Actually Sparse Variational Gaussian Processes

Jake Cunningham, So Takao, Mark van der Wilk†, Marc Deisenroth

University College London, Imperial College London†

Imperial College London

UCL

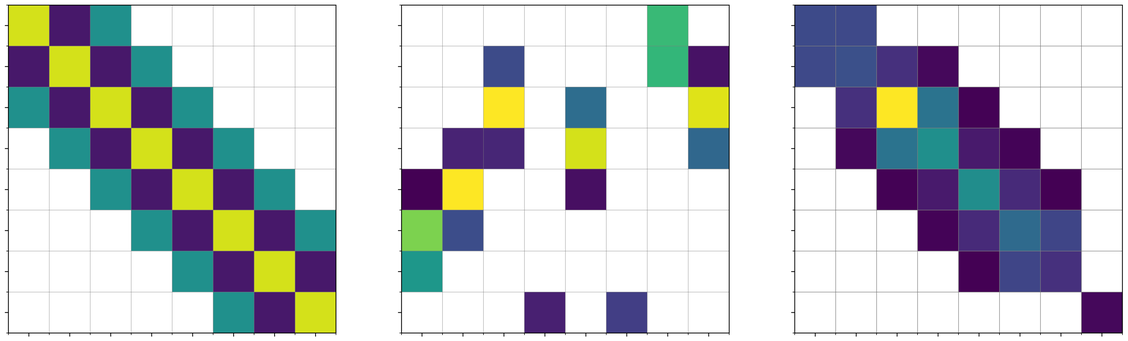
Contribution

- In sparse variational GP inference, **precomputation** (as done in VFF/VISH) is faster than **stochastic optimization** in SVGP
- However, computing the ELBO with VFF requires the **Cholesky factor** of the **dense matrix** $\left(\mathbf{K_{uu}} - \sigma^{-2}\mathbf{K_{uf}}\mathbf{K_{fu}}\right)$
- We overcome this by using **B-spline features** to make this matrix **sparse,** reducing the computational complexity to **linear in the number of inducing points**

Algorithm	Pre-computation	Computational complexity	Storage
SGPR (Titsias, 2009)	✗	$\mathcal{O}(NM^2 + M^3)$	$\mathcal{O}(NM)$
SVGP (Hensman et al, 2013)	✗	$\mathcal{O}(N_bM^2 + M^3)$	$\mathcal{O}(M^2 + N_bM)$
VFF (Hensman et al, 2017)	$\mathcal{O}(NM^2)$	$\mathcal{O}(M^3)$	$\mathcal{O}(M^2 + NM)$
VISH (Dutordoir et al, 2020)	$\mathcal{O}(NM^2)$	$\mathcal{O}(M^3)$	$\mathcal{O}(M^2 + NM)$
AS-VGP (Ours)	$\mathcal{O}(N)$	$\mathcal{O}(Mk^2)$	$\mathcal{O}(Mk)$

Sparse Matrices

$\mathbf{K_{uu}} = [\langle \phi_i, \phi_j \rangle_{\mathcal{H}}]_{i,j=1}^{M}$ $\mathbf{K_{uf}} = [\phi_i(x_j)]_{i=1,j=1}^{M,N}$ $\mathbf{K_{uf}}\mathbf{K_{fu}}$



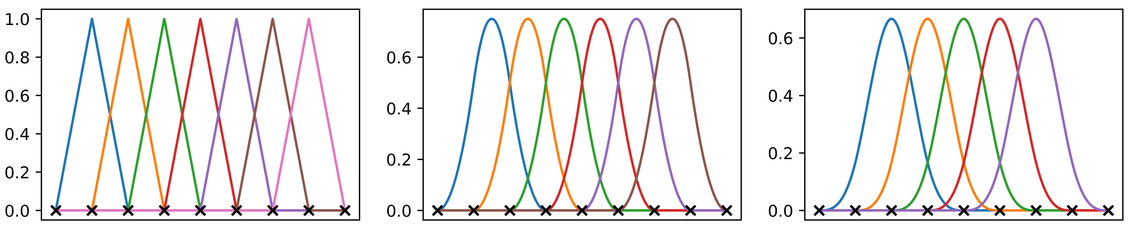
- Projecting onto B-spline basis, $\left(\mathbf{K_{uu}} - \sigma^{-2}\mathbf{K_{uf}}\mathbf{K_{fu}}\right)$ is a **band diagonal matrix**

B-Spline Inducing Features

- Construct inter-domain inducing variables by projecting GP onto a set of compactly supported basis functions

$$u_m = \langle f, \phi_m \rangle_{\mathcal{H}}$$

- ϕ_m are B-Spline basis functions of order k



$k = 1$ $k = 2$ $k = 3$

Inference

- Cholesky factor of a band diagonal matrix computed in $\mathcal{O}(Mk^2)$ where k is the bandwidth of the matrix

Fit a GP with $N = 2,049,279$ and $M = 1000$ in under 4 seconds!

Method	$M = 1000$	$M = 10,000$	$M = 25,000$
AS-VGP (MSE $\times 10^{-1}$)	8.67	4.53	2.94
SVGP (MSE $\times 10^{-1}$)	8.98	/	/
AS-VGP (NLPD)	1.35	1.04	0.86
SVGP (NLPD)	1.37	/	/
AS-VGP (Time in s)	3.90	18.62	51.69
SVGP (Time in s)	932	/	/

Correspondence to *jake.cunningham.21@ucl.ac.uk*