A. Kernel Fourier Transform

We have shown that the Fourier series of length T form the harmonic kernel decomposition with T kernels. Intuitively, if $T \to \infty$, we obtain a "continuous" frequency representation of the kernel, which would be akin to a Fourier transform.

Consider the transformation $G^{\mathbf{s}} : \mathcal{X} \to \mathcal{X}, \mathbf{s} \in \mathbb{R}^J$, corresponding to J-way transformations. We assume the transformation G is $\mathbf{1}$-periodic: $G^{\mathbf{0}}(\mathbf{x}) = G^{\mathbf{1}}(\mathbf{x}) = \mathbf{x}, G^{\mathbf{s}_1+\mathbf{s}_2}(\mathbf{x}) = G^{\mathbf{s}_1}(G^{\mathbf{s}_2}(\mathbf{x})), \forall \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^J$. A kernel is G-invariant if for any $\mathbf{s} \in \mathbb{R}^J$, $k(G^{\mathbf{s}}(\mathbf{x}), G^{\mathbf{s}}(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}')$.

Given the inputs $\mathbf{x}, \mathbf{x}'$, we consider the space of kernel values: $k(\mathbf{x}, G^{\mathbf{s}}(\mathbf{x}')), \mathbf{s} \in \mathbb{R}^J$. For $\mathbf{t} \in \mathbb{R}^J$, we define the complex-valued function $k_{\mathbf{t}} : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ using the Fourier transform,

$$k_{\mathbf{t}}(\mathbf{x}, \mathbf{x}') = \int_{\mathbb{R}^J} e^{-2\pi i \mathbf{s}^\top \mathbf{t}} k(\mathbf{x}, G^{\mathbf{s}}(\mathbf{x}')) \mathrm{d}\mathbf{s}, \tag{25}$$

In this way, $k_{\mathbf{t}}(\mathbf{x}, \mathbf{x}')$ captures the frequency of $\mathbf{t}$ in the function $\mathbf{s} \to k(\mathbf{x}, G^{\mathbf{s}}(\mathbf{x}'))$. Similar to the harmonic kernel decomposition, we show an alternative representation of the kernel using k_t.

Theorem A.1 (Harmonic Kernel Representation)**.**

$$k(\mathbf{x}, \mathbf{x}') = \int_{\mathbb{R}^J} k_{\mathbf{t}}(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{t}. \tag{26}$$

Moreover, $k_{\mathbf{t}}$ is a kernel for all $\mathbf{t} \in \mathbb{R}^J$.

Proof of Theorem A.1. We prove this theorem by the following derivation,

$$\begin{aligned}\int_{\mathbb{R}^J} k_{\mathbf{t}}(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{t} &= \int_{\mathbb{R}^J} \int_{\mathbb{R}^J} e^{-2\pi i \mathbf{s}^\top \mathbf{t}} k(\mathbf{x}, G^{\mathbf{s}}(\mathbf{x}')) \mathrm{d}\mathbf{s} \mathrm{d}\mathbf{t} = \int_{\mathbb{R}^J} k(\mathbf{x}, G^{\mathbf{s}}(\mathbf{x}')) \int_{\mathbb{R}^J} e^{-2\pi i \mathbf{s}^\top \mathbf{t}} \mathrm{d}\mathbf{t} \mathrm{d}\mathbf{s} \\ &= \int_{\mathbb{R}^J} k(\mathbf{x}, G^{\mathbf{s}}(\mathbf{x}')) \delta_{\mathbf{s}} \mathrm{d}\mathbf{s} = k(\mathbf{x}, \mathbf{x}').\end{aligned}$$

where we used the property that the Fourier transform of the constant function is the delta function.

To show that k_t is a kernel, we prove the following equality,

$$\begin{aligned}\int_{\mathbb{R}^J} \int_{\mathbb{R}^J} e^{-2\pi i \mathbf{t}^\top (\mathbf{s}_2 - \mathbf{s}_1)} k(G^{\mathbf{s}_1}(\mathbf{x}), G^{\mathbf{s}_2}(\mathbf{x}')) \mathrm{d}\mathbf{s}_1 \mathrm{d}\mathbf{s}_2 &= \int_{\mathbb{R}^J} \int_{\mathbb{R}^J} e^{-2\pi i \mathbf{t}^\top \mathbf{s}_2} k(\mathbf{x}, G^{\mathbf{s}_2}(\mathbf{x}')) \mathrm{d}\mathbf{s}_1 \mathrm{d}\mathbf{s}_2 \\ &= \int_{\mathbb{R}^J} e^{-2\pi i \mathbf{t}^\top \mathbf{s}_2} k(\mathbf{x}, G^{\mathbf{s}_2}(\mathbf{x}')) \mathrm{d}\mathbf{s}_2 = k_{\mathbf{t}}(\mathbf{x}, \mathbf{x}').\end{aligned}$$

□

We demonstrate the Kernel Fourier Transform by considering a stationary kernel on the unit circle. We denote the input x as the angle, then the kernel admits the form $k(x, x') = \kappa(x - x')$, where κ is a periodic function of period 2π. Let $\kappa_0(t) = \kappa(t)\mathbb{I}[0 \le t < 2\pi]$, then

$$k(x, x') = \kappa(x - x') = \sum_{n \in \mathbb{Z}} \kappa_0(x - x' - 2\pi n),$$

Let $G^s(x) = x + 2\pi s$, we obtain,

$$\begin{aligned}k_t(x, x') &= \sum_{n \in \mathbb{Z}} \int_{\mathbb{R}} e^{-2\pi i s t} \kappa_0(x - x' - 2\pi s - 2\pi n) \mathrm{d}s = \frac{1}{2\pi} \sum_{n \in \mathbb{Z}} \int_{\mathbb{R}} e^{-i(x - x' - 2\pi n - w)t} \kappa_0(w) \mathrm{d}w \\ &= e^{-it(x-x')} \left[\frac{1}{2\pi} \int_{\mathbb{R}} e^{iwt} \kappa_0(w) \mathrm{d}w \right] \sum_{n \in \mathbb{Z}} e^{2\pi i n t} = e^{-it(x-x')} \hat{\kappa}_0(t) \sum_{n \in \mathbb{Z}} \delta(t - n).\end{aligned}$$

where $\hat{\kappa}_0$ is the inverse Fourier transform of κ_0. Then we have the Fourier series,

$$k(x, x') = \int_{\mathbb{R}} k_t(x, x') dt = \sum_{n \in \mathbb{Z}} \hat{\kappa}_0(n) e^{-in(x-x')}.$$

B. Inter-domain Inducing Points Formulation

We present an inter-domain inducing points interpretation of the harmonic kernel decomposition. An inter-domain inducing point is a function $w : \mathcal{X} \to \mathbb{C}$ whose inducing variable is defined as,

$$u_w = \int f(\mathbf{x})w(\mathbf{x})d\mathbf{x}, \tag{27}$$

We introduce T kinds of inter-domain inducing points. For $t = 0, ..., T-1$, given $\mathbf{z}_t \in \mathcal{X}$, the inter-domain inducing point of the t-th kind is,

$$w_t = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H \delta_{G^s(\mathbf{z}_t)}, \tag{28}$$

$$u_{w_t} = \int f(\mathbf{x})w_t(\mathbf{x})d\mathbf{x} = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f(G^s(\mathbf{z}_t)), \tag{29}$$

Therefore, we can generalize the kernel function to include inter-domain inputs,

$$k(\mathbf{x}, w_t) = \mathbb{E}[f(\mathbf{x})u_{w_t}^H] = \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(\mathbf{x}, G^s(\mathbf{z}_t)), \tag{30}$$

$$k(w_t, w_t') = \mathbb{E}[u_{w_t} u_{w_t'}^H] = \sum_{s=0}^{T-1}\sum_{s'=0}^{T-1} \mathbf{F}_{t,s}^H \mathbf{F}_{t,s'} k(G^s(\mathbf{z}_t), G^{s'}(\mathbf{z}_t')) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(\mathbf{z}_t, G^s(\mathbf{z}_t')), \tag{31}$$

where the last equality is based on Lemma D.1. Furthermore, for $0 \leq t \neq t' \neq T-1$,

$$k(w_t, w_{t'}) = \sum_{s=0}^{T-1}\sum_{s'=0}^{T-1} \mathbf{F}_{t,s} \mathbf{F}_{t',s'}^H k(G^s(\mathbf{z}_t), G^{s'}(\mathbf{z}_{t'}')) = 0.$$

where the last equality is based on Lemma D.2. We observe,

$$k(\mathbf{x}, w_t) = k_t(\mathbf{x}, \mathbf{z}_t);\ k(w_t, w_t') = k_t(\mathbf{z}_t, \mathbf{z}_t'). \tag{32}$$

Now we find that *the proposed inter-domain inducing points formulation is equivalent to the kernel Fourier series.* Furthermore, the equivalence also reinterprets HVGPs as standard SVGPs using inter-domain inducing points while enforcing block diagonal posterior covariances.

C. More Experiments and Details

C.1. Toy Visualization

HVGPs are based on the decomposed GP formulation and assume independent variational posteriors. Therefore, the predictions on a target location $\mathbf{x}^\star$ can be decomposed as the combination of independent elements,

$$\mathcal{N}(\mathbf{0}, \mathbf{K}_{\star\star} - \sum_{t=0}^{T-1} \mathbf{K}_{t,\star\mathbf{u}} \mathbf{K}_{t,\mathbf{uu}}^{-1} \mathbf{K}_{t,\mathbf{u}\star})) + \sum_{t=0}^{T-1} \mathcal{N}(\mathbf{K}_{t,\star\mathbf{u}_t} \mathbf{K}_{t,\mathbf{u}_t\mathbf{u}_t}^{-1} \boldsymbol{\mu}_t, \mathbf{K}_{t,\star\mathbf{u}_t} \mathbf{K}_{t,\mathbf{u}_t\mathbf{u}_t}^{-1} \mathbf{S}_t \mathbf{K}_{t,\mathbf{u}_t\mathbf{u}_t}^{-1} \mathbf{K}_{t,\star\mathbf{u}_t})$$

where we use $\mathbf{K}_{t,\cdot}$ to represent the kernel k_t. The first term in the prediction represents the error of the Nyström approximation, and the remaining terms contain the predictions from all subprocesses.

In this section we conduct a Snelson's 1D toy experiment to visualize the posterior predictions and each term. We set $T = 2, G(x) = -x, m = 5$, which results in HVGP (2×5). Because the original training inputs are positive, we preprocess it by subtracting the inputs by the mean. The results are shown in Figure 9. We find that using 2×5 inducing points fit the training data well, and generate reasonable predictive uncertainty as well. The predictions for the two GPs correspond to the symmetric and the antisymmetric fraction, respectively.

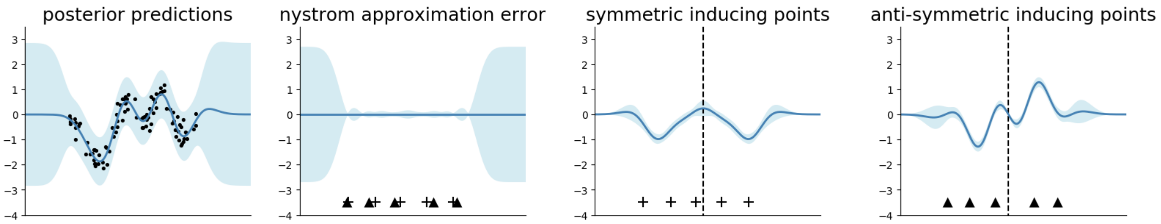

Figure 9. Posterior predictions on the Snelson dataset, where shaded bands correspond to intervals of ±3 standard deviation. The leftmost *posterior prediction* plot is the combination of the right three plots, plus the observation variance. We also visualize the associated inducing points for each plot. We observe that the HVGP predictions are separated as the symmetric fraction and the antisymmetric fraction.

C.2. Regression Benchmarks

We use the Matérn 3/2 kernel with shared lengthscales across input dimensions. For HVGPs, the transformation G is by negating over PCA directions. We split the PCA directions into J subsets, then apply negations over which results in a $2^J \times M$ model. We let the j-th subset contain the directions with the jth large, $(J+j)$th large, ... eigenvalues, so that the principal subspace is covered well. Except for the *year* dataset which has a standard train/test split, each dataset is randomly split into 64% training, 16% validating, 20% testing sets and is averaged over 3 random splits. We initialize the inducing points using K-means and initialize the kernel lengthscale using the median heuristic. The Gaussian likelihood variance is initialized at 0.1. For all experiments, we optimize for 30k iterations with the Adam optimizer using learning rate 0.003 and batch size 256. We visualize the results for test RMSEs in Figure 10, and how each criterion evolves along training in Figure 11.

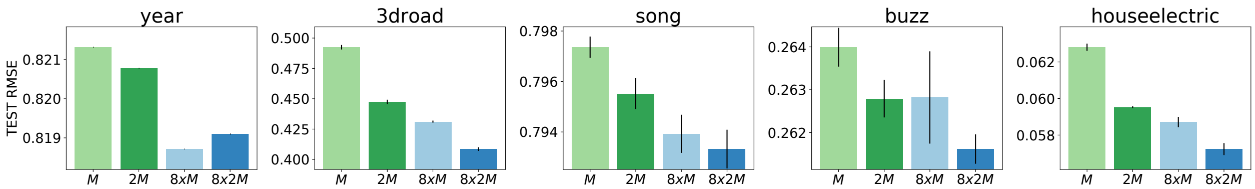

Figure 10. Test RMSEs on regression benchmarks. We compare SVGPs using $M, 2M$ inducing points and HVGPs using $8 \times M, 8 \times 2M$ inducing points, for $M = 1000$.

C.3. CIFAR-10 Classification

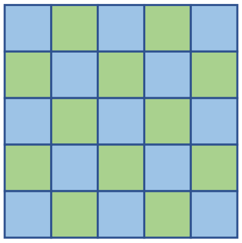

Figure 12. Group Splitting for the HVGP ($4 \times M$).

	1-layer	2-layer	3-layer	4-layer
filter size	5	5, 4	5,4,5	5,4,5,4
stride size	1	1,2	1,2,1	1,1,1,1
channel num	-	10	10,10	16,16,16
pooling	-	-	-	mean
pooling size	-	-	-	1,2,1
padding	SAME	SAME	SAME	SAME
M	384x0, 1K	384x1, 1K	384x2, 1K	384x3, 1K

Table 3. Model Configurations for Deep Convolutional Gaussian processes.

For deep Gaussian processes, we let M_l, h_l be the number of inducing points and the number of input units in the l-th layer, respectively. The variational posterior for the inducing points $\mathbf{U}^l, \mathbf{U}^l \in \mathbb{R}^{M_l \times h_{l+1}}$ in the l-th layer is usually a multivariate Gaussian,

$$q(\mathrm{vec}(\mathbf{U}^l)) = \mathcal{N}(\mathrm{vec}(\mathbf{M}^l), \mathbf{\Sigma}^l), \tag{33}$$

where $\mathbf{M}^l \in \mathbb{R}^{M_l \times h_{l+1}}, \mathbf{\Sigma}^l \in \mathbb{R}^{(M_l h_{l+1}) \times (M_l h_{l+1})}$ are the mean and the covariance, respectively. A commonly-used

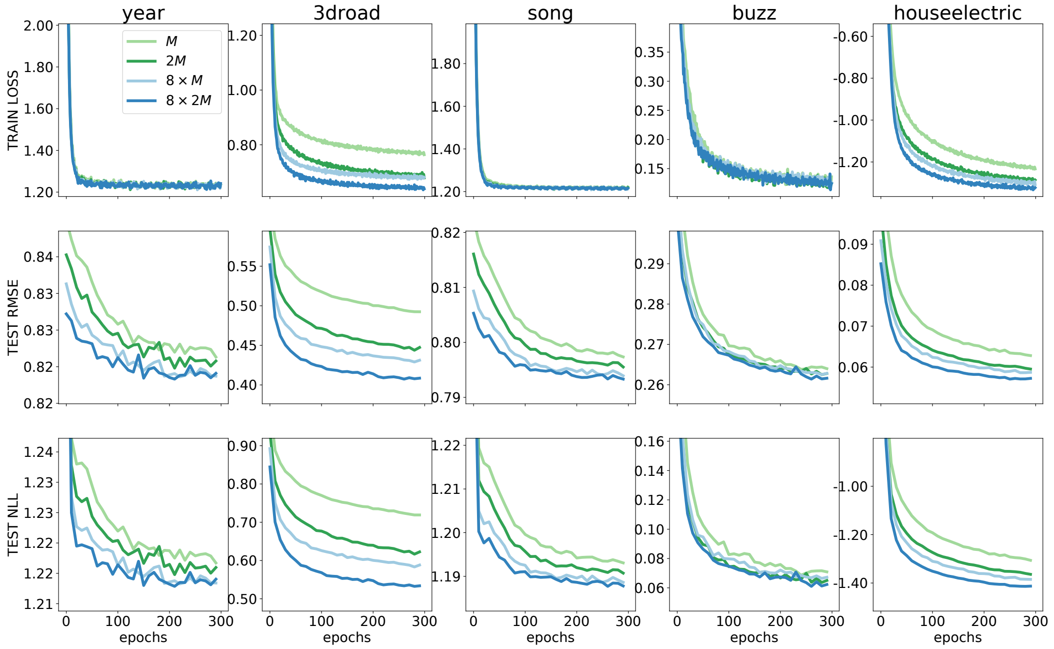

Figure 11. How *train loss*, *test rmse*, and *test nll* evolve during training. We compare SVGPs using $M, 2M$ inducing points and HVGPs using $8 \times M, 8 \times 2M$ inducing points, for $M = 1000$.

structure for $\mathbf{\Sigma}^l$ is the block-diagonal covariance (Salimbeni & Deisenroth, 2017) , i.e., assuming independence between output channels. However, the true posterior is not independent. Moreover, such covariance involves h_{l+1} covariances of shape $M_l \times M_l$, which are both memory intensive and computation intensive. Therefore, following Park et al. (2018), we use the Kronecker-factored structure for the covariance, i.e., $\mathbf{\Sigma}^l = \mathbf{\Sigma}^l_o \otimes \mathbf{\Sigma}^l_i$, where $\mathbf{\Sigma}_o \in \mathbb{R}^{h_{l+1} \times h_{l+1}}$, $\mathbf{\Sigma}_i \in \mathbb{R}^{M_l \times M_l}$ correspond to the output covariance and the input covariance, respectively.

Following Shi et al. (2020), all models were optimized using 270k iterations with the Adam optimizer using a learning rate 0.003 and a batch size 64. We anneal the learning rate by 0.25 every 50k iterations to ensure convergence. Unlike Shi et al. (2020) which used a zero mean function, we used a convolution mean function whose filter is 1 for the center pixel and 0 everywhere else, since we observe it with a better performance. We used the robust multi-class classification likelihood. For lower layers in the deep convolutional GP, we used multi-output GPs for each input patch (Blomqvist et al., 2019); for the output layer, we used the TICK kernel (Dutordoir et al., 2019). The patch kernels are RBF kernels with shared lengthscales, whose lengthscales and variances are initialized at 5. The TICK location kernel is a Matérn 3/2 kernel whose lengthscales and variances are initialized at 1 and 3, respectively. To initialize the inducing filters, we use K-means samples from $\min(100 * M, 10000)$ random input patches, while the inputs in all layers are obtained by forwarding the image through a random Xavier convnet.

For the HVGP ($2 \times M$) we use the negation transformation on the inducing points $G(\mathbf{z}) = -\mathbf{z}$. For the HVGP ($4 \times M$), we also use the negation transformation over two groups that are determined by pixel locations, as shown in Figure 12.

D. Proofs

D.1. Lemmas

Lemma D.1. *For any* $t = 0, ..., T-1$, $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$,

$$\sum_{s=0}^{T-1}\sum_{s'=0}^{T-1} \mathbf{F}_{t,s}^H \mathbf{F}_{t,s'} k(G^s(\mathbf{x}), G^{s'}(\mathbf{x}')) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(\mathbf{x}, G^s(\mathbf{x}')).$$

Proof. We prove the equality by the expression of $\mathbf{F}$,

$$\frac{1}{T^2}\sum_{s=0}^{T-1}\sum_{s'=0}^{T-1} e^{-i\frac{2\pi t}{T}(s'-s)} k(G^s(\mathbf{x}), G^{s+(s'-s)}(\mathbf{x}')) = \frac{1}{T}\sum_{s_0=0}^{T-1} e^{-i\frac{2\pi t}{T}s_0} k(\mathbf{x}, G^{s_0}(\mathbf{x}')),$$

where we used the kernel invariance to G. Also, since G is T-cyclic, we changed the variable $s'-s$ to s_0 and s_0 still ranges from 0 to $T-1$. □

Lemma D.2. *For any* $0 \le t_1 \ne t_2 \le T-1$, $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$,

$$\sum_{s_1=0}^{T-1}\sum_{s_2=0}^{T-1} \mathbf{F}_{t_1,s_1}^H \mathbf{F}_{t_2,s_2} k(G^{s_1}(\mathbf{x}), G^{s_2}(\mathbf{x}')) = 0, \tag{34}$$

$$\sum_{s_1=0}^{T-1}\sum_{s_2=0}^{T-1} \mathbf{F}_{t_1,s_1} \mathbf{F}_{t_2,s_2}^H k(G^{s_1}(\mathbf{x}), G^{s_2}(\mathbf{x}')) = 0, \tag{35}$$

Proof. Below we prove (34). The proof of (35) follows similarly.

$$\begin{aligned}\sum_{s_1=0}^{T-1}\sum_{s_2=0}^{T-1} \mathbf{F}_{t_1,s_1}^H \mathbf{F}_{t_2,s_2} k(G^{s_1}(\mathbf{x}), G^{s_2}(\mathbf{x}')) &= \sum_{s_1=0}^{T-1}\sum_{s_2=0}^{T-1} \mathbf{F}_{t_1,s_1}^H \mathbf{F}_{t_2,s_1+s_2} k(G^{s_1}(\mathbf{x}), G^{s_1+s_2}(\mathbf{x}')) \\ = \sum_{s_2=0}^{T-1} k(\mathbf{x}, G^{s_2}(\mathbf{x}')) \sum_{s_1=0}^{T-1} \mathbf{F}_{t_1,s_1}^H \mathbf{F}_{t_2,s_1+s_2} &= \sum_{s_2=0}^{T-1} k(\mathbf{x}, G^{s_2}(\mathbf{x}')) e^{-i\frac{2\pi t s_2}{T}} \sum_{s_1=0}^{T-1} \mathbf{F}_{t_1,s_1}^H \mathbf{F}_{t_2,s_1} = 0.\end{aligned}$$

In the last step, $\sum_{s_1=0}^{T-1} \mathbf{F}_{t_1,s_1}^H \mathbf{F}_{t_2,s_1} = 0$ whenever $t_1 \ne t_2$. This is because the columns of $\mathbf{F}$ form an orthogonal basis over the set of T-dimensional complex vectors. □

Lemma D.3. *Under the harmonic formulation, for* $\mathbf{x} \in \mathcal{X}$,

$$f_t(\mathbf{x}) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f(G^s(\mathbf{x})). \tag{36}$$

Proof. We consider a marginal distribution on a subset of function values,

$$p(\{f(G^s(\mathbf{x}))\}_{s=0}^{T-1}, \{f_0(G^s(\mathbf{x}))\}_{s=0}^{T-1}, ..., \{f_{T-1}(G^s(\mathbf{x}))\}_{s=0}^{T-1}),$$

The distribution can be represented as,

$$f(G^s(\mathbf{x})) = \sum_{t=0}^{T-1} f_t(G^s(\mathbf{x})), s = 0, ..., T-1,$$

$$\mathbf{f}_t(G^{0:T-1}(\mathbf{x})) \sim \mathcal{N}(0, \mathbf{K}_t(G^{0:T-1}(\mathbf{x}), G^{0:T-1}(\mathbf{x}))),$$

We first investigate the structure of the kernel matrix $\mathbf{K}_t$,

$$\begin{aligned} k_t(G^j(\mathbf{x}), G^{j'}(\mathbf{x})) &= \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(G^j(\mathbf{x}), G^{s+j'}(\mathbf{x})) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(\mathbf{x}, G^{s+j'-j}(\mathbf{x})) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s+j-j'} k(\mathbf{x}, G^s(\mathbf{x})) \\ &= e^{-\frac{2\pi i t(j-j')}{T}} \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(\mathbf{x}, G^s(\mathbf{x})) = e^{-\frac{2\pi i t(j-j')}{T}} k_t(\mathbf{x}, \mathbf{x}), \end{aligned}$$

Therefore, the matrix $\mathbf{K}_t = [k_t(\mathbf{x}, \mathbf{x}) e^{-\frac{2\pi i t(j-j')}{T}}]_{j,j'=0}^{T-1}$. Let $\epsilon_t \in \mathbb{R}$ be a random Gaussian noise, then the random vector of f_t can be written as,

$$\mathbf{f}_t(G^{0:T-1}(\mathbf{x})) = [\sqrt{k_t(\mathbf{x}, \mathbf{x})} e^{-\frac{2\pi i t j}{T}} \epsilon_t]_{j=0}^{T-1}, \tag{37}$$

Now we can compute the RHS in the lemma,

$$\sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f(G^s(\mathbf{x})) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H \sum_{t'=0}^{T-1} f_{t'}(G^s(\mathbf{x})) = \sum_{t'=0}^{T-1} \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f_{t'}(G^s(\mathbf{x})),$$

If $t' = t$,

$$\sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f_t(G^s(\mathbf{x})) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H \sqrt{k_t(\mathbf{x}, \mathbf{x})} e^{-\frac{2\pi i t s}{T}} \epsilon_t = \sqrt{k_t(\mathbf{x}, \mathbf{x})} \epsilon_t = f_t(\mathbf{x}). \tag{38}$$

If $t' \neq t$,

$$\sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f_{t'}(G^s(\mathbf{x})) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H \sqrt{k_{t'}(\mathbf{x}, \mathbf{x})} e^{-\frac{2\pi i t' s}{T}} \epsilon_t = \sqrt{k_{t'}(\mathbf{x}, \mathbf{x})} \epsilon_t \sum_{s=0}^{T-1} e^{\frac{2\pi i (t-t')}{T}} = 0. \tag{39}$$

Therefore,

$$\sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f(G^s(\mathbf{x})) = f_t(\mathbf{x}). \tag{40}$$

Because this holds for all marginal distributions, it holds as well for the function samples from Gaussian processes. □

D.2. Proofs for Sec 3

Proof of Proposition 3.3. The equality can be directly proven,

$$\begin{aligned} k_t(\mathbf{x}, G(\mathbf{x}')) &= \sum_{s=0}^{T-1} \mathbf{F}_{t,s} k(\mathbf{x}, G^{s+1}(\mathbf{x}')) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s-1} k(\mathbf{x}, G^s(\mathbf{x}')) \\ &= \sum_{s=0}^{T-1} \frac{1}{T} e^{-i\frac{2\pi t(s-1)}{T}} k(\mathbf{x}, G^s(\mathbf{x}')) = e^{i\frac{2\pi t}{T}} k_t(\mathbf{x}, \mathbf{x}'). \end{aligned} \tag{41}$$

□

Proof of Theorem 3.5. The equality can be directly proven,

$$\sum_{t=0}^{T-1} k_t(\mathbf{x}, \mathbf{x}') = \sum_{t=0}^{T-1} \sum_{s=0}^{T-1} \frac{1}{T} e^{-i\frac{2\pi s t}{T}} k(\mathbf{x}, G^s(\mathbf{x}')) = \frac{1}{T} \sum_{s=0}^{T-1} k(\mathbf{x}, G^s(\mathbf{x}')) \sum_{t=0}^{T-1} e^{-i\frac{2\pi s t}{T}} = \frac{1}{T} \sum_{s=0}^{T-1} k(\mathbf{x}, G^s(\mathbf{x}')) T \delta_s = k(\mathbf{x}, \mathbf{x}').$$

To prove that k_t is a kernel, we observe from Lemma D.1 that $k_t(\mathbf{x}, \mathbf{x}') = \mathbf{F}_{t,:}^H \mathbf{K} \mathbf{F}_{t,:}$, where $\mathbf{K} = [k(G^{s_1}(\mathbf{x}), G^{s_2}(\mathbf{x}'))]_{s_1,s_2=0}^{T-1}$. Since k is a kernel, we conclude that $k_t(\mathbf{x}, \mathbf{x}') = \mathbf{F}_{t,:}^H \mathbf{K} \mathbf{F}_{t,:}$ is a kernel as well.

□

Proof of Lemma 3.6. We firstly prove that, for all $t_1 \neq t_2$ and $\mathbf{x}, \mathbf{x}'$, $\langle k_{t_1}(\cdot, \mathbf{x}), k_{t_2}(\cdot, \mathbf{x}')\rangle_{\mathcal{H}_k} = 0$. The RKHS inner product can be computed as,

$$\langle k_{t_1}(\cdot, \mathbf{x}), k_{t_2}(\cdot, \mathbf{x}')\rangle_{\mathcal{H}_k} = \sum_{s_1=0}^{T-1}\sum_{s_2=0}^{T-1} \mathbf{F}_{t_1,s_1}\mathbf{F}^H_{t_2,s_2} k(G^{s_1}(\mathbf{x}), G^{s_2}(\mathbf{x}')) = 0, \tag{42}$$

where the last equality is due to Lemma D.2.

Moreover, if the functions f, g can be written as linear combinations of the corresponding kernels,

$$f(\mathbf{x}) = \sum_s a_s k_{t_1}(\mathbf{x}, \mathbf{x}^s_{t_1});\; g(\mathbf{x}) = \sum_s b_s k_{t_2}(\mathbf{x}, \mathbf{x}^s_{t_2})$$

Following that $\langle k_{t_1}(\cdot, \mathbf{x}), k_{t_2}(\cdot, \mathbf{x}')\rangle_{\mathcal{H}_k}$ for all $\mathbf{x}, \mathbf{x}'$, $\langle f, g\rangle_{\mathcal{H}_k} = 0$ as well.

Based on Moore-Aronszajn Theorem (Aronszajn, 1950; Berlinet & Thomas-Agnan, 2011), the RKHS spaces $\mathcal{H}_{k_{t_1}}$ and $\mathcal{H}_{k_{t_1}}$ are the set of functions which are pointwise limits of Cauchy sequences in the form $f_n(\mathbf{x}) = \sum_s a_s k_{t_1}(\mathbf{x}, \mathbf{x}^s_{t_1})$ and $g_n(\mathbf{x}) = \sum_s b_s k_{t_2}(\mathbf{x}, \mathbf{x}^s_{t_2})$, respectively. Moreover, based on the Berlinet & Thomas-Agnan (2011, Lemma 5), the inner product of two pointwisely convergent Cauchy sequences also converges. We conclude that for any $f \in \mathcal{H}_{k_{t_1}}, g \in \mathcal{H}_{k_{t_2}}$, $\langle f, g\rangle_{\mathcal{H}_k} = 0$. □

Proof of Proposition 3.7. Without loss of generality, we only need to prove that,

$$\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\},$$

Firstly, $0 \in \mathcal{H}_1, 0 \in \mathcal{H}_2$ because $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces. Then we assume another function $f \neq 0$ and $f \in \mathcal{H}_1 \cap \mathcal{H}_2$. By Lemma 3.6, $\|f\|_{\mathcal{H}_k} = \langle f, f\rangle_{\mathcal{H}_k} = 0$, which is contradictory to $f \neq 0$ and $\mathcal{H}_k$ being a Hilbert space. □

Proof of Theorem 3.8. We use $\mathcal{H}_t$ to represent the RKHS corresponding to the kernel k_t. Given a function $f \in \mathcal{H}_k$, we firstly assume f can be written as a linear combination of the kernel functions,

$$f(\mathbf{x}) = \sum_s a_s k(\mathbf{x}, \mathbf{x}^s),$$

Based on the kernel sum decomposition, we can rewrite f,

$$f(\mathbf{x}) = \sum_s a_s \sum_{t=0}^{T-1} k_t(\mathbf{x}, \mathbf{x}^s) = \sum_{t=0}^{T-1} \underbrace{\sum_s a_s k_t(\mathbf{x}, \mathbf{x}^s)}_{:=f_t(\mathbf{x})},$$

Because f_t is a linear combination of k_t, $f_t \in \mathcal{H}_t$, for $t = 0, ..., T-1$. Proposition 3.7 states that the RKHSs $\mathcal{H}_{t_1}, \mathcal{H}_{t_2}$ are disjoint except the zero function, thus $f = \sum_{t=0}^{T-1} f_t$ is a unique expansion of f to these RKHSs. Moreover, we can represent the function f alternatively,

$$f(\mathbf{x}) = \langle f, k(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k} = \sum_{t=0}^{T-1} \langle f, k_t(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k} = \sum_{t=0}^{T-1} \langle \sum_{t'=0}^{T-1} f_{t'}, k_t(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k} = \sum_{t=0}^{T-1} \langle f_t, k_t(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k},$$

where the last equality uses the orthogonality between RKHSs. By using the orthogonality again, we also show that,

$$\langle f_t, k_t(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k} = \langle f_t, \sum_{t'=0}^{T_1} k_{t'}(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k} = \langle f_t, k(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k} = f_t(\mathbf{x}), \tag{43}$$

Therefore, $f(\mathbf{x}) = \sum_{t=0}^{T-1} \langle f, k_t(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k}$ uniquely separates f into these RKHSs. More generally, if f is the pointwise limits of Cauchy sequences of functions in the form of linear combinations of the kernel function. Based on the Berlinet &

Thomas-Agnan (2011, Lemma 5), the inner product of two pointwisely convergent Cauchy sequences also converges. We conclude that for any $f \in \mathcal{H}_k$,

$$f(\mathbf{x}) = \sum_{t=0}^{T-1} \langle f, k_t(\mathbf{x}, \cdot)\rangle_{\mathcal{H}_k}, \tag{44}$$

uniquely decomposes the function f into the RKHSs $\mathcal{H}_t, t = 0, ..., T-1$.

Based on Berlinet & Thomas-Agnan (2011, Theorem 5), the squared RKHS norm of f can be written as the sum of squared RKHS norms,

$$\|f\|_{\mathcal{H}_k}^2 = \sum_{t=0}^{T-1} \|f_t\|_{\mathcal{H}_t}^2.$$

□

D.3. Proof of Sec 4

Proof of Theorem 4.1. Under the HVGP formulation, the inducing variable $u_t = f_t(\mathbf{z})$. From Lemma D.3, we have $f_t(\mathbf{z}) = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f(G^s(\mathbf{z}))$.

Under the inter-domain formulation, let w_t be the inter-domain inducing point corresponding to $\mathbf{z}$ in k_t,

$$w_t = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H \delta_{G^s(\mathbf{z})},$$

Then the inducing variable corresponding to w_t is,

$$u_{w_t} = \int f(\mathbf{x}) w_t(\mathbf{x}) d\mathbf{x} = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H f(G^s(\mathbf{z})), \tag{45}$$

Therefore, the two inducing variables are the same,

$$u_{w_t} = u_t. \tag{46}$$

Therefore, the variational posterior under the harmonic formulation can be rewritten in an inter-domain SVGP form,

$$q^{inter}(f, \mathbf{U}) = p(f|\mathbf{U}; \{\mathbf{w}_t\}_{t=0}^{T-1}) q(\text{vec}(\mathbf{U})), \tag{47}$$

where $\mathbf{U} := [\mathbf{u}_0, ..., \mathbf{u}_{T-1}]^\top$ and p is the inter-domain Gaussian process.

Now we connect the inter-domain SVGP to the standard SVGP using $\{G^t(\mathbf{Z})\}_{t=0}^{T-1}$. For the standard SVGP, the inducing variables are $\mathbf{v}_t = f(G^t(\mathbf{Z}))$, and $\mathbf{V} := [\mathbf{v}_0, ..., \mathbf{v}_{T-1}]^\top \in \mathbb{C}^{T \times m}$. For the inter-domain SVGP, the inducing variables are $\mathbf{u}_t$. As shown in Eq. (45), $\mathbf{u}_t = \sum_{s=0}^{T-1} \mathbf{F}_{t,s}^H \mathbf{v}_s$, then we have the equality,

$$\mathbf{U} = \mathbf{F}^H \mathbf{V}, \tag{48}$$

Because of the bijective linearity,

$$p(f|\mathbf{U}; \{\mathbf{w}_t\}_{t=0}^{T-1}) = p(f|\mathbf{V}; \{G^t(\mathbf{Z})\}_{t=0}^{T-1}), \tag{49}$$

Furthermore, the variational posterior for $\mathbf{V}$ is $\mathcal{N}(\text{vec}(\mathbf{V})|\text{vec}(\mathbf{M}_v), \mathbf{S}_v)$, which is equivalent to the variational posterior for $\mathbf{U}$,

$$q(\text{vec}(\mathbf{U})) = \mathcal{N}(\text{vec}(\mathbf{U})|\text{vec}(\mathbf{F}^H\mathbf{M}_v), (\mathbf{I} \otimes \mathbf{F}^H)\mathbf{S}_v(\mathbf{I} \otimes \mathbf{F})).$$

So the argument has been proved.

□

Lemma D.4 (Variational Gaussian Approximations)**.** *Let $\mathcal{N}(\boldsymbol{\mu}, \mathbf{S})$ be a Gaussian variational posterior for a SVGP, then the optimal $\mathbf{S}^\star$ is in the form of,*

$$\mathbf{S}^\star = \mathbf{K_{uu}}\left(\mathbf{K_{uu}} + \mathbf{K_{uf}}\boldsymbol{\Lambda}\mathbf{K_{fu}}\right)^{-1}\mathbf{K_{uu}}. \tag{50}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}([\lambda_n]_{n=1}^N)$ is diagonal,

$$\lambda_n = -2\nabla_{\sigma_n^2}\mathbb{E}_{q(f_n)}[\log p(y_n|f_n)], \tag{51}$$

where σ_n^2 is the predictive variance of f_n under the variational posterior.

Proof. Given the variational posterior, the predictive distribution of f_n can be computed as,

$$\mathcal{N}(\mathbf{k}_{f\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}\boldsymbol{\mu}, k_{ff} + \mathbf{k}_{f\mathbf{u}}\mathbf{K}_{\mathbf{uu}}^{-1}(\mathbf{S} - \mathbf{K_{uu}})\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f}),$$

where we denote the predictive variance as σ_n^2. The variational posterior is optimized by maximizing the ELBO, which can be computed as,

$$\mathcal{L} = \sum_{n=1}^N \mathbb{E}_{q(f_n)}[\log p(y_n|f_n)] - \mathrm{KL}\left(\mathcal{N}(\boldsymbol{\mu}, \mathbf{S})||\mathcal{N}(\mathbf{0}, \mathbf{K_{uu}})\right),$$

We compute the derivatives of $\mathcal{L}$ towards $\mathbf{S}$,

$$\begin{aligned}\nabla_{\mathbf{S}}\mathcal{L} &= -\frac{1}{2}\sum_{n=1}^N \lambda_n \nabla_{\mathbf{S}}\sigma_n^2 - \frac{1}{2}(\mathbf{K}_{\mathbf{uu}}^{-1} - \mathbf{S}^{-1}) = -\frac{1}{2}\mathbf{K}_{\mathbf{uu}}^{-1}(\sum_{n=1}^N \lambda_n \mathbf{k}_{\mathbf{u}f_n}\mathbf{k}_{f_n\mathbf{u}})\mathbf{K}_{\mathbf{uu}}^{-1} - \frac{1}{2}(\mathbf{K}_{\mathbf{uu}}^{-1} - \mathbf{S}^{-1}) \\ &= -\frac{1}{2}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{K_{uf}}\boldsymbol{\Lambda}\mathbf{K_{fu}}\mathbf{K}_{\mathbf{uu}}^{-1} - \frac{1}{2}\mathbf{K}_{\mathbf{uu}}^{-1} + \frac{1}{2}\mathbf{S}^{-1},\end{aligned} \tag{52}$$

Let the derivative be zero, we obtain the optimal $\mathbf{S}^\star$,

$$\mathbf{S}^\star = \mathbf{K_{uu}}(\mathbf{K_{uu}} + \mathbf{K_{uf}}\boldsymbol{\Lambda}\mathbf{K_{fu}})^{-1}\mathbf{K_{uu}}.$$

□

Proof of Theorem 4.2. Based on Lemma D.4, the optimal posterior covariance is

$$\mathbf{S}^\star = \mathbf{K_{uu}}\left(\mathbf{K_{uu}} + \mathbf{K_{uf}}\boldsymbol{\Lambda}\mathbf{K_{fu}}\right)^{-1}\mathbf{K_{uu}}, \tag{53}$$

Given that $\mathbf{K_{uu}}$ is block diagonal, by the continuous mapping theorem, it remains to prove that $\mathbf{K_{uf}}\boldsymbol{\Lambda}\mathbf{K_{fu}}$ approaches block diagonal.

Firstly we assume that Hermitian kernels are not resolved, thus $\mathbf{K_{fu}} = \mathbf{K}_{\mathbf{uf}}^H$. Because λ_n only depends on $(\mathbf{x}_n, y_n)$, for the $(\mathbf{z}_t, \mathbf{z}_{t'})$ off-diagonal element in $\mathbf{K_{uf}}\boldsymbol{\Lambda}\mathbf{K_{fu}}$,

$$\frac{1}{N}\sum_{n=1}^N \lambda_n k_t(\mathbf{z}_t, \mathbf{x}_n)k_{t'}^H(\mathbf{z}_{t'}, \mathbf{x}_n) \to \mathbb{E}_{p(\mathbf{x})p(y|\mathbf{x})}[\lambda(\mathbf{x}, y)k_t(\mathbf{z}_t, \mathbf{x})k_{t'}^H(\mathbf{z}_{t'}, \mathbf{x})] = \mathbb{E}_{p(\mathbf{x})}[\mathbb{E}_{p(y|\mathbf{x})}[\lambda(\mathbf{x}, y)]k_t(\mathbf{z}_t, \mathbf{x})k_{t'}^H(\mathbf{z}_{t'}, \mathbf{x})],$$

We let $\hat{\lambda}(\mathbf{x}) := \mathbb{E}_{p(y|\mathbf{x})}[\lambda(\mathbf{x}, y)]$, then the formula can be further computed as,

$$\begin{aligned}&\mathbb{E}_{p(\mathbf{x})}[\hat{\lambda}(\mathbf{x})k_t(\mathbf{z}_t, \mathbf{x})k_{t'}^H(\mathbf{z}_{t'}, \mathbf{x})] \\ &= \mathbb{E}_{p(\mathbf{x})}[\hat{\lambda}(\mathbf{x})\sum_{s=0}^{T-1}\sum_{s'=0}^{T-1}\mathbf{F}_{t,s}\mathbf{F}_{t',s'}^H k(\mathbf{x}, G^s(\mathbf{z}_t))k^H(\mathbf{x}, G^{s'}(\mathbf{z}_{t'}))] \\ &= \mathbb{E}_{p(\mathbf{x})}[\hat{\lambda}(\mathbf{x})\sum_{s=0}^{T-1}\sum_{s'=0}^{T-1}\mathbf{F}_{t,s}\mathbf{F}_{t',s+s'}^H k(\mathbf{x}, G^s(\mathbf{z}_t))k^H(\mathbf{x}, G^{s+s'}(\mathbf{z}_{t'}))] \\ &= \mathbb{E}_{p(\mathbf{x})}[\hat{\lambda}(G^s(\mathbf{x}))\sum_{s=0}^{T-1}\sum_{s'=0}^{T-1}\mathbf{F}_{t,s}\mathbf{F}_{t',s+s'}^H k(G^s(\mathbf{x}), G^s(\mathbf{z}_t))k^H(G^s(\mathbf{x}), G^{s+s'}(\mathbf{z}_{t'}))] \\ &= \mathbb{E}_{p(\mathbf{x})}[\hat{\lambda}(G^s(\mathbf{x}))k(\mathbf{x}, \mathbf{z}_t)\sum_{s'=0}^{T-1}k^H(\mathbf{x}, G^{s'}(\mathbf{z}_{t'}))\sum_{s=0}^{T-1}\mathbf{F}_{t,s}\mathbf{F}_{t',s+s'}^H] = 0,\end{aligned}$$

In the second equality we used the periodicity of G; In the third equality we used the assumption that $G^s(\mathbf{x})$ has the same distribution as $\mathbf{x}$; In the last equality we used the property that $\sum_{s=0}^{T-1} \mathbf{F}_{t,s}\mathbf{F}^H_{t',s+s'} = 0$ for all $t \neq t'$.

Furthermore, if the Hermitian kernels are resolved in HVGP, let T be the period, then $\mathbf{K_{uf}}\mathbf{\Lambda}\mathbf{K_{fu}}$ is a matrix of $(1 + \lfloor T/2 \rfloor) \times (1 + \lfloor T/2 \rfloor)$. For any off-diagonal element at (t, t'), $[\mathbf{K_{uf}}\mathbf{\Lambda}\mathbf{K_{fu}}]_{t,t'}$ equals to,

$$\frac{1}{N}\sum_{n=1}^{N} \lambda_n \left(k_t(\mathbf{z}_t, \mathbf{x}_n) + k_{T-t}(\mathbf{z}_t, \mathbf{x}_n)\right)\left(k^H_{t'}(\mathbf{z}_{t'}, \mathbf{x}_n) + k^H_{T-t'}(\mathbf{z}_{t'}, \mathbf{x}_n)\right)$$

Given previous results, because $t, T - t$ are both different with $t', T - t'$, the formula becomes 0 as well, as $N \to \infty$. $\square$