Imperial College London

Actually Sparse Variational Gaussian Processes

Harry Jake Cunningham[1] Daniel Augusto de Souza[1] So Takao[1] Mark van der Wilk[2] Marc Peter Deisenroth[1]

University College London[1], Imperial College London[2]

Motivation

- Sparse variational Gaussian processes (GPs) approximate the GP posterior with a variational distribution conditioned on a set of inducing points
- In practice however, for large datasets with low lengthscales even sparse GPs can become computationally expensive, limited by the number of inducing variables one can use
- Inter-domain inducing variables condition the approximate posterior on linear transformations of the true GP to construct efficient matrix structures

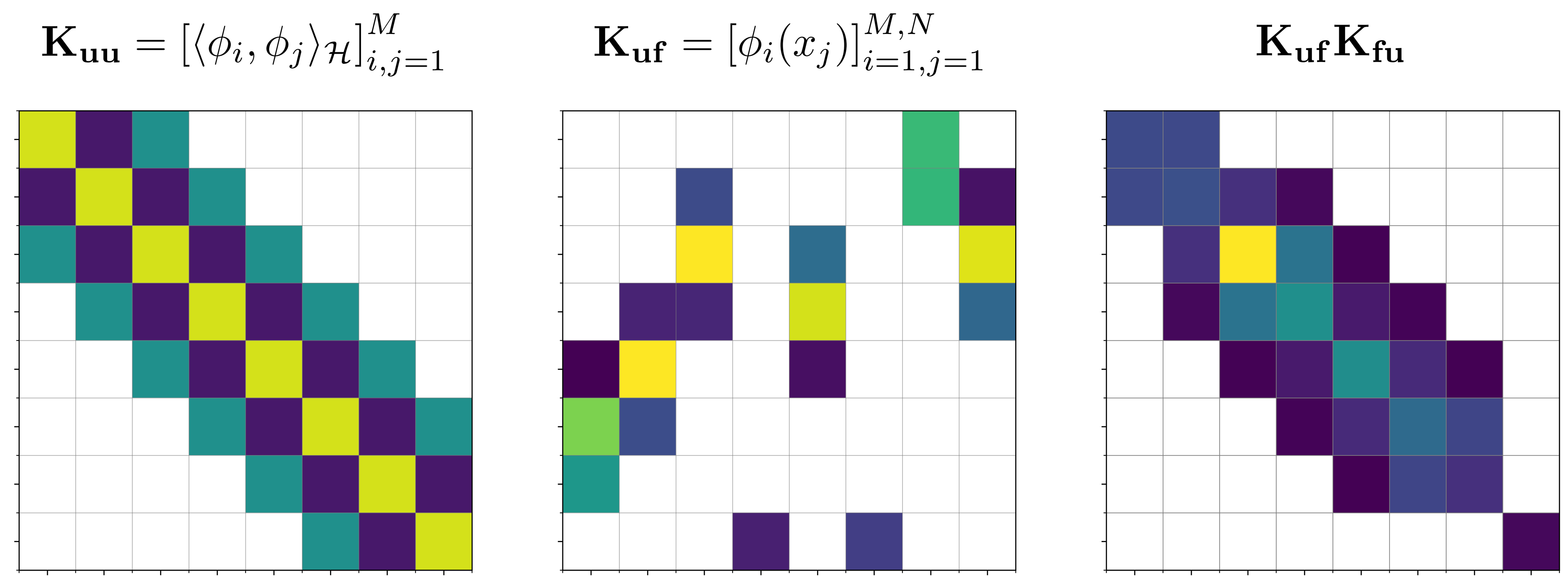
Sparse Linear Algebra

By projecting onto a B-spline basis $(\mathbf{K_{uu}} - \sigma^{-2}\mathbf{K_{uf}}\mathbf{K_{fu}})$ is a band-diagonal matrix

1) **One-time sparse** matrix product $\mathbf{K_{uf}}\mathbf{K_{fu}}$
2) **Band-diagonal** Cholesky of $\mathbf{K_{uu}}$
3) **Band-diagonal** Cholesky of $(\mathbf{K_{uu}} - \sigma^{-2}\mathbf{K_{uf}}\mathbf{K_{fu}})$

Precomputation $\mathcal{O}(N)$

Optimization $\mathcal{O}(M(k+1)^2)$



Contributions

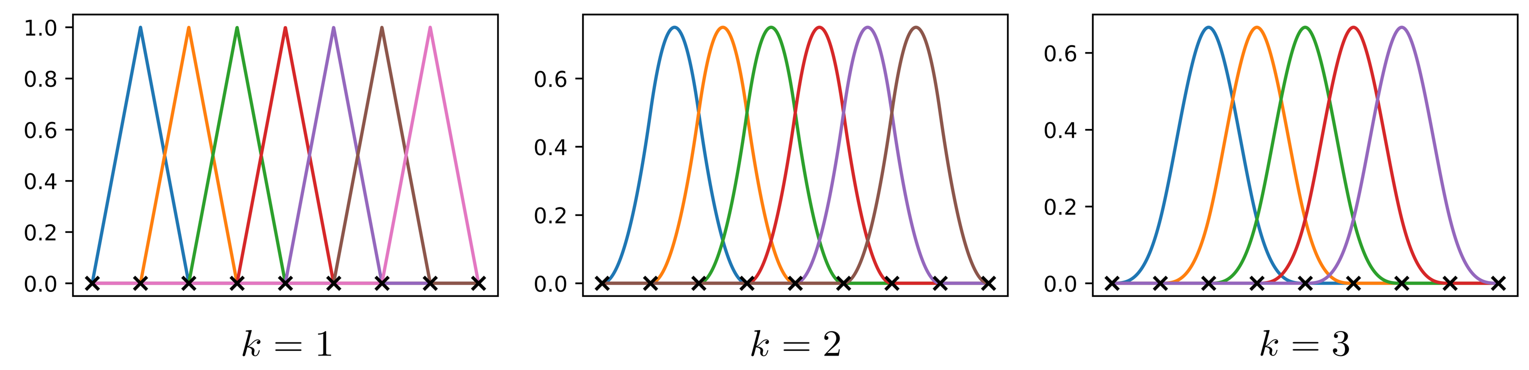
We propose Actually Sparse Variational Gaussian Processes (AS-VGP) that:

- Construct inter-domain inducing variables by projecting the GP onto a **compactly supported B-spline basis**
- Use banded-matrices to reduce per iteration computational complexity to **linear in the number of inducing points**
- Avoid ever having to instantiate a dense matrix reducing memory requirements to **linear in the number of data points**

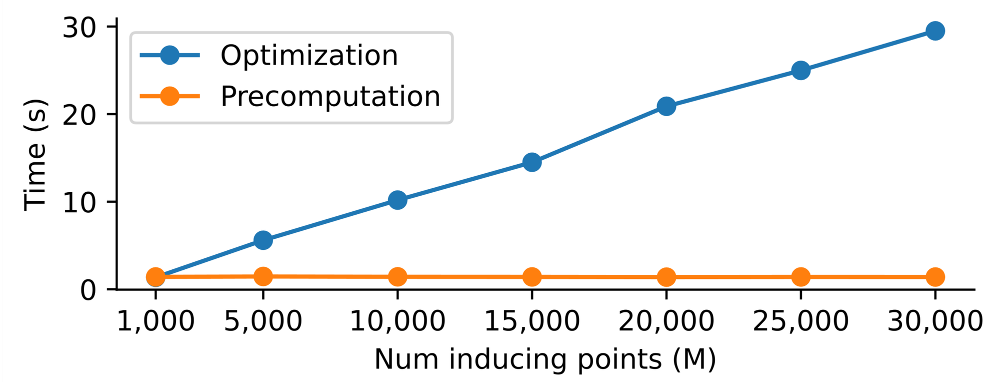
B-Spline Inducing Features

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

where ϕ_m are B-spline basis functions

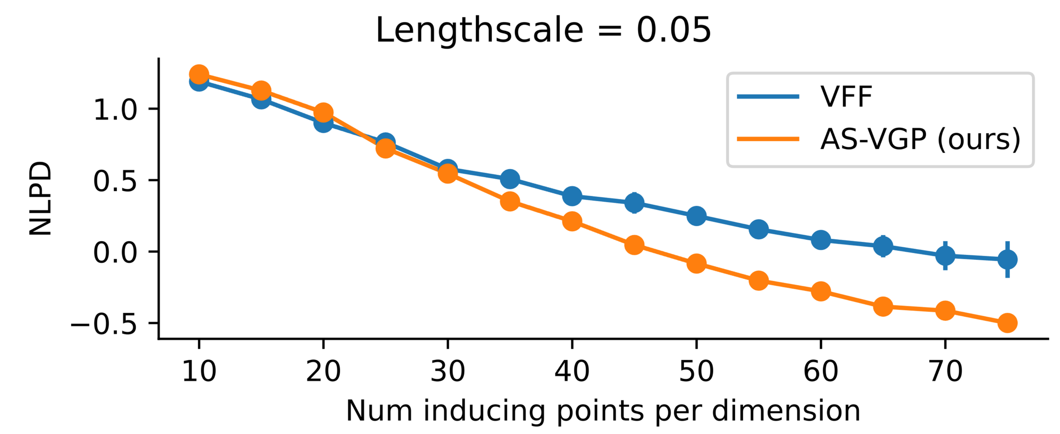


Linear Scaling



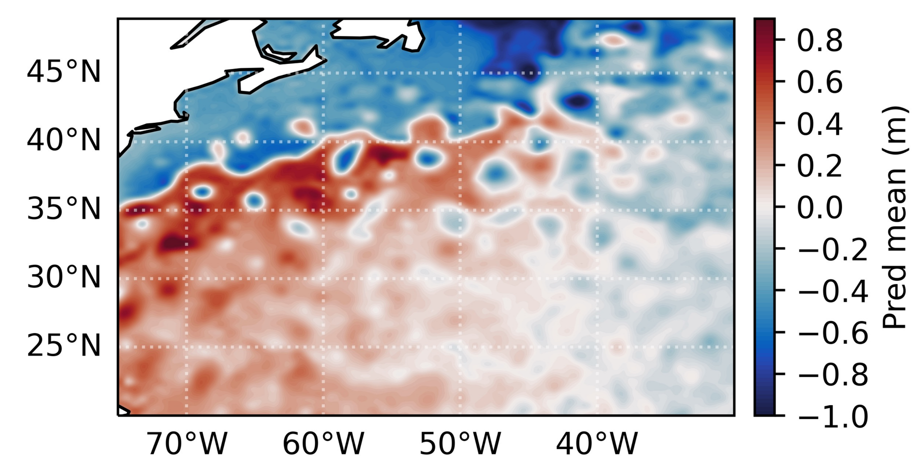
Precomputation of $\mathbf{K_{uf}}\mathbf{K_{fu}}$ is independent of number of inducing points and optimization scales linearly

Low Lengthscales



Our locally supported basis functions are better at modelling fast varying processes than globally supported ones

Spatial Data



AS-VGP is well suited to modelling low-dimensional problems with low lengthscales

Algorithm	Pre-computation	Computational complexity	Storage
SGPR (Titsias, 2009)	✗	$\mathcal{O}(NM^2 + M^3)$	$\mathcal{O}(NM)$
SVGP (Hensman et al, 2013)	✗	$\mathcal{O}(N_b M^2 + M^3)$	$\mathcal{O}(M^2 + N_b M)$
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)	$\boldsymbol{\mathcal{O}(N)}$	$\boldsymbol{\mathcal{O}(M(k+1)^2)}$	$\boldsymbol{\mathcal{O}(M(k+1)+N)}$

How We Compare

In low-dimensions AS-VGP is both faster and more memory efficient than prior inter-domain inducing point approximations

[1] Harry Jake Cunningham, Daniel Augusto de Souza, So Takao, Mark van der Wilk, Marc Deisenroth. Actually Sparse Variational Gaussian Processes. AISTATS, 2023.
[2] Michalis Titsias. Variational learning of inducing variables in sparse Gaussian processes. AISTATS, 2009.
[3] James Hensman, Nicol`o Fusi, and Neil D Lawrence. Gaussian processes for big data. AISTATS, 2013.
[4] James Hensman, Nicolas Durrande, and Arno Solin. Variational Fourier features for Gaussian processes. JMLR, 18(1):5537– 5588, 2017.
[5] Vincent Dutordoir, Nicolas Durrande, and James Hensman. Sparse Gaussian processes with spherical harmonic features. ICML, 2020.

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

https://github.com/HJakeCunningham/ASVGP

27th April 2023