Efficient Reinforcement Learning for Robots using Informative Simulated Priors -Additional Material-

Mark Cutler and Jonathan P. How

I. Additional Material

In this addition to the regular paper, we derive the required derivatives required to implement the informative prior from a simulator in PILCO [1]. First, for completeness, we repeat the derivation of the mean, covariance, and input-output covariance of the predictive mean of a Gaussian process (GP) when the prior mean is a radial basis function (RBF) network. Then, we detail the partial derivatives of the predictive distribution with respect to the input distribution.

A. Predictive Distribution

Following the outline of the derivations in [1] and [2] the predictive mean of uncertain input $\boldsymbol{x}_* \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ is given by

$$\begin{aligned}\mu_* &= \mathrm{E}_{\boldsymbol{x}_*,f} f[f(\boldsymbol{x}_*)] = \mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f(\boldsymbol{x}_*)]] \\ &= \mathrm{E}_{\boldsymbol{x}_*}[k(\boldsymbol{x}_*, X)\boldsymbol{\beta} + m(\boldsymbol{x}_*)]. \qquad (1)\end{aligned}$$

We assume the prior mean function $m(\boldsymbol{x}_*)$ is the mean of a GP that is trained using data from a simulator. Thus,

$$m(\boldsymbol{x}_*) = k_p(\boldsymbol{x}_*, X_p)\boldsymbol{\beta}_p$$

where $\{X_p, \boldsymbol{y_p}\}$ are the simulated data, $\boldsymbol{\beta}_p = (K_p + \sigma_{n_p}^2 I)^{-1}(\boldsymbol{y_p} - m(X_p))$, $K_p = k_p(X_p, X_p)$, and $\sigma_{n_p}^2$ is the noise variance parameter of the simulated data. Note that we assume that the prior mean is trained using a zero-prior GP. Substituting the form of the mean function into Eq. (1) yields

$$\mu_* = \boldsymbol{\beta}^T \boldsymbol{q} + \boldsymbol{\beta}_p^T \boldsymbol{q}_p, \qquad (2)$$

where $q_i = \alpha^2 |\Sigma\Lambda^{-1} + I|^{-1/2} \exp(-\frac{1}{2}\boldsymbol{\nu}_i^T(\Sigma+\Lambda)^{-1}\boldsymbol{\nu}_i)$ with $\boldsymbol{\nu}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$. The corresponding prior terms are similar with $q_{p_i} = \alpha_p^2|\Sigma\Lambda_p^{-1} + I|^{-1/2} \exp(-\frac{1}{2}\boldsymbol{\nu}_{p_i}^T(\Sigma + \Lambda_p)^{-1}\boldsymbol{\nu}_{p_i})$ and $\boldsymbol{\nu}_{p_i} = \boldsymbol{x}_{p_i} - \boldsymbol{\mu}$.

Multi-output regression problems can be solved by training a separate GP for each output dimension. When the inputs are uncertain, these output dimensions covary. We now compute the covariance for different output dimensions a and b as

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{x}_*,f}[f_a(\boldsymbol{x}_*), f_b(\boldsymbol{x}_*)] &= \mathrm{E}_{\boldsymbol{x}_*}[\mathrm{Cov}_f[f_a(\boldsymbol{x}_*), f_b(\boldsymbol{x}_*)]] \\ &+ \mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f_a(\boldsymbol{x}_*)]\mathrm{E}_f[f_b(\boldsymbol{x}_*)]] \\ &- \mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f_a(\boldsymbol{x}_*)]]\mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f_b(\boldsymbol{x}_*)]]. \qquad (3)\end{aligned}$$

Mark Cutler and Jonathan How are with the Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA, USA {cutlerm, jhow}@mit.edu

As noted in [2], due to the independence assumptions of the GPs, the first term in Eq. (3) is zero when $a \neq b$. Also, for a given output dimension, $\mathrm{Cov}_f[f_a(\boldsymbol{x}_*), f_b(\boldsymbol{x}_*)]$ does not depend on the prior mean function. Therefore, using the results of [1], the first term in Eq. (3) becomes

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{x}_*}[\mathrm{Cov}_f[f_a(\boldsymbol{x}_*), f_b(\boldsymbol{x}_*)]] = \\ &\qquad \delta_{ab}(\alpha_a^2 - \mathrm{tr}((K_a + \sigma_{\epsilon_a}^2 I)^{-1} Q)), \qquad (4)\end{aligned}$$

where δ_{ab} is 1 when $a = b$ and 0 otherwise, and

$$\begin{aligned}Q &= \int k_a(\boldsymbol{x}_*, X)^T k_b(\boldsymbol{x}_*, X) p(\boldsymbol{x}_*) d\boldsymbol{x}_* \\ Q_{ij} &= |R|^{-1/2} k_a(\boldsymbol{x}_i, \boldsymbol{\mu}) k_b(\boldsymbol{x}_j, \boldsymbol{\mu}) \exp(\tfrac{1}{2}\boldsymbol{z}_{ij}^T T^{-1} \boldsymbol{z}_{ij}) \qquad (5) \\ R &= \Sigma(\Lambda_a^{-1} + \Lambda_b^{-1}) + I \\ T &= \Lambda_a^{-1} + \Lambda_b^{-1} + \Sigma^{-1} \\ \boldsymbol{z}_{ij} &= \Lambda_a^{-1}\boldsymbol{\nu}_i + \Lambda_b^{-1}\boldsymbol{\nu}_j.\end{aligned}$$

The third term in Eq. (3) is computed using Eq. (2) as

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f_a(\boldsymbol{x}_*)]]\mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f_b(\boldsymbol{x}_*)]] = \\ &\qquad \left(\boldsymbol{\beta}_a^T \boldsymbol{q}_a + \boldsymbol{\beta}_{p_a}^T \boldsymbol{q}_{p_a}\right)\left(\boldsymbol{\beta}_b^T \boldsymbol{q}_b + \boldsymbol{\beta}_{p_b}^T \boldsymbol{q}_{p_b}\right). \qquad (6)\end{aligned}$$

Finally, we compute the second term in Eq. (3) as

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{x}_*}[\mathrm{E}_f[f_a(\boldsymbol{x}_*)]\mathrm{E}_f[f_b(\boldsymbol{x}_*)]] = \\ &\quad \mathrm{E}_{\boldsymbol{x}_*}[k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_a k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_b + m_a(\boldsymbol{x}_*)m_b(\boldsymbol{x}_*) + \\ &\quad m_a(\boldsymbol{x}_*)k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_b + k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_a m_b(\boldsymbol{x}_*)]. \qquad (7)\end{aligned}$$

As above, we will compute each term separately. Using Eq. (5), the first term in Eq. (7) becomes

$$\mathrm{E}_{\boldsymbol{x}_*}[k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_a k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_b] = \boldsymbol{\beta}_a^T Q \boldsymbol{\beta}_b. \qquad (8)$$

Similarly, the second term in Eq. (7) is

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{x}_*}[m_a(\boldsymbol{x}_*)m_b(\boldsymbol{x}_*)] = \\ &\quad \mathrm{E}_{\boldsymbol{x}_*}[k_p(\boldsymbol{x}_*, X_p)\boldsymbol{\beta}_{p_a} k_p(\boldsymbol{x}_*, X_p)\boldsymbol{\beta}_{p_b}] = \boldsymbol{\beta}_{p_a}^T Q_p \boldsymbol{\beta}_{p_b}, \qquad (9)\end{aligned}$$

where Q_p is defined analogously to Eq. (5) but using the prior rather than the current data. The third term in Eq. (7) is

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{x}_*}[m_a(\boldsymbol{x}_*)k(\boldsymbol{x}_*, X)\boldsymbol{\beta}_b] = \\ &\quad \boldsymbol{\beta}_{p_a}^T \mathrm{E}_{\boldsymbol{x}_*}[k_p(X_p, \boldsymbol{x}_*)k(\boldsymbol{x}_*, X)]\boldsymbol{\beta}_b = \boldsymbol{\beta}_{p_a}^T ({}^p\hat{Q})\boldsymbol{\beta}_b, \qquad (10)\end{aligned}$$

where ${}^p\hat{Q}$ is defined as

$$
\begin{aligned}
{}^p\hat{Q} &= \int k_{p_a}(\boldsymbol{x}_*, X_p)^T k_b(\boldsymbol{x}_*, X) p(\boldsymbol{x}_*) d\boldsymbol{x}_* \\
{}^p\hat{Q}_{ij} &= |{}^p\hat{R}|^{-1/2} k_{p_a}(\boldsymbol{x}_{p_i}, \boldsymbol{\mu}) k_b(\boldsymbol{x}_j, \boldsymbol{\mu}) \times \\
&\qquad \exp(\tfrac{1}{2}({}^p\hat{\boldsymbol{z}}_{ij})^T ({}^p\hat{T})^{-1} ({}^p\hat{\boldsymbol{z}}_{ij})) \qquad (11) \\
{}^p\hat{R} &= \Sigma(\Lambda_{p_a}^{-1} + \Lambda_b^{-1}) + I \\
{}^p\hat{T} &= \Lambda_{p_a}^{-1} + \Lambda_b^{-1} + \Sigma^{-1} \\
{}^p\hat{\boldsymbol{z}}_{ij} &= \Lambda_{p_a}^{-1}\boldsymbol{\nu}_{p_i} + \Lambda_b^{-1}\boldsymbol{\nu}_j.
\end{aligned}
$$

The forth term in Eq. (7) is analogously defined as $\boldsymbol{\beta}_a^T \hat{Q}^p \boldsymbol{\beta}_{p_b}$, where

$$
\begin{aligned}
\hat{Q}^p &= \int k_a(\boldsymbol{x}_*, X)^T k_{p_b}(\boldsymbol{x}_*, X_p) p(\boldsymbol{x}_*) d\boldsymbol{x}_* \\
\hat{Q}_{ij}^p &= |\hat{R}^p|^{-1/2} k_a(\boldsymbol{x}_i, \boldsymbol{\mu}) k_{p_b}(\boldsymbol{x}_{p_j}, \boldsymbol{\mu}) \times \\
&\qquad \exp(\tfrac{1}{2}(\hat{\boldsymbol{z}}_{ij}^p)^T (\hat{T}^p)^{-1} \hat{\boldsymbol{z}}_{ij}^p) \qquad (12) \\
\hat{R}^p &= \Sigma(\Lambda_a^{-1} + \Lambda_{p_b}^{-1}) + I \\
\hat{T}^p &= \Lambda_a^{-1} + \Lambda_{p_b}^{-1} + \Sigma^{-1} \\
\hat{\boldsymbol{z}}_{ij}^p &= \Lambda_a^{-1}\boldsymbol{\nu}_i + \Lambda_{p_b}^{-1}\boldsymbol{\nu}_{p_j}.
\end{aligned}
$$

Combining Eq. (4)-(12) we obtain the covariance for an uncertain input with multiple outputs. Writing this covariance element-wise we obtain

$$
\begin{aligned}
\sigma_{ab}^2 = &\delta_{ab}(\alpha_a^2 - \mathrm{tr}((K_a + \sigma_{\epsilon_a}^2 I)^{-1} Q)) + \boldsymbol{\beta}_a^T Q \boldsymbol{\beta}_b + \\
&\boldsymbol{\beta}_{p_a}^T Q_p \boldsymbol{\beta}_{p_b} + \boldsymbol{\beta}_{p_a}^T {}^p\hat{Q} \boldsymbol{\beta}_b + \boldsymbol{\beta}_a^T \hat{Q}^p \boldsymbol{\beta}_{p_b} - \\
&\left(\boldsymbol{\beta}_a^T \boldsymbol{q}_a + \boldsymbol{\beta}_{p_a}^T \boldsymbol{q}_{p_a}\right)\left(\boldsymbol{\beta}_b^T \boldsymbol{q}_b + \boldsymbol{\beta}_{p_b}^T \boldsymbol{q}_{p_b}\right). \qquad (13)
\end{aligned}
$$

The final derivation needed for propagating uncertain inputs through the GP transition model in the PILCO algorithm is the covariance between the uncertain test input $\boldsymbol{x}_* \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and the predicted output $f(\boldsymbol{x}_*) \sim \mathcal{N}(\mu_*, \Sigma_*)$. This covariance is calculated as

$$
\begin{aligned}
\Sigma_{\boldsymbol{x}_*, f_*} = &\mathrm{E}_{\boldsymbol{x}_*, f}[\boldsymbol{x}_* f(\boldsymbol{x}_*)^T] - \mathrm{E}_{\boldsymbol{x}_*}[\boldsymbol{x}_*]\mathrm{E}_{\boldsymbol{x}_*, f}[f(\boldsymbol{x}_*)]^T \\
= &\mathrm{E}_{\boldsymbol{x}_*, f}[\boldsymbol{x}_* k(\boldsymbol{x}_*, X)\boldsymbol{\beta}] - \\
&\mathrm{E}_{\boldsymbol{x}_*}[\boldsymbol{x}_*]\mathrm{E}_{\boldsymbol{x}_*}[k(\boldsymbol{x}_*, X)\boldsymbol{\beta}]^T + \\
&\mathrm{E}_{\boldsymbol{x}_*, f}[\boldsymbol{x}_* k_p(\boldsymbol{x}_*, X_p)\boldsymbol{\beta}_p] - \\
&\mathrm{E}_{\boldsymbol{x}_*}[\boldsymbol{x}_*]\mathrm{E}_{\boldsymbol{x}_*}[k_p(\boldsymbol{x}_*, X_p)\boldsymbol{\beta}_p]^T.
\end{aligned}
$$

Here we have separated the input-output covariance into a part that comes from the current data and a part that comes from the prior data. Therefore, we can directly apply the results from [1] to obtain

$$
\begin{aligned}
\Sigma_{\boldsymbol{x}_*, f_*} = &\Sigma(\Sigma + \Lambda)^{-1} \sum_{i=1}^{n} \beta_i q_i (\boldsymbol{x}_i - \boldsymbol{\mu}) + \\
&\Sigma(\Sigma + \Lambda_p)^{-1} \sum_{i=1}^{n_p} \beta_{p_i} q_{p_i} (\boldsymbol{x}_{p_i} - \boldsymbol{\mu}). \qquad (14)
\end{aligned}
$$

Note that in the derivation above we do not assume that there are the same number of data points in the prior GP and the current GP. Thus, the matrices ${}^p\hat{Q}$ and $\hat{Q}^p$ need not be square.

B. Partial Derivatives

Given the predictive distribution $\mathcal{N}(\mu_*, \Sigma_*)$ from Section I-A, we first compute the partial derivative of the predictive mean μ_* with respect to the input mean $\boldsymbol{\mu}$. Using the mean derived in Eq. (2) we get

$$
\begin{aligned}
\frac{\partial \mu_*}{\partial \boldsymbol{\mu}} = &\sum_{i=1}^{n} \beta_i \frac{\partial q_i}{\partial \boldsymbol{\mu}} + \sum_{i=1}^{n_p} \beta_{p_i} \frac{\partial q_{p_i}}{\partial \boldsymbol{\mu}} \\
= &\sum_{i=1}^{n} \beta_i q_i (\boldsymbol{x}_i - \boldsymbol{\mu})^T (\Sigma + \Lambda)^{-1} + \\
&\sum_{i=1}^{n_p} \beta_{p_i} q_{p_i} (\boldsymbol{x}_{p_i} - \boldsymbol{\mu})^T (\Sigma + \Lambda_p)^{-1}. \qquad (15)
\end{aligned}
$$

The derivative of the predictive mean with respect to the input covariance is written as

$$
\frac{\partial \mu_*}{\partial \Sigma} = \sum_{i=1}^{n} \beta_i \frac{\partial q_i}{\partial \Sigma} + \sum_{i=1}^{n_p} \beta_{p_i} \frac{\partial q_{p_i}}{\partial \Sigma}, \qquad (16)
$$

where, as in Eq. (15), the derivative consists of two distinct parts, one from the current data and one from the prior data. Using results from [1], we obtain

$$
\begin{aligned}
\frac{\partial \mu_*}{\partial \Sigma} = &\sum_{i=1}^{n} \beta_i q_i \left(-\tfrac{1}{2}((\Lambda^{-1}\Sigma + I)^{-1}\Lambda^{-1})^T \right. \\
&\left. -\tfrac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^T \frac{\partial (\Lambda + \Sigma)^{-1}}{\partial \Sigma} (\boldsymbol{x}_i - \boldsymbol{\mu})\right) + \\
&\sum_{i=1}^{n_p} \beta_{p_i} q_{p_i} \left(-\tfrac{1}{2}((\Lambda_p^{-1}\Sigma + I)^{-1}\Lambda_p^{-1})^T \right. \\
&\left. -\tfrac{1}{2}(\boldsymbol{x}_{p_i} - \boldsymbol{\mu})^T \frac{\partial (\Lambda_p + \Sigma)^{-1}}{\partial \Sigma} (\boldsymbol{x}_{p_i} - \boldsymbol{\mu})\right), \qquad (17)
\end{aligned}
$$

where, for D input dimensions and E output dimensions and $u, v = 1, \ldots, D + E$

$$
\begin{aligned}
\frac{\partial (\Lambda + \Sigma)^{-1}}{\partial \Sigma_{(uv)}} = &-\tfrac{1}{2}\left((\Lambda + \Sigma)^{-1}_{(:,u)}(\Lambda + \Sigma)^{-1}_{(v,:)} \right. \\
&\left. +(\Lambda + \Sigma)^{-1}_{(:,v)}(\Lambda + \Sigma)^{-1}_{(u,:)}\right), \qquad (18)
\end{aligned}
$$

and the corresponding prior term

$$
\begin{aligned}
\frac{\partial (\Lambda_p + \Sigma)^{-1}}{\partial \Sigma_{(uv)}} = &-\tfrac{1}{2}\left((\Lambda_p + \Sigma)^{-1}_{(:,u)}(\Lambda_p + \Sigma)^{-1}_{(v,:)} \right. \\
&\left. +(\Lambda_p + \Sigma)^{-1}_{(:,v)}(\Lambda_p + \Sigma)^{-1}_{(u,:)}\right). \qquad (19)
\end{aligned}
$$

Next, we derive the partial derivatives of the predictive covariance Σ_* with respect to the input mean and covariance. We take these derivatives element-wise for output dimensions a and b using Eq. (13). For the derivative with respect to the input mean we get

$$\begin{aligned}\frac{\partial\sigma_{ab}^2}{\partial\boldsymbol{\mu}} =&\delta_{ab}\left(-(K_a+\sigma_{\epsilon_a}^2 I)^{-1}\frac{\partial Q}{\partial\boldsymbol{\mu}}\right)+\\ &\boldsymbol{\beta}_a^T\left(\frac{\partial Q}{\partial\boldsymbol{\mu}}-\frac{\partial\boldsymbol{q}_a}{\partial\boldsymbol{\mu}}\boldsymbol{q}_b^T-\boldsymbol{q}_a\frac{\partial\boldsymbol{q}_b^T}{\partial\boldsymbol{\mu}}\right)\boldsymbol{\beta}_b+\\ &\boldsymbol{\beta}_{p_a}^T\left(\frac{\partial Q_p}{\partial\boldsymbol{\mu}}-\frac{\partial\boldsymbol{q}_{p_a}}{\partial\boldsymbol{\mu}}\boldsymbol{q}_{p_b}^T-\boldsymbol{q}_{p_a}\frac{\partial\boldsymbol{q}_{p_b}^T}{\partial\boldsymbol{\mu}}\right)\boldsymbol{\beta}_{p_b}+\\ &\boldsymbol{\beta}_a^T\left(\frac{\partial\hat{Q}^p}{\partial\boldsymbol{\mu}}-\frac{\partial\boldsymbol{q}_a}{\partial\boldsymbol{\mu}}\boldsymbol{q}_{p_b}^T-\boldsymbol{q}_a\frac{\partial\boldsymbol{q}_{p_b}^T}{\partial\boldsymbol{\mu}}\right)\boldsymbol{\beta}_{p_b}+\\ &\boldsymbol{\beta}_{p_a}^T\left(\frac{\partial({}^p\hat{Q})}{\partial\boldsymbol{\mu}}-\frac{\partial\boldsymbol{q}_{p_a}}{\partial\boldsymbol{\mu}}\boldsymbol{q}_b^T-\boldsymbol{q}_{p_a}\frac{\partial\boldsymbol{q}_b^T}{\partial\boldsymbol{\mu}}\right)\boldsymbol{\beta}_b, \quad (20)\end{aligned}$$

where, from [1],

$$\begin{aligned}\frac{\partial Q_{ij}}{\partial\boldsymbol{\mu}} = &Q_{ij}((\Lambda_a+\Lambda_b)^{-1}(\Lambda_b\boldsymbol{x}_i+\Lambda_a\boldsymbol{x}_j)-\\ &\boldsymbol{\mu})((\Lambda_a+\Lambda_b)^{-1}+\Sigma)^{-1} \quad (21)\end{aligned}$$

and similarly

$$\begin{aligned}\frac{\partial Q_{p_{ij}}}{\partial\boldsymbol{\mu}} =& Q_{p_{ij}}((\Lambda_{p_a}+\Lambda_{p_b})^{-1}(\Lambda_{p_b}\boldsymbol{x}_{p_i}+\Lambda_{p_a}\boldsymbol{x}_{p_j})-\\ &\boldsymbol{\mu})((\Lambda_{p_a}+\Lambda_{p_b})^{-1}+\Sigma)^{-1} \quad (22)\\ \frac{\partial({}^p\hat{Q}_{ij})}{\partial\boldsymbol{\mu}} =& {}^p\hat{Q}_{ij}((\Lambda_{p_a}+\Lambda_b)^{-1}(\Lambda_b\boldsymbol{x}_{p_i}+\Lambda_{p_a}\boldsymbol{x}_j)-\\ &\boldsymbol{\mu})((\Lambda_{p_a}+\Lambda_b)^{-1}+\Sigma)^{-1} \quad (23)\\ \frac{\partial\hat{Q}_{ij}^p}{\partial\boldsymbol{\mu}} =& \hat{Q}_{ij}^p((\Lambda_a+\Lambda_{p_b})^{-1}(\Lambda_{p_b}\boldsymbol{x}_i+\Lambda_a\boldsymbol{x}_{p_j})-\\ &\boldsymbol{\mu})((\Lambda_a+\Lambda_{p_b})^{-1}+\Sigma)^{-1}. \quad (24)\end{aligned}$$

Note that $\frac{\partial\boldsymbol{q}}{\partial\boldsymbol{\mu}}$ and $\frac{\partial\boldsymbol{q}_p}{\partial\boldsymbol{\mu}}$ are given in Eq. (15).

The derivative of the predictive covariance with respect to the input covariance is

$$\begin{aligned}\frac{\partial\sigma_{ab}^2}{\partial\Sigma} =&\delta_{ab}\left(-(K_a+\sigma_{\epsilon_a}^2 I)^{-1}\frac{\partial Q}{\partial\Sigma}\right)+\\ &\boldsymbol{\beta}_a^T\left(\frac{\partial Q}{\partial\Sigma}-\frac{\partial\boldsymbol{q}_a}{\partial\Sigma}\boldsymbol{q}_b^T-\boldsymbol{q}_a\frac{\partial\boldsymbol{q}_b^T}{\partial\Sigma}\right)\boldsymbol{\beta}_b+\\ &\boldsymbol{\beta}_{p_a}^T\left(\frac{\partial Q_p}{\partial\Sigma}-\frac{\partial\boldsymbol{q}_{p_a}}{\partial\Sigma}\boldsymbol{q}_{p_b}^T-\boldsymbol{q}_{p_a}\frac{\partial\boldsymbol{q}_{p_b}^T}{\partial\Sigma}\right)\boldsymbol{\beta}_{p_b}+\\ &\boldsymbol{\beta}_a^T\left(\frac{\partial\hat{Q}^p}{\partial\Sigma}-\frac{\partial\boldsymbol{q}_a}{\partial\Sigma}\boldsymbol{q}_{p_b}^T-\boldsymbol{q}_a\frac{\partial\boldsymbol{q}_{p_b}^T}{\partial\Sigma}\right)\boldsymbol{\beta}_{p_b}+\\ &\boldsymbol{\beta}_{p_a}^T\left(\frac{\partial({}^p\hat{Q})}{\partial\Sigma}-\frac{\partial\boldsymbol{q}_{p_a}}{\partial\Sigma}\boldsymbol{q}_b^T-\boldsymbol{q}_{p_a}\frac{\partial\boldsymbol{q}_b^T}{\partial\Sigma}\right)\boldsymbol{\beta}_b, \quad (25)\end{aligned}$$

where, from [1],

$$\begin{aligned}\frac{\partial Q_{ij}}{\partial\Sigma} =& -\tfrac{1}{2}Q_{ij}\left[(\Lambda_a^{-1}+\Lambda_b^{-1})R^{-1}-\boldsymbol{y}_{ij}^T\Xi\boldsymbol{y}_{ij}\right] \quad (26)\\ \boldsymbol{y}_{ij} =&\Lambda_b(\Lambda_a+\Lambda_b)^{-1}\boldsymbol{x}_i+\\ &\Lambda_a(\Lambda_a+\Lambda_b)^{-1}\boldsymbol{x}_j-\boldsymbol{\mu}\\ \Xi_{(uv)} =&\tfrac{1}{2}(\Phi_{(uv)}+\Phi_{(vu)})\\ \Phi_{(uv)} =& \Big(((\Lambda_a^{-1}+\Lambda_b^{-1})^{-1}+\Sigma)^{-1}_{(:,u)}\times\\ &((\Lambda_a^{-1}+\Lambda_b^{-1})^{-1}+\Sigma)^{-1}_{(v,:)}\Big).\end{aligned}$$

As before, the terms containing the prior data are similar as

$$\begin{aligned}\frac{\partial Q_{p_{ij}}}{\partial\Sigma} =& -\tfrac{1}{2}Q_{p_{ij}}\times\\ &\left[(\Lambda_{p_a}^{-1}+\Lambda_{p_b}^{-1})R_p^{-1}-\boldsymbol{y}_{p_{ij}}^T\Xi_p\boldsymbol{y}_{p_{ij}}\right] \quad (27)\\ \boldsymbol{y}_{p_{ij}} =&\Lambda_{p_b}(\Lambda_{p_a}+\Lambda_{p_b})^{-1}\boldsymbol{x}_{p_i}+\\ &\Lambda_{p_a}(\Lambda_{p_a}+\Lambda_{p_b})^{-1}\boldsymbol{x}_{p_j}-\boldsymbol{\mu}\\ \Xi_{p_{(uv)}} =&\tfrac{1}{2}(\Phi_{p_{(uv)}}+\Phi_{p_{(uv)}})\\ \Phi_{p_{(uv)}} =& \Big(((\Lambda_{p_a}^{-1}+\Lambda_{p_b}^{-1})^{-1}+\Sigma)^{-1}_{(:,u)}\times\\ &((\Lambda_{p_a}^{-1}+\Lambda_{p_b}^{-1})^{-1}+\Sigma)^{-1}_{(v,:)}\Big)\\ \frac{\partial({}^p\hat{Q}_{ij})}{\partial\Sigma} =& -\tfrac{1}{2}({}^p\hat{Q}_{ij})\times\\ &\left[(\Lambda_{p_a}^{-1}+\Lambda_b^{-1})({}^pR)^{-1}-({}^p\boldsymbol{y}_{ij})^T({}^p\Xi)({}^p\boldsymbol{y}_{ij})\right] \quad (28)\\ {}^p\boldsymbol{y}_{ij} =&\Lambda_b(\Lambda_{p_a}+\Lambda_b)^{-1}\boldsymbol{x}_{p_i}+\\ &\Lambda_{p_a}(\Lambda_{p_a}+\Lambda_b)^{-1}\boldsymbol{x}_j-\boldsymbol{\mu}\\ {}^p\Xi_{(uv)} =&\tfrac{1}{2}({}^p\Phi_{(uv)}+{}^p\Phi_{(uv)})\\ {}^p\Phi_{(uv)} =& \Big(((\Lambda_{p_a}^{-1}+\Lambda_b^{-1})^{-1}+\Sigma)^{-1}_{(:,u)}\times\\ &((\Lambda_{p_a}^{-1}+\Lambda_b^{-1})^{-1}+\Sigma)^{-1}_{(v,:)}\Big)\\ \frac{\partial\hat{Q}_{ij}^p}{\partial\Sigma} =& -\tfrac{1}{2}\hat{Q}_{ij}^p\times\\ &\left[(\Lambda_a^{-1}+\Lambda_{p_b}^{-1})(R^p)^{-1}-(\boldsymbol{y}_{ij}^p)^T\Xi^p\boldsymbol{y}_{ij}^p\right] \quad (29)\\ \boldsymbol{y}_{ij}^p =&\Lambda_{p_b}(\Lambda_a+\Lambda_{p_b})^{-1}\boldsymbol{x}_i+\\ &\Lambda_a(\Lambda_a+\Lambda_{p_b})^{-1}\boldsymbol{x}_{p_j}-\boldsymbol{\mu}\\ \Xi_{(uv)}^p =&\tfrac{1}{2}(\Phi_{(uv)}^p+\Phi_{(uv)}^p)\\ \Phi_{(uv)}^p =& \Big(((\Lambda_a^{-1}+\Lambda_{p_b}^{-1})^{-1}+\Sigma)^{-1}_{(:,u)}\times\\ &((\Lambda_a^{-1}+\Lambda_{p_b}^{-1})^{-1}+\Sigma)^{-1}_{(v,:)}\Big).\end{aligned}$$

Note that $\frac{\partial\boldsymbol{q}}{\partial\Sigma}$ and $\frac{\partial\boldsymbol{q}_p}{\partial\Sigma}$ are given in Eq. (17).

The final derivatives are the partial derivatives of the input-output covariance with respect to the input mean and covariance. From Eq. (14) we see that $\Sigma_{\boldsymbol{x}_*,f_*}$ consists of two distinct but similar parts, one from the current data and one from the prior data. Thus, applying results from [1], we get

$$\frac{\partial \Sigma_{\boldsymbol{x}_*,f_*}}{\partial \boldsymbol{\mu}} =$$
$$\Sigma(\Sigma+\Lambda)^{-1}\sum_{i=1}^{n}\beta_i\left((\boldsymbol{x}_i-\boldsymbol{\mu})\frac{\partial q_i}{\partial \boldsymbol{\mu}}-q_i I\right)+$$
$$\Sigma(\Sigma+\Lambda_p)^{-1}\sum_{i=1}^{n_p}\beta_{p_i}\left((\boldsymbol{x}_{p_i}-\boldsymbol{\mu})\frac{\partial q_{p_i}}{\partial \boldsymbol{\mu}}-q_{p_i} I\right) \tag{30}$$
$$\frac{\partial \Sigma_{\boldsymbol{x}_*,f_*}}{\partial \Sigma} =$$
$$\left((\Sigma+\Lambda)^{-1}+\Sigma\frac{\partial(\Sigma+\Lambda)^{-1}}{\partial \Sigma}\right)\sum_{i=1}^{n}\beta_i q_i(\boldsymbol{x}_i-\boldsymbol{\mu})+$$
$$\Sigma(\Sigma+\Lambda)^{-1}\sum_{i=1}^{n}\beta_i(\boldsymbol{x}_i-\boldsymbol{\mu})\frac{\partial q_i}{\partial \Sigma}+$$
$$\left((\Sigma+\Lambda_p)^{-1}+\Sigma\frac{\partial(\Sigma+\Lambda_p)^{-1}}{\partial \Sigma}\right)\sum_{i=1}^{n_p}\beta_{p_i} q_{p_i}(\boldsymbol{x}_{p_i}-\boldsymbol{\mu})+$$
$$\Sigma(\Sigma+\Lambda_p)^{-1}\sum_{i=1}^{n_p}\beta_{p_i}(\boldsymbol{x}_{p_i}-\boldsymbol{\mu})\frac{\partial q_{p_i}}{\partial \Sigma}, \tag{31}$$

where $\partial(\Sigma+\Lambda)^{-1}/\partial\Sigma$ and $\partial(\Sigma+\Lambda_p)^{-1}/\partial\Sigma$ are defined in Eq. (18) and Eq. (19), respectively.

This concludes the derivation of the partial derivatives needed to implement PILCO with a prior mean function that is an RBF network.

REFERENCES

[1] M. Deisenroth, D. Fox, and C. Rasmussen, "Gaussian processes for data-efficient learning in robotics and control," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PP, no. 99, 2014.

[2] B. Bischoff, D. Nguyen-Tuong, H. van Hoof, A. McHutchon, C. Rasmussen, A. Knoll, J. Peters, and M. Deisenroth, "Policy search for learning robot control using sparse data," in *IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong: IEEE, June 2014.