

# How to update a process so that it obeys the marginal defined by another external processes

Reed Essick  
essick@cita.utoronto.ca

April 24, 2023

## Derivation

Assume we have an existing Gaussian process that defines a measure for two vectors ( $f_a$  and  $f_b$  with  $N_a$  and  $N_b$  elements, respectively) and can be written as

$$p_O(f_a, f_b) \sim \mathcal{N} \left( [\mu_a, \mu_b], \begin{bmatrix} C_{aa} & C_{ab} \\ C_{ba} & C_{bb} \end{bmatrix} \right) \quad (1)$$

with mean vectors  $\mu_a, \mu_b$  and covariance matrix  $C$  decomposed into  $C_{aa}$  ( $N_a \times N_a$  elements),  $C_{ab}$  ( $N_a \times N_b$ ),  $C_{ba}$  ( $N_b \times N_a$ ), and  $C_{bb}$  ( $N_b \times N_b$ ). We note that

$$C_{aa} = C_{aa}^T \quad (2)$$

$$C_{ab} = C_{ba}^T \quad (3)$$

$$C_{bb} = C_{bb}^T \quad (4)$$

We wish to update this process so that the marginal distribution for  $f_b$  follows another process, namely

$$p_E(f_b) = \mathcal{N}(y_b, \Sigma_{bb}) \quad (5)$$

while maintaining the rest of the covariance structure encoded in  $C$ . We do this by constructing a new process

$$p_N(f_a, f_b) = p_O(f_a | f_b) p_E(f_b) \quad (6)$$

where  $p_O(f_a | f_b)$  can be derived from  $p_O(f_a, f_b)$  as

$$p_O(f_a | f_b) = \mathcal{N}(\mu_a + C_{ab} C_{bb}^{-1} (f_b - \mu_b), C_{aa} - C_{ab} C_{bb}^{-1} C_{ba}) \quad (7)$$

Expanding the contractions, grouping like terms, and dropping those that do not depend on either  $f_a$  or  $f_b$ , we obtain

$$\begin{aligned} -2 \ln p_N(f_a, f_b) &= f_a^T (C_{aa} - C_{ab} C_{bb}^{-1} C_{ba})^{-1} f_a \\ &\quad - 2 f_a^T (C_{aa} - C_{ab} C_{bb}^{-1} C_{ba})^{-1} (\mu_a - C_{ab} C_{bb}^{-1} \mu_b) \\ &\quad - 2 f_a^T \left[ (C_{aa} - C_{ab} C_{bb}^{-1} C_{ba})^{-1} C_{ab} C_{bb}^{-1} \right] f_b \\ &\quad + 2 f_b^T \left[ C_{bb}^{-1} C_{ba}^T (C_{aa} - C_{ab} C_{bb}^{-1} C_{ba})^{-1} (\mu_a - C_{ab} C_{bb}^{-1} \mu_b) - \Sigma_{bb}^{-1} y_b \right] \\ &\quad + f_b^T \left[ C_{bb}^{-1} C_{ba} (C_{aa} - C_{ab} C_{bb}^{-1} C_{ba})^{-1} C_{ab} C_{bb}^{-1} + \Sigma_{bb}^{-1} \right] f_b \end{aligned} \quad (8)$$

This is still Gaussian in both  $f_a$  and  $f_b$ , and we obtain direct relations for the new mean vectors and (inverse) covariance defined by

$$p_N(f_a, f_b) = \mathcal{N} \left( [m_a, m_b], \begin{bmatrix} \Gamma_{aa} & \Gamma_{ab} \\ \Gamma_{ba} & \Gamma_{bb} \end{bmatrix}^{-1} \right) \quad (9)$$

as follows:

$$\Gamma_{aa} = (C_{aa} - C_{ab}C_{bb}^{-1}C_{ba})^{-1} \quad (10)$$

$$\Gamma_{ab} = -(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba})^{-1} C_{ab}C_{bb}^{-1} \quad (11)$$

$$\Gamma_{bb} = C_{bb}^{-1}C_{ba} (C_{aa} - C_{ab}C_{bb}^{-1}C_{ba})^{-1} C_{ab}C_{bb}^{-1} + \Sigma_{bb}^{-1} \quad (12)$$

and

$$\Gamma_{aa}m_a + \Gamma_{ab}m_b = (C_{aa} - C_{ab}C_{bb}^{-1}C_{ba})^{-1} (\mu_a C_{ab}C_{bb}^{-1}\mu_b) \quad (13)$$

$$\Gamma_{ba}m_a + \Gamma_{bb}m_b = \Sigma_{bb}^{-1}y_b - C_{bb}^{-1}C_{ba} (C_{aa} - C_{ab}C_{bb}^{-1}C_{ba})^{-1} (\mu_a - C_{ab}C_{bb}^{-1}\mu_b) \quad (14)$$

which simply to

$$m_a = \mu_a + C_{ab}C_{bb}^{-1}(y_b - \mu_b) \quad (15)$$

$$m_b = y_b \quad (16)$$

Finally, we can solve for

$$\gamma = \Gamma^{-1} = \begin{bmatrix} \gamma_{aa} & \gamma_{ab} \\ \gamma_{ba} & \gamma_{bb} \end{bmatrix} \quad (17)$$

by recognizing that

$$\begin{bmatrix} \gamma_{aa} & \gamma_{ab} \\ \gamma_{ba} & \gamma_{bb} \end{bmatrix} \begin{bmatrix} \Gamma_{aa} & \Gamma_{ab} \\ \Gamma_{ba} & \Gamma_{bb} \end{bmatrix} = \mathbb{1} \quad (18)$$

and therefore

$$\gamma_{aa}\Gamma_{aa} + \gamma_{ab}\Gamma_{ba} = \mathbb{1} \quad (19)$$

$$\gamma_{aa}\Gamma_{ab} + \gamma_{ab}\Gamma_{bb} = \mathbb{0} \quad (20)$$

$$\gamma_{ba}\Gamma_{aa} + \gamma_{bb}\Gamma_{ba} = \mathbb{0} \quad (21)$$

$$\gamma_{ba}\Gamma_{ab} + \gamma_{bb}\Gamma_{bb} = \mathbb{1} \quad (22)$$

Further simplification yields

$$\gamma_{aa} = [\Gamma_{aa} - \Gamma_{ab}\Gamma_{bb}^{-1}\Gamma_{ba}]^{-1} \quad (23)$$

$$\gamma_{ab} = C_{ab}C_{bb}^{-1}\Sigma_{bb} \quad (24)$$

$$\gamma_{ba} = \Sigma_{bb}C_{bb}^{-1}C_{ba} \quad (25)$$

$$\gamma_{bb} = \Sigma_{bb} \quad (26)$$

where we've left  $\gamma_{aa}$  in terms of  $\Gamma$  because of the length of the expression but have substituted and simplified the rest of the terms. Note that the marginal distribution  $p_N(f_b) = \mathcal{N}(y_b, \Sigma_{bb}) = p_E(f_b)$ , as desired.

## Modifications for numerical stability

In general, we find that  $m_a$  and  $\gamma_{aa}$  can suffer from issues associated with numerical stability. This is because they involve the inversion of (possibly) high-dimensional matrices that may be ill-conditioned. While the preceding is exact, we therefore implement two additional approximations to help better control the calculations.

## Damping $C_{ab}$ , $C_{ba}$ , and $C_{bb}$ to make them easier to invert

One issue we have found is that strong correlations in  $C_{bb}$  can make numerical inversion difficult. Given that we wish to replace  $C_{bb}$  with  $\Sigma_{bb}$  anyway, and really only wish there to be a relatively smooth transition between  $f_b$  and  $f_a$ , we modify  $C_{ab}$ ,  $C_{ba}$ , and  $C_{bb}$  in order to damp the off-diagonal elements (and therefore make them easier to invert).

Specifically, we define a squared-exponential damping term

$$D(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{l^2}\right) \quad (27)$$

and a white noise contribution that modify  $C$  so that

$$(C_{ab})_{ij} \rightarrow (C_{ab})_{ij} D(x_i, x_j) \quad (28)$$

$$(C_{bb})_{ij} \rightarrow (C_{bb})_{ij} D(x_i, x_j) + \sigma_W^2 \delta_{ij} \quad (29)$$

$$(30)$$

We then use these modified  $C_{ab}$  and  $C_{bb}$  within the expressions in the previous section.

This modifies the original process, but as long as  $l$  is relatively large and  $\sigma_W$  is relatively small, the modifications will be minor over the transition between  $f_b$  and  $f_a$ . Empirically, we find that

$$l = 5.0 \quad (31)$$

$$\sigma_W = 0.01 \quad (32)$$

work well when updating our our model-agnostic priors.

## Approximation for $\gamma_{aa}$ when $\Sigma_{bb}$ is small

Finally, we note that it will often be the case that  $\Sigma_{bb}$  will be much smaller than  $C_{bb}$  with respect to an appropriate matrix norm. That is, we wish to update a process to restrict the marginals of certain covariates to be more tightly constrained than they otherwise would be.

By repeated use of the approximation

$$(A + X)^{-1} \approx A^{-1} - A^{-1} X A^{-1} \quad (33)$$

we can see that this limit corresponds to

$$\Gamma_{bb}^{-1} \approx \Sigma_{bb} - \Sigma_{bb} C_{bb}^{-1} C_{ba} (C_{aa} - C_{ab} C_{bb}^{-1} C_{ba})^{-1} C_{ba} C_{bb}^{-1} \Sigma_{bb} \quad (34)$$

and (retaining terms linear in  $\Sigma_{bb}$ )

$$\begin{aligned} \gamma_{aa} &\approx C_{aa} - C_{ab} C_{bb}^{-1} C_{ba} + C_{ab} C_{bb}^{-1} \Gamma_{bb}^{-1} C_{bb}^{-1} C_{ba} \\ &\approx C_{aa} - C_{ab} C_{bb}^{-1} C_{ba} + C_{ab} C_{bb}^{-1} \Sigma_{bb} C_{bb}^{-1} C_{ba} \end{aligned} \quad (35)$$

If we examine only terms up to linear order in  $\Sigma_{bb}$ , we obtain the relatively intuitive expression

$$\gamma_{aa} \approx C_{aa} - C_{ab} C_{bb}^{-1} (C_{bb} - \Sigma_{bb}) C_{bb}^{-1} C_{ba} \quad (36)$$

This makes sense in two limiting cases

- $\Sigma_{bb} = 0$  : we know  $f_b$  exactly and obtain the standard expression for the covariance for  $f_a | f_b$
- $\Sigma_{bb} = C_{bb}$  : we do not update the original process, and as such we obtain  $\gamma_{aa} = C_{aa}$ .

**To Do:** Interpret this final expression in terms of the measurement uncertainty we would need to use if we were to condition on the external process as if it was just a noisy observation.