A Computational complexity

As discussed in the Experiments, we used the high-dimensional synthetic data to consider our method's computational complexity. The effect of the dimensionality d was negligible in practice, because the main calculations rely only on an $n \times n$ Gram matrix whose calculation is relatively fast even for high dimensions. Our method's time complexity scales as $\mathcal{O}(n^2)$ as shown in the Appendix in Fig. A5 (but as discussed in Section 4.2, a primal representation is available which would give linear scaling.) where we used $s = 200$ sample points to estimate $\tilde{k}$. While a small s worked well in practice, we investigated much larger values of s. As shown in the Appendix in Fig. A6 the time complexity scaled as $\mathcal{O}(s^2)$ where the number of points was fixed to be 150; note that we fixed the rank of the eigendecomposition to be 20.

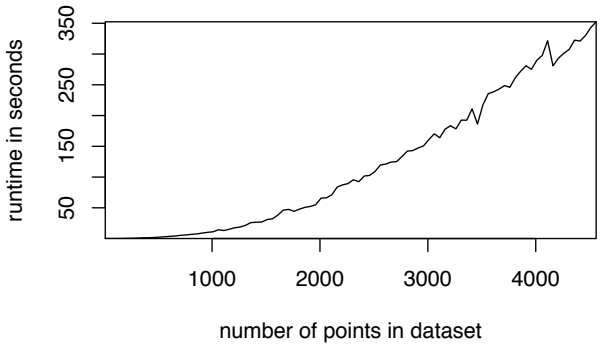

Figure A5: Run-time of our method versus number of points in the point pattern dataset.

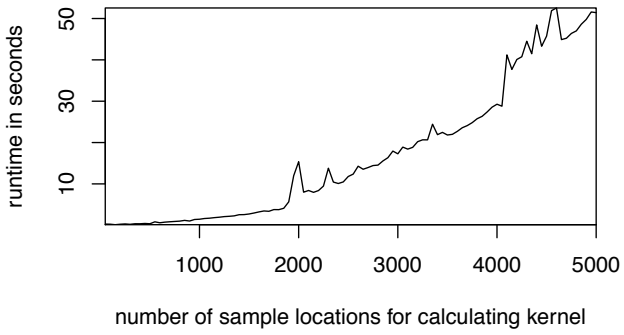

Figure A6: Run-time of our method versus number of sample points used to calculate $\tilde{k}$.

B Supplementary results

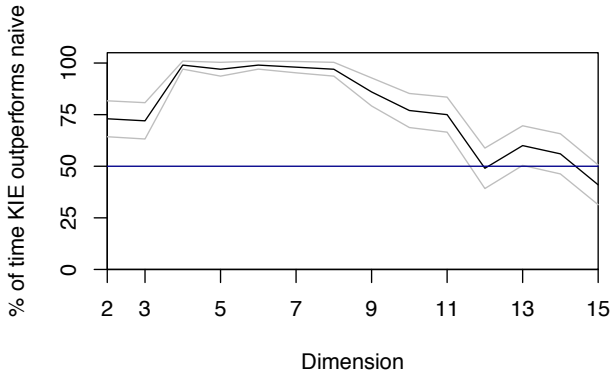

Figure A7: Comparison of kernel intensity estimation versus the naïve RKHS approach, as the dimensionality grows, with 95% confidence intervals shown in gray based on 100 random intensities for each dimension. The number of points in the point pattern was fixed to be between 190 and 210. Kernel intensity estimation almost always outperforms the naïve approach.

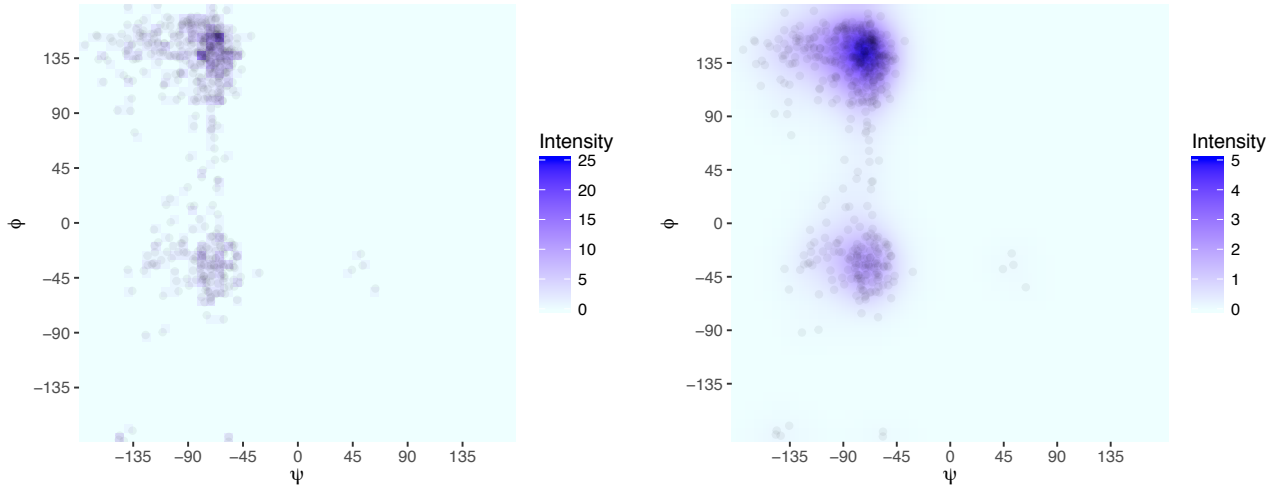

Figure A8: Standard Ramachandran plot (left) based on a two-dimensional histogram versus our proposed Ramachandran plot based on an intensity estimate with a two-dimensional Sobolev kernel

C Kernels with Explicit Mercer Expansions

C.1 Sobolev space on $[0,1]$ with a periodic boundary condition

We consider domain $S = [0,1]$. The kernel is given by:

$$\begin{aligned} k(x,y) &= 1 + \sum_{m=1}^{\infty} \frac{2\cos\left(2\pi m\left(x-y\right)\right)}{(2\pi m)^{2s}} \\ &= 1 + \sum_{m=1}^{\infty} \frac{2}{(2\pi m)^{2s}} \left[\cos\left(2\pi mx\right)\cos\left(2\pi my\right) + \sin\left(2\pi mx\right)\sin\left(2\pi my\right)\right], \\ &= 1 + \frac{(-1)^{s-1}}{(2s)!} B_{2s}(\{x-y\}), \end{aligned}$$

where $s = 1, 2, \ldots$ denotes the order of the Sobolev space and $B_{2s}(\{x-y\})$ is the Bernoulli polynomial of degree $2s$ applied to the fractional part of $x-y$. The corresponding RKHS is the space of functions on $[0,1]$ with absolutely continuous $f, f', \ldots, f^{(s-1)}$ and square integrable $f^{(s)}$ satisfying a periodic boundary condition $f^{(l)}(0) = f^{(l)}(1)$, $l = 0, \ldots, s-1$. For more details, see (Wahba, 1990, Chapter 2) Bernoulli polynomials admit a simple form for low degrees. In particular,

$$\begin{aligned} B_2(t) &= t^2 - t + \frac{1}{6}, \\ B_4(t) &= t^4 - 2t^3 + t^2 - \frac{1}{30}, \\ B_6(t) &= t^6 - 3t^5 + \frac{5}{2}t^4 - \frac{1}{2}t^2 + \frac{1}{42}. \end{aligned}$$

If we consider the Mercer expansion where the underlying measure ρ is uniform on $[0,1]$: $d\rho(x) = dx$, we have

$$\begin{aligned} \int_0^1 2\cos\left(2\pi mx\right)\sin\left(2\pi m'x\right)dx &= 0 \\ \int_0^1 2\cos\left(2\pi mx\right)\cos\left(2\pi m'x\right)dx &= \delta(m-m') \\ \int_0^1 2\sin\left(2\pi mx\right)\sin\left(2\pi m'x\right)dx &= \delta(m-m'). \end{aligned}$$

Thus, the desired Mercer expansion $k(x,y) = \sum_{m\in\mathbb{Z}} \eta_m e_m(x) e_m(y)$ has eigenfunctions $e_0(x) = 1$ and for $m = \{1, 2, \ldots\}$, $e_m(x) = \sqrt{2}\cos\left(2\pi mx\right)$, $e_{-m}(x) = \sqrt{2}\sin\left(2\pi mx\right)$ and corresponding eigenvalues $\eta_0 = 1$, $\eta_m = \eta_{-m} = (2\pi m)^{-2s}$.

- $\tilde{k}(x,y)$ is the kernel of $T_k(T_k + cI)^{-1}$ and has form

$$\begin{aligned} \tilde{k}(x,y) &= \sum_{m\in\mathbb{Z}} \frac{\eta_m}{\eta_m + c} e_m(x) e_m(y) \\ &= \frac{1}{1+c} + \sum_{m=1}^{\infty} \frac{2\cos\left(2\pi m\left(x-y\right)\right)}{1 + c(2\pi m)^{2s}} \end{aligned}$$

- To compute $\int_0^1 f^2(x)dx$ for $f = \sum_i \alpha_i \tilde{k}(\cdot, x_i)$ we have

$$\begin{aligned} \int_0^1 f^2(x)dx &= \sum_{i,j} \alpha_i\alpha_j \int_0^1 \tilde{k}(x_i,u)\tilde{k}(u,x_j)du \\ &= \sum_{i,j} \alpha_i\alpha_j \sum_{m,m'} \frac{\eta_m\eta'_m}{(\eta_m + c)(\eta_{m'} + c)} e_m(x_i) e_{m'}(x_j) \int_0^1 e_m(u) e_{m'}(u)du \\ &= \sum_{i,j} \alpha_i\alpha_j \sum_{m} \frac{\eta_m^2}{(\eta_m + c)^2} e_m(x_i) e_m(x_j) \\ &= \alpha^\top \tilde{R}\alpha, \end{aligned}$$

where kernel matrix $\tilde{R}$ is computed using kernel $\tilde{r}$ of $T_k^2(T_k+cI)^{-2}$, i.e.

$$\begin{aligned}\tilde{r}(x,y) &= \sum_{m\in\mathbb{Z}} \frac{\eta_m^2}{(\eta_m+c)^2} e_m(x)e_m(y) \\ &= \frac{1}{(1+c)^2} + \sum_{m=1}^{\infty} \frac{2\cos\left(2\pi m\left(x-y\right)\right)}{(1+c(2\pi m)^{2s})^2}.\end{aligned}$$

- To generate a function $f \in \mathcal{H}_k$ of unit norm $\|f\|_{\mathcal{H}_k} = 1$, one takes

$$f(x) = a_0 + \sqrt{2}\sum_{m=1}^{M}\left(a_m\cos(2\pi mx) + a_{-m}\sin(2\pi mx)\right), \tag{23}$$

for which the norm is given by

$$\|f\|_{\mathcal{H}_k}^2 = a_0^2 + \sum_{m=1}^{M}(a_m^2 + a_{-m}^2)(2\pi m)^{2s}. \tag{24}$$

Thus we can simply generate $\mathbf{z} = (z_{-M},\ldots,z_0,\ldots,z_M) \sim \mathcal{N}(0, I_{2M+1})$, set $\tilde{\mathbf{z}} = \mathbf{z}/\|\mathbf{z}\|$ and then $a_0 = \tilde{z}_0$, $a_m = \tilde{z}_m(2\pi|m|)^{-s}$, for $m \neq 0$.

C.2 Squared exponential kernel

A Mercer expansion for the squared exponential kernel was proposed in Zhu et al. (1997) and refined in Fasshauer and McCourt (2012). However, this expansion is with respect to a Gaussian measure on $\mathbb{R}$, i.e., it consists of eigenfunctions which form an orthonormal set in $L^2(\mathbb{R},\nu)$ where $\nu = \mathcal{N}(0,\ell^2 I)$. The formalism can therefore be used to estimate Poisson intensity functions with respect to such Gaussian measure. In the classical framework, where the intensity is with respect to a Lebesgue measure, numerical approximations of Mercer expansion, as described in Section 4.2 are needed. Following the exposition in (Rasmussen and Williams, 2006, section 4.3.1) and the relevant errata[3] we parameterize the kernel as:

$$k(x,x') = \exp(-\frac{\|x-x'\|^2}{2\sigma^2}) \tag{25}$$

The Mercer expansion with respect to $\nu = \mathcal{N}(0,\ell^2 I)$ then has the following eigenvalues:

$$\eta_i = \sqrt{\frac{2a}{A}}B^i \tag{26}$$

And eigenfunctions:

$$e_i(x) = \frac{1}{\sqrt{\sqrt{a/c}\;2^i i!}}\exp(-(c-a)x^2)H_i(\sqrt{2c}x) \tag{27}$$

where H_i is the i-th order (physicist's) Hermite polynomial, $a = \frac{1}{4\sigma^2}$, $b = \frac{1}{2\ell^2}$, $c = \sqrt{a^2+2ab}$, $A = a+b+c$, and $B = b/A$. Thus we have the following eigenvalues for $\tilde{k}$:

$$\widetilde{\eta}_i = \frac{\eta_i}{a\eta_i+\gamma} = \frac{1}{a+\gamma\sqrt{\frac{A}{2a}}B^{-i}} \tag{28}$$

while the eigenfunctions remain the same.

C.3 Brownian Bridge kernel

This is the kernel on $[0,1]$, given by

$$k(x,y) = \min(x,y) - xy = \sum_{m=1}^{\infty}\frac{2\sin(\pi mx)\sin(\pi my)}{\pi^2 m^2},$$

with the eigenvalues and eigenfunctions in the Mercer expansion with respect to Lebesgue measure

$$\eta_m = \frac{1}{\pi^2 m^2},\; e_m(x) = \sqrt{2}\sin\left(\pi mx\right), \quad m = 1,2,\ldots. \tag{29}$$

[3] http://www.gaussianprocess.org/gpml/errata.html

Thus one can form

$$\begin{aligned}\tilde{k}(x,y) &= \sum_{m=1}^{\infty} \frac{\eta_m}{\eta_m + c} e_m(x) e_m(y) \\ &= \sum_{m=1}^{\infty} \frac{2\sin(\pi m x)\sin(\pi m y)}{1 + c\pi^2 m^2}.\end{aligned}$$

The functions in the corresponding RKHS are pinned to zero at both ends of the segment.

C.4 Extending the Mercer expansion to multiple dimensions

The extension of any kernel to higher dimensions can be constructed by considering tensor product spaces: $\mathcal{H}_{k_1 \otimes k_2}$ (where k_1 and k_2 could potentially be different kernels with different hyperparameters). If k_1 has eigenvalues η_i and eigenfunctions e_i and k_2 has eigenvalues δ_j and eigenfunctions f_j, then the eigenvalues of the product space are then given by the Cartesian product $\eta_i \delta_j, \forall i, j$, and similarly the eigenfunctions are given by $e_i(x) f_j(y)$. Our regularized kernel has the following Mercer expansion:

$$\widetilde{k_1 \otimes k_2}((x,y),(x',y')) = \sum_{ij} \frac{\eta_i \delta_j}{a\eta_i \delta_j + \gamma} e_i(x) e_i(x') f_j(y) f_j(y') \tag{30}$$

Notice that $\widetilde{k_1 \otimes k_2}$ is the kernel corresponding to the integral operator $(T_{k_1} \otimes T_{k_2})(aT_{k_1} \otimes T_{k_2} + \gamma I)^{-1}$ which is different than $\tilde{k}_1 \otimes \tilde{k}_2$.

D Proof of the Representer Theorem

We decompose $f \in \mathcal{H}_{\tilde{k}}$ as the sum of two functions:

$$f(\cdot) = \sum_{j=1}^{N} \alpha_j \tilde{k}(x_j, \cdot) + v \tag{31}$$

where v is orthogonal to the span of $\{\tilde{k}(x_j, \cdot)\}_j$. We prove that the first term in the objective $J[f]$ given in Eq. (13), $-\sum_{i=1}^{N} \log(af^2(x_i))$, is independent of v. It depends on f only through the evaluations $f(x_i)$ for all i. Using the reproducing property we have:

$$f(x_i) = \langle f, \tilde{k}(x_i, \cdot)\rangle = \sum_j \alpha_j \tilde{k}(x_j, x_i) + \langle v, \tilde{k}(x_i, \cdot)\rangle = \sum_j \alpha_j \tilde{k}(x_j, x_i) \tag{32}$$

where the last step is by orthogonality. Next we substitute into the regularization term:

$$\gamma \| \sum_j \alpha_j \tilde{k}(x_j, \cdot) + v\|^2_{\mathcal{H}_{\tilde{k}}} = \gamma \| \sum_j \alpha_j \tilde{k}(x_j, \cdot)\|^2_{\mathcal{H}_{\tilde{k}}} + \|v\|^2_{\mathcal{H}_{\tilde{k}}} \geq \gamma \| \sum_j \alpha_j \tilde{k}(x_j, \cdot)\|^2_{\mathcal{H}_{\tilde{k}}}. \tag{33}$$

Thus, the choice of v has no effect on the first term in $J[f]$ and a non-zero v can only increase the second term $\|f\|^2_{\mathcal{H}_{\tilde{k}}}$, so we conclude that $v = 0$ and that $f^* = \sum_{j=1}^{N} \alpha_j \tilde{k}(x_j, \cdot)$ is the minimizer.

E Numerical evaluation of kernel approximations

Here we present an evaluation of the numerical approximation to $\tilde{k}$ described in 4.2 on the case of the Sobolev kernel where Mercer expansion is also available so that truncated Mercer expansion representation of $\tilde{k}$ can be treated as a ground truth. As Figure A9, demonstrates, good approximation is possible with a fairly coarse grid $\mathbf{u} = (u_1, \ldots, u_m)$ as well as with a low-rank approximation.

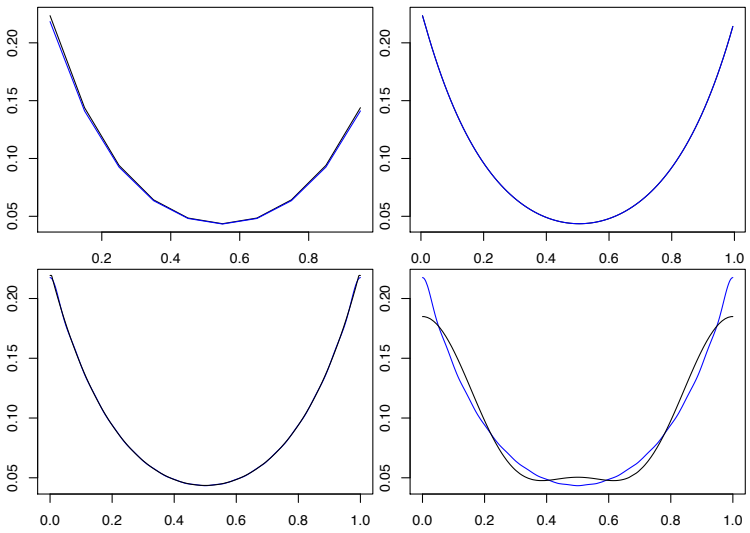

Figure A9: We compared the exact calculation of $\tilde{K}_{\mathbf{uu}}$ with $s = 1$, $a = 10$, and $\gamma = .5$ to our approximate calculation. For illustration we tried a coarse grid of size 10 on the unit interval (top left) to a finer grid of size 100 (top right). The RMSE was 2E-3 for the coarse grid and 1.6E-5 for the fine grid. We compare the exact calculation of $\tilde{K}_{\mathbf{xx}}$ with $s = 1$, $a = 10$, and $\gamma = .5$ to our Nyström-based approximation, where $x_1, \ldots, x_{400} \sim$ Beta$(.5, .5)$ distribution (bottom left). The RMSE was 0.98E-3. A low-rank approximation using only the top 5 eigenvalues gives the RMSE of 1.6E-2 (bottom right).