Appendix

Proof of Lemma 2.1. Since unitary transformations preserve dot-products, i.e., $\langle T(x), T(y)\rangle = \langle x, y\rangle$, we need to show that a group element acting on the image $I : \mathbb{R}^2 \mapsto \mathbb{R}$ as $T_g[I(x)] = |J_g|^{-1/2} I(T_g^{-1}(x)), \forall\, x$ is a unitary transformation.

Let J_g be the Jacobian of the transformation T_g, with determinant $|J_g|$. We have

$$\begin{aligned} \|I(T_g^{-1}(\cdot))\|^2 &= \int I^2(T_g^{-1}(x))dx \\ &= \int I^2(z)\,|J_g|\,dz, \quad \text{substituting } z = T_g^{-1}(x) \Rightarrow dx = |J_g|\,dz \\ &= |J_g|\,\|I(\cdot)\|^2 \end{aligned}$$

Hence the transformation given as $T_g[I(\cdot)] = |J_g|^{-1/2} I(T_g^{-1}(\cdot))$ is unitary and thus $\langle T_g(I), T_g(I')\rangle = \langle I, I'\rangle$ for two images I and I'. □

Proof of Theorem 3.1. We first define the notion of *U-statistics* [21].

U-statistics - Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a symmetric function of its arguments. Given an i.i.d. sequence $X_1, X_2 \cdots X_k$ of $k(\geq 2)$ random variables, the quantity $U := \frac{1}{n(n-1)} \sum_{i\neq j, i,j=1}^{n} g(X_i, X_j)$ is known as a pairwise **U-statistics**. If $\theta(P) = \mathbb{E}_{X_1,X_2\sim P}\, g(X_1, X_2)$ then U is an unbiased estimate of $\theta(P)$.

Our goal is to bound

$$\sup_{x,y\in X}\left|\left\langle \psi_{RF}(x), \psi_{RF}(y)\right\rangle - k_{q,G}(x,y)\right|$$

where

$$\psi_{RF}(x) = \frac{1}{r}\sum_{i=1}^{r} z(g_i x), x \in X \subset \mathbb{R}^d.$$

We work with $z(\cdot) = \sqrt{2/s}[\cos(\langle\omega_1, \cdot\rangle + b_1), \ldots, cos(\langle\omega_s, \cdot\rangle + b_s] \in \mathbb{R}^s$ with $b_i \sim \text{Unif}(0, 2\pi)$ as in [33].

Let $\widehat{k}_{q,G}(x,y) := \frac{1}{r^2}\sum_{i,j=1}^{r^2} k(g_i x, g_j y)$ and $\widetilde{k}_{q,G}(x,y) := \frac{1}{r(r-1)}\sum_{i\neq j,i,j=1}^{r^2} k(g_i x, g_j y)$.

Using the triangle inequality we have

$$\begin{aligned} \sup_{x,y\in X}\left|\left\langle \psi_{RF}(x), \psi_{RF}(v)\right\rangle - k_{q,G}(x,y)\right| &\leq \underbrace{\sup_{x,y\in X}\left|\left\langle \psi_{RF}(x), \psi_{RF}(y)\right\rangle - \widehat{k}_{q,G}(x,y)\right|}_{A} \\ &+ \underbrace{\sup_{x,y\in X}\left|\widetilde{k}_{q,G}(x,y) - k_{q,G}(x,y)\right|}_{B} + \underbrace{\sup_{x,y\in X}\left|\widehat{k}_{q,G}(x,y) - \widetilde{k}_{q,G}(x,y)\right|}_{C} \end{aligned}$$

Bounding A.

$$A := \sup_{x,y\in X}\left|\frac{1}{r^2}\sum_{i,j}\left(\langle z(g_i x), z(g_j y)\rangle - k(g_i x, g_j y)\right)\right|$$

Let us define $f_{ij}(x,y) := \langle z(g_i x), z(g_j y)\rangle - k(g_i x, g_j y)$, and $f(x,y) = 1/r^2\sum_{i,j} f_{ij}(x,y)$. Since each of the s independent random variables in the summand of $1/r^2\sum_{i,j}\langle z(g_i x), z(g_j y)\rangle = \frac{1}{s}\sum_{k=1}^{s}\Big(\frac{1}{r^2}$

$\sum_{i,j} 2\cos(\langle\omega_k, g_i x\rangle + b_k)\cos(\langle\omega_k, g_j y\rangle + b_k)\Big)$ is bounded by $[-2,2]$, using Hoeffding's inequality for a given pair x, y, we have

$$\Pr[|f(x,y)| \geq \varepsilon/4] \leq 2\exp(-s\varepsilon^2/128).$$

To obtain a uniform convergence guarantee over X, we follow the arguments in [33], relying on covering the space with an ε-net and Lipschitz continuity of the function $f(x,y)$.

Since X is compact, we can find an ε-net that covers X with $N_X = \left(\frac{2\,\text{diam(X)}}{\eta}\right)^d$ balls of radius η [12]. Let $\{c_k\}_{k=1}^{N_X}$ be the centers of these balls, and let L_f denote the Lipschitz constant of $f(\cdot,\cdot)$, i.e., $|f(x,y) - f(c_k, c_l)| \leq L_f \|\binom{x}{y} - \binom{c_k}{c_l}\|$ for all $x, y, c_k, c_l \in X$. For any $x, y \in X$, there exists a pair of centers c_k, c_l such that $\|\binom{x}{y} - \binom{c_k}{c_l}\| < \sqrt{2}\eta$. We will have $|f(x,y)| < \varepsilon/2$ for all x, y if (i) $|f(c_k, c_l)| < \frac{\varepsilon}{4}$, $\forall c_k, c_l$, and (ii) $L_f < \frac{\varepsilon}{4\sqrt{2}\eta}$.

We immediately get the following by applying union bound for all the center pairs (c_k, c_l)

$$\Pr\left[\cup_{k,l}|f(c_k,c_l)| \geq \varepsilon/4\right] \leq 2\,N_X^2\exp(-s\varepsilon^2/128). \tag{9}$$

We use Markov inequality to bound the Lipschitz constant of f. By definition, we have $L_f = \sup_{x,y}\|\nabla_{x,y} f(x,y)\| = \|\nabla_{x,y} f(x^*, y^*)\|$, where $\nabla_{x,y} f(x,y) = \begin{pmatrix}\nabla_x f(x,y)\\ \nabla_y f(x,y)\end{pmatrix}$. We also have $\mathbb{E}_{\omega\sim p}\nabla_{x,y}\langle z(g_i x), z(g_j y)\rangle = \nabla_{x,y} k(g_i x, g_j y)$. It follows that

$$\begin{aligned}
\mathbb{E}_{\omega\sim p}\|\nabla_{x,y} f(x^*,y^*)\|^2 &= \mathbb{E}_{\omega\sim p}\left\|\frac{1}{r^2}\sum_{i,j=1}^{r}\nabla_{x,y}\langle z(g_i x^*), z(g_j y^*)\rangle\right\|^2 - \left\|\frac{1}{r^2}\sum_{i,j=1}^{r}\nabla_{x,y} k(g_i x^*, g_j y^*)\right\|^2\\
&\leq \mathbb{E}_{\omega\sim p}\left\|\frac{1}{r^2}\sum_{i,j=1}^{r}\nabla_{x,y}\langle z(g_i x^*), z(g_j y^*)\rangle\right\|^2\\
&\leq \mathbb{E}_{\omega\sim p}\left(\frac{1}{r^2}\sum_{i,j=1}^{r}\|\nabla_{x,y}\langle z(g_i x^*), z(g_j y^*)\rangle\|\right)^2\\
&\leq 2\mathbb{E}_{\omega\sim p}\sup_{x,y,g_i,g_j}\|\nabla_x\langle z(g_i x), z(g_j y)\rangle\|^2\\
&\leq 2\mathbb{E}_{\omega\sim p}\sup_{x,g}\left(\frac{1}{s}\sum_{k=1}^{s}\|\nabla_x T_g(x)\omega_k\|\right)^2\\
&\leq 2\mathbb{E}_{\omega\sim p}\sup_{x,g}\left(\frac{1}{s}\sum_{k=1}^{s}\|\nabla_x T_g(x)\|_2\|\omega_k\|\right)^2\\
&= 2\mathbb{E}_{\omega\sim p}\sup_{x,g}\|\nabla_x T_g(x)\|_2^2\frac{1}{s^2}\sum_{k=1}^{s}\sum_{l=1}^{s}\|\omega_k\|\|\omega_l\|\\
&= 2\sup_{x,g}\|\nabla_x T_g(x)\|_2^2\frac{1}{s^2}\sum_{k=1}^{s}\sum_{l=1}^{s}\mathbb{E}_{\omega\sim p}\|\omega_k\|\|\omega_l\|\\
&= 2\sup_{x,g}\|\nabla_x T_g(x)\|_2^2\frac{1}{s^2}\left(s\mathbb{E}_{\omega\sim p}\|\omega\|^2 + \sum_{k,l=1,k\neq l}^{s}(\mathbb{E}_{\omega\sim p}\|\omega\|)^2\right) \quad (\omega_k \text{ i.i.d.})\\
&\leq 2\sup_{x,g}\|\nabla_x T_g(x)\|_2^2\frac{1}{s^2}\left(s\mathbb{E}_{\omega\sim p}\|\omega\|^2 + \sum_{k,l=1,k\neq l}^{s}\mathbb{E}_{\omega\sim p}\|\omega\|^2\right) \quad (\text{Jensen's inequality})
\end{aligned}$$

$$\leq 2\sigma_p^2 \sup_{x \in X, g \in G} \|\nabla_x T_g(x)\|_2^2,$$

where $\sigma_p^2 = \mathbb{E}(\omega^\top \omega)$, and $T_g(x) = gx$ denotes the transformation corresponding to the group action. If we assume the group action to be linear, i.e., $T_g(x+y) = T_g(x) + T_g(y)$ and $T_g(\alpha x) = \alpha T_g(x)$, which holds for all group transformations considered in this work (e.g., rotation, translation, scaling or general affine transformations on image x; permutations of x), we can bound $\|\nabla_x T_g(x)\|_2$ as

$$\begin{aligned}
\|\nabla_x T_g(x)\|_2 &= \sup_{u:\|u\|=1} \|\nabla_x T_g(x) u\| \\
&= \sup_{u:\|u\|=1} \left\| \lim_{h \to 0} \frac{T_g(x+hu) - T_g(x)}{h} \right\| \quad \text{(directional derivative of vector valued function } T_g(\cdot)) \\
&= \sup_{u:\|u\|=1} \|T_g(u)\| = 1 \\
&\quad \text{(since } T_g(\cdot) \text{ is either unitary or is converted to unitary by construction (see Lemma 2.1))}
\end{aligned}$$

Using Markov inequality, $\Pr[L_f^2 \geq \varepsilon] \leq \mathbb{E}(L_f^2)/\varepsilon$, hence we get

$$\Pr\left[L_f \geq \frac{\varepsilon}{4\sqrt{2}\eta}\right] \leq \frac{64\sigma_p^2\eta^2}{\varepsilon^2}.$$

Combining Eq. (9) with the above result on Lipschitz continuity, we get

$$\Pr\left[\sup_{x,y} |f(x,y)| \leq \varepsilon/2\right] \geq 1 - 2\,N_X^2 \exp(-s\varepsilon^2/128) - \frac{64\sigma_p^2\eta^2}{\varepsilon^2}. \tag{10}$$

Bounding B.

As defined earlier, $\widetilde{k}_{q,G}(x,y) := \frac{1}{r(r-1)} \sum_{i \neq j, i,j=1}^{r} k(g_i x, g_j y)$. From the result of U-statistics literature [21], it is easy to see that $\mathbb{E}(\widetilde{k}_{q,G}(x,y)) = k_{q,G}(x,y)$.
Since $g_1, g_2 \cdots g_r$ are i.i.d samples, we can consider $\widetilde{k}_{q,G}(x,y)$ as function of r random variables $(g_1, g_2, \cdots g_r)$. Denote $\widetilde{k}_{q,G}(x,y)$ as $f(g_1, g_2, \cdots g_r)$. Now if a variable g_p is changed to g'_p then we can bound the absolute difference of the changed and the original function. For the rbf kernel, $|k(g_p x, g_j y) - k(g'_p x, g_j y)| \leq 1$

$$\begin{aligned}
|f(g_1, g_2, \cdots g_p, \cdots g_r) - f(g_1, \cdots g_{p-1}, g'_p, g_{p+1} \cdots g_r)| &= \frac{1}{r(r-1)} \Big| \sum_{j=1, j \neq p}^{r} k(g_p x, g_j y) - k(g'_p x, g_j y) \Big| \\
&\leq \frac{1}{r(r-1)} \sum_{j=1, j \neq p}^{r} |k(g_p x, g_j y) - k(g'_p x, g_j y)| \\
&\leq \frac{(r-1)}{r(r-1)} = \frac{1}{r}
\end{aligned}$$

Using bounded difference inequality

$$Pr\Big[\big|f(g_1, g_2, \cdots g_r) - \mathbb{E}[f(g_1, g_2 \cdots g_r)]\big| \geq \frac{\varepsilon}{2}\Big] \leq 2 \exp\Big(\frac{-r\varepsilon^2}{2}\Big).$$

The above bound holds for a given pair x, y. Similar to the earlier segment for bounding the first term A, we use the ε-net covering of X and Lipschitz continuity arguments to get a uniform convergence guarantee. Using a union bound on all pairs of centers, we have

$$Pr\Big[\cup_{k,\ell=1}^{N_X} \Big|\mathbb{E}[k(gc_k, g'c_\ell)] - \frac{1}{r(r-1)} \sum_{i,j=1, i \neq j}^{r} k(g_i c_k, g_j c_\ell)\Big| > \frac{\varepsilon}{2}\Big] \leq 2N_X^2 \exp\Big(\frac{-r\varepsilon^2}{2}\Big). \tag{11}$$

In order to extend the bound from the centers c_i to all $x \in X$, we use the Lipschitz continuity argument. Let

$$h(x,y) = k_{q,G}(x,y) - \widetilde{k}_{q,G}(x,y).$$

Let L_h denote the Lipschitz constant of $h(\cdot,\cdot)$, i.e., $|h(x,y) - h(c_k,c_l)| \leq L_h \|\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) - \left(\begin{smallmatrix} c_k \\ c_l \end{smallmatrix}\right)\|$ for all $x, y, c_k, c_l \in X$. By the definition of ε-net, for any $x, y \in X$, there exists a pair of centers c_k, c_l such that $\|\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) - \left(\begin{smallmatrix} c_k \\ c_l \end{smallmatrix}\right)\| < \sqrt{2}\eta$. We will have $|h(x,y)| < \varepsilon/2$ for all x, y if (i) $|h(c_k,c_l)| < \frac{\varepsilon}{4}$, $\forall c_k, c_l$, and (ii) $L_h < \frac{\varepsilon}{4\sqrt{2}\eta}$.

We will again use Markov inequality to bound the Lipschitz constant of h. By definition, we have $L_h = \sup_{x,y} \|\nabla_{x,y} h(x,y)\| = \|\nabla_{x,y} h(x^*,y^*)\|$, where $\nabla_{x,y} h(x,y) = \left(\begin{smallmatrix} \nabla_x h(x,y) \\ \nabla_y h(x,y) \end{smallmatrix}\right)$. We also have $\mathbb{E}_{\omega \sim p} \nabla_{x,y} \widetilde{k}_{q,G}(x,y) = \nabla_{x,y} k_{q,G}(x,y)$. It follows that

$$\begin{aligned}
\mathbb{E}_{g_1,\ldots,g_r} \|\nabla_{x,y} h(x^*,y^*)\|^2 &= \mathbb{E}_{g_1,\ldots,g_r} \|\nabla_{x,y} \widetilde{k}_{q,G}(x^*,y^*)\|^2 - \|\nabla_{x,y} k_{q,G}(x^*,y^*)\|^2 \\
&\leq \mathbb{E}_{g_1,\ldots,g_r} \|\nabla_{x,y} \widetilde{k}_{q,G}(x^*,y^*)\|^2 \\
&= \mathbb{E}_{g_1,\ldots,g_r} \left\| \frac{1}{r(r-1)} \sum_{i \neq j} \nabla_{x,y} k(g_i x^*, g_j y^*) \right\|^2.
\end{aligned}$$

Noting $T_{g_i}(x) = g_i x$, and $k(x,y) = \exp -\frac{1}{2\sigma^2} \left|\left| x - y \right|\right|^2$, we have

$$\begin{aligned}
\nabla_x k(g_i x, g_j y) &= \nabla_x k(T_{g_i}(x), T_{g_j}(y)) \\
&= -\frac{1}{\sigma^2} \nabla_x T_{g_i}(x)(g_i x - g_j y) \exp\left(-\frac{1}{2\sigma^2} \left|\left| g_i x - g_j y \right|\right|^2\right).
\end{aligned}$$

Continuing

$$\begin{aligned}
\left|\left| \frac{1}{r(r-1)} \sum_{i \neq j} \nabla_{x,y} k(g_i x, g_j y) \right|\right| &\leq \frac{1}{r(r-1)} \sum_{i \neq j} \left|\left| \nabla_{x,y} k(g_i x, g_j y) \right|\right| \\
&\leq \frac{\sqrt{2}}{r(r-1)} \sup_x \sum_{i \neq j} \left|\left| \nabla_x k(g_i x, g_j y) \right|\right| \quad \text{(using symmetry of } k(\cdot,\cdot)\text{)} \\
&= \frac{\sqrt{2}}{r(r-1)\sigma^2} \sup_x \sum_{i \neq j} k(g_i x, g_j y) \left|\left| \nabla_x T_{g_i}(x)(g_i x - g_j y) \right|\right| \\
&\leq \frac{\sqrt{2}}{r(r-1)\sigma^2} \sum_{i \neq j} k(g_i x, g_j y) ||\nabla_x T_{g_i}(x)||_2 ||(g_i x - g_j y)|| \\
&\leq \frac{\sqrt{2} e^{-1/2}}{\sigma} \sup_{x \in X, g \in G} ||\nabla_x T_g(x)||_2 \quad \text{(using } \sup_{z \geq 0} z e^{-z^2/(2\sigma^2)} = \sigma e^{-1/2}\text{)} \\
&\leq \frac{\sqrt{2} e^{-1/2}}{\sigma} \quad \text{(using linearity and unitariy of } T_g(\cdot) \text{ as before)}
\end{aligned}$$

It follows that

$$\mathbb{E}({L_h}^2) \leq \frac{2}{\sigma^2 e}.$$

Now using Markov inequality we have

$$\mathbb{P}\left[L_h > \sqrt{t}\right] \leq \frac{\mathbb{E}({L_h}^2)}{t},$$

Hence we have for $t = \left(\frac{\varepsilon}{4\sqrt{2}\eta}\right)^2$,

$$\mathbb{P}\left[L_h > \frac{\varepsilon}{4\sqrt{2}\eta}\right] \leq \frac{32\eta^2\mathbb{E}((L_h)^2)}{\varepsilon^2} \leq \frac{64\eta^2}{e\sigma^2\varepsilon^2},$$

Hence

$$\Pr[B \leq \varepsilon/2)] \geq 1 - 2(N_X)^2 \exp\left(\frac{-r\varepsilon^2}{2}\right) - \frac{64\eta^2}{e\sigma^2\varepsilon^2}.$$

Bounding C.

$$\begin{aligned}
\left|\widetilde{k}_{q,G}(x,y) - \widehat{k}_{q,G}(x,y)\right| &= \Big|\frac{1}{r(r-1)}\sum_{i,j=1,i\neq j}^{r} k(g_ix, g_jy) - \frac{1}{r^2}\sum_{i,j=1}^{r} k(g_ix, g_jy)\Big| \\
&= \Big|\Big(\frac{1}{r(r-1)} - \frac{1}{r^2}\Big)\sum_{i,j=1,i\neq j}^{r} k(g_ix, g_jy) - \frac{1}{r^2}\sum_{i,j=1,i=j}^{r} k(g_ix, g_jy)\Big| \\
&\leq \max\left(\frac{1}{r^2(r-1)}\sum_{i,j=1,i\neq j}^{r} k(g_ix, g_jy), \frac{1}{r^2}\sum_{i,j=1,i=j}^{r} k(g_ix, g_jy)\right) \quad (\text{since } k(\cdot,\cdot) \geq 0) \\
&\leq \frac{1}{r} \quad (\text{as Gaussian kernel } k(\cdot,\cdot) \leq 1)
\end{aligned}$$

Finally we have

$$\sup_{x,y\in X}\left|\Big\langle\psi_{RF}(x), \psi_{RF}(y)\Big\rangle - k_{q,G}(x,y)\right| \leq A + B + C \leq \varepsilon + \frac{1}{r}$$

with a probability at least $1 - 2{N_X}^2\exp\left(\frac{-s\varepsilon^2}{128}\right) - 2{N_X}^2\exp\left(\frac{-r\varepsilon^2}{2}\right) - \left(\frac{64\eta^2 d}{\varepsilon^2\sigma^2}\right) - \left(\frac{64\eta^2}{e\varepsilon^2\sigma^2}\right)$, noting that $\sigma_p^2 = d/\sigma^2$ for the Gaussian kernel $k(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$.

Let

$$\begin{aligned}
p &= 1 - 2{N_X}^2\exp\left(\frac{-s\varepsilon^2}{128}\right) - 2{N_X}^2\exp\left(\frac{-r\varepsilon^2}{2}\right) - \left(\frac{64\eta^2 d}{\varepsilon^2\sigma^2}\right) - \left(\frac{64\eta^2}{e\varepsilon^2\sigma^2}\right) \\
&= 1 - 2\Big(\frac{2diam(X)}{\eta}\Big)^{2d}\exp\left(\frac{-s\varepsilon^2}{128}\right) - 2\Big(\frac{2diam(X)}{\eta}\Big)^{2d}\exp\left(\frac{-r\varepsilon^2}{2}\right) - \Big(\frac{64\eta^2 d}{\varepsilon^2\sigma^2}\Big) - \Big(\frac{64\eta^2}{e\varepsilon^2\sigma^2}\Big) \\
&\geq 1 - 2\eta^{-2d}\Big(\big(2diam(X)\big)^{2d}\exp\left(\frac{-r\varepsilon^2}{2}\right) + \big(2diam(X)\big)^{2d}\exp\left(\frac{-s\varepsilon^2}{128}\right)\Big) - \eta^2\Big(\frac{64(d+1)}{\varepsilon^2\sigma^2}\Big).
\end{aligned}$$

The above probability is of the form of $1 - (\kappa_1 + \kappa_2)\eta^{-2d} - \kappa_3\eta^2$ where $\kappa_1 = 2\big(2diam(X)\big)^{2d}\exp\left(\frac{-r\varepsilon^2}{2}\right)$, $\kappa_2 = 2\big(2diam(X)\big)^{2d}\exp\left(\frac{-s\varepsilon^2}{128}\right)$ and $\kappa_3 = \left(\frac{64(d+1)}{\varepsilon^2\sigma^2}\right)$. Choose $\eta = \left(\frac{\kappa_1+\kappa_2}{\kappa_3}\right)^{\frac{1}{2(d+1)}}$

Hence $p \geq 1 - 2(\kappa_1+\kappa_2)^{\frac{1}{d+1}}\kappa_3^{\frac{d}{d+1}}$.

For given $\delta_1, \delta_2 \in (0,1)$, we conclude that for fixed constants C_1, C_2, for

$$r \geq \frac{C_1 d}{\varepsilon^2}\log(\mathrm{diam(X)}/\delta_1),$$

$$s \geq \frac{C_2 d}{\varepsilon^2}\left(\log(\mathrm{diam(X)}/\delta_2\right),$$

we have

$$\sup_{x,y\in X}\left|\Big\langle\psi_{RF}(x), \psi_{RF}(y)\Big\rangle - k_{q,G}(x,y)\right| \leq \varepsilon + \frac{1}{r},$$

with probability $1 - \left(\frac{64(d+1)}{\varepsilon^2\sigma^2}\right)^{\frac{d}{d+1}}(\delta_1+\delta_2)^{\frac{2d}{d+1}}$.

□

Proof of Theorem 3.2. We give here the proof of Theorem 3.2.

Lemma A.1 (Lemma 4 [35])**.** - *Let* $\boldsymbol{X} = \{x_1, x_2 \cdots x_K\}$ *be iid random variables in a ball* $\mathcal{H}$ *of radius* M *centered around the origin in a Hilbert space. Denote their average by* $\overline{\boldsymbol{X}} = \frac{1}{K}\sum_{i=1}^{K} x_i$*. Then for any* $\delta > 0$*, with probability at least* $1 - \delta$*,*

$$\|\overline{\boldsymbol{X}} - \mathbb{E}\overline{\boldsymbol{X}}\| \leq \frac{M}{\sqrt{K}}\Big(1 + \sqrt{2\log\frac{1}{\delta}}\Big)$$

Proof. For proof, see [35]. □

Now consider a space of functions,

$$\mathcal{F}_p \equiv \Big\{f(x) = \int_\Omega \alpha(\omega)\int_G \phi(gx,\omega)q(g)d\nu(g)d\omega \Big| |\alpha(\omega)| \leq Cp(\omega)\Big\},$$

and also consider another space of functions,

$$\hat{\mathcal{F}}_p \equiv \Big\{\hat{f}(x) = \sum_{k=1}^{s}\alpha_k\frac{1}{r}\sum_{i=1}^{r}\phi(g_i x,\omega_k)\Big| |\alpha_k| \leq \frac{C}{s}\Big\},$$

where $\phi(gx,\omega) = e^{-i\langle gx,\omega\rangle}$.

Lemma A.2. *Let* μ *be a measure defined on* X*, and* $f^\star$ *a function in* $\mathcal{F}_p$*. If* $\omega_1, \omega_2 \dots \omega_s$ *are iid samples from* $p(\omega)$*, then for* $\delta_1, \delta_2 > 0$*, there exists a function* $\hat{f} \in \hat{\mathcal{F}}_p$ *such that*

$$\Big\|f^\star - \hat{f}\Big\|_{\mathcal{L}_2(X,\mu)} \leq \frac{C}{\sqrt{s}}\Big(1 + \sqrt{2\ \log\frac{1}{\delta_1}}\Big) + \frac{C}{\sqrt{r}}\Big(1 + \sqrt{2\log\frac{1}{\delta_2}}\Big),$$

with probability at least $1 - \delta_1 - \delta_2$*.*

Proof. Consider $\psi(x;\omega_k) = \int_G \phi(gx,\omega_k)q(g)d\nu(g)$. Let $\tilde{f}_k = \beta_k\ \psi(.;\omega_k), k = 1\cdots s$, with $\beta_k = \frac{\alpha(\omega_k)}{p(\omega_k)}$. Hence $\mathbb{E}_{\omega_k\sim p}\tilde{f}_k = f^\star$.
Define $\tilde{f}(x) = \frac{1}{s}\sum_{k=1}^{s}\tilde{f}_k$. Let $\hat{f}_k(x) = \beta_k\hat{\psi}(x;\omega_k)$, where $\hat{\psi}(x;\omega_k) = \frac{1}{r}\sum_{i=1}^{r}\phi(g_i x,\omega_k)$ is the empirical estimate of $\psi(x;\omega_k)$. Define $\hat{f}(x) = \frac{1}{s}\sum_{k=1}^{s}\hat{f}_k(x)$. We have $\mathbb{E}_{g_i\sim q}\hat{f}(x) = \tilde{f}(x)$.

$$\Big\|f^\star - \hat{f}\Big\|_{\mathcal{L}_2(X,\mu)} \leq \Big\|f^\star - \tilde{f}\Big\|_{\mathcal{L}_2(X,\mu)} + \Big\|\tilde{f} - \hat{f}\Big\|_{\mathcal{L}_2(X,\mu)}$$

From Lemma 1 of [35], with probability $1 - \delta_1$,

$$\Big\|f^\star - \tilde{f}\Big\|_{\mathcal{L}_2(X,\mu)} \leq \frac{C}{\sqrt{s}}\Big(1 + \sqrt{2\ \log\frac{1}{\delta_1}}\Big).$$

Since $\hat{f}(x) = \frac{1}{r}\sum_{i=1}^{r}\sum_{k=1}^{s}\frac{\beta_k}{s}\phi(g_i x,\omega_k)$ and $\mathbb{E}_{g_i\sim q}\hat{f}(x) = \tilde{f}(x)$ with g_i iid (and $\{\omega_k\}_{k=1}^{s}$ fixed beforehand), we can apply Lemma A.1 with

$$M = \Bigg\|\sum_{k=1}^{s}\frac{\beta_k}{s}\phi(g_i x,\omega_k)\Bigg\| \leq \sum_{k=1}^{s}\Bigg|\frac{\beta_k}{s}\Bigg|\,\|\phi(g_i x,\omega_k)\| \leq \sum_{k=1}^{s}\Bigg|\frac{\beta_k}{s}\Bigg| \leq C.$$

We conclude that with a probability at least $1 - \delta_2$,

$$\Big\|\tilde{f} - \hat{f}\Big\|_{\mathcal{L}_2(X,\mu)} \leq \frac{C}{\sqrt{r}}\Big(1 + \sqrt{2\log\frac{1}{\delta_2}}\Big).$$

Hence, with probability at least $1-\delta_1-\delta_2$, we have

$$\left\|f^\star - \hat{f}\right\|_{\mathcal{L}_2(X,\mu)} \leq \frac{C}{\sqrt{s}}\Big(1+\sqrt{2\ \log\frac{1}{\delta_1}}\Big) + \frac{C}{\sqrt{r}}\Big(1+\sqrt{2\log\frac{1}{\delta_2}}\Big)$$

□

Theorem A.3 (Estimation error [35])**.** *Let $\mathcal{F}$ be a bounded class of functions, $\sup_{x\in X}|f(x)| \leq C$ for all $f \in \mathcal{F}$. Let $V(y_i f(x_i))$ be an L-Lipschitz loss. Then with probability $1-\delta$, with respect to training samples $\{x_i, y_i\}_{i=1,2\cdots N}$ (iid $\sim P$), every f satisfies*

$$\mathcal{E}_V(f) \leq \hat{\mathcal{E}}_V(f) + 4L\mathcal{R}_N(\mathcal{F}) + \frac{2|V(0)|}{\sqrt{N}} + LC\sqrt{\frac{1}{2N}\log\frac{1}{\delta}},$$

where $\mathcal{R}_N(\mathcal{F})$ is the Rademacher complexity of the class $\mathcal{F}$:

$$\mathcal{R}_N(\mathcal{F}) = \mathbb{E}_{x,\sigma}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{i=1}^{N}\sigma_i f(x_i)\right|\right],$$

and σ_i are iid symmetric Bernoulli random variables taking value in $\{-1,1\}$, with equal probability and are independent form x_i.

Proof. See in [35]. □

Let $f \in \mathcal{F}_p$ and $\hat{f} \in \hat{\mathcal{F}}_p$ then the *approximation error* is bounded as

$$\begin{aligned}
\mathcal{E}_V(\hat{f}) - \mathcal{E}_V(f) &\leq \mathbb{E}_{(x,y)\sim P}\Big|V(y\hat{f}(x)) - V(yf(x))\Big| \\
&\leq L\mathbb{E}\big|\hat{f}(x) - f(x)\big| \\
&\leq L\sqrt{\mathbb{E}\big(\hat{f}(x)-f(x)\big)^2} \quad \text{(Jensen's inequaity for } \sqrt{\cdot} \text{ concave function)} \\
&\leq LC\bigg(\frac{1}{\sqrt{s}}\Big(1+\sqrt{2\ \log\frac{1}{\delta_1}}\Big) + \frac{1}{\sqrt{r}}\Big(1+\sqrt{2\log\frac{1}{\delta_2}}\Big)\bigg),
\end{aligned}$$

with probability at least $1-\delta_1-\delta_2$. Now let $f_N^\star = \arg\min_{f\in\hat{\mathcal{F}}_p}\hat{\mathcal{E}}_V(f)$ and $\tilde{f} = \arg\min_{f\in\hat{\mathcal{F}}_p}\mathcal{E}_V(f)$. We have

$$\begin{aligned}
\mathcal{E}_V(f_N^\star) - \min_{f\in\mathcal{F}_P}\mathcal{E}_V(f) &= \hat{\mathcal{E}_V}(f_N^\star) - \mathcal{E}_V(\tilde{f}) + \mathcal{E}_V(\tilde{f}) - \min_{f\in\mathcal{F}_P}\mathcal{E}_V(f) \\
&\leq 2\sup_{\tilde{f}\in\hat{\mathcal{F}}_P}\Big|\mathcal{E}_V(\tilde{f}) - \hat{\mathcal{E}_V}(\tilde{f})\Big| + L\bigg(\frac{C}{\sqrt{s}}\Big(1+\sqrt{2\ \log\frac{1}{\delta_1}}\Big) + \frac{C}{\sqrt{r}}\Big(1+\sqrt{2\log\frac{1}{\delta_2}}\Big)\bigg) \\
&\leq 2\bigg(4L\mathcal{R}_N(\mathcal{F}) + \frac{2|V(0)|}{\sqrt{N}} + LC\sqrt{\frac{1}{2N}\log\frac{1}{\delta}}\bigg) + \\
&\qquad LC\bigg(\frac{1}{\sqrt{s}}\Big(1+\sqrt{2\ \log\frac{1}{\delta_1}}\Big) + \frac{1}{\sqrt{r}}\Big(1+\sqrt{2\log\frac{1}{\delta_2}}\Big)\bigg),
\end{aligned}$$

with probability at least $1-\delta-\delta_1-\delta_2$. It is easy to show that $\mathcal{R}_N(\mathcal{F}) \leq \frac{C}{\sqrt{N}}$. Taking $\delta=\delta_1=\delta_2$ yields the statement of the theorem.

□