A Basic lemmas and claims

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

Claim 13 (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*

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

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

It is not hard to see that the Fourier transform of a δ_d is the constant function which is 1 everywhere:

$$(\mathcal{F}\delta_d)(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} e^{-2\pi i \mathbf{t}^\top \boldsymbol{\xi}} \cdot \delta_d(\mathbf{t})\, d\mathbf{t} = e^{-2\pi i \cdot 0^\top \cdot \boldsymbol{\xi}} = 1$$

for all $\boldsymbol{\xi}$. Similarly, the Fourier transform of a shifted delta function is as follows:

$$\begin{aligned}(\mathcal{F}\delta(\cdot - \mathbf{a}))(\boldsymbol{\xi}) &= \int_{\mathbb{R}^d} e^{-2\pi i \mathbf{t}^\top \boldsymbol{\xi}} \cdot \delta_d(\mathbf{t} - \mathbf{a})\, d\mathbf{t} \\ &= e^{-2\pi i \mathbf{a}^\top \boldsymbol{\xi}}.\end{aligned}$$

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

Claim 14. *Given a function $f : \mathbb{R}^d \to \mathbb{C}$, we have*

$$\begin{aligned}(\mathcal{F}f(\cdot - \mathbf{a}))(\boldsymbol{\xi}) &= (\mathcal{F}(f * \delta_d(\cdot - \mathbf{a})))(\boldsymbol{\xi}) \\ &= \hat{f}(\boldsymbol{\xi}) \cdot e^{-2\pi i \mathbf{a}^\top \boldsymbol{\xi}}.\end{aligned}$$

Similarly,

Claim 15. *Given a function $f : \mathbb{R}^d \to \mathbb{C}$, we have*

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

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

Definition 16. For any $\mathbf{T} \in \mathbb{R}^d_+$ the d-dimensional *Dirac comb function* with period $\mathbf{T}$ is defined as f satisfying

$$f(\mathbf{x}) = \sum_{\mathbf{j} \in \mathbb{Z}^d} \delta(\mathbf{x} - \mathbf{j}\mathbf{T}),$$

where $\mathbf{j}\mathbf{T} = (j_1 T_1, j_2 T_2, \cdots j_d T_d)^\top$.

Claim 17. *For any function $g : \mathbb{R}^d \to \mathbb{R}$, and any $\mathbf{w} \in \mathbb{R}^d_+$ the following holds,*

$$\mathcal{F}\left[g(\frac{\cdot}{\mathbf{w}})\right](\boldsymbol{\xi}) = \left(\prod_{l=1}^d w_l\right) \widehat{g}(\mathbf{w}\boldsymbol{\xi}).$$

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 18. *Let*

$$f(\mathbf{x}) = \sum_{\mathbf{j} \in \mathbb{Z}^d} \delta(\mathbf{x} - \mathbf{j}\mathbf{T})$$

be the d-dimensional Dirac comb function with period $\mathbf{T}$. Then,

$$(\mathcal{F}f)(\boldsymbol{\xi}) = \prod_{l=1}^d \frac{1}{T_l} \cdot \sum_{\mathbf{j} \in \mathbb{Z}^d} \delta\left(\boldsymbol{\xi} - \frac{\mathbf{j}}{\mathbf{T}}\right),$$

where $\frac{\mathbf{j}}{\mathbf{T}} = (j_1/T_1, j_2/T_2, \ldots, j_d/T_d)^\top$.

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

Claim 19 (Nyquist-Shannon)**.** *Given a function $f : \mathbb{R}^d \to \mathbb{C}$, we have:*

$$\begin{aligned}&\mathcal{F}\left(f(\cdot) \sum_{\mathbf{j} \in \mathbb{Z}^d} \delta_d(\cdot - \mathbf{w} \cdot \mathbf{j})\right)(\boldsymbol{\xi}) \\ &\quad = \left(\prod_{i=1}^d w_i^{-1}\right) \sum_{\mathbf{j} \in \mathbb{Z}^d} \mathcal{F}(f)(\boldsymbol{\xi} - \mathbf{j}/\mathbf{w}).\end{aligned}$$

B Omitted claims and proofs from Section 3.1

We use the following basic claim about Fourier transform of Nyquist-Shannon sampling of functions.

Claim 20. *For any $\mathbf{w} \in \mathbb{R}^d_+$, every sequences $\{c_\mathbf{j}\}_{\mathbf{j} \in \mathbb{Z}^d}$ and $\{b_\mathbf{j}\}_{\mathbf{j} \in \mathbb{Z}^d}$ such that $\sum_{\mathbf{j} \in \mathbb{Z}^d} |c_\mathbf{j}|^2 < \infty$ and $\sum_{\mathbf{j} \in \mathbb{Z}^d} |b_\mathbf{j}|^2 < \infty$, if $g(\cdot) = \sum_{\mathbf{j} \in \mathbb{Z}^d} c_\mathbf{j} \delta(\cdot - \mathbf{j}\mathbf{w})$ and $h(\cdot) = \sum_{\mathbf{j} \in \mathbb{Z}^d} b_\mathbf{j} \delta(\cdot - \mathbf{j}\mathbf{w})$, then the following conditions hold.*

(1) $\widehat{g}(\boldsymbol{\xi}) = \sum_{\mathbf{j} \in \mathbb{Z}^d} c_\mathbf{j} \exp(-2\pi i \boldsymbol{\xi}^\top (\mathbf{j}\mathbf{w}))$ *for every $\boldsymbol{\xi} \in \mathbb{R}^d$.*

(2) $c_\mathbf{j} = \left(\prod_{l=1}^d w_l\right) \int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi}) \exp(2\pi i (\mathbf{j}\mathbf{w})^\top \boldsymbol{\xi})\, d\boldsymbol{\xi}$ *for every $\mathbf{j} \in \mathbb{Z}^d$.*

(3) $\sum_{\mathbf{j} \in \mathbb{Z}^d} c_\mathbf{j}^* b_\mathbf{j} = \left(\prod_{l=1}^d w_l\right) \int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi})^* \widehat{h}(\boldsymbol{\xi})\, d\boldsymbol{\xi}$.

Proof. We have

$$\begin{aligned}\widehat{g}(\boldsymbol{\xi}) &= \int_{\mathbb{R}^d} \left(\sum_{\mathbf{j} \in \mathbb{Z}^d} c_\mathbf{j} \delta(\mathbf{x} - \mathbf{j}\mathbf{w})\right) \exp(-2\pi i \boldsymbol{\xi}^\top \mathbf{x})\, d\mathbf{x} \\ &= \sum_{\mathbf{j} \in \mathbb{Z}^d} c_\mathbf{j} \int_{\mathbb{R}^d} \delta(\mathbf{x} - \mathbf{j}\mathbf{w}) \exp(-2\pi i \boldsymbol{\xi}^\top \mathbf{x})\, d\mathbf{x} \\ &= \sum_{\mathbf{j} \in \mathbb{Z}^d} c_\mathbf{j} \exp(-2\pi i \boldsymbol{\xi}^\top (\mathbf{w}\mathbf{j})),\end{aligned}$$

which gives the first claim. The second claim can be verified directly:

$$\begin{aligned}
&\int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi}) \exp(2\pi i(\mathbf{wj})^\top \boldsymbol{\xi})\, d\xi \\
&= \int_{[0,1/\mathbf{w}]} \sum_{\mathbf{l}\in\mathbb{Z}^d} c_{\mathbf{l}} e^{2\pi i(\mathbf{wj})^\top \boldsymbol{\xi} - 2\pi i \boldsymbol{\xi}^\top(\mathbf{wl})}\, d\boldsymbol{\xi} \\
&= \sum_{\mathbf{l}\in\mathbb{Z}^d} c_{\mathbf{l}} \int_{[0,1/w]^d} e^{2\pi i(\mathbf{wj})^\top \boldsymbol{\xi} - 2\pi i \boldsymbol{\xi}^\top(\mathbf{wl})}\, d\boldsymbol{\xi} \\
&= \left(\prod_{i=1}^d \frac{1}{w_i}\right) c_{\mathbf{j}},
\end{aligned}$$

proving the second claim.

For the third claim we have

$$\begin{aligned}
&\int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi})^* \widehat{h}(\boldsymbol{\xi})\, d\xi \\
&= \int_{[0,1/\mathbf{w}]} \sum_{\mathbf{j}\in\mathbb{Z}^d} \sum_{\mathbf{l}\in\mathbb{Z}^d} c_{\mathbf{j}}^* b_{\mathbf{l}} \exp(-2\pi i \boldsymbol{\xi}^\top(\mathbf{w}(\mathbf{l}-\mathbf{j})))\, d\boldsymbol{\xi} \\
&= \sum_{\mathbf{j}\in\mathbb{Z}^d} \sum_{\mathbf{l}\in\mathbb{Z}^d} c_{\mathbf{j}}^* b_{\mathbf{l}} \int_{[0,1/\mathbf{w}]} \exp(-2\pi i \boldsymbol{\xi}^\top(\mathbf{w}(\mathbf{l}-\mathbf{j})))\, d\boldsymbol{\xi} \\
&= \left(\prod_{i=1}^d \frac{1}{w_i}\right) \sum_{\mathbf{j}\in\mathbb{Z}^d} c_{\mathbf{j}}^* b_{\mathbf{j}},
\end{aligned}$$

as required. □

The properties of the WLSH estimator are best understood using the means of Fourier transform. Therefore, we express the WLSH estimator in the Fourier domain. The following lemma expresses the WLSH estimator in the spectral domain.

Lemma 21 (Spectral Representation of WLSH Estimator)**.** *For any* $\mathbf{w} \in \mathbb{R}_+^d$*, any* $\mathbf{z} \in [0, \mathbf{w}]$*, any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$*, if the WLSH estimator* $\widetilde{k}_{f,p}(\mathbf{x}, \mathbf{y})$ *is defined as in* (5)*, then the following holds,*

$$\tilde{k}_{f,p}(\mathbf{x}, \mathbf{y}) = \left(\prod_{l=1}^d w_l\right) \int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi})^* \widehat{h}(\boldsymbol{\xi})\, d\boldsymbol{\xi}, \tag{6}$$

where $\widehat{g}(\boldsymbol{\xi}) = \sum_{\mathbf{j}\in\mathbb{Z}^d} e^{-2\pi i(\mathbf{x}-\mathbf{z})^\top(\boldsymbol{\xi}-\frac{\mathbf{j}}{\mathbf{w}})} \cdot \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}-\mathbf{j})$ *and* $\widehat{h}(\boldsymbol{\xi}) = \sum_{\mathbf{j}\in\mathbb{Z}^d} e^{-2\pi i(\mathbf{y}-\mathbf{z})^\top(\boldsymbol{\xi}-\frac{\mathbf{j}}{\mathbf{w}})} \cdot \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}-\mathbf{j})$.

Proof. By (5) and part **(3)** of Claim 20 we have,

$$\begin{aligned}
\widetilde{k}_{f,p}(\mathbf{x}, \mathbf{y}) &= \sum_{\mathbf{j}\in\mathbb{Z}^d} f^{\otimes d}(\mathbf{j} + \frac{\mathbf{z}-\mathbf{x}}{\mathbf{w}}) \cdot f^{\otimes d}(\mathbf{j} + \frac{\mathbf{z}-\mathbf{y}}{\mathbf{w}}) \\
&= \left(\prod_{l=1}^d w_l\right) \int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi})^* \widehat{h}(\boldsymbol{\xi})\, d\boldsymbol{\xi}, \qquad (7)
\end{aligned}$$

where $g(\mathbf{t}) = (\delta_d(\cdot - \mathbf{x}+\mathbf{z}) * f^{\otimes d}(\frac{\cdot}{\mathbf{w}})) \cdot \sum_{\mathbf{j}\in\mathbb{Z}^d} \delta(\mathbf{t}-\mathbf{jw})$ and $h(\mathbf{t}) = (\delta_d(\cdot - \mathbf{y}+\mathbf{z}) * f^{\otimes d}(\frac{\cdot}{\mathbf{w}})) \cdot \sum_{\mathbf{j}\in\mathbb{Z}^d} \delta(\mathbf{t}-\mathbf{jw})$. We now apply Claim 19 to $g(\cdot)$ and $h(\cdot)$, obtaining the following for every $\boldsymbol{\xi} \in \mathbb{R}^d$,

$$\widehat{g}(\boldsymbol{\xi}) = \prod_{l=1}^d \frac{1}{w_l} \sum_{\mathbf{j}\in\mathbb{Z}^d} \mathcal{F}\left[f^{\otimes d}\left(\frac{\cdot - \mathbf{x}+\mathbf{z}}{\mathbf{w}}\right)\right](\boldsymbol{\xi} - \mathbf{j}/\mathbf{w}). \tag{8}$$

Similarly,

$$\widehat{h}(\boldsymbol{\xi}) = \prod_{l=1}^d \frac{1}{w_l} \sum_{\mathbf{j}\in\mathbb{Z}^d} \mathcal{F}\left[f^{\otimes d}\left(\frac{\cdot - \mathbf{y}+\mathbf{z}}{\mathbf{w}}\right)\right](\boldsymbol{\xi} - \mathbf{j}/\mathbf{w}).$$

Now by Claim 13 and Claim 17,

$$\begin{aligned}
&\mathcal{F}\left[f^{\otimes d}(\frac{\cdot - \mathbf{x}+\mathbf{z}}{\mathbf{w}})\right](\boldsymbol{\xi}) \\
&\qquad = \prod_{l=1}^d w_l \cdot e^{-2\pi i(\mathbf{x}-\mathbf{z})^\top \boldsymbol{\xi}} \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}).
\end{aligned}$$

And similarly $\mathcal{F}[f^{\otimes d}\left(\frac{\cdot - \mathbf{y}+\mathbf{z}}{\mathbf{w}}\right)](\boldsymbol{\xi}) = \prod_{l=1}^d w_l \cdot e^{-2\pi i(\mathbf{y}-\mathbf{z})^\top \boldsymbol{\xi}} \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi})$. Substituting these into (7), we get

$$\widetilde{k}_{\mathbf{w},\mathbf{z}}(\mathbf{x}, \mathbf{y}) = \left(\prod_{l=1}^d w_l\right) \int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi})^* \widehat{h}(\boldsymbol{\xi})\, d\boldsymbol{\xi},$$

where

$$\widehat{g}(\boldsymbol{\xi}) = \sum_{\mathbf{j}\in\mathbb{Z}^d} e^{-2\pi i(\mathbf{x}-\mathbf{z})^\top(\boldsymbol{\xi}-\frac{\mathbf{j}}{\mathbf{w}})} \cdot \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}-\mathbf{j})$$

and

$$\widehat{h}(\boldsymbol{\xi}) = \sum_{\mathbf{j}\in\mathbb{Z}^d} e^{-2\pi i(\mathbf{y}-\mathbf{z})^\top(\boldsymbol{\xi}-\frac{\mathbf{j}}{\mathbf{w}})} \cdot \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}-\mathbf{j}).$$

□

Proof of Claim 7:

We first take the expectation of the WLSH estimator $\widetilde{k}(\mathbf{x}, \mathbf{y})$ with respect to $\mathbf{z} \sim \mathrm{Unif}([0, \mathbf{w}])$. By (6) we have,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{z}\sim\mathrm{Unif}([0,\mathbf{w}])}\left[\widetilde{k}(\mathbf{x}, \mathbf{y})\right] \\
&= \mathbb{E}_{\mathbf{z}}\left[\prod_{l=1}^d w_l \int_{[0,1/\mathbf{w}]} \widehat{g}(\boldsymbol{\xi})^* \widehat{h}(\boldsymbol{\xi})\, d\boldsymbol{\xi}\right] \\
&= \prod_{l=1}^d w_l \int_{[0,1/\mathbf{w}]} \sum_{\mathbf{j},\mathbf{j}'\in\mathbb{Z}^d} \mathbb{E}_{\mathbf{z}}\left[e^{+2\pi i \mathbf{z}^\top(\frac{\mathbf{j}-\mathbf{j}'}{\mathbf{w}})}\right] \\
&\qquad \cdot e^{2\pi i \mathbf{x}^\top(\boldsymbol{\xi}-\frac{\mathbf{j}}{\mathbf{w}})} \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}-\mathbf{j}) \\
&\qquad \cdot e^{-2\pi i \mathbf{y}^\top(\boldsymbol{\xi}-\frac{\mathbf{j}'}{\mathbf{w}})} \cdot \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi}-\mathbf{j})\, d\boldsymbol{\xi}.
\end{aligned}$$

Now if you take the expectation with respect to $\mathbf{z} \sim \mathrm{Unif}([0, \mathbf{w}])$, by orthogonality, the only non-zero terms in the sum will correspond to the case when $\mathbf{j} = \mathbf{j}'$. Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{z}}\left[\widetilde{k}(\mathbf{x}, \mathbf{y})\right] \\
&= \prod_{l=1}^{d} w_l \int_{[0,1/\mathbf{w}]} \sum_{\mathbf{j} \in \mathbb{Z}^d} e^{+2\pi i \mathbf{x}^\top (\boldsymbol{\xi} - \frac{\mathbf{j}}{\mathbf{w}})} \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi} - \mathbf{j}) \\
&\quad \cdot e^{-2\pi i \mathbf{y}^\top (\boldsymbol{\xi} - \frac{\mathbf{j}}{\mathbf{w}})} \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi} - \mathbf{j}) d\boldsymbol{\xi} \\
&= \prod_{l=1}^{d} w_l \int_{\mathbb{R}^d} e^{2\pi i (\mathbf{x}-\mathbf{y})^\top \boldsymbol{\xi}} \left| \widehat{f}^{\otimes d}(\mathbf{w}\boldsymbol{\xi} - \mathbf{j}) \right|^2 d\boldsymbol{\xi}.
\end{aligned}$$

□

Claim 22 (WLSH is Unbiased)**.** *For any PDF $p(\cdot)$ with non-negative support and any even function $f : \mathbb{R} \to \mathbb{R}$ with support $[-1/2, 1/2]$ and $\|f\|_2 = 1$ if $\mathcal{H}$ is the LSH family as per Definition 5 and $k(\cdot)$ is the WLSH kernel as in Definition 8 then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ the following holds for the expectation of WLSH estimator (see Definition 6),*

$$\mathbb{E}_{h_{\mathbf{w},\mathbf{z}} \sim \mathcal{H}}\left[\widetilde{k}(\mathbf{x}, \mathbf{y})\right] = k(\mathbf{x} - \mathbf{y}).$$

Proof. The proof follows from Claim 7 and Definition 8. □

C Omitted claims and proofs from Section 3.2

The following lemma follows from Theorem 5 of Ghosal et al. (2006).

Lemma 23. *For any shift-invariant kernel $k(\cdot)$, which has bounded mixed partial derivatives of up to fourth order, if $\eta : \mathbb{R} \to \mathbb{R}$ is a sample path from the Gaussian process $\mathrm{GP}(0, k(\mathbf{x} - \mathbf{y}))$, then for any $j \in [d]$, the derivative process $D_j \eta(x)$ is a Gaussian process with zero mean, i.e., $\mathbb{E}[D_j \eta(\mathbf{x})] = 0$ for every $\mathbf{x} \in \mathbb{R}^d$, and the covariance $\mathbb{E}[D_j \eta(\mathbf{x}) \cdot D_j \eta(\mathbf{y})] = -D_j^2 k(\mathbf{x} - \mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.*

The above lemma can be applied multiple times and extend to higher order derivative of GP.

Corollary 24. *For any positive integer q, any shift-invariant kernel $k(\cdot)$ which has bounded mixed partial derivatives of order up to $2q + 2$, if $\eta : \mathbb{R} \to \mathbb{R}$ is a sample path from the Gaussian process $\mathrm{GP}(0, k(\mathbf{x} - \mathbf{y}))$, then for any $j \in [d]$ the qth order partial derivative process $D_j^q \eta(\mathbf{x})$ is a Gaussian process with zero mean and covariance $\mathbb{E}[D_j^q \eta(\mathbf{x}) \cdot D_j^q \eta(\mathbf{y})] = (-1)^q D_j^{2q} k(\mathbf{x} - \mathbf{y})$.*

The following lemma gives the derivatives of our WLSH kernel from Definition 8.

Lemma 25. *For any positive integer q, any $i \in [d]$, any function $f : \mathbb{R} \to \mathbb{R}$ with support $[-1/2, 1/2]$ and norm $\|f\|_2 = 1$ which is $\lceil q/2 \rceil$ times differentiable and any probability density function $p(\cdot)$ with non-negative support, if $k(\cdot)$ is the WLSH kernel as in Definition 8, then for any $i \in [d]$ the following holds,*

$$\begin{aligned}
&D_i^q k(\mathbf{x}) \qquad (9) \\
&= \int_{\mathbb{R}_+^d} \frac{p^{\otimes d}(\mathbf{w})}{w_i^q} \left(D_i^{\lceil q/2 \rceil} f^{\otimes d} * D_i^{\lfloor q/2 \rfloor} f^{\otimes d} \right) (\frac{\mathbf{x}}{\mathbf{w}}) d\mathbf{w},
\end{aligned}$$

where $D_i^j f^{\otimes d}(\mathbf{x}) = \prod_{\substack{l \in [d] \\ l \neq i}} f(x_l) \cdot f^{(j)}(x_i)$ for any integer $j \leq \lceil q/2 \rceil$ and $p^{\otimes d}(\mathbf{w}) = \prod_{l=1}^{d} p(w_l)$.

Therefore if the function $f(\cdot)$ is $\lceil q/2 \rceil$ times differentiable then $k(\cdot)$ will be q times partially differentiable with respect to any coordinate. Hence, the LSH-able kernel $k(\cdot)$ inherits certain smoothness properties from the band-limited function $f(\cdot)$.

Now we use the result of Corollary 24 to show that a $\mathrm{GP}(0, k(\mathbf{x} - \mathbf{y}))$ with WLSH covariance kernel $k(\cdot)$ defined as in Definition 8 inherits its smoothness from the band-limited function $f(\cdot)$.

Lemma 26. *For any positive integer q, any even function $f : \mathbb{R} \to \mathbb{R}$ with support $[-1/2, 1/2]$ which has bounded derivatives of order up to $q + 1$, if $\eta : \mathbb{R}^d \to \mathbb{R}$ is a sample path from a $\mathrm{GP}(0, k(\mathbf{x} - \mathbf{y}))$, where $k(\cdot)$ is the WLSH kernel as in Definition 8, then for any $j \in [d]$, $D_j^q \eta(\mathbf{x})$ is a Gaussian process with zero mean and covariance*

$$\begin{aligned}
&\mathbb{E}[D_j^q \eta(x) \cdot D_j^q \eta(y)] \\
&= (-1)^q \int_{\mathbb{R}_+^d} \frac{p^{\otimes d}(\mathbf{w})}{w_j^{2q}} \left(D_j^q f^{\otimes d} * D_j^q f^{\otimes d} \right) \left(\frac{\mathbf{x} - \mathbf{y}}{\mathbf{w}} \right) d\mathbf{w}
\end{aligned}$$

where $D_j^q f^{\otimes d}(\mathbf{x}) = \prod_{\substack{l \in [d] \\ l \neq j}} f(x_l) \cdot f^{(q)}(x_j)$.

Proof. The proof follows from Corollary 24 and Lemma 25. □

One can use Cauchy-Schwarz inequality to bound the covariance of $D^q \eta(\mathbf{x})$ by the following,

$$\left| \mathbb{E}[D_j^q \eta(\mathbf{x}) \cdot D_j^q \eta(\mathbf{y})] \right| \leq \left\| D_j^q f^{\otimes d} \right\|_2^2 \cdot \int_{\mathbb{R}_+} \frac{p(w)}{w^{2q}} dw.$$

In particular, the derivative Gaussian process $D_j^q \eta(\mathbf{x})$ has the following variance, as long as the band-limited filter f is a normalized function ($\|f\|_2 = 1$),

$$\mathbb{E}\left[|D_j^q \eta(\mathbf{x})|^2 \right] = \left\| f^{(q)} \right\|_2^2 \cdot \int_0^\infty \frac{p(w)}{w^{2q}} dw.$$

Now we are ready to prove Lemma 9

Proof of Lemma 9: It follows from multiple application of Lemma 26 that the derivative process $\mathbf{D}\eta(x)$ is a Gaussian process with zero mean and the covariance of $\mathbf{D}\eta(x)$ is the following,

$$\begin{aligned}
&\mathbb{E}[\mathbf{D}\eta(\mathbf{x})\cdot\mathbf{D}\eta(\mathbf{y})]\\
&=(-1)^q\int_{\mathbb{R}_+^d}\prod_{l=1}^d\frac{p(w_l)}{w_l^{2q_l}}\left(\mathbf{D}f^{\otimes d}*\mathbf{D}f^{\otimes d}\right)\left(\frac{\mathbf{x}-\mathbf{y}}{\mathbf{w}}\right)d\mathbf{w}\\
&=(-1)^q\prod_{l=1}^d\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2q_l}}\left(D_l^{q_l}f*D_l^{q_l}f\right)\left(\frac{x_l-y_l}{w_l}\right)dw_l
\end{aligned}$$

In order to show that the supremum of the Gaussian process $\mathbf{D}\eta(\mathbf{x})$ has sub-Gaussian tail bound we use Proposition A.2.7 of Van Der Vaart and Wellner (1996). Let $\|\cdot\|_\rho$ denote the intrinsic semi-metric of the process $\mathbf{D}\eta(\mathbf{x})$ which is defined as follows:

$$\begin{aligned}
&\|\mathbf{x}-\mathbf{y}\|_\rho^2=\mathbb{E}\left[|\mathbf{D}\eta(\mathbf{x})-\mathbf{D}\eta(\mathbf{y})|^2\right]\\
&=2\prod_{l=1}^d\left(\left\|f^{(q_l)}\right\|_2^2\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2q_l}}\,dw_l\right)\\
&-2(-1)^q\prod_{l=1}^d\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2q_l}}\left(D_l^{q_l}f*D_l^{q_l}f\right)\left(\frac{x_l-y_l}{w_l}\right)dw_l.
\end{aligned}$$

Since f is an even function with bounded derivatives of order up to $q+1$, we have that

$$D_j\|\mathbf{x}\|_\rho^2|_{\mathbf{x}=0}=0,$$

for every $j\in[d]$ and also

$$D_jD_k\|\mathbf{x}\|_\rho^2|_{\mathbf{x}=0}=0$$

for every $j\neq k\in[d]$. Therefore, by Taylor's theorem we have the following,

$$\begin{aligned}
&\|\mathbf{x}-\mathbf{y}\|_\rho^2\le\sup_{\substack{\mathbf{z}\in[0,1]^d\\ j\in[d]}}\left|D_j^2\|\mathbf{z}\|_\rho^2\right|\cdot\|\mathbf{x}-\mathbf{y}\|_2^2\\
&=2\sup_{j\in[d]}\left|\prod_{\substack{l\in[d]\\ l\neq j}}\left\|f^{(q_l'+\delta_{l,j})}\right\|_2^2\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2(q_l+\delta_{l,j})}}\right|\|\mathbf{x}-\mathbf{y}\|_2^2,
\end{aligned}$$

where $\delta_{l,j}=1$ iff $l=j$, and $\delta_{l,j}=0$ otherwise. Therefore, the covering number of $[0,1]^d$ with respect to ρ is bounded as follows:

$$N\left(\epsilon,[0,1]^d,\rho\right)\le\left(\frac{C}{\epsilon}\right)^d,$$

where

$$C=2\sqrt{d}\cdot\sup_{j\in[d]}\left|\prod_{\substack{l\in[d]\\ l\neq j}}\left\|f^{(q_l')}\right\|_2^2\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2q_l'}}\,dw_l\right|.$$

Now using Proposition A.2.7 of Van Der Vaart and Wellner (1996), we have that

$$\Pr\left[\sup_{x\in[0,1]^d}|\mathbf{D}\eta(x)|>M\right]\le\left(\frac{LM}{\sigma^2}\right)^d e^{-\frac{M^2}{\sigma^2}},$$

where $\sigma^2=\prod_{l=1}^d\left(\left\|f^{(q_l)}\right\|_2^2\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2q_l}}\,dw_l\right)$ and $L=O\left(\sup_{j\in[d]}\left|\prod_{\substack{l\in[d]\\ l\neq j}}\left\|f^{(q_l+\delta_{l,j})}\right\|_2^2\int_{\mathbb{R}_+}\frac{p(w_l)}{w_l^{2(q_l+\delta_{l,j})}}\,dw_l\right|\right)$.

□

D Omitted lemmas and proofs from Section 4.1

Lemma 27 (Running time and Memory of WLSH Kernel Matrix)**.** *For any positive integers n,d and any dataset $\mathbf{x}^1,\mathbf{x}^2,\ldots,\mathbf{x}^n\in\mathbb{R}^d$, if $k_{f,p}(\cdot)$ is the WLSH estimator as in Definition 6 and $\widetilde{K}\in\mathbb{R}^{n\times n}$ is its corresponding kernel matrix then there exists an algorithm which using $O(dn)$ pre-processing time forms a data structure which can be stored using $O(n)$ memory words such that using this data structure, the product $\widetilde{K}\beta$ can be computed in time $O(n)$ for an arbitrary vector $\beta\in\mathbb{R}^n$.*

Proof of Claim 10: First note that for any $\beta\in\mathbb{R}^n$, we can write the quadratic form as,

$$\beta^\top\widetilde{K}\beta=\sum_{\mathbf{j}\in\mathbb{Z}^d}\left(\sum_{i=1}^n\beta_i\cdot f^{\otimes d}\left(\mathbf{j}+\frac{\mathbf{z}-\mathbf{x}^i}{\mathbf{w}}\right)\right)^2\ge0$$

Also by Cauchy-Schwarz inequality we have,

$$\begin{aligned}
\beta^\top\widetilde{K}\beta&=\sum_{\mathbf{j}\in\mathbb{Z}^d}\left(\sum_{i=1}^n\beta_i\cdot f^{\otimes d}\left(\mathbf{j}+\frac{\mathbf{z}-\mathbf{x}^i}{\mathbf{w}}\right)\right)^2\\
&\le\|f^{\otimes d}\|_\infty^2\sum_{\mathbf{j}\in\mathbb{Z}^d}\left(\sum_{i:h_{\mathbf{w},\mathbf{z}}(x^i)=\mathbf{j}}|\beta_i|\right)^2\\
&\le\|f^{\otimes d}\|_\infty^2\|\beta\|_1^2\\
&\le n\|f^{\otimes d}\|_\infty^2\|\beta\|_2^2.
\end{aligned}$$

□

Lemma 28. *(Matrix Chernoff, Tropp (2012)) Let $A_i\in\mathbb{R}^{n\times n}$ be independent random positive semi-definite matrices satisfying $\mathbf{E}[\sum_i A_i]=I$ and for all i, $\|A_i\|_{op}\le\alpha$ with probability 1. Then for any $0<\epsilon<1$, $\Pr\left[\|\sum_i A_i-I\|_{op}\le\epsilon\right]\ge1-2n\cdot\exp\left(-\frac{\epsilon^2}{3\alpha}\right)$.*

Lemma 29 (Slud's Inequality Slud (1977))**.** *Let $X_1,X_2,\cdots X_n$ be iid Bernoulli random variables with $\Pr[X_i=1]=p$. If $p\le1/4$ and $np\le r\le n$ or*

$np \le r \le (1-p)n$ then the following holds,

$$\Pr\left[\sum_{i\in[n]} X_i \ge r\right] \ge \Pr\left[Z \ge \frac{r-np}{\sqrt{np(1-p)}}\right],$$

where Z is a normal random variable with zero mean and variance one.

Therefore it follows from the above that, if $p \le 1/4$ and $t \ge 0$ or $p \le 1/2$ and $0 \le t \le n(1-2p)$ then,

$$\Pr\left[\sum_{i\in[n]} X_i - \mu \ge t\right] \ge \Pr\left[Z \ge \frac{t}{\sqrt{np(1-p)}}\right],$$

where $\mu = \mathbb{E}[\sum_{i\in[n]} X_i]$. This probability can be further lower bounded as follows:

$$\Pr\left[\sum_{i\in[n]} X_i - \mu \ge t\right] \ge \frac{1}{4}e^{-2t^2/\mu}.$$

Proof of Theorem 12: Let the points $\{x^i\}_{i=1}^n \subseteq \mathbb{R}^d$ be positioned as $x^1 = \cdots = x^{n/2} = (-\lambda/n, 0, 0, \cdots 0)^\top$ and $x^{n/2+1} = \cdots = x^n = (\lambda/n, 0, 0, \cdots 0)^\top$. Let the vector $\beta \in \mathbb{C}^n$ be defined as, $\beta_1 = \beta_2 = \cdots = \beta_{n/2} = -1$ and $\beta_{n/2+1} = \cdots = \beta_n = 1$. The proof proceeds by showing that in order to preserve the quadratic form corresponding to this β, one needs to set $m = \Omega\left(\frac{1}{\epsilon^2}\cdot\frac{n}{\lambda}\cdot\log n\right)$ and hence the lower bound follows for achieving an OSE. Let us compute the expectation of the quadratic form,

$$\begin{aligned}\mathbb{E}_{h_{\mathbf{w},\mathbf{z}}\sim\mathcal{H}}[\beta^\top \widetilde{K}^s\beta] &= \beta^\top K\beta \\ &= n^2\left(1-\exp(-2\lambda/n)\right)/2.\end{aligned}$$

Now we compute the second moment of the quadratic form as follows:

$$\begin{aligned}&\mathbb{E}_{h_{\mathbf{w},\mathbf{z}}\sim\mathcal{H}}\left[(\beta^\top\widetilde{K}^s\beta)^2\right] \\ &= \mathbb{E}_{\mathbf{w},\mathbf{z}}\left[\left(\sum_{\mathbf{j}\in\mathbb{Z}^d}\left(\sum_{i=1}^n \beta_i \operatorname{rect}^{\otimes d}\left(\frac{\mathbf{j}\mathbf{w}-x^i+\mathbf{z}}{\mathbf{w}}\right)\right)^2\right)^2\right] \\ &= \mathbb{E}_{w,z}\left[\frac{n^4}{4}\cdot\mathbf{1}_{\left\{|z|>\frac{w}{2}-\frac{\lambda}{n}\text{ or } w\le\frac{2\lambda}{n}\right\}}\right] \\ &= \mathbb{E}_w\left[\frac{n^4}{4}\cdot\min\left(1,\frac{2\lambda}{nw}\right)\right] = \frac{n^4}{4}\cdot\left(1-e^{-2\lambda/n}\right).\end{aligned}$$

Hence, we have the following for the ratio of second moment to the square of the first moment,

$$\frac{\mathbb{E}\left[(\beta^\top\widetilde{K}^s\beta)^2\right]}{\mathbb{E}\left[\beta^\top\widetilde{K}^s\beta\right]^2} = \frac{1}{1-e^{-2\lambda/n}} \ge \frac{n}{2\lambda}.$$

Note that the LSH estimator $\beta^\top\widetilde{K}^s\beta$ for this particular dataset $x_1, \cdots, x_n$ and vector β take in two possible values, zero and $\frac{n^2}{2}$. Therefore,

$$\beta^\top\widetilde{K}^s\beta = \begin{cases}\frac{n^2}{2} & \text{with probability } p \le \frac{2\lambda}{n} \\ 0 & \text{with probability } 1-p\end{cases}.$$

Note that $(1+\epsilon)\beta^\top K\beta + \epsilon\lambda\|\beta\|_2^2 \le (1+3\epsilon)\beta^\top K\beta$. Since $p \le 1/4$, by using Slud's inequality, Lemma 29, the probability of guaranteeing that $\frac{1}{m}\sum_{s=1}^m \beta^\top\widetilde{K}^s\beta \le (1+\epsilon)\beta^\top K\beta + \epsilon\lambda\|\beta\|_2^2$ is bounded as follows:

$$\begin{aligned}&\Pr\left[\frac{1}{m}\sum_{s=1}^m \beta^\top\widetilde{K}^s\beta > (1+\epsilon)\beta^\top K\beta + \epsilon\lambda\|\beta\|_2^2\right] \\ &\ge \Pr\left[\frac{1}{m}\sum_{s=1}^m \beta^\top\widetilde{K}^s\beta > (1+3\epsilon)\beta^\top K\beta\right] \\ &= \Pr\left[\frac{1}{m}\sum_{s=1}^m \beta^\top\widetilde{K}^s\beta > (1+3\epsilon)\cdot\mathbb{E}\left[\beta^\top\widetilde{K}\beta\right]\right] \\ &\ge \frac{1}{4}e^{-\frac{2(3\epsilon pm)^2}{pm}} = \frac{1}{4}e^{-18\epsilon^2 pm} \ge \frac{1}{4}e^{-36\epsilon^2\frac{\lambda}{n}m}.\end{aligned}$$

Therefore, in order to have $\Pr[\frac{1}{m}\sum_{s=1}^m \beta^\top\widetilde{K}^s\beta > (1+\epsilon)\beta^\top K\beta + \epsilon\lambda\|\beta\|_2^2] < \frac{1}{n}$, we need to average at least $m = \Omega(\frac{n}{\lambda}\cdot\log n/\epsilon^2)$ independent instances of WLSH estimator. □

E Risk bound of approximate KRR via LSH-Estimator

We use risk bounds to analyze the quality our approximate KRR estimator. It is common to bound the expected in-sample predication error of the KRR estimator as an empirical estimate of the statistical risk Avron et al. (2017b); Bach (2013); Alaoui and Mahoney (2015); Musco and Musco (2017). Formally, the *empirical risk* of an estimator η is defined as,

$$\mathcal{R}(\eta) = \mathbb{E}_{\{\epsilon_i\}}\left[\frac{1}{n}\sum_{i=1}^n\left(\eta(\mathbf{x}^i) - \eta^*(\mathbf{x}^i)\right)^2\right].$$

Suppose $\eta(\cdot)$ is the exact KRR estimator using kernel function $k(\cdot)$. Also suppose that $\widetilde{\eta}(\cdot)$ is the regressor obtained by solving the approximate KRR problem using the approximate kernel function $\widetilde{k}(\cdot)$. The following Lemma bounds the excess risk of approximate KRR estimator $\widetilde{\eta}(\cdot)$.

Lemma 30 (Approximate KRR Empirical Risk Bound)**.** *Let $\eta(\cdot)$ be the exact KRR estimator using the WLSH kernel function $k(\cdot)$ (Definition 8). Suppose $\widetilde{k}^s(\cdot)$ are independent instances of WLSH estimator for all $s \in [m]$. Let $\widetilde{\eta}$ be the approximate KRR estimator obtained by using the approximate*

kernel function $\widetilde{k}(\cdot) := \frac{1}{m}\sum_{s=1}^{m}\widetilde{k}^s(\cdot)$ *and let* $\widetilde{K}$ *be the corresponding kernel matrix to* $\widetilde{k}(\cdot)$*. If* $m = \Omega\left(\frac{\|f^{\otimes d}\|_\infty^2}{\epsilon^2}\cdot(n/\lambda)\cdot\log n\right)$ *then the following holds*[4]*,*

$$\Pr\left[\mathcal{R}(\widetilde{\eta}) \leq \frac{\mathcal{R}(\eta)}{1-\epsilon} + \frac{\epsilon\sigma_\epsilon^2\cdot\operatorname{rank}(\widetilde{K})}{(1+\epsilon)n}\right] \geq 1 - \frac{1}{\operatorname{poly}(n)}.$$

Proof. First note that Theorem 11 implied that with probability $1-\frac{1}{\operatorname{poly}(n)}$ the approximate kernel matrix satisfies following spectral guarantee,

$$(1-\epsilon)(K+\lambda I_n) \preceq \widetilde{K}+\lambda I_n \preceq (1+\epsilon)(K+\lambda I_n).$$

Therefore the lemma follows directly from invoking Lemma 2 of Avron et al. (2017b). □

[4]When we hash n points using LSH, we expect the number of non-empty buckets to grow at a lower rate than n. Therefore, we expect to have $\frac{\operatorname{rank}(\widetilde{K})}{n} \to 0$ as n grows.