Conditioning Variational Gaussian Processes for Scalable Online Decision-Making

Wesley Maddox, Sam Stanton, Andrew Gordon Wilson

Paper: https://arxiv.org/abs/2110.15172
Code: https://github.com/wjmaddox/online_vargp

Methods

Sparse Gaussian processes [1,2] approximate GP posteriors using inducing points, $\mathbf{u}$ and the variational parameters, $\mathbf{m_u}, S_\mathbf{u}$

Their optimal values are:

$$\mathbf{m_u} := K_{uu}(K_{uu} + C)^{-1}c \text{ and}$$
$$S_\mathbf{u} := K_{uu}(K_{uu} + C)^{-1}K_{uu}.$$

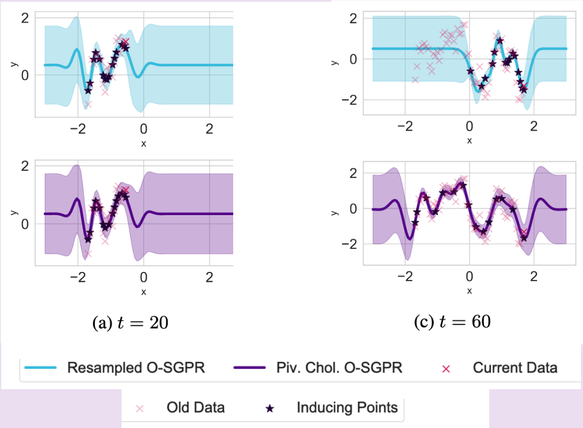
For: $c = K_{ux}\Sigma_x^{-1}y$ and $C = K_{ux}\Sigma_x^{-1}K_{xu}$

To **update the variational parameterization**, we use a recursive formula, simplifying [3]:

$$c^* = K_{ux^*}\Sigma_{x^*}^{-1}y^* + K_{uu'}K_{u'u'}^{-1}c \text{ and}$$
$$C^* = K_{ux^*}\Sigma_{x^*}^{-1}K_{x^*u} + K_{uu'}(K_{u'u'}CK_{u'u'})K_{u'u}$$

x^*, y^* is the new data and u' is the old inducing points.

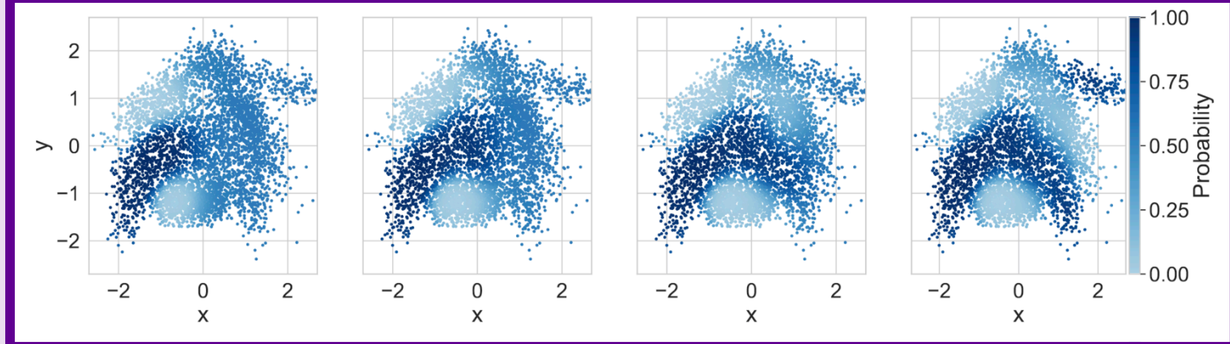
To **update the inducing points**, we use a pivoted cholesky decomposition on the old inducing points and the newly observed data.



References:
[1] Variational Learning of Inducing Variables in Sparse Gaussian Processes, Titsias, AISTATS, '09
[2] Gaussian Processes for Big Data, Hensman et al, UAI, '13
[3] Streaming Sparse Gaussian Process Approximations, Bui et al, NeurIPS, '17
[4]Trust Region Bayesian Optimization, Eriksson et al, NeurIPS, '19

NEW YORK UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS



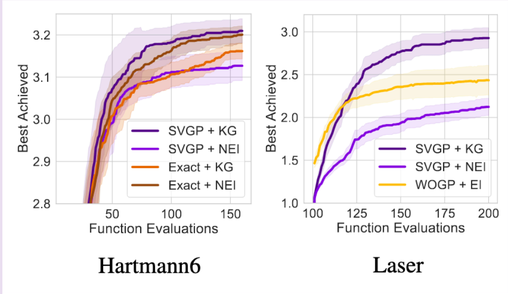
As the SVGP sees more data points, it becomes more confident about its predictions because the SVGP is updating itself.

Challenge: Stochastic variational GPs (SVGPs) scale well to many data points but cannot be updated efficiently on seeing a new data point.

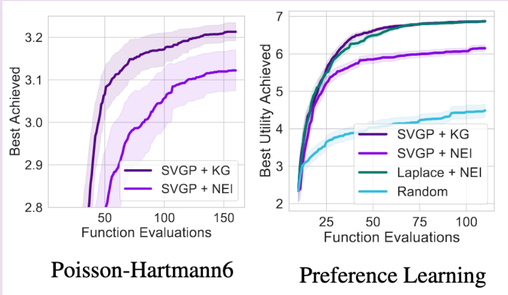
Approach: Online variational conditioning (OVC) which updates the inducing points and the variational distribution.

Enables SVGPs to be used for ***lookahead acquisition functions*** such as the knowledge gradient out of the box.

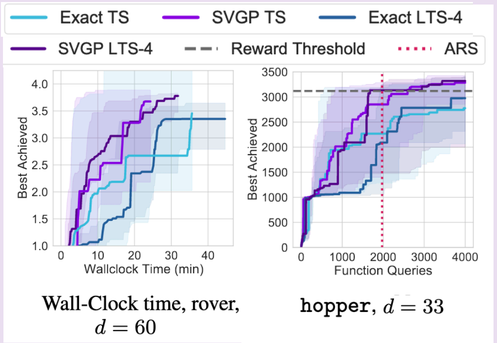
Experiments



Directly applying SVGPs with OVC enables using the Knowledge Gradient (KG), and produces competitive results on standard BO test problems.



We can also use SVGPs + KG on non-Gaussian responses via Laplace approximations.



For large scale problems, we use TrBO [3] with SVGPs and develop a lookahead version of Thompson sampling (LTS). SVGPs are faster than GPs, while reaching better rewards.