Checking the correctness of the implementation of the variational lower bound of Bayesian GP-LVM in GPflow

Joaquín Rapela*

January 16, 2022

Here I verify that the right-hand side of Eq. 14 in Titsias and Lawrence (2010) (reproduced here as Eq 1) is correctly implemented in the file `gplvm.py` of GPflow (Listing 1).

$$\tilde{F}_d(q) \geq \log\left[\frac{(\beta)^{\frac{N}{2}}|K_{MM}|^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}|\beta\Psi_2+K_{MM}|^{\frac{1}{2}}}e^{-\frac{1}{2}\mathbf{y}_d^\intercal W\mathbf{y}_d}\right] - \frac{\beta\psi_0}{2}+\frac{\beta}{2}\operatorname{Tr}(K_{MM}^{-1}\Psi_2) \quad (1)$$

where

$$W=\beta I_N-\beta^2\Psi_1(\beta\Psi_2+K_{MM})^{-1}\Psi_1^\intercal \quad (2)$$

Listing 1: Calculation of the variational lower bound in GPflow's gplvm.py

```
        cov_uu = covariances.Kuu(self.inducing_variable, self.kernel, jitter=default_jitter())
        L = tf.linalg.cholesky(cov_uu)
        sigma2 = self.likelihood.variance
        sigma = tf.sqrt(sigma2)

        # Compute intermediate matrices
        A = tf.linalg.triangular_solve(L, tf.transpose(psi1), lower=True) / sigma
        tmp = tf.linalg.triangular_solve(L, psi2, lower=True)
        AAT = tf.linalg.triangular_solve(L, tf.transpose(tmp), lower=True) / sigma2
        B = AAT + tf.eye(num_inducing, dtype=default_float())
        LB = tf.linalg.cholesky(B)
        log_det_B = 2.0 * tf.reduce_sum(tf.math.log(tf.linalg.diag_part(LB)))
        c = tf.linalg.triangular_solve(LB, tf.linalg.matmul(A, Y_data), lower=True) / sigma

        # compute log marginal bound
        ND = to_default_float(tf.size(Y_data))
        bound = -0.5 * ND * tf.math.log(2 * np.pi * sigma2)
        bound += -0.5 * D * log_det_B
        bound += -0.5 * tf.reduce_sum(tf.square(Y_data)) / sigma2
        bound += 0.5 * tf.reduce_sum(tf.square(c))
        bound += -0.5 * D * (tf.reduce_sum(psi0) / sigma2 -
                             tf.reduce_sum(tf.linalg.diag_part(AAT)))
```

*j.rapela@ucl.ac.uk

Expanding the logarithm and grouping terms, Eq. 1 can we rewritten as:

$$\begin{aligned}\tilde{F}(q) \geq &- \frac{ND}{2}\log 2\pi\beta^{-1} \\ &- \frac{D}{2}(\log|\beta\Psi_2 + K_{MM}| - \log|K_{MM}|) \\ &- \frac{1}{2}\sum_{d=1}^{D}\mathbf{y}_d^\intercal \, W \, \mathbf{y}_d \\ &+ \frac{D}{2}\frac{1}{\beta^{-1}}\operatorname{Tr}(K_{MM}^{-1}\Psi_2) \\ &- \frac{D}{2}\frac{\Psi_0}{\beta^{-1}}\end{aligned}$$

and replacing W (Eq. 2) we obtain:

$$\begin{aligned}\tilde{F}(q) \geq &- \frac{ND}{2}\log 2\pi\beta^{-1} \\ &- \frac{D}{2}(\log|\frac{\Psi_2}{\beta^{-1}} + K_{MM}| - \log|K_{MM}|) \\ &- \frac{1}{2}\frac{1}{\beta^{-1}}\sum_{d=1}^{D}||\mathbf{y}_d||_2^2 \\ &+ \frac{1}{2}\frac{1}{\beta^{-2}}\sum_{d=1}^{D}\mathbf{y}_d^\intercal\Psi_1\left(\frac{\Psi_2}{\beta^{-1}} + K_{MM}\right)^{-1}\Psi_1^\intercal\mathbf{y}_d \\ &+ \frac{D}{2}\frac{1}{\beta^{-1}}\operatorname{Tr}(K_{MM}^{-1}\Psi_2) \\ &- \frac{D}{2}\frac{\Psi_0}{\beta^{-1}}\end{aligned} \tag{3}$$

Define:

$$\tilde{B} \doteq \frac{\Psi_2}{\beta^{-1}} + K_{MM}$$

and replace it in Eq. 3

$$
\begin{aligned}
\tilde{F}(q) \geq & -\frac{ND}{2}\log 2\pi\beta^{-1} \\
& -\frac{D}{2}(\log|\tilde{B}| - \log|K_{MM}|) \\
& -\frac{1}{2}\frac{1}{\beta^{-1}}\sum_{d=1}^{D}||\mathbf{y}_d||_2^2 \\
& +\frac{1}{2}\frac{1}{\beta^{-2}}\sum_{d=1}^{D}\mathbf{y}_d^\intercal\Psi_1\tilde{B}^{-1}\Psi_1^\intercal\mathbf{y}_d \\
& +\frac{D}{2}\frac{1}{\beta^{-1}}\operatorname{Tr}(\frac{\Psi_2}{\beta^{-1}}K_{MM}^{-1}`) \\
& -\frac{D}{2}\frac{\Psi_0}{\beta^{-1}} \qquad (4)
\end{aligned}
$$

As outlined in Rasmussen and Williams (2006, Section 3.4.3), to invert $\tilde{B}$ more reliably, we re-write it in the following form (using the Cholesky decomposition $K_{MM} = LL^\intercal$):

$$
\begin{aligned}
\tilde{B} &\doteq \frac{\Psi_2}{\beta^{-1}} + K_{MM} \\
&= \frac{\Psi_2}{\beta^{-1}} + LL^\intercal \\
&= LL^{-1}\left(\frac{\Psi_2}{\beta^{-1}} + LL^\intercal\right)(L^\intercal)^{-1}L^\intercal \\
&= L\left(L^{-1}\frac{\Psi_2}{\beta^{-1}}(L^\intercal)^{-1} + I\right)L^\intercal \\
&= L\left(\mathrm{AAT} + I\right)L^\intercal \qquad (5) \\
&= LBL^\intercal \qquad (6)
\end{aligned}
$$

where AAT and B in Eqs. 5 and 6 are defined as:

$$
\mathrm{AAT} \doteq \frac{1}{\beta^{-1}}L^{-1}\Psi_2(L^\intercal)^{-1} \qquad (7)
$$

$$
B \doteq \mathrm{AAT} + I \qquad (8)
$$

Calling L_B the Cholesky factor of B (i.e., $B = L_B L_B^\intercal$) we have

$$\begin{aligned}
\frac{1}{\beta^{-2}}\mathbf{y}_d^\intercal \Psi_1 \tilde{B}^{-1}\Psi_1^T \mathbf{y}_d &= \frac{1}{\beta^{-2}}\mathbf{y}_d^\intercal \Psi_1 (L^\intercal)^{-1} B^{-1} L^{-1} \Psi_1^T \mathbf{y}_d \\
&= \frac{1}{\beta^{-2}}\mathbf{y}_d^\intercal \Psi_1 (L^\intercal)^{-1} (L_B^\intercal)^{-1} L_B^{-1} L^{-1} \Psi_1^T \mathbf{y}_d \\
&= \mathbf{c}_d^\intercal \mathbf{c}_d && (9) \\
\text{with} \quad \mathbf{c}_d &= \frac{1}{\beta^{-1}} L_B^{-1} L^{-1} \Psi_1^T \mathbf{y}_d && (10)
\end{aligned}$$

Then

$$\begin{aligned}
\log|B| &= \log|\frac{1}{\beta^{-1}} L^{-1}\Psi_2 (L^\intercal)^{-1} + I| \\
&= \log|\frac{1}{\beta^{-1}} L^{-1}\Psi_2 (L^\intercal)^{-1} + L^{-1}L| \\
&= \log|L^{-1}(\frac{\Psi_2}{\beta^{-1}}(L^\intercal)^{-1} + L)| \\
&= \log|L^{-1}(\frac{\Psi_2}{\beta^{-1}} + LL^\intercal)(L^\intercal)^{-1}| \\
&= \log(|L^{-1}||\frac{\Psi_2}{\beta^{-1}} + LL^\intercal||(L^\intercal)^{-1}|) \\
&= \log(|L^{-1}||(L^\intercal)^{-1}||\frac{\Psi_2}{\beta^{-1}} + LL^\intercal|) \\
&= \log(|L^{-1}(L^\intercal)^{-1}||\frac{\Psi_2}{\beta^{-1}} + LL^\intercal|) \\
&= \log(|K_{MM}^{-1}||\tilde{B}|) \\
&= \log|\tilde{B}| + \log|K_{MM}^{-1}| \\
&= \log|\tilde{B}| - \log|K_{MM}| && (11)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Tr}(\mathrm{AAT}) &= \mathrm{Tr}(\frac{1}{\beta^{-1}} L^{-1}\Psi_2 (L^\intercal)^{-1}) \\
&= \mathrm{Tr}(\frac{\Psi_2}{\beta^{-1}}(L^\intercal)^{-1}L^{-1}) \\
&= \mathrm{Tr}(\frac{\Psi_2}{\beta^{-1}} K_{MM}^{-1}) && (12)
\end{aligned}$$

Replacing Eqs. 9, 11 and 12 in Eq. 4 we obtain

$$
\begin{aligned}
\tilde{F}(q) \geq & -\frac{ND}{2}\log 2\pi\beta^{-1} \\
& -\frac{D}{2}\log|B| \\
& -\frac{1}{2}\frac{1}{\beta^{-1}}\sum_{d=1}^{D}||\mathbf{y}_d||_2^2 \\
& +\frac{1}{2}\sum_{d=1}^{D}||\mathbf{c}_d||_2^2 \\
& +\frac{D}{2}\frac{1}{\beta^{-1}}\,\mathrm{Tr}(\mathrm{AAT}) \\
& -\frac{D}{2}\frac{\Psi_0}{\beta^{-1}}
\end{aligned} \tag{13}
$$

Lines 1-6 in Eq. 13 corresponds to lines 17-22 in Lising 1, taking $\beta^{-1} = \sigma^2$.
From line 8 in Listing 1:

$$
\begin{aligned}
L\ \mathrm{tmp} &= \Psi_2 \\
\mathrm{tmp} &= L^{-1}\Psi_2 \\
\mathrm{tmp}^\intercal &= \Psi_2(L^{-1})^\intercal
\end{aligned}
$$

and from line 9 in Listing 1:

$$
\begin{aligned}
\beta^{-1}L\ \mathrm{AAT} &= \mathrm{tmp}^\intercal \\
\mathrm{AAT} &= \frac{1}{\beta^{-1}}L^{-1}\Psi_2(L^{-1})^\intercal
\end{aligned} \tag{14}
$$

Eq. 14 agrees with Eq. 7.
From line 7 in Listing 1:

$$
\begin{aligned}
\sqrt{\beta^{-1}}LA &= \Psi_1^\intercal \\
A &= \frac{1}{\sqrt{\beta^{-1}}}L^{-1}\Psi_1^\intercal
\end{aligned}
$$

and from line 13 in Listing 1

$$\sqrt{\beta^{-1}}L_B\,\mathbf{c} = A\mathbf{y} = \frac{1}{\sqrt{\beta^{-1}}}L^{-1}\Psi_1^\intercal\mathbf{y} \tag{15}$$

$$\mathbf{c} = \frac{1}{\beta^{-1}}L_B^{-1}L^{-1}\Psi_1^\intercal\mathbf{y} \tag{16}$$

Eq. 16 agrees with Eq. 10.

To compute $\log|B|$, line 12 in Listing 1 uses the fact that if $C = LL^\intercal$ then $\log|C| = 2\,\mathrm{Tr}(\log(L))$.

References

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT press Cambridge, MA.

Titsias, M. and Lawrence, N. D. (2010). Bayesian gaussian process latent variable model. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 844–851. JMLR Workshop and Conference Proceedings.