Structured Variational Inference for coupled Gaussian processes

Vincent Adam
Gatsby Computational Neuroscience Unit, UCL

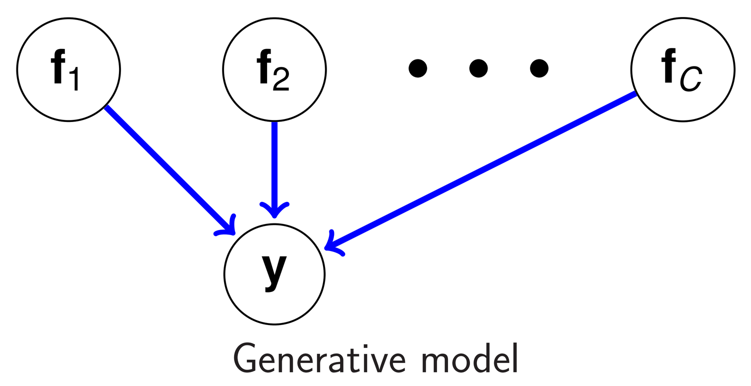
UCL

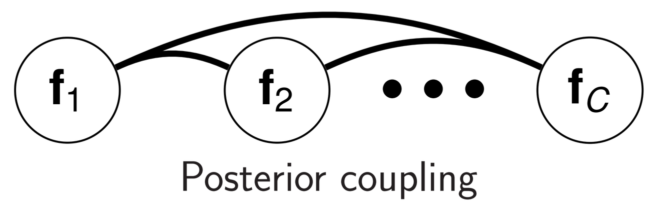
Motivation: Bayesian structured non-linear Regression

We consider a regression model where multiple functions $f_1, f_2, ..., f_C$ interact in the predictor through $\phi : \mathbb{R}^C \to \mathbb{R}$ and a factorized likelihood

$$p(y|f_{1..C}, X) = \prod_{i=1}^{N} p(y_i|\phi(f_1(X_i), ..., f_C(X_i))$$

We use GPs as priors. These GPs are a posteriori coupled.





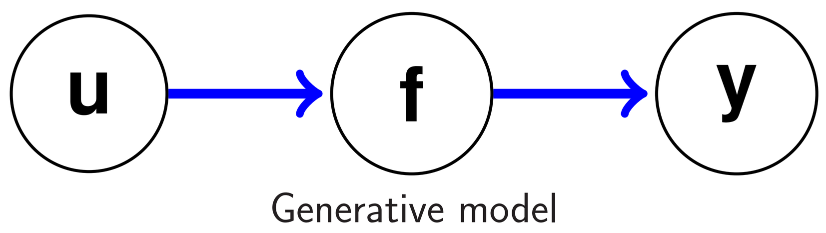
Desiderata for an approximate inference algorithm

- scalable
- represent posterior dependencies
- applies to arbitrary predictor/likelihood

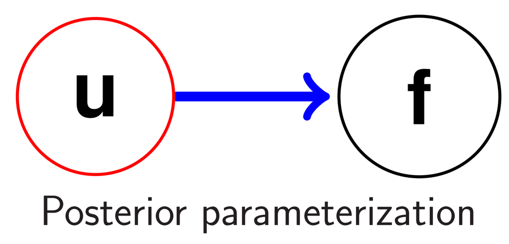
Background: Sparse Variational GP Regression [1]

Single GP Regression

- GP prior $p(f) \sim \mathcal{GP}(0, k)$
- Likelihood $p(y|f, X)$.
- Pseudo point: $u = f(Z)$



Sparse variational posterior



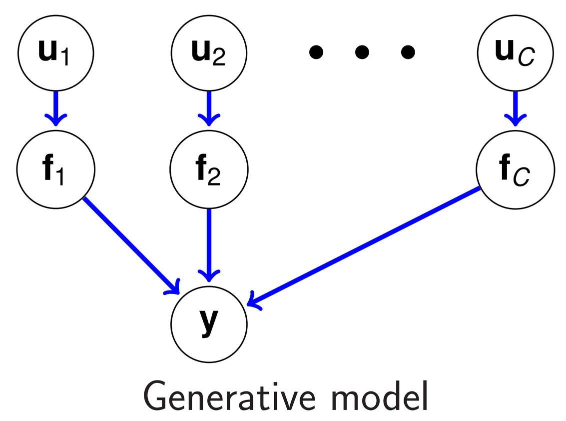
$$q(f) = q(u)p(f_{\neg u}|u)$$
$$q(u) = \mathcal{N}(\mu, \Sigma)$$

Prediction $\mathbf{O}(\mathbf{NM^3})$: $q(f) = \int du\, q(u)p(f|u)$

Variational Lower bound:

$$\log p(y|X) \geq E_{q(f)} \log p(y|f) - KL[q(u)|p(u)]$$

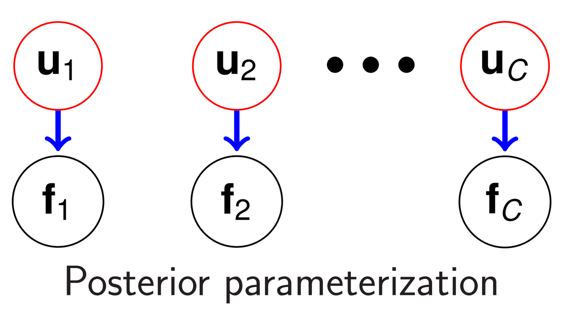
Multiple Sparse GPs



Each GP is 'augmented' with its set of inducing points Z_c. Its associated values are $u_c = f_c(Z_c)$

Previous approach: Mean Field [2]

In the context of multi-GP models, a mean field posterior parameterization has been proposed



$$q(f_{1..C}) = \prod_c q(u_c)p(f_{c\neg u_c}|u_c)$$
$$q(u_c) = \mathcal{N}(\mu_c, \Sigma_c)$$

Marginal predictions $\mathbf{O}(\mathbf{NM^3C})$: $q(f_c) = \int du_c\, q(u_c)p(f_c|u_c)$

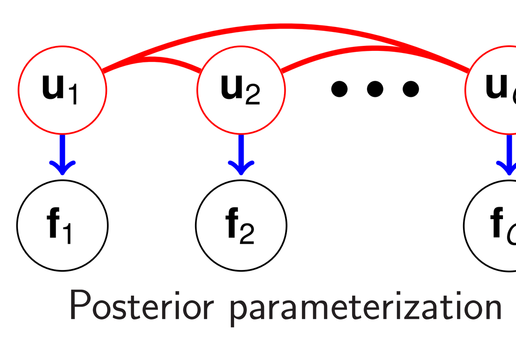
Variational Lower bound:

$$\log p(y|X) \geq E_{\prod_c q(f_c)} \log p(y|f_{1..C}) - \sum_c KL[q(u_c)|p(u_c)]$$

It is scalable but it suffers from the usual variance underestimation of mean field methods which in turn may bias learning

Coupling through inducing points

We propose a coupled parameterization:



$$q(f_{1..C}) = q(u) \prod_c p(f_{c\neg u_c}|u_c)$$
$$q(u) = \mathcal{N}(\mu, \Sigma)$$

Joint marginal predictions $\mathbf{O}(\mathbf{NM^3C^3})$:

$$q(f_{1..C}) = \int du\, q(u) \prod_c p(f_c|u_c)$$

Variational Lower bound:

$$\log p(y|X) \geq E_{q(f_{1..C})} \log p(y|f_{1..C}) - KL[q(u)|\prod_c p(u_c)]$$

Optimization

We optimize a stochastic objective with

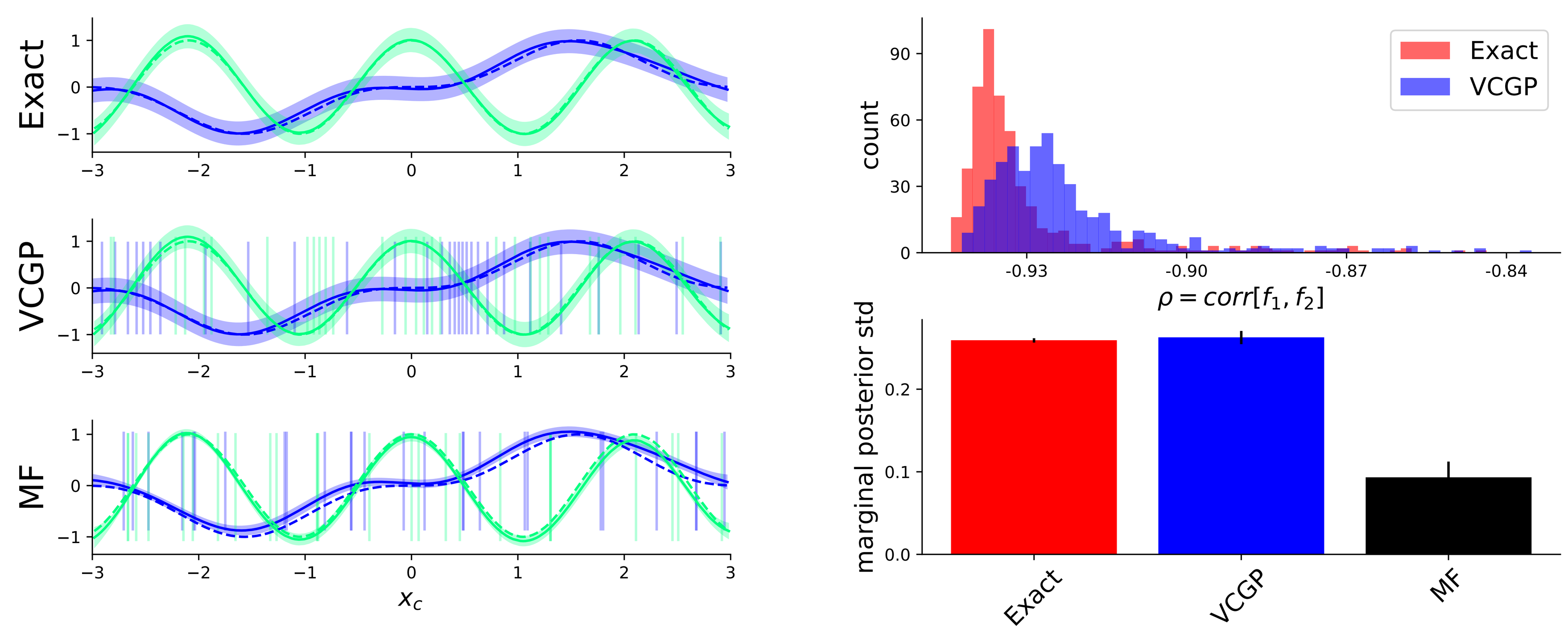
- analytical KL
- reparameterization trick for the likelihood expectations [3]:

For each expectation $E_{q(v)}f(v)$ under $v \sim \mathcal{N}(\mu, \Sigma = LL^T)$, we reparameterize the Gaussian distribution as $v = L\epsilon + \mu$ with $\epsilon \sim \mathcal{N}(0, I)$.

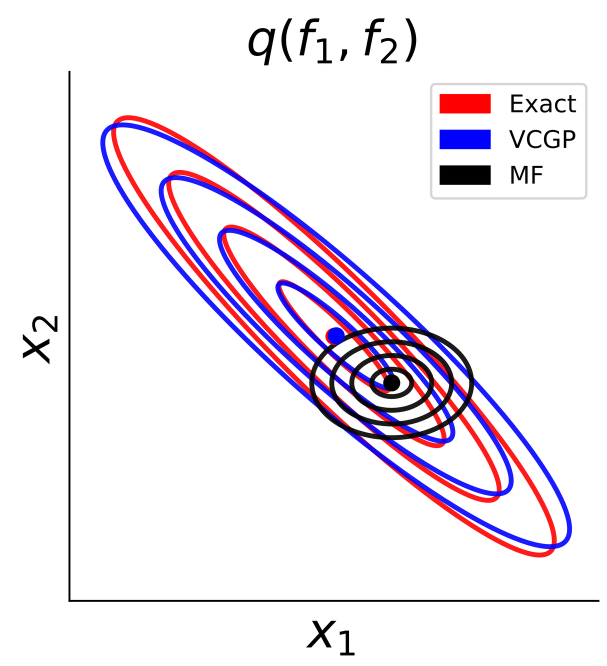
(Toy) Experiment

Additive Regression with conjugate likelihood

$$y|f_1, f_2, X = f_1(x_1) + f_2(x_2) + \sigma\epsilon \quad \epsilon \sim \mathcal{N}(0, 1)$$



Approximation recovers posterior coupling and correct marginal variance estimates.



Model	$\mathbb{E}_{\mathcal{D}}[\rho_{post}(f_1, f_2)]$	$-\mathcal{L}(q)$
Exact	-0.9306	381.924
VCGP[10]	-0.892	388.667
VCGP[30]	-0.9265	383.714
MF[10]	0	398.376
MF[30]	0	409.01

Summary

- Sparse Variational approximations for GPs can readily be extended to models with multiple GPs to capture the posterior dependencies.
- The proposed method offers a scalable treatment of structured non-linear regression model and can be applied to arbitrary likelihoods.

References

1. Titsias M. Variational learning of inducing variables in sparse Gaussian processes. *AISTATS 2009*
2. Saul A et al. Chained Gaussian Processes. *AISTATS 2016*
3. Kingma D et al. Auto-encoding variational Bayes. *ICLR 2014*

Contact: vincenta@gatsby.ucl.ac.uk