On Consistency of Non-convex RKDES

Anonymous Author(s)
Affiliation
Address
email

1 In progress

Definition 1. Let H be a Hilbert space and $S \subset H$. We call S *postive* if, for all $s, s' \in S$, $\langle s, s' \rangle \geq 0$.

Proposition 1. *If S is a positive set then* $\operatorname{conv} S$ *is also a positive set.*

The proof of this is obvious.

Proposition 2. *If S is a positive set then* $\overline{\operatorname{conv} S}$ *is a positive set.*

Proof. Let $s, s' \in \overline{\operatorname{conv} S}$ and $(s_i)_{i=1}^{\infty}, (s_i')_{i=1}^{\infty}$ such that $s_i \to s$ and $s_i' \to s'$. This will be easy. ☐

Proposition 3. *Let S, T be positive sets and μ a probability measure composed of a finite number of atoms,ie $\mu = \sum_{i=1}^{m} \alpha_i \delta_{s_i}$ with $\|s_i\| = C$ on S and $q : S \to T$ and $x, x' \in S$. Then*

$$\begin{aligned}
&\left\| \int q(s) \left[\left(2 \langle x, s \rangle - \|x\|^2 \right)^+ - \left(2 \langle x', s \rangle - \|x'\|^2 \right)^+ \right] d\mu(s) \right\| \\
&\leq \left(\left| \|x\|^2 - \|x'\|^2 \right| + 2C \|x - x'\| \right) \left\| \int q(s) d\mu(s) \right\|.
\end{aligned}$$

Proof of Proposition 3. We have

$$\begin{aligned}
&\left\| \int q(s) \left[\left(2 \langle x, s \rangle - \|x\|^2 \right)^+ - \left(2 \langle x', s \rangle - \|x'\|^2 \right)^+ \right] d\mu(s) \right\| \\
&\leq \left\| \int q(s) \left| 2 \langle x, s \rangle - \|x\|^2 - 2 \langle x', s \rangle + \|x'\|^2 \right| d\mu(s) \right\| \\
&\leq \left\| \int q(s) \left| 2 \langle x, s \rangle - 2 \langle x', s \rangle \right| d\mu(s) \right\| \\
&\quad + \left\| \int q(s) \left| \|x\|^2 - \|x'\|^2 \right| d\mu(s) \right\|.
\end{aligned}$$

For the second summand we have

$$\left\| \int q(s) \left| \|x\|^2 - \|x'\|^2 \right| d\mu(s) \right\| \quad = \quad \left| \|x\|^2 - \|x'\|^2 \right| \left\| \int q(s) d\mu(s) \right\|.$$

And for the first summand we have

$$\begin{aligned}
\left\| \int q(s) \left| 2 \langle x, s \rangle - 2 \langle x', s \rangle \right| d\mu(s) \right\| &= \left\| \sum_{i=1}^{n} \alpha_i q(s_i) \left| 2 \langle x, s_i \rangle - 2 \langle x', s_i \rangle \right| \right\| \\
&= \max_{\delta \in \{-1,1\}^m} \left\| \sum_{i=1}^{n} \delta_i \alpha_i q(s_i) \left[2 \langle x, s_i \rangle - 2 \langle x', s_i \rangle \right] \right\| \\
&= 2 \max_{\delta \in \{-1,1\}^m} \left\| \sum_{i=1}^{n} \delta_i \alpha_i q(s_i) \langle x - x', s_i \rangle \right\| \\
&\leq 2 \max_{\delta \in \{-1,1\}^m} \left\| \sum_{i=1}^{n} \delta_i \alpha_i q(s_i) \|x - x'\| \|s_i\| \right\| \\
&= 2 \|x - x'\| C \max_{\delta \in \{-1,1\}^m} \left\| \sum_{i=1}^{n} \delta_i \alpha_i q(s_i) \right\| \\
&\leq 2 \|x - x'\| C \left\| \sum_{i=1}^{n} \alpha_i q(s_i) \right\| \\
&= 2 \|x - x'\| C \left\| \int q(s) d\mu(s) \right\|
\end{aligned}$$

¡++¿ □

Proposition 4. *If* $n\sigma^{2d} \to 0$ *then with probability going to one the following holds*

$$\max_{\delta \in \{-1,1\}^n} \left\| \frac{1}{n} \sum_{i=1}^{n} \delta_i \Phi_\sigma(X_i) \langle \Phi_\sigma(X_i), \cdot \rangle \right\|_{op} < 2 \|f\|_\infty .$$

Proof of Proposition 4. Let $T_{\sigma,n}$ be an operator which maximizes the previous epxression. First we will show that $T_{\sigma,n}$ is Hermetian. Let $g, h \in \mathcal{H}_\sigma$ and $\delta \in \{-1,1\}^n$ be arbitrary,

$$\begin{aligned}
\left\langle \frac{1}{n} \sum_{i=1}^{n} \delta_i \Phi_\sigma(X_i) \langle \Phi_\sigma(X_i), g \rangle, h \right\rangle &= \frac{1}{n} \sum_{i=1}^{n} \delta_i \langle \Phi_\sigma(X_i) \langle \Phi_\sigma(X_i), g \rangle, h \rangle \\
&= \frac{1}{n} \sum_{i=1}^{n} \delta_i \langle \Phi_\sigma(X_i), h \rangle \langle \Phi_\sigma(X_i), g \rangle \\
&= \frac{1}{n} \sum_{i=1}^{n} \delta_i \langle \langle \Phi_\sigma(X_i), g \rangle \Phi_\sigma(X_i), h \rangle \\
&= \frac{1}{n} \sum_{i=1}^{n} \delta_i \langle \Phi_\sigma(X_i) \langle \Phi_\sigma(X_i), h \rangle, g \rangle \\
&= \left\langle \frac{1}{n} \sum_{i=1}^{n} \delta_i \Phi_\sigma(X_i) \langle \Phi_\sigma(X_i), h \rangle, g \right\rangle
\end{aligned}$$

Because $T_{\sigma,n}$ is Hermetian, we have that $\|T_{\sigma,n}\|_{op} = \max_{v\in\mathcal{H}_\sigma:\|v\|=1}\langle Tv, v\rangle$ (Proposition 7.36 in [1]). Using this we have

$$\begin{aligned}
\|T_{\sigma,n}\|_{op} &= \max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^m \delta_i\Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),\cdot\rangle\right\|_{op}\\
&= \max_{v:\|v\|=1}\max_{\delta\in\{-1,1\}^n}\left|\left\langle v, \frac{1}{n}\sum_{i=1}^n \delta_i\Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),v\rangle\right\rangle\right|\\
&= \max_{v:\|v\|=1}\max_{\delta\in\{-1,1\}^n}\left|\sum \frac{1}{n}\delta_i\langle v,\Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),v\rangle\rangle\right|\\
&= \max_{v:\|v\|=1}\max_{\delta\in\{-1,1\}^n}\left|\sum \frac{1}{n}\delta_i\langle\Phi_\sigma(X_i),v\rangle^2\right|\\
&= \max_{v:\|v\|=1}\sum \frac{1}{n}\langle\Phi_\sigma(X_i),v\rangle^2\\
&= \max_{v:\|v\|=1}\left|\left\langle v, \frac{1}{n}\sum_{i=1}^m \Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),v\rangle\right\rangle\right|\\
&= \left\|\frac{1}{n}\sum_{i=1}^m \Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),\cdot\rangle\right\|_{op}.
\end{aligned}$$

We will simply let $T_{\sigma,n}$ be the operator in the last line. We will now consider the expected value of $T_{\sigma,n}$ as a Boncher integral. First we must show that this is in fact Bochner integrable. First we will bound $\|\Phi_\sigma(x)\langle\Phi_\sigma(x),\cdot\rangle\|_{op}$ for arbitary x. We have

$$\begin{aligned}
\|\Phi_\sigma(x)\langle\Phi_\sigma(x),\cdot\rangle\|_{op} &= \max_{v\in\mathcal{H}_\sigma:\|v\|=1}\|\Phi_\sigma(x)\langle\Phi_\sigma(x),v\rangle\|_{\mathcal{H}_\sigma}\\
&\le \max_{v\in\mathcal{H}_\sigma:\|v\|=1}\|\Phi_\sigma(x)\|\,\|\Phi_\sigma(x)\|\,\|v\|\\
&= \|\Phi_\sigma\|_{\mathcal{H}_\sigma}^2.
\end{aligned}$$

Now we can show that the expected value of $T_{\sigma,n}$ is Bochner integrable,

$$\begin{aligned}
\mathbb{E}_{X_1,\dots X_n \overset{iid}{\sim} f}\left[\|T_{\sigma,n}\|\right] &= \mathbb{E}_{X_1,\dots X_n \overset{iid}{\sim} f}\left[\left\|\frac{1}{n}\sum_{i=1}^m \Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),\cdot\rangle\right\|_{op}\right]\\
&\le \mathbb{E}_{X_1,\dots X_n \overset{iid}{\sim} f}\left[\frac{1}{n}\sum_{i=1}^m \|\Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),\cdot\rangle\|_{op}\right]\\
&\le \mathbb{E}_{X_1,\dots X_n \overset{iid}{\sim} f}\left[\frac{1}{n}\sum_{i=1}^m \|\Phi_\sigma\|_{\mathcal{H}_\sigma}^2\right]\\
&= \mathbb{E}_{X_1,\dots X_n \overset{iid}{\sim} f}\left[\|\Phi_\sigma\|_{\mathcal{H}_\sigma}^2\right]\\
&= \|\Phi_\sigma\|_{\mathcal{H}_\sigma}^2.
\end{aligned}$$

Evaluating the Bocher integral of $T_{\sigma,n}$ we get

$$\left\|\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^m \Phi_\sigma(X_i)\langle\Phi_\sigma(X_i),\cdot\rangle\right]\right\|_{op} = \left\|\int f(x)\Phi_\sigma(x)\langle\Phi_\sigma(x),\cdot\rangle\right\|_{op}.$$

We will call the operator in the last line T_σ. T_σ is also Hermetian: let $g, h \in \mathcal{H}_\sigma$ be arbitrary,

$$\begin{aligned}
\langle T_\sigma g, h\rangle &= \left\langle \int f(x)\Phi_\sigma(x)\left\langle \Phi_\sigma(x), g\right\rangle dx, h\right\rangle \\
&= \int f(x)\left\langle \Phi_\sigma(x)\left\langle \Phi_\sigma(x), g\right\rangle, h\right\rangle dx \\
&= \int f(x)g(x)h(x)dx \\
&= \left\langle \int f(x)\Phi_\sigma(x)\left\langle \Phi_\sigma(x), h\right\rangle dx, g\right\rangle \\
&= \left\langle g, T_\sigma h\right\rangle .
\end{aligned}$$

From this we have have that $\|T_\sigma\| = \max_{g:\|g\|=1}\langle Tg, g\rangle$. Note that from this we have

$$\begin{aligned}
\|T\|_{op} &= \langle T_\sigma g, g\rangle \\
&= \int f(x)g^2(x)dx \\
&\leq \|f\|_\infty \int g^2(x)dx.
\end{aligned}$$

we have that $\|T_\sigma\|_{op} \leq \|f\|_\infty \left\|\int \Phi_\sigma(x)\left\langle \Phi_\sigma(x), \cdot\right\rangle dx\right\|_{op}$. We denote the operator on the right as S_σ. We will now find the operator norm of S_σ.

Lemma 1. $\|S_\sigma\| \leq 1$.

Proof of Lemma 1. First observe that, for any L2 THING, NEED TO BE DELICATE AND FINISH L2

Let $x, y \in \mathbb{R}^d$ be arbitrary, we have Let $g \in \operatorname{span}\left(\left\{\Phi_\sigma(x) : x \in \mathbb{R}^d\right\}\right)$, with $g = \sum_{i=1}^n a_i \Phi_\sigma(x_i)$. We have that

$$\begin{aligned}
\|g\|^2_{\mathcal{H}_\sigma} &= \sum_{i,j} a_i a_j \left\langle \Phi_\sigma(x_i), \Phi_\sigma(x_j)\right\rangle \\
&= \sum_{i,j} a_i a_j k_\sigma\left(x_i, x_j\right) \\
&= \sum_{i,j} a_i a_j k_{\frac{\sigma}{\sqrt{2}}} * k_{\frac{\sigma}{\sqrt{2}}}(x_i, \cdot)(x_j) \\
&= \sum_{i,j} a_i a_j \left\langle k_{\frac{\sigma}{\sqrt{2}}}(x_i, \cdot), k_{\frac{\sigma}{\sqrt{2}}}(x_j, \cdot)\right\rangle \\
&= \left\|\sum_{i=1}^n a_i k_{\frac{\sigma}{\sqrt{2}}}(x_i, \cdot)\right\|^2 \\
&\geq \left\|\sum_i a_i k_{\sigma\sqrt{\frac{1}{2}}}(x_i, \cdot) * k_{\frac{\sigma}{\sqrt{2}}}\right\|^2 \\
&= \left\|\sum_i a_i k_\sigma\left(x_i, \cdot\right)\right\|^2 \\
&= \|g\|^2
\end{aligned}$$

where the inequality follows form Young's Inequality. Let $g \in \mathcal{H}_\sigma$ and let $g_1, g_2, \ldots \in \operatorname{span}\left(\left\{\Phi_\sigma(x) : x \in \mathbb{R}^d\right\}\right)$ such that $g_i \to g$. We have that g_i is a Cauchy sequence in $\mathcal{H}_\sigma$ and since $\|g_i - g_j\|_{L^2} \leq \|g_i - g_j\|_{\mathcal{H}_\sigma}$ it follows that it is also a Cauchy sequence in L^2. Let this sequence converge to g^*. We also have that

$$\begin{aligned}
\|g - g_i\|_\infty &= \max_x |g(x) - g_i(x)| \\
&= \max_x |\langle g, \Phi_\sigma(x)\rangle - \langle g_i, \Phi_\sigma(x)\rangle| \\
&= \max_x |\langle g - g_i, \Phi_\sigma(x)\rangle| \\
&\leq \|g - g_i\|_{\mathcal{H}_\sigma} \|\Phi_\sigma\|_{\mathcal{H}_\sigma}
\end{aligned}$$

so g_i converges to g uniformly. We will now show that $g = g^*$ almost everywhere. Suppose that this weren't true and there existed a set A of positive measure and $\varepsilon > 0$ such that $|g(a) - g^*(a)| > \varepsilon$ for all $a \in A$. Then it would follow that $\|g^* - g_i\|_2 \not\to 0$, a contradiction thus we have that $g = g^*$ and $\|g_i - g\|_2 \to 0$.

Since $\|g_i\|_2 \to \|g\|_2, \|g_i\|_{\mathcal{H}_\sigma} \to \|g\|_{\mathcal{H}_\sigma}$ and $\|g_i\|_2 \leq \|g_i\|_{\mathcal{H}_\sigma}$ it follows that $\|g\|_2 \leq \|g\|_{\mathcal{H}_\sigma}$. Thus for all $g \in \mathcal{H}_\sigma$, $\|g\|_2 \leq \|g\|_{\mathcal{H}_\sigma}$. To finish up,

$$\begin{aligned}
\|S\| &= \max_{g \in \mathcal{H}_\sigma : \|g\|_{\mathcal{H}_\sigma} = 1} \langle g, Sg\rangle \\
&= \max_{g \in \mathcal{H}_\sigma : \|g\|_{\mathcal{H}_\sigma} = 1} \|g\|_2 \\
&\leq \max_{g \in \mathcal{H}_\sigma : \|g\|_{\mathcal{H}_\sigma} = 1} \|g\|_{\mathcal{H}_\sigma} \\
&= 1.
\end{aligned}$$

□

□

2 Basics

L^q will always refer to the standard space induced by the Lebesgue measure. Let $\mathbb{R}_+$ be the set of nonnegative reals. Let $(\cdot)^+$ denote the postive component of whatever is in the parenthesis. All integrals and spaces will be defined with respect to the d-dimensional Lebesgue measure. Let $f \in L^1\left(\mathbb{R}^d\right) \cap L^\infty\left(\mathbb{R}^d\right)$ be a pdf and $X_1, \ldots, X_n$ be iid samples from f.

Let $k_\sigma(x, x')$ be a radial smoothing kernel of the form $k_\sigma(x, x') = \sigma^{-d} q\left(\|x - x'\|_2 / \sigma\right)$ for some function $q \geq 0$ such that $q\left(\|\cdot\|_2\right)$ is a pdf on $\mathbb{R}^d$. Then

$$\bar{f}_{\sigma,n} := \frac{1}{n} \sum_{i=1}^n k_\sigma(\cdot, X_i)$$

is the standard KDE. We will assume k_σ corresponds to a PSD kernel so that k_σ is an element of a RKHS with feature map Φ_σ. Furthermore we will define

$$\bar{f}_\sigma = \int_{\mathbb{R}^d} \Phi_\sigma(x) f(x)\, dx.$$

Let $\mathcal{D}$ be the space of square-integrable pdfs on $\mathbb{R}^d$.

Let

$$\mathcal{D}_\sigma = \left\{ \int \Phi_\sigma(x) d\nu(x) \middle| \nu \text{ is a probability measure} \right\}.$$

Let

$$\mathcal{D}_{\sigma,n} = \left\{ \sum_{i=1}^n \Phi_\sigma(X_i) w_i : \sum_i w_i = 1, w_i \geq 0 \forall i \right\}.$$

Note that this is not necessarily well defined as it may be $0/0$. The $\frac{1}{n}$ coefficients are unneccary for the defninition but will be handy when performing analysis.

Let $R:\mathcal{D}\to\mathcal{D}$ be defined as

$$\begin{aligned}R(g)(x) &= \frac{N(g)(x)}{D(g)}\\ N(g)(x) &= f(x)\left(2g(x)-\|g\|^2_{L^2(\mathbb{R}^d)}\right)^+\\ D(g) &= \int f(y)\left(2g(y)-\|g\|^2_{L^2(\mathbb{R}^d)}\right)^+ dy.\end{aligned}$$

For $m\geq 0$ we define the following

$$R^m(g) = \underbrace{R(R(\cdots(R}_{m\text{ times}}(g))\cdots))$$

We define $R_\sigma:\mathcal{D}_\sigma\to\mathcal{D}_\sigma$ as

$$\begin{aligned}R_\sigma(g) &= \frac{N_\sigma(g)}{D_\sigma(g)}\\ D_\sigma(g) &= \int f(x)\,\Phi_\sigma(x)\left(2g(x)-\|g\|^2_{\mathcal{H}_\sigma}\right)^+ dx\\ N_\sigma(g) &= \int f(y)\left(2g(y)-\|g\|^2_{\mathcal{H}_\sigma}\right)^+ dy\end{aligned}$$

For $m\geq 0$ we define the following

$$R^m_\sigma(g) = \underbrace{R_\sigma(R_\sigma(\cdots(R_\sigma}_{m\text{ times}}(g))\cdots))$$

Define $R_{\sigma,n}:\mathcal{D}_\sigma\to\mathcal{D}_\sigma$

$$R_{\sigma,n}(g) = \frac{N_{\sigma,n}(g)}{D_{\sigma,n}(g)} \quad (1)$$

$$N_{\sigma,n}(g) = \frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(2g(X_i)-\|g\|^2_{\mathcal{H}_\sigma}\right)^+ \quad (2)$$

$$D_{\sigma,n}(g) = \frac{1}{n}\sum_{j=1}^n \left(2g(X_j)-\|g\|^2_{\mathcal{H}_\sigma}\right)^+. \quad (3)$$

$$\|f\|^2_{L^2} = \|ff\|_{L^1} \leq \|f\|_{L^1}\|f\|_{L^\infty} = \|f\|_{L^\infty}.$$

We will define R^0 for whatever flavor of R to just be the identity function. FIX AT SOME POINT!

3 Main Results

Theorem 1. *For all $m\geq 0$, $\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m(f)\right\| \xrightarrow{p} 0$.*

Proof. By repeated application of the triangle inequality, we know that

$$\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m(f)\right\|$$

is less than or equal to

$$\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|+\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)-R^m_\sigma\left(\bar{f}_\sigma\right)\right\|+\left\|R^m_\sigma\left(\bar{f}_\sigma\right)-R^m(f)\right\|.$$

We will show all the summands in the previous expression converge to zero in probability. ☐

Though this may seem contrived it will be convenient if, for the remainer of the paper we define

$$\begin{aligned} N_\sigma\left(R_\sigma^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq \bar{f}_\sigma \\ N\left(R^{-1}\left(f\right)\right) &\triangleq f \\ D_\sigma\left(R_\sigma^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq 1 \\ D\left(R^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq 1 \end{aligned}$$

Lemma 2. *For all $m \geq 0$, we have*

1. $\left\|R_\sigma^m\left(\bar{f}_\sigma\right) - R^m\left(f\right)\right\|_{L^2} \to 0$
2. $\left|\|R^m(f)\|_{L^2} - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}\right| \to 0$
3. $\left\|N_\sigma\left(R_\sigma^{m-1}\left(\bar{f}_\sigma\right)\right) - N\left(R^{m-1}\left(f\right)\right)\right\|_{L^2} \to 0$
4. $\left|D_\sigma\left(R_\sigma^{m-1}\left(\bar{f}_\sigma\right)\right) - D\left(R^{m-1}\left(f\right)\right)\right| \to 0$

as $\sigma \to 0$.

Proof of Lemma 2. We will proceed by induction. The base case $m = 0$ clearly holds for (1), (3), and (4). For (2) we have

$$\begin{aligned} \left\|\bar{f}_\sigma\right\|^2_{\mathcal{H}_\sigma} &= \left\langle \int f(x)\Phi_\sigma(x)dx, \int f(y)\Phi_\sigma(y)dy \right\rangle_{\mathcal{H}_\sigma} \\ &= \left\langle \int f(x)\Phi_\sigma(x)dx, \int f(y)\Phi_\sigma(y)dy \right\rangle_{\mathcal{H}_\sigma} \\ &= \int\int f(x)f(y)k_\sigma\left(x,y\right)dxdy \\ &= \int\int f(x)\left(f * k_\sigma\right)(x)dx \\ &= \langle f, f * k_\sigma\rangle_{L^2} \end{aligned}$$

where the last line goes to 0 due to Proposition 6.

For the induction step suppose that the Lemma holds for m. Let $R_\sigma^m\left(\bar{f}_\sigma\right) = f_\sigma^m$ and $R^m\left(f\right) = f^m$ Since $D\left(f^m\right)$ and $N\left(f^m\right)$ are fixed as σ varies, to (1) $\left\|R_\sigma^m\left(\bar{f}_\sigma\right) - R^m\left(f\right)\right\|_2 \to 0$ it is sufficient to show (4) $\left|D_\sigma\left(f_\sigma^m\right) - D\left(f^m\right)\right| \to 0$ and (3) $\left\|N_\sigma^m\left(\bar{f}_\sigma\right) - N^m\left(f\right)\right\|_2 \to 0$. First We will take care of (3).

$$\begin{aligned} &\left\|N_\sigma\left(f_\sigma^m\right) - N\left(f^m\right)\right\|_2 \\ &= \left\|\int \Phi_\sigma\left(x\right)f(x)\left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|^2_{\mathcal{H}_\sigma}\right)^+ dx - f(\cdot)\left(2f^m\left(\cdot\right) - \|f^m\|^2_{L^2}\right)^+\right\|_{L^2} \\ &\leq \left\|\int \Phi_\sigma\left(x\right)f(x)\left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|^2_{\mathcal{H}_\sigma}\right)^+ dx - \int \Phi_\sigma\left(y\right)f(y)\left(2f^m\left(y\right) - \|f^m\|^2_{L^2}\right)^+ dy\right\|_{L^2} \\ &\quad + \left\|\int \Phi_\sigma\left(x\right)f(x)\left(2f^m\left(x\right) - \|f^m\|^2_{L^2}\right)^+ dx - f(\cdot)\left(2f^m\left(\cdot\right) - \|f^m\|^2_{L^2}\right)^+\right\|_{L^2}. \end{aligned}$$

The second summand goes to zero by standard kern to zero (Folland). Now we can take care of the first summand. Using Corollaries 1 and 2, we can do the following.

$$\begin{aligned}
&\left\|\int \Phi_\sigma\left(x\right) f(x) \left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|_{\mathcal{H}_\sigma}^2\right)^+ dx - \int \Phi_\sigma\left(y\right) f(y) \left(2f^m\left(y\right) - \|f^m\|_{L^2}^2\right)^+ dy\right\|_{L^2} \\
&= \left\|\int \Phi_\sigma\left(x\right) f(x) \left[\left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|_{\mathcal{H}_\sigma}^2\right)^+ - \left(2f^m\left(x\right) + \|f^m\|_{L^2}^2\right)^+\right] dx\right\|_{L^2} \\
&\leq \left\|\int \Phi_\sigma\left(x\right) f(x) \left|2f_\sigma^m\left(x\right) - \|f_\sigma^m\|_{\mathcal{H}_\sigma}^2 - 2f^m\left(x\right) + \|f^m\|_{L^2}^2\right| dx\right\|_{L^2} \\
&\leq \left\|\int \Phi_\sigma\left(x\right) f(x) \left[\left|2f_\sigma^m\left(x\right) - 2f^m\left(x\right)\right| + \left|\|f_\sigma^m\|_{\mathcal{H}_\sigma}^2 - \|f^m\|_{L^2}^2\right|\right] dx\right\|_{L^2} \\
&\leq \left\|\int \Phi_\sigma\left(x\right) f(x) \left|2f_\sigma^m\left(x\right) - 2f^m\left(x\right)\right| dx\right\|_{L^2} \\
&\quad \ldots + \left\|\int \Phi_\sigma\left(x\right) f(x) \left|\|f_\sigma^m\|_{\mathcal{H}_\sigma}^2 - \|f^m\|_{L^2}^2\right| dx\right\|_{L^2}.
\end{aligned}$$

The second summand is easy to dispatch

$$\left\|\int \Phi_\sigma\left(x\right) f(x) \left|\|f_\sigma^m\|_{\mathcal{H}_\sigma}^2 - \|f^m\|_{L^2}^2\right| dx\right\|_{L^2} \quad \leq \quad \left\|\int \Phi_\sigma\left(x\right) f(x) dx\right\|_{L^2} \left|\|f_\sigma^m\|_{\mathcal{H}_\sigma}^2 - \|f^m\|_{L^2}^2\right|,$$

we have already shown that the left factor goes to $\|f\|_{L^2}$ and the right factor goes to zero by the inductive hypothesis. Returning to the first summand we have, by Young's inequality (and some other stuff)…

$$\begin{aligned}
\left\|\int \Phi_\sigma\left(x\right) f(x) \left|2f_\sigma^m\left(x\right) - 2f^m\left(x\right)\right| dx\right\|_{L^2} &= \left\|k_\sigma * \left[f\left|2f_\sigma^m - 2f^m\right|\right]\right\|_{L^2} \\
&= \left\|f\left|2f_\sigma^m - 2f^m\right|\right\|_{L^2} \\
&= \|f\|_{L^\infty} \left\|\left|2f_\sigma^m - 2f^m\right|\right\|_{L^2}.
\end{aligned}$$

which goes to zero by the inductive hypothesis.

Now to take care of (4) we have

$$\begin{aligned}
&\left|D_\sigma\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right) - D\left(R^m\left(f\right)\right)\right| \\
&= \left|\int f(x) \left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|_{\mathcal{H}_\sigma}^2\right)^+ dx - \int f(y) \left(2f^m\left(y\right) - \|f^m\|_{L^2}^2\right)^+ dy\right| \\
&= \left|\int f(x) \left[\left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|_{\mathcal{H}_\sigma}^2\right)^+ - \left(2f^m\left(x\right) - \|f^m\|_{L^2}^2\right)^+\right] dx\right| \\
&\leq \left|\int f(x) \left|\left(2f_\sigma^m\left(x\right) - \|f_\sigma^m\|_{\mathcal{H}_\sigma}^2\right) - \left(2f^m\left(x\right) - \|f^m\|_{L^2}^2\right)\right| dx\right| \\
&\leq \left|\int f(x) \left|2f_\sigma^m\left(x\right) - 2f^m\left(x\right)\right| dx\right| \\
&\quad + \left|\int f(x) \left|\|f_\sigma^m\|_{\mathcal{H}_\sigma}^2 - \|f^m\|_{L^2}^2\right| dx\right|.
\end{aligned}$$

The second summand goes to zero by the inductive hypothesis. To take care of of the first summand

$$\begin{aligned}
&\left| \int f(x) \left| 2f_\sigma^m(x) - 2f^m(x) \right| dx \right| \\
&= \mathbb{E}_{X\sim f}\left[\left| 2f_\sigma^m(X) - 2f^m(X) \right| \right] \\
&= \mathbb{E}_{X\sim f}\left[\sqrt{\left(2f_\sigma^m(X) - 2f^m(X)\right)^2} \right] \\
&\leq \sqrt{\mathbb{E}_{X\sim f}\left[\left(2f_\sigma^m(X) - 2f^m(X)\right)^2 \right]} \\
&\leq 2\sqrt{f(x)\left(f_\sigma^m(x) - f^m(x)\right)^2} \\
&\leq 2\sqrt{\|f\|_\infty \left\|f_\sigma^m - f^m\right\|_2^2}
\end{aligned}$$

which goes to zero by the inductive hypothesis.

The convergence of (3) and (4) implies (1). (1) clearly implies (2) so we are done. □

We define the following

$$\begin{aligned}
N_{\sigma,n}\left(R_{\sigma,n}^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq \bar{f}_{\sigma,n} \\
N_\sigma\left(R_\sigma^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq \bar{f}_\sigma \\
D_{\sigma,n}\left(R_{\sigma,n}^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq 1 \\
D_\sigma\left(R_\sigma^{-1}\left(\bar{f}_\sigma\right)\right) &\triangleq 1.
\end{aligned}$$

Lemma 3. *If $n \to 0$, with $\sigma \to 0$ and something with the rate of bandwitdh then for all $m \geq 0$ we have,*

$$\begin{aligned}
\left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right) - R_\sigma^m\left(\bar{f}_\sigma\right)\right\| &\xrightarrow{p} 0 \\
\left\|N_{\sigma,n}\left(R_{\sigma,n}^{m-1}\left(\bar{f}_\sigma\right)\right) - N_\sigma\left(R_\sigma^{m-1}\left(\bar{f}_\sigma\right)\right)\right\| &\xrightarrow{p} 0 \\
\left|D_{\sigma,n}\left(R_{\sigma,n}^{m-1}\left(\bar{f}_\sigma\right)\right) - D_\sigma\left(R_\sigma^{m-1}\left(\bar{f}_\sigma\right)\right)\right| &\xrightarrow{p} 0.
\end{aligned}$$

Proof. We will proceed by induction. We have that the base case $m = 0$ is trivial. For the inductive step, suppose the lemma statement holds for m. We will first prove the lemma for the last two limits and then use those to prove the first. First we will show that $\left\|N_{\sigma,n}\left(R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right) - N_\sigma\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right)\right\| \xrightarrow{p} 0$. We can decompose this norm as

$$\begin{aligned}
&\left\|N_{\sigma,n}\left(R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right) - N_\sigma\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right)\right\| \\
&\leq \left\|N_{\sigma,n}\left(R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right) - N_{\sigma,n}\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right)\right\| + \left\|N_{\sigma,n}\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right) - N_\sigma\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right)\right\|
\end{aligned}$$

First we will demonstrate that the right summand goes to zero. To do this we will need the following lemma

Lemma 4. *For all $l \geq 0$*

$$\left\|R_\sigma^l\left(\bar{f}_\sigma\right)\right\|_\infty < \infty.$$

Proof of Lemma 4. Will proceed by induction. For the base case we have that $\left\|\bar{f}_\sigma\right\|_\infty < \infty$ by assumption. Now let $R_\sigma^m\left(\bar{f}_\sigma\right) = C < \infty$. We have the following

$$R_\sigma\left(R_\sigma^m\left(\bar{f}_\sigma\right)\right) = \frac{\int \Phi_\sigma(x) f(x) \left(2R_\sigma^m\left(\bar{f}_\sigma\right)(x) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ dx}{\int f(y) \left(2R_\sigma^m\left(\bar{f}_\sigma\right)(y) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ dy}.$$

The denominator is just some positive value so it is sufficient to show that the numerator is bounded. With this in mind we have

$$\begin{aligned}
\left\|\int \Phi_\sigma\left(x\right) f(x) \left(2R_\sigma^m\left(\bar{f}_\sigma\right)(x) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ dx\right\|_\infty &= \left\|k_\sigma * f(\cdot)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(\cdot) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ dx\right\|_\infty \\
&\leq \left\|f(\cdot)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(\cdot) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+\right\|_\infty \\
&\leq \|f\|_\infty \left\|\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(\cdot) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+\right\|_\infty \\
&\leq \|f\|_\infty \left\|2R_\sigma^m\left(\bar{f}_\sigma\right)(\cdot)\right\|_\infty \\
&< \infty,
\end{aligned}$$

and we are done. □

Returning to the right summand we have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ - \int_x \Phi_\sigma(x)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(x) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ dx\right\| \\
&= \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|^2\right)^+ - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|^2\right)^+\right]\right\|
\end{aligned}$$

which goes to zero by standard kde convergence (NEED to show). Now for the left hand side,

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(2R_{\sigma,n}^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ - \frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+\right\| \\
&= \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left(\left(2R_{\sigma,n}^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+ - \left(2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right)^+\right)\right\| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left|2R_{\sigma,n}^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2 - 2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) + \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right|\right\| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left|2R_{\sigma,n}^m\left(\bar{f}_\sigma\right)(X_i) - \left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2 - 2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i) + \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right|\right\| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left|2R_{\sigma,n}^m\left(\bar{f}_\sigma\right)(X_i) - 2R_\sigma^m\left(\bar{f}_\sigma\right)(X_i)\right|\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left|\left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2 - \left\|R_\sigma^m\left(\bar{f}_\sigma\right)\right\|_{\mathcal{H}_\sigma}^2\right|\right\|
\end{aligned}$$

the second linek goes to zero by standard KDE consistency, now to the first line, we will remove the 2 factor for convience.

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left|R_{\sigma,n}^m\left(\bar{f}_\sigma\right)(X_i) - R_\sigma^m\left(\bar{f}_\sigma\right)(X_i)\right|\right\| \\
&= \left\|\frac{1}{n}\sum_{i=1}^n \Phi_\sigma(X_i)\left|\left\langle R_{\sigma,n}^m\left(\bar{f}_\sigma\right) - R_\sigma^m\left(\bar{f}_\sigma\right), \Phi_\sigma(X_i)\right\rangle\right|\right\| \\
&\leq \max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^n \delta_i\Phi_\sigma(X_i)\left\langle R_{\sigma,n}^m\left(\bar{f}_\sigma\right) - R_\sigma^m\left(\bar{f}_\sigma\right), \Phi_\sigma(X_i)\right\rangle\right\| \\
&\leq \max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^m \delta_i\Phi_\sigma\left(X_i\right)\left\langle\Phi_\sigma(X_i),\cdot\right\rangle\right\|_{op}\left\|R_{\sigma,n}^m\left(\bar{f}_\sigma\right) - R_\sigma^m\left(\bar{f}_\sigma\right)\right\|.
\end{aligned}$$

The RHS of the last line goes to zero by the inductive hypothesis, now we need only show that the LHS remains bounded whp. Returning to our proof we have

$$\max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^{m}\delta_i\Phi_\sigma\left(X_i\right)\langle\Phi_\sigma(X_i),\cdot\rangle\right\|_{op}\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)-R^m_\sigma\left(\bar{f}_\sigma\right)\right\|$$

where the LHS goes to 1 and the RHS goes to zero thus completing the convergence of the numerator. Lets now take care of the denominator

$$\begin{aligned}&\left|D_{\sigma,n}\left(R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right)-D_\sigma\left(R^m_\sigma\left(\bar{f}_\sigma\right)\right)\right|\\ \leq&\left|D_{\sigma,n}\left(R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right)-D_{\sigma,n}\left(R^m_\sigma\left(\bar{f}_\sigma\right)\right)\right|+\left|D_{\sigma,n}\left(R^m_\sigma\left(\bar{f}_\sigma\right)\right)-D_\sigma\left(R^m_\sigma\left(\bar{f}_\sigma\right)\right)\right|\end{aligned}$$

Lets take care of the first summand. From Proposition 3 we have

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\left(2R^m_{\sigma,n}\left(\bar{f}_\sigma\right)(X_i)-\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2\right)^+-\frac{1}{n}\sum_{j=1}^{n}\left(2R^m_\sigma\left(\bar{f}_\sigma\right)(X_j)-\left\|R^m_\sigma\left(\bar{f}_\sigma\right)\right\|^2\right)^+\right|\\ &\qquad\leq\left|\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2-\left\|R^m_\sigma\left(\bar{f}_\sigma\right)\right\|^2\right|+2\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)-R^m_\sigma\left(\bar{f}_\sigma\right)\right\|\left\|\bar{f}_{\sigma,n}\right\|.\end{aligned}$$

The left summand goes to zero by the inductive hypotheseis and the RHS goes to zero by the inductive hypothesis and KDE consistency. Note we now have that $D_{\sigma,n}\left(R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right)\xrightarrow{p}D\left(R^m\left(f\right)\right)$. Using this fact with the convergence of the numerator and the denominator completes the lemma. □

For the following lemma we define

$$\begin{aligned}N_{\sigma,n}\left(R^{-1}_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right)&\triangleq\bar{f}_{\sigma,n}\\ D_{\sigma,n}\left(R^{-1}_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right)&\triangleq 1\end{aligned}$$

Lemma 5. *For all $m\geq 0$,we have*

$$\begin{aligned}\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|&\xrightarrow{p}0\\ \left\|N_{\sigma,n}\left(R^{m-1}_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right)-N_{\sigma,n}\left(R^{m-1}_{\sigma,n}\left(\bar{f}_\sigma\right)\right)\right\|&\xrightarrow{p}0\\ \left|D_{\sigma,n}\left(R^{m-1}_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right)-D_{\sigma,n}\left(R^{m-1}_{\sigma,n}\left(\bar{f}_\sigma\right)\right)\right|&\xrightarrow{p}0\end{aligned}$$

Proof. The base case is trivial (NEED TO DO ANYWAYS). Suppose the lemma holds for m, we will show it holds for $m+1$. We will show the numerator and denominator converge separately. First the numerator.

$$\begin{aligned}&\left\|N_{\sigma,n}\left(R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right)-N_{\sigma,n}\left(R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right)\right\|\\ =&\left\|\frac{1}{n}\sum_{i=1}^{m}\Phi_\sigma(X_i)\left[\left(2R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)(X_i)-\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right\|^2_{\mathcal{H}_\sigma}\right)^+-\left(2R^m_{\sigma,n}\left(\bar{f}_\sigma\right)(X_i)-\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2_{\mathcal{H}_\sigma}\right)^+\right]\right\|\\ \leq&\left\|\frac{1}{n}\sum_{i=1}^{m}\Phi_\sigma(X_i)\left|2R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)(X_i)-\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right\|^2_{\mathcal{H}_\sigma}-2R^m_{\sigma,n}\left(\bar{f}_\sigma\right)(X_i)+\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2_{\mathcal{H}_\sigma}\right|\right\|\\ \leq&\left\|\frac{1}{n}\sum_{i=1}^{m}\Phi_\sigma(X_i)\left|2R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)(X_i)-2R^m_{\sigma,n}\left(\bar{f}_\sigma\right)(X_i)\right|\right\|\\ &+\left\|\frac{1}{n}\sum_{i=1}^{m}\Phi_\sigma(X_i)\left|\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right\|^2_{\mathcal{H}_\sigma}-\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2_{\mathcal{H}_\sigma}\right|\right\|\end{aligned}$$

The second summand goes to zero by the inductive hypothesis and standard KDE consistency. Going to the first summand and removing a 2 factor for convencience

$$
\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^{m}\Phi_\sigma(X_i)\left|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)(X_i)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right)(X_i)\right|\right\| \\
&=\left\|\frac{1}{n}\sum_{i=1}^{m}\Phi_\sigma(X_i)\left|\left\langle R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right),\Phi_\sigma(X_i)\right\rangle\right|\right\| \\
&=\max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^{m}\delta_i\Phi_\sigma(X_i)\left[\left\langle R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right),\Phi_\sigma(X_i)\right\rangle\right]\right\| \\
&=\max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^{m}\delta_i\Phi_\sigma(X_i)\left\langle R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right),\Phi_\sigma(X_i)\right\rangle\right\| \\
&\leq\max_{\delta\in\{-1,1\}^n}\left\|\frac{1}{n}\sum_{i=1}^{m}\delta_i\Phi_\sigma(X_i)\left\langle\cdot,\Phi_\sigma(X_i)\right\rangle\right\|\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|
\end{aligned}
$$

which goes to zero by uniform convergence when $n\sigma^{2d}\to 0$.

Now to take care of the denominator,

$$
\begin{aligned}
&\left|D_{\sigma,n}\left(R^{m-1}_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right)-D_{\sigma,n}\left(R^{m-1}_{\sigma,n}\left(\bar{f}_\sigma\right)\right)\right| \\
=&\left|\frac{1}{n}\sum_{i=1}^{m}\left[\left(2R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)(X_i)-\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right\|^2_{\mathcal{H}_\sigma}\right)^+-\left(2R^m_{\sigma,n}\left(\bar{f}_\sigma\right)(X_i)-\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2_{\mathcal{H}_\sigma}\right)^+\right]\right| \\
&\left|\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)\right\|^2_{\mathcal{H}_\sigma}-\left\|R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|^2_{\mathcal{H}_\sigma}\right|+2\left\|R^m_{\sigma,n}\left(\bar{f}_{\sigma,n}\right)-R^m_{\sigma,n}\left(\bar{f}_\sigma\right)\right\|\left\|\bar{f}_{\sigma,n}\right\|
\end{aligned}
$$

which goes to zero in probability. □

4 Auxiliary Results

Lemma 6. *Let $g\in L^2\left(\mathbb{R}^d\right)$, then $\left\|\int\Phi_\sigma(x)g(x)dx\right\|_{\mathcal{H}_\sigma}\leq\|g\|_{L^2}$.*

Proof of Lemma 6. Using the Cauchy-Schwarz Inequality and Young's Inequality we have

$$
\begin{aligned}
\left\|\int\Phi_\sigma(x)g(x)dx\right\|^2_{\mathcal{H}_\sigma} &= \left\langle\int\Phi_\sigma(x)g(x)dx,\int\Phi_\sigma(y)g(y)dy\right\rangle \\
&= \int g(x)g(y)\left\langle\Phi_\sigma(x),\int\Phi_\sigma(y)dy\right\rangle dx \\
&= \int g(x)\left\langle\Phi_\sigma(x),\int\Phi_\sigma(y)g(y)dy\right\rangle dx \\
&= \int g(x)(k_\sigma * g)(x)dx \\
&= \langle g, g*k_\sigma\rangle_{L^2} \\
&\leq \|g\|_{L^2}\|g*k_\sigma\|_{L^2} \\
&\leq \|g\|_{L^2}\|g\|_{L^2}\|k_\sigma\|_{L^1} \\
&= \|g\|^2_{L^2}.
\end{aligned}
$$

□

Lemma 7. *Let $\alpha_1,\ldots,\alpha_m\in\mathbb{R}$ and $x_1,\ldots,x_m\in\mathbb{R}^d$, then*

$$\left\|\sum_{i=1}^{m}\alpha_i\Phi_\sigma\left(x_i\right)\right\|\leq\left\|\sum_{i=1}^{m}|\alpha_i|\,\Phi_\sigma\left(x_i\right)\right\|,$$

in both $\mathcal{H}_\sigma$ and L^2 norms.

Proof of Lemma 7. We have

$$\begin{aligned}\left\|\sum_{i=1}^{m}\alpha_i\Phi_\sigma\left(x_i\right)\right\|^2 &= \sum_{i=1}^{m}\sum_{j=1^m}\alpha_i\alpha_j\left\langle\Phi_\sigma(x_i),\Phi_\sigma(x_j)\right\rangle \\ &\leq \sum_{i=1}^{m}\sum_{j=1^m}\left|\alpha_i\right|\left|\alpha_j\right|\left\langle\Phi_\sigma(x_i),\Phi_\sigma(x_j)\right\rangle \\ &= \left\|\sum_{i=1}^{m}\left|\alpha_i\right|\Phi_\sigma\left(x_i\right)\right\|^2.\end{aligned}$$

$\square$

Corollary 1. *Let $f \in L^1\left(\mathbb{R}^d\right)$, then*

$$\left\|\int_{\mathbb{R}^d} f(x)\Phi_\sigma(x)dx\right\| \leq \left\|\int_{\mathbb{R}^d}\left|f(x)\right|\Phi_\sigma(x)dx\right\|,$$

in both $\mathcal{H}_\sigma$ and L^2 norms.

Proof of Corollary 1.

$$\begin{aligned}\left\|\int f(x)\Phi_\sigma(x)dx\right\|^2 &= \left\langle\int f(x)\Phi_\sigma\left(x\right)dx,\int f(y)\Phi_\sigma\left(y\right)dy\right\rangle \\ &= \int\int f(x)f(y)\left\langle\Phi_\sigma\left(x\right),\Phi_\sigma\left(y\right)\right\rangle dydx \\ &\leq \int\int\left|f(x)\right|\left|f(y)\right|\left\langle\Phi_\sigma\left(x\right),\Phi_\sigma\left(y\right)\right\rangle dydx \\ &= \left\langle\int\left|f(x)\right|\Phi_\sigma\left(x\right)dx,\int\left|f(y)\right|\Phi_\sigma\left(y\right)dy\right\rangle \\ &= \left\|\int\left|f(x)\right|\Phi_\sigma(x)dx\right\|^2\end{aligned}$$

$\square$

Lemma 8. *Let $\alpha_1,\ldots,\alpha_m \geq 0$, $\beta_1,\ldots,\beta_m \in \mathbb{R}$ such that $\beta_i \geq \alpha_i$ for all i and $x_1,\ldots,x_m \in \mathbb{R}^d$, then*

$$\left\|\sum_{i=1}^{m}\alpha_i\Phi_\sigma\left(x_i\right)\right\| \leq \left\|\sum_{i=1}^{m}\beta_i\Phi_\sigma\left(x_i\right)\right\|,$$

in both $\mathcal{H}_\sigma$ and L^2 norms.

Proof of Lemma 8. This proof is similar to the proof of Lemma 7, so we will omit it. $\square$

Corollary 2. *Let $0 \leq f \leq g$. Then*

$$\left\|\int f(x)\Phi_\sigma(x)dx\right\| \leq \left\|\int g(x)\Phi_\sigma(x)dx\right\|,$$

in both $\mathcal{H}_\sigma$ and L^2 norms.

Proof of Corollary 2.

$$\begin{aligned}
\left\|\int f(x)\Phi_\sigma(x)dx\right\|^2 &= \left\langle \int f(x)\Phi_\sigma(x)\,dx, \int f(y)\Phi_\sigma(y)\,dy \right\rangle \\
&= \int\int f(x)f(y)\left\langle \Phi_\sigma(x), \Phi_\sigma(y)\right\rangle dydx \\
&\leq \int\int g(x)g(y)\left\langle \Phi_\sigma(x), \Phi_\sigma(y)\right\rangle dydx \\
&= \left\langle \int g(x)\Phi_\sigma(x)\,dx, \int g(y)\Phi_\sigma(y)\,dy \right\rangle \\
&= \left\|\int g(x)\Phi_\sigma(x)dx\right\|^2
\end{aligned}$$

□

Lemma 9. *Let* $T_{\sigma,n} = \sum_{i=1}^m \Phi_\sigma(X_i)\left\langle \Phi_\sigma(X_i), \cdot\right\rangle$ *and* $T_{\sigma,n} = \int \Phi_\sigma(x)\left\langle \Phi_\sigma(x), \cdot\right\rangle f(x)dx$. *We have that* $\|T_{\sigma,n} - T_{\sigma,n}\| \xrightarrow{p} 0$.

The following is a classic result in functional analysis, but it is worth mentioning expliticly (Folland ?).

Proposition 5 (Young's Inequality)**.** *Let* $1 \leq p,q,r \leq \infty$ *with* $p^{-1} + q^{-1} = r^{-1} + 1$. *Let* $g \in L^p\left(\mathbb{R}^d\right)$ *and* $h \in L^q\left(\mathbb{R}^d\right)$. *We have*

$$\|g * h\|_{L^r} \leq \|g\|_{L^p}\|h\|_{L^q}.$$

We will find the case where $r = p = 2$ and $q = 1$ to be useful in particular. (MAYBE the infinity version too…)

Proposition 6 (Folland)**.** *CONVOLUTION CONVERG*

Lemma 10. *Let* $g \in L^2\left(\mathbb{R}^d\right)$ *then* $\left\|\int \Phi_\sigma(x)g(x)dx\right\|_{\mathcal{H}_\sigma} \to \|g\|_{L^2}$ *as* $\sigma \to 0$.

Proof of Lemma 10. By direct evaluation we have

$$\begin{aligned}
\left\|\int \Phi_\sigma(x)g(x)dx\right\|^2_{\mathcal{H}_\sigma} &= \left\langle \int \Phi_\sigma(x)g(x)dx, \int \Phi_\sigma(y)g(y)dy \right\rangle_{\mathcal{H}_\sigma} \\
&= \int\int g(x)g(y)\left\langle \Phi_\sigma(x), \Phi_\sigma(y)\right\rangle_{\mathcal{H}_\sigma} dxdy \\
&= \int\int g(x)g(y)k_\sigma(x,y)dxdy \\
&= \int g(y)g * k_\sigma(y)dy \\
&= \left\langle g, g * k_\sigma\right\rangle_{L^2}
\end{aligned}$$

which we know goes to $\|g\|^2_{L^2}$ by Proposition 6. □

We need a couple other technical results before moving forward.

Lemma 11. *Let* $g \in L^2\left(\mathbb{R}^d\right)$. *Then* $\int g(x)\Phi_\sigma(x)dx = g * k_\sigma$. *DO WE NEED L1?*

Proof of Lemma 11. Let $y \in \mathbb{R}^d$ be arbitrary. By direct evaluation we get

$$\begin{aligned}
\left(\int g(x)\Phi_\sigma(x)dx\right)(y) &= \left\langle \int g(x)\Phi_\sigma(x)dx, \Phi_\sigma(y) \right\rangle_{\mathcal{H}_\sigma} \\
&= \int g(x)\left\langle \Phi_\sigma(x), \Phi_\sigma(y)\right\rangle_{\mathcal{H}_\sigma} dx \\
&= \int g(x)k_\sigma(x,y)\,dx \\
&= g * k_\sigma(y).
\end{aligned}$$

$\square$

Lemma 12. *Let f be a pdf, $\varepsilon > 0$, and $y \in \mathbb{R}^d$. There exists $r > 0$ such that*

$$\int_{B(y,r)} f(x)dx \geq 1 - \varepsilon.$$

or equivalently

$$\int_{B(y,r)^C} f(x)dx < \varepsilon.$$

Proof. We will prove the second statement. Consider the following, where $i \in \mathbb{N}$,

$$\int_{B(y,i)^C} f(x)\, dx \quad = \int \chi_{B(y,i)^C}(x)\, f(x)\, dx.$$

Clearly as $i \to \infty$, $\chi_{B(y,i)^C} f \to 0$ pointwise. Since $\chi_{B(y,i)^C} f$ is dominated by f, $\int \chi_{B(y,i)^C}(x)\, f(x)\, dx \to \int 0dx = 0$ by the dominated convergence theorem. Thus there exists $n \in \mathbb{N}$ where $\int_{B(y,n)^C} f(x)\, dx < \varepsilon$. $\square$

Lemma 13. *Let $S \in \mathbb{R}^d$ be a set with finite Lebesgue measure and $g \in \mathcal{H}_\sigma$. Then*

$$\int_S |g(x)|\, dx \leq 2\sqrt{\lambda(S)}\, \|g\|_{\mathcal{H}_\sigma}.$$

Proof. **of Lemma 13** Let $S^+ = \{s | s \in S, g(s) \geq 0\}$ and $S^- = S \setminus S^+$. We have

$$\begin{aligned}
\int_S |g(x)|\, dx \quad &= \textstyle\int_{S^+} g(x)dx + \int_{S^-} -g(x')dx' \\
&= \textstyle\int_{S^+} \langle g, \Phi_\sigma(x)\rangle_{\mathcal{H}_\sigma} dx + \int_{S^-} \langle -g, \Phi_\sigma(x')\rangle_{\mathcal{H}_\sigma} dx' \\
&= \textstyle\left\langle g, \int_{S^+} \Phi_\sigma(x)dx\right\rangle_{\mathcal{H}_\sigma} + \left\langle -g, \int_{S^-} \Phi_\sigma(x')dx'\right\rangle_{\mathcal{H}_\sigma} \\
&\leq \|g\|_{\mathcal{H}_\sigma} \left(\left\|\textstyle\int_{S^+} \Phi_\sigma(x)dx\right\|_{\mathcal{H}_\sigma} + \left\|\textstyle\int_{S^-} \Phi_\sigma(x')dx'\right\|_{\mathcal{H}_\sigma} \right).
\end{aligned} \tag{4}$$

Now consider

$$\begin{aligned}
\left\| \int_{S^+} \Phi_\sigma(x)dx \right\|_{\mathcal{H}_\sigma}^2 \quad &= \textstyle\left\langle \int_{S^+} \Phi_\sigma(x)dx, \int_{S^+} \Phi_\sigma(x')dx' \right\rangle_{\mathcal{H}_\sigma} \\
&= \textstyle\int_{S^+} \int_{S^+} \langle \Phi_\sigma(x), \Phi_\sigma(x')\rangle_{\mathcal{H}_\sigma} dxdx' \\
&= \textstyle\int_{S^+} \int_{S^+} k_\sigma(x, x')\, dxdx' \\
&\leq \textstyle\int_{S^+} 1dx' \\
&= \lambda(S^+)
\end{aligned}$$

and a similar result can be shown for S^-. Plugging back into (4) we get

$$\begin{aligned}
\int_S |g(x)|\, dx \quad &\leq \|g\|_{\mathcal{H}_\sigma} \left(\sqrt{\lambda(S^+)} + \sqrt{\lambda(S^-)} \right) \\
&\leq \|g\|_{\mathcal{H}_\sigma} 2\sqrt{\lambda(S)}.
\end{aligned}$$

$\square$

Lemma 14. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a pdf and g_σ^n and h_σ^n be sequences of (possibly random) densities in a sequence of spaces $\mathcal{D}_\sigma$ (again σ is implicitly a function of n). If $\|g_\sigma^n - f\|_1 \xrightarrow{p} 0$ and $\|g_\sigma^n - h_\sigma^n\|_{\mathcal{H}_\sigma} \xrightarrow{p} 0$ then $\|g_\sigma^n - h_\sigma^n\|_1 \xrightarrow{p} 0$.*

Proof. **of Lemma 14** Let $\varepsilon > 0$; by Lemma 12 let $r > 0$ such that $\left\|f\chi_{B(0,r)^C}\right\|_1 < \varepsilon/3$. From Lemma 13 we have

$$\left\|(g_\sigma^n - h_\sigma^n)\chi_{B(0,r)}\right\|_1 \xrightarrow{p} 0.$$

Since $\|g_\sigma^n - f\|_1 \xrightarrow{p} 0$, we have $\left\|g_\sigma^n\chi_{B(0,r)}\right\|_1 \xrightarrow{p} \left\|f\chi_{B(0,r)}\right\|_1$, and therefore

$$\begin{aligned}
\left|\left\|h_\sigma^n\chi_{B(0,r)^C}\right\|_1 - \left\|f\chi_{B(0,r)^C}\right\|_1\right| &= \left|\left(1 - \left\|h_\sigma^n\chi_{B(0,r)}\right\|_1\right) - \left(1 - \left\|f\chi_{B(0,r)}\right\|_1\right)\right| \\
&= \left|\left\|h_\sigma^n\chi_{B(0,r)}\right\|_1 - \left\|f\chi_{B(0,r)}\right\|_1\right| \\
&\leq \left\|\left(h_\sigma^n - f\right)\chi_{B(0,r)}\right\|_1 \\
&\leq \left\|\left(h_\sigma^n - g_\sigma^n\right)\chi_{B(0,r)}\right\|_1 + \left\|\left(g_\sigma^n - f\right)\chi_{B(0,r)}\right\|_1 \\
&\xrightarrow{p} 0.
\end{aligned}$$

Thus, $\left\|h_\sigma^n\chi_{B(0,r)^C}\right\|_1 \xrightarrow{p} \left\|f\chi_{B(0,r)^C}\right\|_1$. Since $\left\|f\chi_{B(0,r)^C}\right\|_1 < \varepsilon/3$, we have

$$\mathbb{P}\left(\left\|h_\sigma^n\chi_{B(0,r)^C}\right\|_1 \geq \varepsilon 5/12\right) \to 0. \tag{5}$$

Now to finish the proof,

$$\begin{aligned}
\mathbb{P}\left(\|h_\sigma^n - g_\sigma^n\|_1 > \varepsilon\right) = {} & \mathbb{P}\left(\left\|(h_\sigma^n - g_\sigma^n)\chi_{B(0,r)}\right\|_1 + \left\|\left(h_\sigma^n - g_\sigma^n\right)\chi_{B(0,r)^C}\right\|_1 > \varepsilon\right) \\
\leq {} & \mathbb{P}\left(\left\|\left(h_\sigma^n - g_\sigma^n\right)\chi_{B(0,r)}\right\|_1 \geq \varepsilon/4\right) + \mathbb{P}\left(\left\|\left(h_\sigma^n - g_\sigma^n\right)\chi_{B(0,r)^C}\right\|_1 > 3\varepsilon/4\right)
\end{aligned}$$

We've already shown the left summand goes to zero, now we take care of the right term

$$\begin{aligned}
\mathbb{P}\left(\left\|\left(h_\sigma^n - g_\sigma^n\right)\chi_{B(0,r)^C}\right\|_1 > 3\varepsilon/4\right) &\leq \mathbb{P}\left(\left\|h_\sigma^n\chi_{B(0,r)^C}\right\|_1 + \left\|g_\sigma^n\chi_{B(0,r)^C}\right\|_1 > 3\varepsilon/4\right) \\
&\leq \mathbb{P}\left(\left\|h_\sigma^n\chi_{B(0,r)^C}\right\|_1 \geq 5\varepsilon/12\right) + \mathbb{P}\left(\left\|g_\sigma^n\chi_{B(0,r)^C}\right\|_1 > \varepsilon/3\right)
\end{aligned}$$

The left summand goes to zero by (5). Since $\left\|g_\sigma^n\chi_{B(0,r)^C} - f\chi_{B(0,r)^C}\right\|_1 \to 0$ and $\left\|f\chi_{B(0,r)^C}\right\|_1 < \frac{\varepsilon}{3}$, with probability going to one, we have $\left\|g_\sigma^n\chi_{B(0,r)^C}\right\|_1 \leq \varepsilon/3$ and the right summand goes to zero. This completes our proof. □

5 Basic Lemmas

Lemma 15. *If $n \to \infty$, $\sigma \to 0$ with $n\sigma^d \to \infty$ then*

$$\left\|\bar{f}_{\sigma,n} - \bar{f}_\sigma\right\|_{\mathcal{H}_\sigma} \xrightarrow{p} 0.$$

Lemma 16. *If $n \to \infty$, $\sigma \to 0$ with $n\sigma^d \to \infty$ then*

$$\left\|\bar{f}_{\sigma,n} - f\right\|_2 \xrightarrow{p} 0$$

A Proofs

Lemma 17. *$R^m(f)$ exists for all $m \geq 0$.*

Proof of Lemma 17. Clearly $\operatorname{supp}\left(R^{m+1}(f)\right) \subseteq \operatorname{supp}\left(R^m(f)\right)$. From Hölder's Inequality we have that

$$\|f\|_{L^2(\mathbb{R}^d)}^2 = \|ff\|_{L^1(\mathbb{R}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)}\|f\|_{L^\infty(\mathbb{R}^d)} = \|f\|_{L^\infty(\mathbb{R}^d)}$$

□

Lemma 18. *Let $g \in \mathcal{D}_{\sigma,n}$. Then $R_{\sigma,n}(g)$ is well defined.*

Proof of Lemma 18. To show that this is well defined we only need to show that $\left(2g(X_i) - \|g\|^2_{\mathcal{H}_\sigma}\right)^+ > 0$ for some i. Let $g = \sum_{i=1}^n w_i \Phi_\sigma(X_i)$. Let q be such that $\sup_i g(X_i) = g(X_q)$. We have

$$\begin{aligned} g(X_q) &= \sum_{i=1}^n w_i g(X_q) \\ &\geq \sum_{i=1}^n w_i g(X_i) \\ &= \sum_{i=1}^n w_i \left\langle \sum_{j=1}^n w_j \Phi_\sigma(X_j), \Phi_\sigma(X_i) \right\rangle_{\mathcal{H}_\sigma} \\ &= \left\langle \sum_{j=1}^n w_j \Phi_\sigma(X_j), \sum_{i=1}^n w_i \Phi_\sigma(X_i) \right\rangle_{\mathcal{H}_\sigma} \\ &= \|g\|^2_{\mathcal{H}_\sigma}. \end{aligned}$$

We know $g(X_q) > 0$ so $2g(X_i) - \|g\|^2_{\mathcal{H}_\sigma} > 0$. □

References

[1] M.J. Fabian. *Functional Analysis and Infinite-Dimensional Geometry*. CMS Books in Mathematics. Springer, 2001.