eta)\cdot|\boldsymbol{\alpha}^*\mathbf{z}(\eta)|^2}{\|\boldsymbol{\Phi}^*\boldsymbol{\alpha}\|^2_{L_2(d\mu)} + \lambda\|\boldsymbol{\alpha}\|_2^2} \approx \frac{(\gamma R)^2p(\eta)}{\gamma^2p(\eta)R + \lambda\gamma R} \approx R\cdot\frac{p(\eta)}{p(\eta)+\lambda/\gamma} \approx R\cdot\frac{p(\eta)}{p(\eta)+2Rn_\lambda^{-1}},$$

which is within a constant factor of the expression in Theorem 14.

Definition 35 (Construction of data points and $\boldsymbol{\alpha}$)**.** We first define the set of data points x_j for $j = 1, 2, \ldots, n$ for odd n as follows:

$$x_j = \left(j - \frac{n+1}{2}\right) \cdot \frac{2R}{n}$$

Thus, the data points are on a grid of width $\frac{2R}{n}$ extending from $-R$ to R.

The vector $\boldsymbol{\alpha}$ is chosen to be the tuple of evaluations of $f_{\eta,b,v}$ at the individual x_j, for some parameters b, v, and η. More specifically, for $1 \leq j \leq n$, we define

$$\begin{aligned} \boldsymbol{\alpha}_j &= f_{\eta,b,v}(x_j) \\ &= 2\cos(2\pi\eta x_j) \int_{x_j - \frac{v}{2}}^{x_j + \frac{v}{2}} \frac{1}{\sqrt{2\pi} b} e^{-t^2/2b^2} \, dt. \end{aligned} \tag{35}$$

G.2. Basic Properties of $f_{\Delta,b,v}$ and $\boldsymbol{\alpha}$

By the Nyquist-Shannon sampling theorem, we have the following lemma.

Lemma 36. *For any parameters $\Delta > 0$, $v > 0$, and $b > 0$, if we define the function $f_{\Delta,b,v}$ as in Definition 33, then for any $w > 0$,*

$$\begin{aligned} \mathcal{F}\left(f_{\Delta,b,v}(\cdot) \cdot \sum_{j=-\infty}^{\infty} \delta(\cdot - jw)\right)(z) = w^{-1} v \sum_{j=-\infty}^{\infty} e^{-2\pi^2 b^2 (z - jw^{-1} - \Delta)^2} \cdot \mathbf{sinc}\left(v(z - jw^{-1} - \Delta)\right) \\ + w^{-1} v \sum_{j=-\infty}^{\infty} e^{-2\pi^2 b^2 (z - jw^{-1} + \Delta)^2} \cdot \mathbf{sinc}\left(v(z - jw^{-1} + \Delta)\right). \end{aligned}$$

Proof. By the Nyquist-Shannon sampling theorem, we have

$$\begin{aligned} \mathcal{F}\left(f_{\Delta,b,v}(\cdot) \sum_{j=-\infty}^{\infty} \delta(\cdot - jw)\right)(z) &= \left(w^{-1} \sum_{j=-\infty}^{\infty} \delta(\cdot - jw^{-1}) * \mathcal{F}(f_{\Delta,b,v})(\cdot)\right)(z) \\ &= \sum_{j=-\infty}^{\infty} w^{-1} \mathcal{F}(f_{\Delta,b,v})(z - jw^{-1}). \end{aligned} \tag{36}$$

Thus, by Lemma 34, we find that (36) can be written as

$$\begin{aligned} \sum_{j=-\infty}^{\infty} w^{-1} \mathcal{F}(f_{\Delta,b,v})(z - jw^{-1}) = w^{-1} \sum_{j=-\infty}^{\infty} e^{-2\pi^2 b^2 (z - jw^{-1} - \Delta)^2} \cdot v \cdot \mathbf{sinc}\left(v(z - jw^{-1} - \Delta)\right) \\ + w^{-1} \sum_{j=-\infty}^{\infty} e^{-2\pi^2 b^2 (z - jw^{-1} + \Delta)^2} \cdot v \cdot \mathbf{sinc}\left(v(z - jw^{-1} + \Delta)\right), \end{aligned}$$

which completes the proof. □

Lemma 37. *For every odd integer $n \geq 3$ and parameters $0 < \lambda \leq \frac{n}{2}$, $\eta > 0$, $v \leq R$, and $b \leq \frac{R}{4\sqrt{\log n_\lambda}}$, if we define the function $f_{\Delta,b,v}$ as in Definition 33, then*

$$\left| \mathcal{F}\left(\sum_{|j| > \frac{n}{2}} f_{\eta,b,v}\left(j \cdot \frac{2R}{n}\right) \cdot \delta\left(\cdot - j \cdot \frac{2R}{n}\right)\right)(z) \right| \leq \sqrt{\lambda n}$$

for all z.

Proof. By definition of $f_{\eta,b,v}$, we have the following for all x:

$$|f_{\eta,b,v}(x)| \le \int_{x-\frac{v}{2}}^{x+\frac{v}{2}} \frac{2}{\sqrt{2\pi}b} e^{-t^2/2b^2}\, dt.$$

Therefore, if $|j| > \frac{n}{2}$, then

$$\begin{aligned}
\left| f_{\eta,b,v}\left(j \cdot \frac{2R}{n}\right)\right| &\le \frac{2}{\sqrt{2\pi}b} \int_{j\cdot\frac{2R}{n}-\frac{v}{2}}^{\infty} e^{-t^2/2b^2}\, dt \\
&\le \frac{2}{\sqrt{2\pi}b} \int_{\frac{jR}{n}}^{\infty} e^{-t^2/2b^2}\, dt \\
&\le \frac{2}{\sqrt{2\pi}} \cdot \frac{nb}{jR} \cdot e^{-\frac{1}{2}\cdot\left(\frac{jR}{nb}\right)^2} \\
&\le \frac{2b}{R} \cdot e^{-\frac{1}{2}\cdot\left(\frac{jR}{nb}\right)^2},
\end{aligned} \tag{37}$$

where we have used the fact that $j \cdot \frac{2R}{n} - \frac{v}{2} \ge j \cdot \frac{2R}{n} - \frac{R}{2} \ge j \cdot \frac{R}{n}$, along with Claim 25. Therefore, again using Claim 25, we have

$$\begin{aligned}
\left| \mathcal{F}\left(\sum_{|j|>\frac{n}{2}} f_{\eta,b,v}\left(j\cdot\frac{2R}{n}\right)\cdot\delta\left(\cdot - j\cdot\frac{2R}{n}\right)\right)(z)\right| &\le \sum_{|j|>\frac{n}{2}} \left| f_{\eta,b,v}\left(j\cdot\frac{2R}{n}\right)\right| \\
&\le \sum_{|j|>\frac{n}{2}} \frac{2b}{R}\cdot e^{-\frac{1}{2}\cdot\left(\frac{jR}{nb}\right)^2} \\
&\le \frac{2b}{R}\cdot\left(\frac{nb}{R}\int_{\frac{(n-1)R}{2nb}}^{\infty} e^{-t^2/2}\, dt\right) \\
&\le \frac{n}{4\log n_\lambda}\cdot\int_{\sqrt{\log n_\lambda}}^{\infty} e^{-t^2/2}\, dt \\
&\le \frac{n}{4\log n_\lambda}\cdot\frac{1}{\sqrt{\log n_\lambda}}\cdot e^{-\frac{1}{2}\cdot\left(\sqrt{\log n_\lambda}\right)^2} \\
&\le \frac{1}{4\log^{3/2}(n_\lambda)}\cdot\sqrt{\lambda n} \\
&\le \sqrt{\lambda n},
\end{aligned}$$

since $n \ge 3$, $R \ge 4b\sqrt{\log n_\lambda}$, and $\lambda \le n/2$. ☐

Lemma 38. *For every odd integer $n \ge 3$, any parameter $0 < \lambda \le \frac{n}{2}$, every frequency η and ξ, and any parameter $v \le R$ and $b \le \frac{R}{4\sqrt{\log n_\lambda}}$, if $\boldsymbol{\alpha}$ is defined as in (35) of Definition 35, then we have,*

$$\begin{aligned}
&\left| \boldsymbol{\alpha}^*\mathbf{z}(\xi) - \frac{nv}{2R}\sum_{j=-\infty}^{\infty}\left(e^{-2\pi^2b^2(\xi-\frac{jn}{2R}-\eta)^2}\,\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right) + e^{-2\pi^2b^2(\xi-\frac{jn}{2R}+\eta)^2}\,\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}+\eta)\right)\right)\right| \\
&\qquad \le \sqrt{\lambda n}.
\end{aligned} \tag{38}$$

Proof. Note that

$$\begin{aligned}
\boldsymbol{\alpha}^*\mathbf{z}(\xi) &= \sum_{j=1}^{n} \boldsymbol{\alpha}_j e^{-2\pi i x_j \xi} \\
&= \sum_{|j|\le \frac{n}{2}} f_{\eta,b,v}(2Rj/n)\cdot e^{-2\pi i(\frac{2Rj}{n})\xi} \\
&= \mathcal{F}\left(\sum_{|j|\le\frac{n}{2}} f_{\eta,b,v}(2Rj/n)\cdot\delta\left(\cdot-\frac{2Rj}{n}\right)\right)(\xi) \\
&= \mathcal{F}\left(\sum_{j=-\infty}^{\infty} f_{\eta,b,v}(\cdot)\cdot\delta\left(\cdot-\frac{2Rj}{n}\right)\right)(\xi) - \mathcal{F}\left(\sum_{|j|>\frac{n}{2}} f_{\eta,b,v}\left(\frac{2Rj}{n}\right)\cdot\delta\left(\cdot-\frac{2Rj}{n}\right)\right)(\xi). \qquad (39)
\end{aligned}$$

By Lemma 36 (applied with $w = 2R/n$), we have the following expression for the first term in (39):

$$\begin{aligned}
\mathcal{F}\left(\sum_{j=-\infty}^{\infty} f_{\eta,b,v}(\cdot)\cdot\delta\left(\cdot-\frac{2Rj}{n}\right)\right)(\xi) = \frac{nv}{2R}\sum_{j=-\infty}^{\infty} e^{-2\pi^2b^2(\xi-\frac{jn}{2R}-\eta)^2}\cdot\mathbf{sinc}\left(v\left(\xi-\frac{jn}{2R}-\eta\right)\right) \\
+\frac{nv}{2R}\sum_{j=-\infty}^{\infty} e^{-2\pi^2b^2(\xi-\frac{jn}{2R}+\eta)^2}\cdot\mathbf{sinc}\left(v\left(\xi-\frac{jn}{2R}+\eta\right)\right). \qquad (40)
\end{aligned}$$

Now, by the assumption that $R \ge 4b\sqrt{\log n_\lambda}$ and $v \le R$, it follows from Lemma 37 that the second term in (39) can be bounded as

$$\left|\mathcal{F}\left(\sum_{|j|>\frac{n}{2}} f_{\eta,b,v}(2Rj/n)\cdot\delta\left(\cdot-\frac{2Rj}{n}\right)\right)(\xi)\right| \le \sqrt{\lambda n}. \qquad (41)$$

Thus, the desired result follows by combining (39), (40), and (41). □

G.3. Bounding $\boldsymbol{\alpha}^*\mathbf{z}(\eta)$

Lemma 39. *For every odd integer $n \ge 17$, any parameter $0 < \lambda \le (\frac{v}{R})^2\cdot n/16$, every frequency $|\eta| \le \frac{n}{10R}$, and any parameter $v \le R$ and $\frac{R}{2\sqrt{n}} \le b \le \frac{R}{4\sqrt{\log(n_\lambda)}}$, if $\boldsymbol{\alpha}$ is defined as in (35) of Definition 35, then we have*

$$|\boldsymbol{\alpha}^*\mathbf{z}(\eta)| \ge \frac{nv}{5R}.$$

Proof. Since $v \le R$ and $b \le \frac{R}{4\sqrt{\log(n_\lambda)}}$ and $\lambda \le n/2$, Lemma 38 implies that

$$\begin{aligned}
&\left|\boldsymbol{\alpha}^*\mathbf{z}(\eta) - \frac{nv}{2R}\sum_{j=-\infty}^{\infty}\left(e^{-2\pi^2b^2(-\frac{jn}{2R})^2}\mathbf{sinc}\left(v(-jn/2R)\right) + e^{-2\pi^2b^2(2\eta-\frac{jn}{2R})^2}\mathbf{sinc}\left(v(2\eta-jn/2R)\right)\right)\right| \\
&\qquad \le \sqrt{\lambda n}. \qquad (42)
\end{aligned}$$

Hence,

$$
\begin{aligned}
|\boldsymbol{\alpha}^* \mathbf{z}(\eta)| \geq \frac{nv}{2R} & \left| \sum_{j=-\infty}^{\infty} \left(e^{-2\pi^2 b^2 (-\frac{jn}{2R})^2} \mathbf{sinc}\left(v(-jn/2R)\right) \right.\right. \\
& \left.\left. + e^{-2\pi^2 b^2 (2\eta - \frac{jn}{2R})^2} \mathbf{sinc}\left(v(2\eta - jn/2R)\right) \right) \right| - \sqrt{\lambda n} \\
& \geq \frac{nv}{2R} e^{-2\pi^2 b^2 (0)^2} \mathbf{sinc}\left(v(0)\right) + \frac{nv}{2R} e^{-2\pi^2 b^2 (2\eta)^2} \mathbf{sinc}\left(v(2\eta)\right) \\
& \quad - \frac{nv}{2R} \sum_{|j| \geq 1} \left(e^{-2\pi^2 b^2 (-\frac{jn}{2R})^2} + e^{-2\pi^2 b^2 (2\eta - \frac{jn}{2R})^2} \right) - \sqrt{\lambda n} \\
& \geq \frac{3}{4} \left(\frac{nv}{2R} \right) - \frac{nv}{2R} \sum_{|j| \geq 1} \left(e^{-2\pi^2 b^2 (-\frac{jn}{2R})^2} + e^{-2\pi^2 b^2 (2\eta - \frac{jn}{2R})^2} \right) - \sqrt{\lambda n}, \qquad (43)
\end{aligned}
$$

since $|\mathbf{sinc}\,(\cdot)\,| \leq 1$ and $\mathbf{sinc}\,(\cdot) \geq -\frac{1}{4}$.

Now we show that $\sum_{|j|\geq 1} \left(e^{-2\pi^2 b^2 (-\frac{jn}{2R})^2} + e^{-2\pi^2 b^2 (2\eta - \frac{jn}{2R})^2} \right)$ is small. Note that by the assumption of $b \geq \frac{R}{2\sqrt{n}}$, we have $e^{-2\pi^2 b^2 (-\frac{jn}{2R})^2} \leq e^{-jn}$ for all $|j| \geq 1$. Also recall that $|\eta| \leq \frac{n}{10R}$, and so, $(2\eta - \frac{jn}{2R})^2 \geq (\frac{jn}{4R})^2$ for all $|j| \geq 1$. Hence, we have

$$
\begin{aligned}
\sum_{|j| \geq 1} \left(e^{-2\pi^2 b^2 (-\frac{jn}{2R})^2} + e^{-2\pi^2 b^2 (2\eta - \frac{jn}{2R})^2} \right) & \leq \sum_{|j| \geq 1} \left(e^{-|j|n} + e^{-\frac{|j|n}{4}} \right) \\
& \leq 5 e^{-\frac{n}{4}} \qquad (44)
\end{aligned}
$$

by assumption $n \geq 17$. The lemma follows by combining (43) and (44). □

G.4. Bounding $\|\boldsymbol{\alpha}\|_2^2$

Lemma 40. *For every odd integer n and parameters $b, \eta, v > 0$, if $\boldsymbol{\alpha}$ is defined as in (35) of Definition 35, then we have*

$$
\|\boldsymbol{\alpha}\|_2^2 \leq 4n.
$$

Now we are ready for the proof of Lemma 40.

Proof of Lemma 40. Let $w = 2R/n$. Then, we observe that

$$
\begin{aligned}
\|\boldsymbol{\alpha}\|_2^2 &= \sum_{j=1}^{n} \alpha_j^2 \\
&\leq \sum_{|j| \leq \frac{n-1}{2}} \left(\frac{2}{\sqrt{2\pi} b} \cos(2\pi j w \eta) \int_{jw - \frac{v}{2}}^{jw + \frac{v}{2}} e^{-x^2/2b^2} \right)^2 \\
&\leq \sum_{|j| \leq \frac{n-1}{2}} \left(\frac{2}{\sqrt{2\pi} b} \cos(2\pi j w \eta) \int_{-\infty}^{\infty} e^{-x^2/2b^2} \right)^2 \\
&\leq \sum_{|j| \leq \frac{n-1}{2}} \left(\frac{2}{\sqrt{2\pi} b} \int_{-\infty}^{\infty} e^{-x^2/2b^2} \right)^2
\end{aligned}
$$

because $|\cos(\cdot)| \leq 1$. Hence,

$$
\begin{aligned}
\|\boldsymbol{\alpha}\|_2^2 &\leq \sum_{|j| \leq \frac{n-1}{2}} 4 \\
&= 4n \qquad (45)
\end{aligned}
$$

as desired. □

G.5. Bounding $\|\mathbf{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}^2$

Note that all the results so far hold for any kernel $p(\eta)$ and are independent of the kernel function. Now, we upper bound $\|\mathbf{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}$. This quantity depends on the particular choice of kernel, which is assumed to be Gaussian.

Lemma 41. *For every odd integer $n \geq 17$, any parameter $\frac{10}{n} < \lambda \leq \frac{n}{2}$, every $|\eta| \leq 100\sqrt{\log n_\lambda}$, and any $1000 \log^{1.5} n_\lambda \leq R \leq \frac{n}{500\sqrt{\log n_\lambda}}$, and $\frac{R}{2\sqrt{n}} \leq b \leq \frac{R}{4\sqrt{\log n_\lambda}}$, if $\boldsymbol{\alpha}$ is defined as in (35) of Definition 35 with parameter $v = R$, then for the Gaussian kernel with $p(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}$, we have:*

$$\|\mathbf{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}^2 \leq 6\frac{n^2}{R} \cdot p(\eta) + 3\lambda n. \tag{46}$$

Proof. Recall from Lemma 38 that:

$$\begin{aligned}|\boldsymbol{\alpha}^*\mathbf{z}(\xi)|^2 &\leq \Bigg|\frac{nv}{2R}\sum_{j=-\infty}^{\infty}\left(e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}-\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right.\\ &\qquad\left.+\,e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}+\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}+\eta)\right)+\sqrt{\lambda n}\right)\Bigg|^2\\ &\leq \frac{n^2}{2}\Bigg|\sum_{j=-\infty}^{\infty}\left(e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}-\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right.\\ &\qquad\left.+\,e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}+\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}+\eta)\right)\right)\Bigg|^2+2\left(\sqrt{\lambda n}\right)^2.\end{aligned}$$

Now, by the definition of the $L_2(d\mu)$ norm, $\|\mathbf{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}^2 = \int_{-\infty}^{\infty} |\boldsymbol{\alpha}^* z(\xi)|^2 p(\xi)\, d\xi$, and so, we have

$$\begin{aligned}\|\mathbf{\Phi}^*\boldsymbol{\alpha}\|_{L_2(d\mu)}^2 &\leq \int_{-\infty}^{+\infty}\frac{n^2}{2}\left(\sum_{j=-\infty}^{\infty}e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}-\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right.\\ &\qquad\left.+\,e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}+\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}+\eta)\right)\right)^2 p(\xi)\,d\xi+\int_{-\infty}^{\infty}2\left(\sqrt{\lambda n}\right)^2 p(\xi)\,d\xi\\ &\leq \int_{-\infty}^{+\infty}n^2\left(\sum_{j=-\infty}^{\infty}e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}-\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 p(\xi)\,d\xi\\ &\qquad+\int_{-\infty}^{+\infty}n^2\left(\sum_{j=-\infty}^{\infty}e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}+\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}+\eta)\right)\right)^2 p(\xi)\,d\xi+2\lambda n\\ &= 2n^2\int_{-\infty}^{\infty}\left(\sum_{j=-\infty}^{\infty}e^{-2\pi^2 b^2(\xi-\frac{jn}{2R}-\eta)^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 p(\xi)\,d\xi+2\lambda n,\end{aligned}\tag{47}$$

where we have used the inequality $(a_1 + a_2)^2 \leq 2a_1^2 + 2a_2^2$ in the second step, and the last equality occurs because the kernel probability distribution function $p(\xi)$ is symmetric in our case, along with the fact that the underlying sum is over all j. Now, the integral in (47) can be split into two integrals as follows:

$$\int_{-\infty}^{\infty} p(\xi)\left(\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\cdot\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 d\xi$$
$$= \int_{-10\sqrt{\log n_\lambda}}^{10\sqrt{\log n_\lambda}} p(\xi)\left(\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\cdot\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 d\xi$$
$$+ \int_{|\xi|\ge 10\sqrt{\log n_\lambda}} p(\xi)\left(\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\cdot\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 d\xi. \tag{48}$$

First, we consider the case in which $|\xi|\le 10\sqrt{\log n_\lambda}$. By the assumption of the lemma, $|\eta|\le 100\sqrt{\log n_\lambda}$, and hence, $|\xi-\eta|\le 110\sqrt{\log n_\lambda}$. This implies that $|\xi-\eta|\le\frac{1}{2}(\frac{n}{2R})$, since we are assuming that $R\le\frac{n}{500\sqrt{\log(n/\lambda)}}$. Therefore, for any integer $j\neq 0$,

$$e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\le e^{-(\frac{jn}{R})^2b^2}.$$

Hence, we have

$$\sum_{|j|\ge1} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\le\sum_{|j|\ge1} e^{-(\frac{jn}{R})^2b^2}$$
$$\le\sum_{|j|\ge1} e^{-j(\frac{n}{R})^2b^2}$$
$$\le 3e^{-n/4}, \tag{49}$$

where we used assumptions $b\ge\frac{R}{2\sqrt{n}}$ and $n\ge 17$.

Now, using (49), we see that the first integral in (48) can be bounded as follows:

$$\int_{-10\sqrt{\log n_\lambda}}^{10\sqrt{\log n_\lambda}} p(\xi)\left(\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\cdot\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 d\xi$$
$$\le 2\int_{-10\sqrt{\log n_\lambda}}^{10\sqrt{\log n_\lambda}} p(\xi)\left(e^{-2\pi^2b^2(\xi-\eta)^2}\mathbf{sinc}\left(v(\xi-\eta)\right)^2\right)^2 d\xi$$
$$+2\int_{-10\sqrt{\log n_\lambda}}^{10\sqrt{\log n_\lambda}} p(\xi)\left(\sum_{|j|\ge1} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2b^2}\mathbf{sinc}\left(v(\xi-\frac{jn}{2R}-\eta)\right)\right)^2 d\xi$$
$$\le 2\int_{-10\sqrt{\log n_\lambda}}^{10\sqrt{\log n_\lambda}} \frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}\left(e^{-2\pi^2b^2(\xi-\eta)^2}\mathbf{sinc}\left(v(\xi-\eta)\right)^2+9e^{-n/2}\right) d\xi$$
$$=2\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}e^{-b^2(\xi-\eta)^2}\cdot\mathbf{sinc}\left(v(\xi-\eta)\right)^2 d\xi+18\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}e^{-n/2}d\xi$$
$$\le\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\xi^2/2}e^{-b^2(\xi-\eta)^2}\cdot\mathbf{sinc}\left(v(\xi-\eta)\right)^2 d\xi+18e^{-n/2}. \tag{50}$$

Next, by Claim 32, we have $e^{-\xi^2/2}\le 3e^{-\eta^2/2}$ for $|\xi-\eta|\le\frac{10\sqrt{\log n_\lambda}}{b}$. Hence,

$$\begin{aligned}
&\int_{\eta-\frac{10\sqrt{\log n_\lambda}}{b}}^{\eta+\frac{10\sqrt{\log n_\lambda}}{b}} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} e^{-b^2(\xi-\eta)^2} \cdot \mathbf{sinc}\left(v(\xi-\eta)\right)^2 \, d\xi \\
&\quad \le 3 \cdot \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2} \int_{-\infty}^{+\infty} e^{-b^2(\xi-\eta)^2} \cdot \mathbf{sinc}\left(v(\xi-\eta)\right)^2 \, d\xi \\
&\quad \le 3 \cdot \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2} \int_{-\infty}^{+\infty} \mathbf{sinc}\left(v(\xi-\eta)\right)^2 \, d\xi \\
&\quad = \frac{3p(\eta)}{v} \qquad (51)
\end{aligned}$$

Note that the last line follows from the fact that $v \cdot \mathbf{sinc}\,(v\eta)$ is the Fourier transform of $\mathrm{rect}_v(x)$, and so, by the convolution theorem,

$$\begin{aligned}
\int_{-\infty}^{\infty} (v \cdot \mathbf{sinc}\,(vx))^2 \, dx &= \left(\mathrm{rect}_v(x) * \mathrm{rect}_v(x)\right)(0) \\
&= v.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\int_{|\xi-\eta| \ge \frac{10\sqrt{\log n_\lambda}}{b}} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} e^{-b^2(\xi-\eta)^2} \cdot \mathbf{sinc}\left(v(\xi-\eta)\right)^2 \, d\xi &\le \left(\frac{\lambda}{n}\right)^{50} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} \, d\xi \\
&= \left(\frac{\lambda}{n}\right)^{50}, \qquad (52)
\end{aligned}$$

since the $\mathbf{sinc}\,(\cdot)$ function is bounded by 1 in absolute value. Thus, (50), (51), and (52) imply that

$$\int_{-10\sqrt{\log n_\lambda}}^{10\sqrt{\log n_\lambda}} p(\xi) \Big(\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2 b^2} \cdot \mathbf{sinc}\left(v(\xi - \frac{jn}{2R} - \eta) \right) \Big)^2 d\xi \;\le\; \frac{3p(\eta)}{v} + \left(\frac{\lambda}{n}\right)^{50} + 18e^{-n/2}. \quad (53)$$

Next, we bound the second integral in (48). Consider ξ satisfying $|\xi| \ge 10\sqrt{\log n_\lambda}$. Note that the following upper bound holds for any ξ and η:

$$\begin{aligned}
\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2 b^2} \cdot \mathbf{sinc}\left(v(\xi - \frac{jn}{2R} - \eta) \right) &\le \sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2 b^2} \\
&= 1 + \sum_{|j| \ge 1} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2 b^2} \\
&\le 1 + \frac{2R}{n} \int_{-\infty}^{\infty} e^{-2\pi^2(\xi-x-\eta)^2 b^2} \, dx \\
&\le 2, \qquad (54)
\end{aligned}$$

where we have used the fact that $\frac{b}{R} \le \frac{1}{4\sqrt{\log n_\lambda}} \le 1/\sqrt{2}$. Thus,

$$\begin{aligned}
&\int_{|\xi| \ge 10\sqrt{\log n_\lambda}} p(\xi) \left(\sum_{j=-\infty}^{\infty} e^{-2\pi^2(\xi-\frac{jn}{2R}-\eta)^2 b^2} \mathbf{sinc}\left(v(\xi - \frac{jn}{2R} - \eta) \right) \right)^2 d\xi \\
&\quad \le 2 \int_{|\xi| \ge 10\sqrt{\log n_\lambda}} p(\xi) \, d\xi \\
&\quad \le \left(\frac{\lambda}{n}\right)^{50}, \qquad (55)
\end{aligned}$$

by Claim 25. Combining (47), (48), (53), and (55) now yields the desired result. □

Proof of Theorem 14. Note that we can choose data points $x_1, x_2, \ldots, x_n$ and the vector $\boldsymbol{\alpha}$ according to the construction in Definition 35 with $v = R$ and $b = \frac{R}{4\sqrt{\log n_\lambda}}$. Thus, Lemmas 39, 40, and 41, as well as (30), imply that

$$\begin{aligned}\tau_\lambda(\boldsymbol{\eta}) &\geq \frac{p(\eta) \cdot |\boldsymbol{\alpha}^* \mathbf{z}(\eta)|^2}{\|\boldsymbol{\Phi}^* \boldsymbol{\alpha}\|^2_{L_2(d\mu)} + \lambda \|\boldsymbol{\alpha}\|_2^2} \\ &\geq \frac{p(\eta) \cdot \left(\frac{n}{5}\right)^2}{\left(6\frac{n^2}{R} \cdot p(\eta) + 3\lambda n\right) + \lambda(4n)} \\ &\geq \frac{R}{150}\left(\frac{p(\eta)}{p(\eta) + 2Rn_\lambda^{-1}}\right),\end{aligned}$$

as desired. □

H. Proof of Corollary 15

First claim of the corollary (upper bound on statistical dimension): Let $t = 10\sqrt{\log n_\lambda}$. We have:

$$s_\lambda(\mathbf{K}) = \int_{\mathbb{R}} \tau(\eta) d\eta = \int_{[-t,t]} \tau(\eta) d\eta + \int_{[-\infty,-t]\cup[t,\infty]} \tau(\eta) d\eta$$

By the naive bound in Proposition 4 and Claim 25 we have:

$$\begin{aligned}\int_{[-\infty,-t]\cup[t,\infty]} \tau(\eta) d\eta &\leq n_\lambda \int_{[-\infty,-t]\cup[t,\infty]} \frac{e^{-\eta^2/2}}{\sqrt{2\pi}} d\eta \\ &\leq n_\lambda \cdot \left(\frac{e^{-t^2/2}}{t}\right) \\ &\leq 1 \end{aligned} \tag{56}$$

Further, by the more refined bound of Theorem 13, for any $\eta \leq 10\sqrt{\log n_\lambda} = t$ we have

$$\begin{aligned}\int_{[-t,t]} \tau(\eta) d\eta &\leq \int_{[-t,t]} 25(R + 3000 \log^{1.5} n_\lambda)\, d\eta \\ &\leq 50t \cdot (R + 3000 \log^{1.5} n_\lambda) \\ &= O\Big(\sqrt{\log n_\lambda} \cdot R + \log^2 n_\lambda\Big). \end{aligned} \tag{57}$$

Combining (56) and (57) gives the lemma.

Second claim of the corollary: Note that for all $|\eta| \leq \sqrt{2 \log \frac{n_\lambda}{R}}$ we have $p(\eta) \geq \frac{R}{\sqrt{2\pi} n_\lambda}$, hence we have:

$$p(\eta) + 2R/n_\lambda \leq 7p(\eta)$$

hence, by Theorem 14, we have:

$$\tau(\eta) \geq \frac{R}{150}\left(\frac{1}{7}\right)$$

And for $|\eta| > \sqrt{2 \log \frac{n_\lambda}{R}}$ we have:

$$p(\eta) + 2R/n_\lambda \leq 3R/n_\lambda$$

therefore,

$$\begin{aligned}s_\lambda(\mathbf{K}) &= \int_{-\infty}^{\infty} \tau(\eta) d\eta \\ &\geq \int_{-\sqrt{2\log\frac{n_\lambda}{R}}}^{\sqrt{2\log\frac{n_\lambda}{R}}} \frac{R}{1050} d\eta + \int_{|\boldsymbol{\eta}| > \sqrt{2\log\frac{n_\lambda}{R}}} \frac{R}{150}\left(\frac{p(\eta)}{3R/n_\lambda}\right) d\eta \\ &= \Omega(R\sqrt{\log\frac{n_\lambda}{R}}) \end{aligned} \tag{58}$$

I. Proof of Theorem 8 and 10

Proof of Theorem 8. We show a lower bound on the number of samples required under the random feature map of Rahimi and Recht by exhibiting a set of data points for which the appropriate number of samples does not suffice.

Our goal is to show that if we take s samples $\eta_1, \eta_2, \ldots, \eta_s$ from the distribution defined by p, for s too small, then there is an $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^n$ such that with at least constant probability,

$$\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{K} + \lambda \mathbf{I}_n)\boldsymbol{\alpha} < \frac{2}{3}\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^* + \lambda \mathbf{I}_n)\boldsymbol{\alpha}. \tag{59}$$

By (3), we have

$$\begin{aligned}
\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} &= \sum_{j,k} \alpha_j \alpha_k \cdot k(x_j, x_k) \\
&= \sum_{j,k} \int_{-\infty}^{\infty} e^{-2\pi i \eta (x_j - x_k)} \alpha_j \alpha_k p(\eta)\, d\eta \\
&= \int_{-\infty}^{\infty} \left(\sum_{j=1}^{n} \alpha_j e^{-2\pi i \eta x_j} \right) \left(\sum_{k=1}^{n} \alpha_k e^{2\pi i \eta x_k} \right) p(\eta)\, d\eta \\
&= \int_{-\infty}^{\infty} p(\eta) \left| \sum_{j=1}^{n} \alpha_j e^{2\pi i \eta x_j} \right|^2 d\eta.
\end{aligned}$$

Also, by the definition of $\mathbf{Z}$ and φ (see Section 2.2), we have

$$\begin{aligned}
\boldsymbol{\alpha}^T \mathbf{Z}\mathbf{Z}^* \boldsymbol{\alpha} &= \left\| \sum_{j=1}^{n} \alpha_j \varphi(x_j) \right\|_2^2 \\
&= \sum_{k=1}^{s} \left| \sum_{j=1}^{n} \alpha_j \cdot \frac{1}{\sqrt{s}} e^{2\pi i \eta_k x_j} \right|^2 \\
&= \frac{1}{s} \sum_{k=1}^{s} \left| \sum_{j=1}^{n} \alpha_j e^{2\pi i \eta_k x_j} \right|^2,
\end{aligned}$$

where $\eta_1, \eta_2, \ldots, \eta_s$ are the s samples from the distribution given by p. Hence, (59) is equivalent to

$$\int_{-\infty}^{\infty} p(\eta) \left| \sum_{j=1}^{n} \alpha_j e^{2\pi i \eta x_j} \right|^2 d\eta + \frac{1}{3}\lambda \|\boldsymbol{\alpha}\|_2^2 < \frac{2}{3} \cdot \frac{1}{s} \sum_{k=1}^{s} \left| \sum_{j=1}^{n} \alpha_j e^{2\pi i \eta_k x_j} \right|^2. \tag{60}$$

We again use the same construction of n data points $x_1, x_2, \ldots, x_n \in \mathbb{R}$, according to the construction in Definition 33. Moreover, we define η^* to be

$$\eta^* = \max_{1 \le j \le s} |\eta_j|$$

and let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ be given by

$$\alpha_j = f_{\eta^*, b, v}(x_j),$$

where $b = R/4\sqrt{\log n_\lambda}$ and $v = \delta$. We will show that this choice of data points and $\boldsymbol{\alpha}$ satisfy (60).

First, we upper bound the first term on the left side of (60). Note that by Claim 26, with probability at least $1/2$ over the samples $z_1, z_2, \ldots, z_s$, we have

$$\begin{aligned}\int_{-\infty}^{\infty} p(\eta) \left|\sum_{j=1}^{n} \alpha_j e^{2\pi i \eta x_j}\right|^2 d\eta &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2} \left|\sum_{j=-m}^{m} \alpha_j e^{2\pi i \eta x_j}\right|^2 d\eta \\ &= \|\mathbf{\Phi}^* \boldsymbol{\alpha}\|_{L_2(d\mu)}^2 \\ &\leq \frac{6n^2}{R} \cdot p(\eta^*) + 3\lambda n \\ &\leq \frac{48n^2}{R} \cdot \frac{\sqrt{\log s}}{s} + 3\lambda n. \end{aligned} \tag{61}$$

where we have let $\eta = \eta^*$ and applied Lemma 41.

Next, we bound the right side of (60) from below. Note that

$$\begin{aligned}\frac{1}{s}\sum_{k=1}^{s} \left|\sum_{j=1}^{n} \alpha_j e^{2\pi i \eta_k x_j}\right|^2 &\geq \frac{1}{s} \left|\sum_{j=1}^{n} \alpha_j e^{2\pi i \eta^* x_j}\right|^2 \\ &= \frac{1}{s} (\boldsymbol{\alpha}^* \mathbf{z}(\eta^*))^2 \\ &\geq \frac{1}{s} \left(\frac{n}{5}\right)^2 = \frac{n^2}{25s}, \end{aligned} \tag{62}$$

by Lemma 39 applied with $\eta = \eta^*$.

We also require the following estimate of $\|\boldsymbol{\alpha}\|_2^2$, which is provided by Lemma 40:

$$\|\boldsymbol{\alpha}\|_2^2 \leq 4n. \tag{63}$$

Note that by combining (61), (62), and (63), we have that with probability at least 1/2,

$$\begin{aligned}\int_{-\infty}^{\infty} p(\eta) \left|\sum_{j=1}^{n} \alpha_j e^{2\pi i \eta x_j}\right|^2 d\eta + \frac{1}{3}\lambda \|\boldsymbol{\alpha}\|_2^2 &\leq \frac{48n^2}{R} \cdot \frac{\sqrt{\log s}}{s} + 3\lambda n + \frac{4}{3}\lambda n \\ &\leq \frac{2n^2}{75s} \\ &\leq \frac{2}{3} \cdot \frac{1}{s} \sum_{k=1}^{s} \left|\sum_{j=1}^{n} \alpha_j e^{2\pi i \eta_k x_j}\right|^2, \end{aligned}$$

since $s \leq n_\lambda/400$ and also because $R \geq 3000 \log^{1.5}(n_\lambda)$. This completes the proof. □

Proof of Theorem 10. By the assumptions of the theorem n is an integer, parameter $0 < \lambda \leq n/2$, and $R > 0$, and all $x_1, ..., x_n \in [-R, R]$ and $p(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}$, therefore all the preconditions of Proposition 4, and Theorem 13 are satisfied and hence the theorem and proposition go through and for every η we have:

$$\tau_\lambda(\eta) \leq \bar{\tau}_R(\eta)$$

Hence applying Lemma 6 with $\tilde{\tau}(\eta) = \bar{\tau}_R(\eta)$ gives the desired spectral approximation with $\frac{8}{3}\Delta^{-2} s_{\bar{\tau}_R} \ln(16 s_{\bar{\tau}_R}/\rho)$ samples where $s_{\bar{\tau}_R} = \int_{\mathbb{R}} \bar{\tau}_R(\eta) d\eta$. Now we show that $s_{\bar{\tau}_R} = O(R\sqrt{\log(n_\lambda)} + \log^2 n_\lambda)$.

Let $t = 10\sqrt{\log n_\lambda}$. We have:

$$s_{\bar{\tau}_R} = \int_{\mathbb{R}} \bar{\tau}_R(\eta) d\eta = \int_{[-t,t]} \bar{\tau}_R(\eta) d\eta + \int_{[-\infty,-t]\cup[t,\infty]} \bar{\tau}_R(\eta) d\eta$$

By Definition 9 and Claim 25 we have:

$$\begin{aligned}\int_{[-\infty,-t]\cup[t,\infty]} \bar{\tau}_R(\eta) d\eta &= n_\lambda \int_{[-\infty,-t]\cup[t,\infty]} \frac{e^{-\eta^2/2}}{\sqrt{2\pi}} d\eta \\ &\leq n_\lambda \cdot \left(2\frac{e^{-t^2/2}}{\sqrt{2\pi} t}\right) \\ &\leq n_\lambda \cdot \left(\frac{e^{-t^2/2}}{t}\right) \\ &\leq 1\end{aligned}$$

Furthermore, for any $\eta \leq 10\sqrt{\log n_\lambda} = t$ we have

$$\begin{aligned}\int_{[-t,t]} \tau(\eta) d\eta &\leq \int_{[-t,t]} 25(R + 3000 \log^{1.5} n_\lambda)\, d\eta \\ &\leq 50t \cdot (R + 3000 \log^{1.5} n_\lambda) \\ &= O\left(\sqrt{\log n_\lambda} \cdot R + \log^2 n_\lambda\right).\end{aligned}$$

Combining the bounds above gives the result.

Sampling from $\bar{p}_R(\cdot)$: Sampling from $\bar{p}_R(\cdot)$ amounts to sampling from a mixture of the uniform distribution on $[-10\sqrt{\log(n_\lambda)}, +10\sqrt{\log(n_\lambda)}]$ and from the tail of the Gaussian distribution: with probability $\frac{25 \max(R, 3000 \log^{1.5} n_\lambda)}{s_{\bar{\tau}_R}} \cdot 20\sqrt{\log(n_\lambda)}$ sample from the uniform distribution and with remaining probability sample from the tail of the Gaussian. Sampling from the tail of the Gaussian can be easily accomplished via rejection sampling at unit expected cost. Indeed, we only need to generate a sample from the tail with probability proportional to the mass of the tail. On the other hand, once we do, the expected cost of obtaining a sample via rejection sampling is inversely proportional to the amount of mass in the tail, leading to unit cost in expectation.

□