# Active Latent Space Shape Model: A Bayesian Treatment of Shape Model Adaptation with an Application to Psoriatic Arthritis Radiographs

## Supplemental Material

Adwaye Rambojun [1] William Tillett [1,2] Tony Shardlow [1] Neill D. F. Campbell [1]

[1] University of Bath [2] Royal National Hospital for Rheumatic Diseases

## A. Numerical Approximation of Final Cost Function

For the sake of what follows, we assume that we are working with an exact marginal likelihood given by

$$\hat{p}(u \,|\, \boldsymbol{t}, \mathcal{T}) = f_u \star \mathcal{N}_{\sigma^2(\boldsymbol{t})+\gamma^2}\big(\mathcal{T} \circ \mu_d(\boldsymbol{t})\big) \ . \tag{1}$$

We approximate the marginal likelihood with the approximate marginal likelihood given by

$$\tilde{p}(u \,|\, \boldsymbol{t}, \mathcal{T}) = \exp\left(-\frac{\sum_{d=0}^{D-1} V_u\big(\mathcal{T} \circ \mu_d(\boldsymbol{t})\big)}{1+\sigma^2(\boldsymbol{t})+\gamma^2}\right). \tag{2}$$

Both of these expressions form a scalar field evaluated on the on the image lattice $\mathcal{X}$ that is generated from a distance transform operation from some edge discriminator output. During model adaptation, we seek to find the pose parameters $\mathcal{T}$ and shape parameters $\boldsymbol{t}$ that generate curve coordinates that sit on some maxima of these scalar fields. We therefore need the approximate marginal likelihood in equation (2) to preserve the local maxima of the marginal likelihood of equation (1). We empirically verify that this is true in Figure 1, where we show values of these two scalar fields at horizontal slices when $\sigma^2(\boldsymbol{t}) = 9$ and $\gamma = 4$.

## B. Effect of Posterior Shape Variance

Recall that the final objective function is given by

$$E_0(\boldsymbol{t}, \mathcal{T}) = \sum_{d=0}^{D-1} V_u(\mathcal{T} \circ \mu_d(\boldsymbol{t})) + \sigma^2(\boldsymbol{t}) \ , \tag{3}$$

We wish to investigate the effect of the posterior shape variance and for this purpose, we also adapt the ALSSM by minimising the following cost function

$$E_1(\boldsymbol{t}, \mathcal{T}) = E_0(\boldsymbol{t}, \mathcal{T}) - \sigma^2(\boldsymbol{t}) \ . \tag{4}$$

We show the results for this minimisation and compare it to the former in Figures 2 and 3. The role of the the posterior shape variance regulariser is to keep shape examples close to the GPLVM training data examples, where the posteior shape variance would be low. At dimension $Q = 2$, we see that using the regulariser results in a lower accuracy. This is because, for low latent space dimensions, latent space positions close to the dataset do not exhibit a rich enough shape variation. At higher dimensions, the latent space is able to generate a richer shape variety at positions close to training data and hence fits better to the true curve outline.

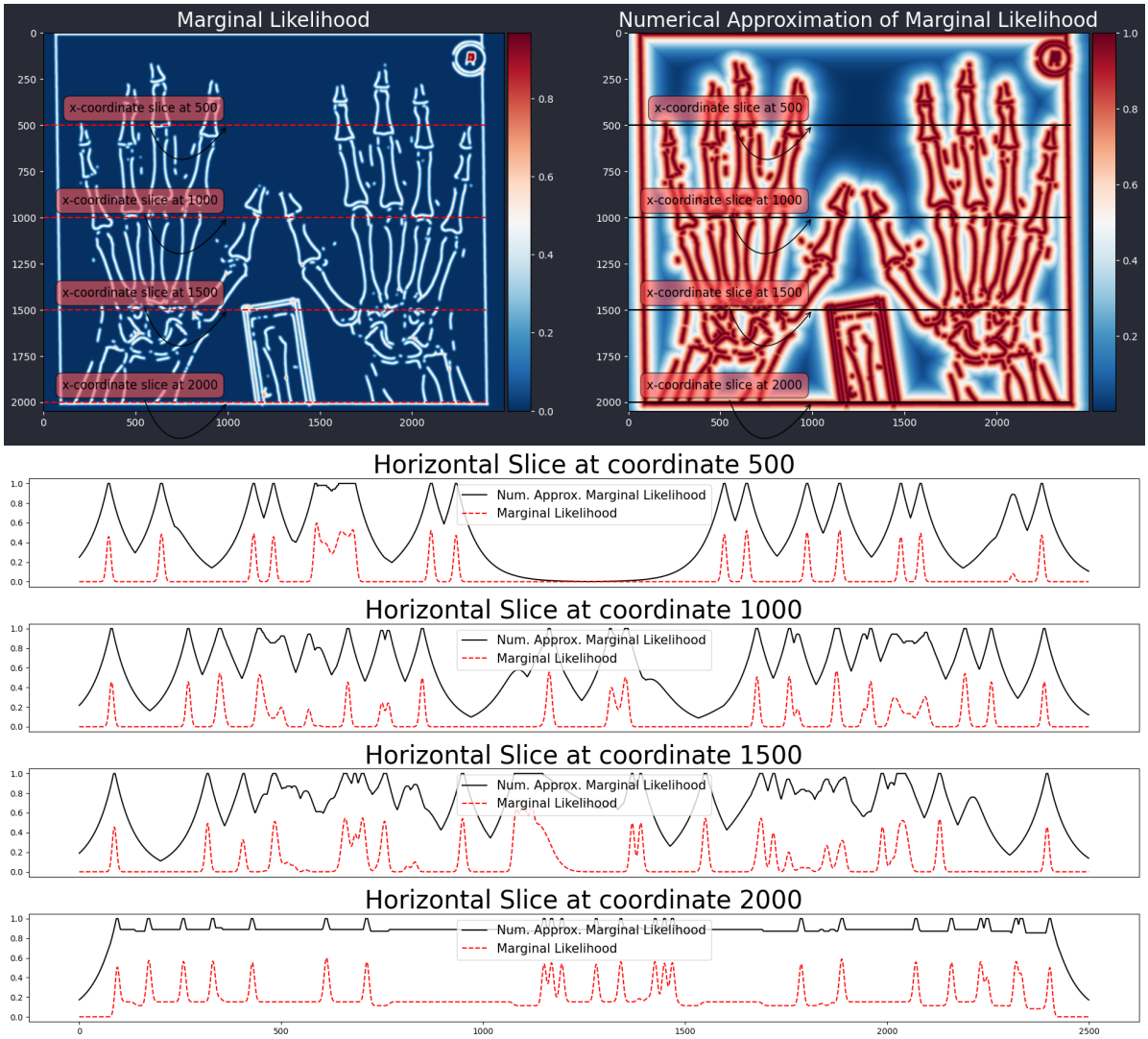


Figure 1. The Figures show slices of the scalar fields induced on the image lattice $\mathcal{X}$ by the marginal likelihood given by equation (1) and the approximate marginal likelihood given by equation (2).

## C. U-net Architecture for Building the Potential Function $V_u(\cdot)$

We train a U-net bone region discriminator on cropped radiographs as shown in Figure 4. We use tensorflow 1.9 for the implementation with a momentum optimizer and we train the architecture for 200 epochs on a dataset of 101 images split into a 80:20 training to test ratio. The batch-size used is 1 so that the architecture can be trained on full sized images without having to resize. We use a momentum optimiser with a momentum of 0.8 and a exponentially decaying learning rate with initial learning rate of 0.01 and a decay rate of 0.95. This combination of momentum optimizer with a high momentum value on small batches has been shown to work well [3].

## D. Radial Basis Function Interpolation

The rest of this supplement is taken from Sections 3.4 and 6.3.3 found in Rambojun et al. [2]. Consider a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ which we wish to approximate using Radial Basis Functions, where $n \in \mathbb{N}$,

$$\psi_h(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi v^2}} \exp\left(-\frac{\|\boldsymbol{x} - \mathcal{X}_h\|^2}{2v^2}\right). \tag{5}$$

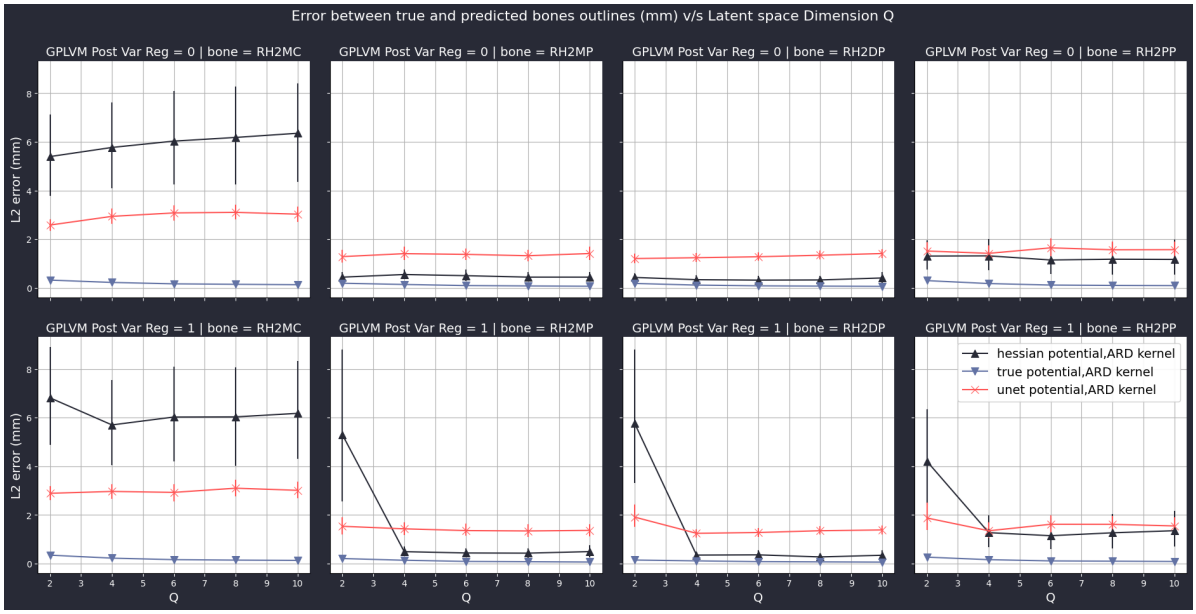


Figure 2. Results for the average L-2 error between the model generated bone outline and the true bone outline against the dimension $Q$ of the GPLVM (ARD kernel) latent space in a 10-fold validation. We compare the performance of three edge potential functions; one built by a U-net discriminator (unet), one built using a hessian based edge finder (hessian) and finally by using the true bone outline to build the potential (true). We also investigate the effect of the posterior shape variance on the result by minimising equation (4) (top row) and equation (3) (bottom row) while fitting the model.

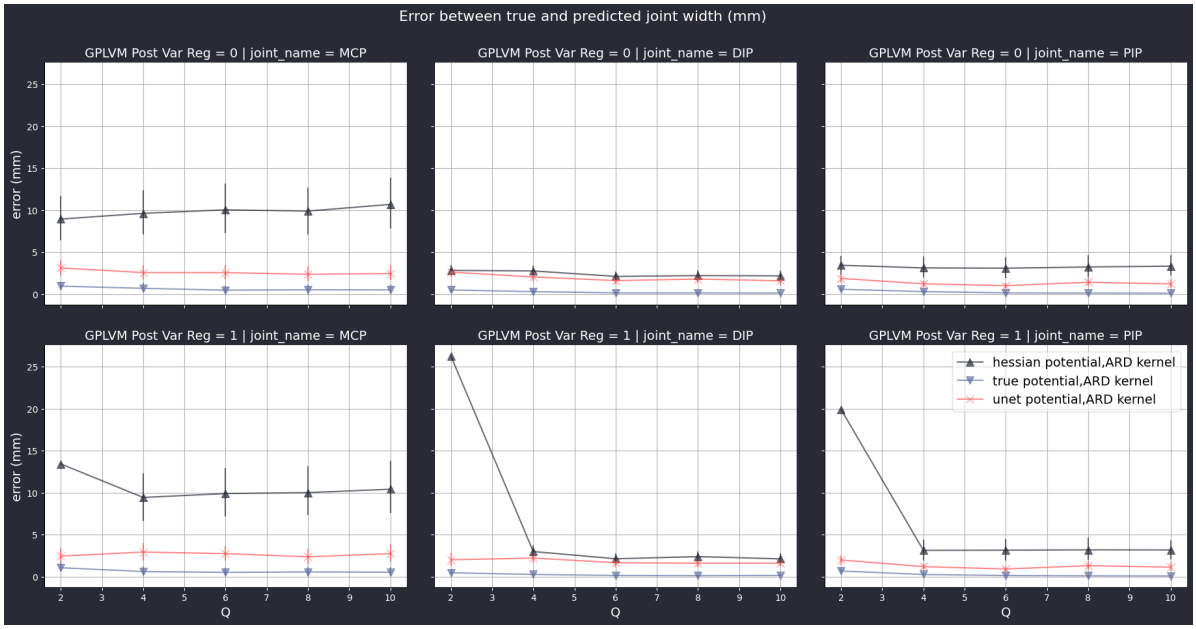


Figure 3. Results for the average error between the model generated joint space width and the true joint space width against the dimension $Q$ of the GPLVM (ARD kernel) latent space in a 10-fold validation. We compare the performance of three edge potential functions; one built by a U-net discriminator (unet), one built using a hessian based edge finder (hessian) and finally by using the true bone outline to build the potential (true). We also investigate the effect of the posterior shape variance on the result by minimising equation (4) (top row) and equation (3) (bottom row) while fitting the model.

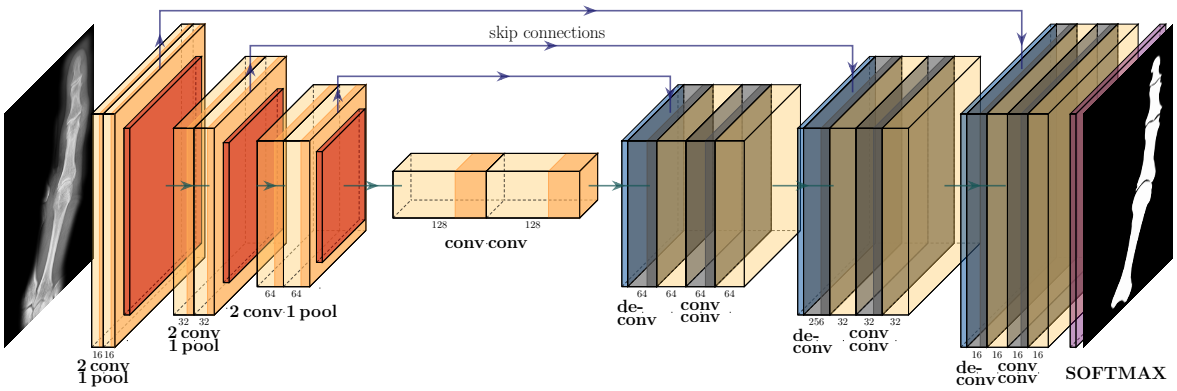


Figure 4. The figure shows the U-net architecture used. Each pooling operation halves the dimensions of its input while each deconvolution layer doubles the dimensions of its input. The training is done for 200 epochs on bacthes of size 1 with a momentum optimizer having momentum of 0.8; and an exponentially decaying learning rate with initial value of 0.01 and decay rate of 0.95.

We assume that we know the value of the function $f$ over a set $\mathcal{S} := \{\mathcal{X}_0, ..., \mathcal{X}_{H-1} : \mathcal{X}_i \neq \mathcal{X}_j \forall i \neq j\}$. Usually, interpolating $f$ using $\psi_h$ would require us to solve the system given by

$$\begin{bmatrix} \psi_0(\mathcal{X}_0) & . & . & . & \psi_{H-1}(\mathcal{X}_0) \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \psi_0(\mathcal{X}_{H-1}) & . & . & . & \psi_{H-1}(\mathcal{X}_{H-1}) \end{bmatrix} \begin{bmatrix} w_0 \\ . \\ . \\ . \\ w_{H-1} \end{bmatrix} = \begin{bmatrix} f(\mathcal{X}_0) \\ . \\ . \\ . \\ f(\mathcal{X}_{H-1}) \end{bmatrix}. \tag{6}$$

The interpolant will then be given by

$$\mathcal{I}_f(\boldsymbol{x}) = \sum_{h=0}^{H-1} \frac{w_h}{\sqrt{2\pi v^2}} \exp\left(-\frac{\|\boldsymbol{x} - \mathcal{X}_h\|^2}{2v^2}\right). \tag{7}$$

### D.1. Interpolation Error Estimate

In Interpolation Theory one is interested in controlling the maximum difference between the interpolant and the function $f$ within a bounded domain $\Omega$, i.e, one wants to bound $\max_{x\in\Omega} |f(\boldsymbol{x}) - \mathcal{I}_f(\boldsymbol{x})|$. We have the following Lemma from Frohlich [1]:

**Lemma D.1.** *Let $\Omega$ be a cube in $\mathbb{R}^d$ and let $\mathcal{I}_f$ be the radial basis function interpolant through a set $\mathcal{S} := \{\mathcal{X}_0, ..., \mathcal{X}_{H-1}\} \subset \Omega$ of a smooth, differentiable function $f$. Then we have that*

$$\max_{x\in\Omega} |f(\boldsymbol{x}) - \mathcal{I}_f(\boldsymbol{x})| \leq \exp\left(-\frac{\log(h_{\Omega,\mathcal{S}})}{h_{\Omega,\mathcal{S}}}\right) \|f\|_{\psi,\Omega}$$

*where*

$$\|f\|^2_{\psi,\Omega} = \frac{1}{(\sqrt{2\pi})^d} \int_\Omega \frac{\hat{f}^2(\omega)}{\hat{\mathcal{I}}_f(\omega)} d\omega$$

*is constant with Where $\hat{f}$ and $\hat{\mathcal{I}}_f$ denote the Fourier transform of $f$ and $\mathcal{I}_f$ respectively, and*

$$h_{\Omega,\mathcal{S}} := \sup_{\boldsymbol{x}\in\Omega} \min_{\mathcal{X}\in\mathcal{S}} \|\boldsymbol{x} - \mathcal{X}\|_2$$

What interests us in Lemma D.1 is the behaviour of $h_{\Omega,\mathcal{S}}$.

**Lemma D.2.** *There exists a sequence $\mathcal{S}_i$ s.t. $\lim_{i\to\infty} h_{\Omega,\mathcal{S}_i} = 0$*

*Proof.* Consider a general $\mathcal{S}$. We define the following subset of $\Omega$

$$B_h(\mathcal{S}) = \{\boldsymbol{x} \in \Omega : \|\boldsymbol{x} - \mathcal{X}_h\|_2 = \min_{\mathcal{X} \in \mathcal{S}} \|\boldsymbol{x} - \mathcal{X}\|_2\} \tag{8}$$

which contains all the elements of $\Omega$ which is in the maximum range of some $\mathcal{X}_h \in \mathcal{S}$. This often referred to as the Voronoi Cell associated with the site $\mathcal{X}_h$. We set

$$\|B_h(\mathcal{S})\| := \sup_{x \in B_h} \|\mathcal{X}_h - \boldsymbol{x}\| \, . \tag{9}$$

Then we have that

$$h_{\Omega,\mathcal{S}} = \max_{h=0,..,H-1} \|B_h(\mathcal{S})\| \, . \tag{10}$$

We define a neighbourhood structure on the elements of $\mathcal{S}$. Let $\mathcal{X}_h \in \mathcal{S}$. Then $\mathcal{X}_k \in \mathcal{S}$ is a neighbour of $\mathcal{X}_h$ if there exists a curve $\boldsymbol{m} : [0,1] \to \mathbb{R}^d$ with $\boldsymbol{m}(0) = \mathcal{X}_h$ and $\boldsymbol{m}(1) = \mathcal{X}_k$ s.t. $\boldsymbol{m} \cap B_h \cap B_k = \boldsymbol{m}$. Equivalently, $\mathcal{X}_h \sim \mathcal{X}_k$ if $B_k \cap B_h \neq \varnothing$. Let $d_i = \max_{p \sim q \in \mathcal{S}_i} \|\mathcal{X} - \boldsymbol{q}\|$. Consider now $\mathcal{X}_h \sim \mathcal{X}_k$. From the definition of a Voronoi cell we have that for $\boldsymbol{x} \in B_k(\mathcal{S}_i)$ and $\boldsymbol{y} \in B_k(\mathcal{S}_i) \cup \{\mathcal{X}_k\}$,

$$\begin{aligned} \|\mathcal{X}_k - \boldsymbol{x}\| &< \|\mathcal{X}_h - \boldsymbol{y}\| \\ &\text{setting } \boldsymbol{y} = \mathcal{X}_k \\ &< d_i \Rightarrow \\ \|B_k(\mathcal{S}_i)\| &< d_i \end{aligned} \tag{11}$$

For each such pair of neighbours, we choose a curve such that $\|\boldsymbol{m}\|_2 = \|\mathcal{X}_h - \mathcal{X}_k\|_2$ and pick a point halfway on the curve. We construct $\mathcal{S}_{i+1}$ to be the union of $\mathcal{S}_i$ and the points that we just created out of the neighbouring structure of $\mathcal{S}_i$.

As $\Omega$ is a box in $\mathbb{R}^d$, we can define $\mathcal{S}_0$ to be the set of points lying on the corners of $\Omega$. Under the above construction rule, the new points are added on lines joining adjacent elements of $\mathcal{S}_0$, as illustrated below.


$\mathcal{S}_0$ $\mathcal{S}_1$

In fact, with $\mathcal{S}_0$ described above, $B_h(\mathcal{S}_i)$ is a square domain for every $l$ and every $i$ and each new point is added on the corners of $B_h(\mathcal{S}_i)$. For such a construction, we have that $d_0$ is the distance between diagonal neighbours. Moreover, for all $p \in \mathcal{S}_0$ we have that:

$$d_0 = \max_{q : q \sim p \in \mathcal{S}_0} \|\mathcal{X} - \boldsymbol{q}\|_2 \, .$$

As we are adding new points that lie halfway between neighbours we necessarily have that $d_{i+1} = \frac{1}{2} d_i$ yielding

$$\|B_h(\mathcal{S}_i)\| \leq \frac{1}{2^i} d_0 \quad \forall l \Rightarrow h_{\Omega,\mathcal{S}_i} \leq \frac{1}{2^i} d_0 \, . \tag{12}$$

Hence we have created a sequence whose limit is zero as $i$ goes to infinity.

□

## E. Marginalisation Error Estimate

Let

$$\begin{aligned} f_u : \Omega \subseteq \mathbb{R}^2 &\to \mathbb{R} \\ \boldsymbol{x} &\mapsto \exp(-V_u(\boldsymbol{x})) \end{aligned}$$

be compactly supported and let $\mathcal{I}$ be the interpolant described in equation (7). Let

$$h_{\Omega,\mathcal{S}} := \sup_{\boldsymbol{x}\in\Omega} \min_{\mathcal{X}\in\mathcal{S}} \|\boldsymbol{x} - \mathcal{X}\|_2$$

where $\mathcal{S}$ is the set of interpolating nodes $\{\mathcal{X}_0, ..., \mathcal{X}_{H-1}\}$. We have the following bound from Frohlich [1]

$$\max_{\boldsymbol{x}\in\Omega} \left|\mathcal{I}(\boldsymbol{x}) - f(\boldsymbol{x})\right| \leq \exp\left(-\frac{\log(h_{\Omega,\mathcal{S}})}{h_{\Omega,\mathcal{S}}}\right) \tag{13}$$

We use this bound to show

(a) that an integral involving the product of $f_u$ and a Gaussian can be approximated by an integral involving the product of $\mathcal{I}$ and the same Gaussian (see equation (14))

(b) that a convolution of $f_u$ with a Gaussian can be approximated by a convolution of $\mathcal{I}$ with the same Gaussian (see equation (15)).

Together with Lemma D.2, this shows that the approximation we make for the marginalisation of model parameters has an error that goes to zero as the image resolution goes to infinity. The error bounds are

$$\begin{aligned} &\left| \int_\Omega \mathcal{I}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_\Omega f(\boldsymbol{x})\mathcal{N}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right| \\ \leq &\int_\Omega \left| \left(\mathcal{I}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\mathcal{N}(\boldsymbol{x}) \right| \mathrm{d}\boldsymbol{x} \\ \leq &\max_{\boldsymbol{x}\in\Omega} \left|\mathcal{I}(\boldsymbol{x}) - f(\boldsymbol{x})\right| \int_\Omega \mathcal{N}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &\text{using Holder's inequality} \\ \leq &C \exp\left(-\frac{\log(h_{\Omega,\mathcal{S}})}{h_{\Omega,\mathcal{S}}}\right) \end{aligned} \tag{14}$$

and

$$\begin{aligned} &\left| f \star \mathcal{N}(\boldsymbol{y}) - \mathcal{I} \star \mathcal{N}(\boldsymbol{y}) \right| \\ = &\left| \int_\Omega \mathcal{I}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x} - \boldsymbol{y})\,\mathrm{d}\boldsymbol{x} - \int_\Omega f(\boldsymbol{x})\mathcal{N}(\boldsymbol{x} - \boldsymbol{y})\,\mathrm{d}\boldsymbol{x} \right| \\ \leq &\int_\Omega \left| \left(\mathcal{I}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\mathcal{N}(\boldsymbol{x} - \boldsymbol{y}) \right| \mathrm{d}\boldsymbol{x} \\ \leq &\max_{\boldsymbol{x}\in\Omega} \left|\mathcal{I}(\boldsymbol{x}) - f(\boldsymbol{x})\right| \int_\Omega \mathcal{N}(\boldsymbol{x} - \boldsymbol{y})\,\mathrm{d}\boldsymbol{x} \\ &\text{using Holder's inequality} \\ \leq &C \exp\left(-\frac{\log(h_{\Omega,\mathcal{S}})}{h_{\Omega,\mathcal{S}}}\right) \end{aligned} \tag{15}$$

where $C$ is a constant depending on $f$ and whose form is given in Lemma D.1. The error estimates for the RBF interpolation of $f_u$ rely heavily on its compact support, which is guaranteed to hold for an image closed image lattice.

**Remark E.1.** *Frohlich [1] requires that the function to be interpolated lies in the so call native space of the interpolant. For our purposes, $f_u$ needs to to satisfy*

$$\int_\Omega \frac{\hat{f_u}^2(\omega)}{\hat{\mathcal{I}}(\omega)} d\omega \in \mathbb{R} \ .$$

*We hence require that the Fourier transform of $f_u$ decays at most as fast as that of a Gaussian. Let $\Omega$ be a square bounded domain on $\mathbb{R}^2$ and let $\Gamma \subset \mathbb{R}^2$ be the boundary of some open set in $\Omega$. Define $V_u$ to be given by*

$$V_u(\boldsymbol{x}) = \min_{\boldsymbol{y} \in \Gamma} \|\boldsymbol{x} - \boldsymbol{y}\| \ . \tag{16}$$

*Then for*

$$\begin{aligned} f_u :& \mathbb{R}^2 \to \mathbb{R} \\ & \boldsymbol{x} \mapsto \exp(-V_u(\boldsymbol{x})) \end{aligned} \tag{17}$$

*we have that $f_u$ behaves like a Gaussian at its tails. This ensures that the Fourier transform decays like a Gaussian at infinity, gauranteeing that $f_u$ lies in the Native space of the RBF interpolant.*

## References


[1] Fabian Fröhlich. Approximation and analysis of probability densities using radial basis functions. Master's thesis, Technische Universität München, Germany, 2013.

[2] Adwaye Rambojun. *Automatic scoring of X-rays in Psoriatic Arthritis*. PhD thesis, University of Bath, 5 2020.

[3] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).