

APPENDIX A TRIGONOMETRIC INTEGRATION

This section gives exact integral equations for trigonometric functions, which are required to implement the discussed algorithms. The following expressions can be found in the book by [1], where $x \sim \mathcal{N}(x|\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2 .

$$\begin{aligned}\mathbb{E}_x[\sin(x)] &= \int \sin(x)p(x) dx = \exp(-\frac{\sigma^2}{2}) \sin(\mu), \\ \mathbb{E}_x[\cos(x)] &= \int \cos(x)p(x) dx = \exp(-\frac{\sigma^2}{2}) \cos(\mu), \\ \mathbb{E}_x[\sin(x)^2] &= \int \sin(x)^2 p(x) dx \\ &= \frac{1}{2}(1 - \exp(-2\sigma^2) \cos(2\mu)), \\ \mathbb{E}_x[\cos(x)^2] &= \int \cos(x)^2 p(x) dx \\ &= \frac{1}{2}(1 + \exp(-2\sigma^2) \cos(2\mu)), \\ \mathbb{E}_x[\sin(x) \cos(x)] &= \int \sin(x) \cos(x) p(x) dx \\ &= \int \frac{1}{2} \sin(2x) p(x) dx \\ &= \frac{1}{2} \exp(-2\sigma^2) \sin(2\mu).\end{aligned}$$

APPENDIX B GRADIENTS

In the beginning of this section, we will give a few derivative identities that will become handy. After that we will detail derivative computations in the context of the moment-matching approximation.

B.1 Identities

Let us start with a set of basic derivative identities [2] that will prove useful in the following:

$$\begin{aligned}\frac{\partial |\mathbf{K}(\boldsymbol{\theta})|}{\partial \boldsymbol{\theta}} &= |\mathbf{K}| \text{tr} \left(\mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}} \right), \\ \frac{\partial |\mathbf{K}|}{\partial \mathbf{K}} &= |\mathbf{K}| (\mathbf{K}^{-1})^\top, \\ \frac{\partial \mathbf{K}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\mathbf{K}^{-1} \frac{\partial \mathbf{K}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{K}^{-1}, \\ \frac{\partial \boldsymbol{\theta}^\top \mathbf{K} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= \boldsymbol{\theta}^\top (\mathbf{K} + \mathbf{K}^\top), \\ \frac{\partial \text{tr}(\mathbf{A} \mathbf{K} \mathbf{B})}{\partial \mathbf{K}} &= \mathbf{A}^\top \mathbf{B}^\top, \\ \frac{\partial |\mathbf{A} \mathbf{K} + \mathbf{I}|^{-1}}{\partial \mathbf{K}} &= -|\mathbf{A} \mathbf{K} + \mathbf{I}|^{-1} ((\mathbf{A} \mathbf{K} + \mathbf{I})^{-1})^\top, \\ \frac{\partial}{\partial B_{ij}} (\mathbf{a} - \mathbf{b})^\top (\mathbf{A} + \mathbf{B})^{-1} (\mathbf{a} - \mathbf{b}) \\ &= -(\mathbf{a} - \mathbf{b})^\top [(\mathbf{A} + \mathbf{B})_{(:,i)}^{-1} (\mathbf{A} + \mathbf{B})_{(j,:)}^{-1}] (\mathbf{a} - \mathbf{b}).\end{aligned}$$

In in the last identity $\mathbf{B}(:, i)$ denotes the i th column of \mathbf{B} and $\mathbf{B}(i, :)$ is the i th row of \mathbf{B} .

B.2 Partial Derivatives of the Predictive Distribution with Respect to the Input Distribution

For an input distribution $\tilde{\mathbf{x}}_{t-1} \sim \mathcal{N}(\tilde{\mathbf{x}}_{t-1} | \tilde{\boldsymbol{\mu}}_{t-1}, \tilde{\boldsymbol{\Sigma}}_{t-1})$, where $\tilde{\mathbf{x}} = [\mathbf{x}^\top \mathbf{u}^\top]^\top$ is the control-augmented state, we detail the derivatives of the predictive mean $\boldsymbol{\mu}_\Delta$, the predictive covariance $\boldsymbol{\Sigma}_\Delta$, and the cross-covariance $\text{cov}[\tilde{\mathbf{x}}_{t-1}, \boldsymbol{\Delta}]$ (in the moment matching approximation) with respect to the mean $\tilde{\boldsymbol{\mu}}_{t-1}$ and covariance $\tilde{\boldsymbol{\Sigma}}_{t-1}$ of the input distribution.

B.2.1 Derivatives of the Predictive Mean with Respect to the Input Distribution

In the following, we compute the derivative of the predictive GP mean $\boldsymbol{\mu}_\Delta \in \mathbb{R}^E$ with respect to the mean and the covariance of the input distribution $\mathcal{N}(\mathbf{x}_{t-1} | \boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1})$. The function value of the predictive mean is given as

$$\boldsymbol{\mu}_\Delta^a = \sum_{i=1}^n \beta_{a_i} q_{a_i}, \quad (1)$$

$$\begin{aligned}q_{a_i} &= \sigma_{f_a}^2 |\mathbf{I} + \boldsymbol{\Lambda}_a^{-1} \tilde{\boldsymbol{\Sigma}}_{t-1}|^{-\frac{1}{2}} \\ &\times \exp \left(-\frac{1}{2} (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}_{t-1})^\top (\boldsymbol{\Lambda}_a + \tilde{\boldsymbol{\Sigma}}_{t-1})^{-1} (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}_{t-1}) \right).\end{aligned} \quad (2)$$

B.2.1.1 Derivative with respect to the Input Mean: Let us start with the derivative of the predictive mean with respect to the mean of the input distribution. From the function value in Eq. (2), we obtain the derivative

$$\frac{\partial \boldsymbol{\mu}_\Delta^a}{\partial \tilde{\boldsymbol{\mu}}_{t-1}} = \sum_{i=1}^n \beta_{a_i} \frac{\partial q_{a_i}}{\partial \tilde{\boldsymbol{\mu}}_{t-1}} \quad (3)$$

$$= \sum_{i=1}^n \beta_{a_i} q_{a_i} (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}_{t-1})^\top (\tilde{\boldsymbol{\Sigma}}_{t-1} + \boldsymbol{\Lambda}_a)^{-1} \quad (4)$$

$\in \mathbb{R}^{1 \times (D+F)}$ for the a th target dimension, where we used

$$\frac{\partial q_{a_i}}{\partial \tilde{\boldsymbol{\mu}}_{t-1}} = q_{a_i} (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}_{t-1})^\top (\tilde{\boldsymbol{\Sigma}}_{t-1} + \boldsymbol{\Lambda}_a)^{-1}. \quad (5)$$

B.2.1.2 Derivative with Respect to the Input Covariance Matrix: For the derivative of the predictive mean with respect to the input covariance matrix $\boldsymbol{\Sigma}_{t-1}$, we obtain

$$\frac{\partial \boldsymbol{\mu}_\Delta^a}{\partial \tilde{\boldsymbol{\Sigma}}_{t-1}} = \sum_{i=1}^n \beta_{a_i} \frac{\partial q_{a_i}}{\partial \tilde{\boldsymbol{\Sigma}}_{t-1}}. \quad (6)$$

By defining

$$\begin{aligned}\eta(\tilde{\mathbf{x}}_i, \tilde{\boldsymbol{\mu}}_{t-1}, \tilde{\boldsymbol{\Sigma}}_{t-1}) \\ = \exp \left(-\frac{1}{2} (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}_{t-1})^\top (\boldsymbol{\Lambda}_a + \tilde{\boldsymbol{\Sigma}}_{t-1})^{-1} (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}_{t-1}) \right)\end{aligned}$$

we obtain

$$\begin{aligned}\frac{\partial q_{a_i}}{\partial \tilde{\boldsymbol{\Sigma}}_{t-1}} &= \sigma_{f_a}^2 \left(\frac{\partial |\mathbf{I} + \boldsymbol{\Lambda}_a^{-1} \tilde{\boldsymbol{\Sigma}}_{t-1}|^{-\frac{1}{2}}}{\partial \tilde{\boldsymbol{\Sigma}}_{t-1}} \eta(\tilde{\mathbf{x}}_i, \tilde{\boldsymbol{\mu}}_{t-1}, \tilde{\boldsymbol{\Sigma}}_{t-1}) \right. \\ &\quad \left. + |\mathbf{I} + \boldsymbol{\Lambda}_a^{-1} \tilde{\boldsymbol{\Sigma}}_{t-1}|^{-\frac{1}{2}} \frac{\partial}{\partial \tilde{\boldsymbol{\Sigma}}_{t-1}} \eta(\tilde{\mathbf{x}}_i, \tilde{\boldsymbol{\mu}}_{t-1}, \tilde{\boldsymbol{\Sigma}}_{t-1}) \right)\end{aligned}$$

for $i = 1, \dots, n$. Here, we compute the two partial derivatives

$$\frac{\partial |I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1}|^{-\frac{1}{2}}}{\partial \tilde{\Sigma}_{t-1}} \quad (7)$$

$$= -\frac{1}{2} |I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1}|^{-\frac{3}{2}} \frac{\partial |I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1}|}{\partial \tilde{\Sigma}_{t-1}} \quad (8)$$

$$= -\frac{1}{2} |I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1}|^{-\frac{3}{2}} |I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1}| \times ((I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1})^{-1} \Lambda_a^{-1})^\top \quad (9)$$

$$= -\frac{1}{2} |I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1}|^{-\frac{1}{2}} ((I + \Lambda_a^{-1} \tilde{\Sigma}_{t-1})^{-1} \Lambda_a^{-1})^\top \quad (10)$$

and for $p, q = 1, \dots, D + F$

$$\frac{\partial}{\partial \tilde{\Sigma}_{t-1}^{(pq)}} (\Lambda_a + \tilde{\Sigma}_{t-1})^{-1} \quad (11)$$

$$= -\frac{1}{2} \left((\Lambda_a + \tilde{\Sigma}_{t-1})_{(:,p)}^{-1} (\Lambda_a + \tilde{\Sigma}_{t-1})_{(q,:)}^{-1} + (\Lambda_a + \tilde{\Sigma}_{t-1})_{(:,q)}^{-1} (\Lambda_a + \tilde{\Sigma}_{t-1})_{(p,:)}^{-1} \right) \in \mathbb{R}^{(D+F) \times (D+F)},$$

where we need to explicitly account for the symmetry of $\Lambda_a + \tilde{\Sigma}_{t-1}$. Then, we obtain

$$\frac{\partial \mu_{\Delta}^a}{\partial \tilde{\Sigma}_{t-1}} = \sum_{i=1}^n \beta_{a_i} q_{a_i} \left(-\frac{1}{2} ((\Lambda_a^{-1} \tilde{\Sigma}_{t-1} + I)^{-1} \Lambda_a^{-1})^\top \underbrace{-\frac{1}{2} \underbrace{(\tilde{x}_i - \tilde{\mu}_{t-1})^\top}_{1 \times (D+F)} \underbrace{\frac{\partial (\Lambda_a + \tilde{\Sigma}_{t-1})^{-1}}{\partial \tilde{\Sigma}_{t-1}}}_{(D+F) \times (D+F)} \underbrace{(\tilde{x}_i - \tilde{\mu}_{t-1})}_{(D+F) \times 1}}_{(D+F) \times (D+F)} \right), \quad (12)$$

where we used a tensor contraction in the last expression inside the bracket when multiplying the difference vectors onto the matrix derivative.

B.2.2 Derivatives of the Predictive Covariance with Respect to the Input Distribution

For target dimensions $a, b = 1, \dots, E$, the entries of the predictive covariance matrix $\Sigma_{\Delta} \in \mathbb{R}^{E \times E}$ are given as

$$\sigma_{\Delta_{ab}}^2 = \beta_a^\top (Q - q_a q_b^\top) \beta_b + \delta_{ab} (\sigma_{f_a}^2 - \text{tr}((K_a + \sigma_{w_a}^2 I)^{-1} Q)) \quad (13)$$

where $\delta_{ab} = 1$ if $a = b$ and 0 otherwise.

The entries of $Q \in \mathbb{R}^{n \times n}$ are given by

$$Q_{ij} = \sigma_{f_a}^2 \sigma_{f_b}^2 |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I|^{-\frac{1}{2}} \times \exp \left(-\frac{1}{2} (\tilde{x}_i - \tilde{x}_j)^\top (\Lambda_a + \Lambda_b)^{-1} (\tilde{x}_i - \tilde{x}_j) \right) \times \exp \left(-\frac{1}{2} (\hat{z}_{ij} - \tilde{\mu}_{t-1})^\top \times ((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})^{-1} (\hat{z}_{ij} - \tilde{\mu}_{t-1}) \right), \quad (14)$$

$$\hat{z}_{ij} := \Lambda_b (\Lambda_a + \Lambda_b)^{-1} \tilde{x}_i + \Lambda_a (\Lambda_a + \Lambda_b)^{-1} \tilde{x}_j, \quad (15)$$

where $i, j = 1, \dots, n$.

B.2.2.1 Derivative with Respect to the Input Mean: For the derivative of the entries of the predictive covariance matrix with respect to the predictive mean, we obtain

$$\frac{\partial \sigma_{\Delta_{ab}}^2}{\partial \tilde{\mu}_{t-1}} = \beta_a^\top \left(\frac{\partial Q}{\partial \tilde{\mu}_{t-1}} - \frac{\partial q_a}{\partial \tilde{\mu}_{t-1}} q_b^\top - q_a \frac{\partial q_b^\top}{\partial \tilde{\mu}_{t-1}} \right) \beta_b + \delta_{ab} \left(-(\Lambda_a + \sigma_{w_a}^2 I)^{-1} \frac{\partial Q}{\partial \tilde{\mu}_{t-1}} \right), \quad (16)$$

where the derivative of Q_{ij} with respect to the input mean is given as

$$\frac{\partial Q_{ij}}{\partial \tilde{\mu}_{t-1}} = Q_{ij} (\hat{z}_{ij} - \tilde{\mu}_{t-1})^\top ((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})^{-1}. \quad (17)$$

B.2.2.2 Derivative with Respect to the Input Covariance Matrix: The derivative of the entries of the predictive covariance matrix with respect to the *covariance matrix of the input distribution* is

$$\frac{\partial \sigma_{\Delta_{ab}}^2}{\partial \tilde{\Sigma}_{t-1}} = \beta_a^\top \left(\frac{\partial Q}{\partial \tilde{\Sigma}_{t-1}} - \frac{\partial q_a}{\partial \tilde{\Sigma}_{t-1}} q_b^\top - q_a \frac{\partial q_b^\top}{\partial \tilde{\Sigma}_{t-1}} \right) \beta_b + \delta_{ab} \left(-(\Lambda_a + \sigma_{w_a}^2 I)^{-1} \frac{\partial Q}{\partial \tilde{\Sigma}_{t-1}} \right). \quad (18)$$

Since the partial derivatives $\partial q_a / \partial \tilde{\Sigma}_{t-1}$ and $\partial q_b / \partial \tilde{\Sigma}_{t-1}$ are known from Eq. (7), it remains to compute $\partial Q / \partial \tilde{\Sigma}_{t-1}$. The entries Q_{ij} , $i, j = 1, \dots, n$ are given in Eq. (14). By defining

$$c := \sigma_{f_a}^2 \sigma_{f_b}^2 \exp \left(-\frac{1}{2} (\tilde{x}_i - \tilde{x}_j)^\top (\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} (\tilde{x}_i - \tilde{x}_j) \right) e_2 := \exp \left(-\frac{1}{2} (\hat{z}_{ij} - \tilde{\mu}_{t-1})^\top ((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})^{-1} \times (\hat{z}_{ij} - \tilde{\mu}_{t-1}) \right)$$

we obtain the desired derivative

$$\frac{\partial Q_{ij}}{\partial \tilde{\Sigma}_{t-1}} = c \left[-\frac{1}{2} |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I|^{-\frac{3}{2}} \times \frac{\partial |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I|}{\partial \tilde{\Sigma}_{t-1}} e_2 + |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I|^{-\frac{1}{2}} \frac{\partial e_2}{\partial \tilde{\Sigma}_{t-1}} \right]. \quad (19)$$

Using the partial derivative

$$\frac{\partial |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I|}{\partial \tilde{\Sigma}_{t-1}} = |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I| \times \left(((\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I)^{-1} (\Lambda_a^{-1} + \Lambda_b^{-1}) \right)^\top \quad (20)$$

$$= |(\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I| \times \text{tr} \left(((\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + I)^{-1} (\Lambda_a^{-1} + \Lambda_b^{-1}) \frac{\partial \tilde{\Sigma}_{t-1}}{\partial \tilde{\Sigma}_{t-1}} \right) \quad (21)$$

the partial derivative of Q_{ij} with respect to the covariance matrix $\tilde{\Sigma}_{t-1}$ is given as

$$\begin{aligned} & \frac{\partial Q_{ij}}{\partial \tilde{\Sigma}_{t-1}} \\ &= c \left[-\frac{1}{2} |(\Lambda_a^{-1} + \Lambda_b)^{-1} \tilde{\Sigma}_{t-1} + \mathbf{I}|^{-\frac{3}{2}} \right. \\ & \quad \times |(\Lambda_a^{-1} + \Lambda_b)^{-1} \tilde{\Sigma}_{t-1} + \mathbf{I}| e_2 \\ & \quad \times \text{tr} \left(((\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + \mathbf{I})^{-1} (\Lambda_a^{-1} + \Lambda_b^{-1}) \frac{\partial \tilde{\Sigma}_{t-1}}{\partial \tilde{\Sigma}_{t-1}} \right) \\ & \quad \left. + |(\Lambda_a^{-1} + \Lambda_b)^{-1} \tilde{\Sigma}_{t-1} + \mathbf{I}|^{-\frac{1}{2}} \frac{\partial e_2}{\partial \tilde{\Sigma}_{t-1}} \right] \end{aligned} \quad (22)$$

$$\begin{aligned} &= c |(\Lambda_a^{-1} + \Lambda_b)^{-1} \tilde{\Sigma}_{t-1} + \mathbf{I}|^{-\frac{1}{2}} \\ & \quad \times \left[-\frac{1}{2} \left(((\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + \mathbf{I})^{-1} (\Lambda_a^{-1} + \Lambda_b^{-1}) \right)^\top e_2 \right. \\ & \quad \left. + \frac{\partial e_2}{\partial \tilde{\Sigma}_{t-1}} \right], \end{aligned} \quad (23)$$

where the partial derivative of e_2 with respect to the entries $\Sigma_{t-1}^{(p,q)}$ is given as

$$\begin{aligned} \frac{\partial e_2}{\partial \tilde{\Sigma}_{t-1}^{(p,q)}} &= -\frac{1}{2} (\hat{z}_{ij} - \tilde{\mu}_{t-1})^\top \frac{\partial ((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})^{-1}}{\partial \tilde{\Sigma}_{t-1}^{(p,q)}} \\ & \quad \times (\hat{z}_{ij} - \tilde{\mu}_{t-1}) e_2. \end{aligned} \quad (25)$$

The missing partial derivative in Eq. (25) is given by

$$\frac{\partial ((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})^{-1}}{\partial \tilde{\Sigma}_{t-1}^{(p,q)}} = -\Xi_{(pq)}, \quad (26)$$

where we define

$$\Xi_{(pq)} = \frac{1}{2} (\Phi_{(pq)} + \Phi_{(qp)}) \in \mathbb{R}^{(D+F) \times (D+F)}, \quad (27)$$

$p, q = 1, \dots, D + F$ with

$$\begin{aligned} \Phi_{(pq)} &= \left(((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})_{(:,p)}^{-1} \right. \\ & \quad \left. \times ((\Lambda_a^{-1} + \Lambda_b^{-1})^{-1} + \tilde{\Sigma}_{t-1})_{(q,:)}^{-1} \right). \end{aligned} \quad (28)$$

This finally yields

$$\begin{aligned} \frac{\partial Q_{ij}}{\partial \tilde{\Sigma}_{t-1}} &= c e_2 |(\Lambda_a^{-1} + \Lambda_b)^{-1} \tilde{\Sigma}_{t-1} + \mathbf{I}|^{-\frac{1}{2}} \\ & \quad \times \left[\left(((\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + \mathbf{I})^{-1} (\Lambda_a^{-1} + \Lambda_b^{-1}) \right)^\top \right. \\ & \quad \left. - (\hat{z}_{ij} - \tilde{\mu}_{t-1})^\top \Xi (\hat{z}_{ij} - \tilde{\mu}_{t-1}) \right] \end{aligned} \quad (29)$$

$$\begin{aligned} &= -\frac{1}{2} Q_{ij} \\ & \quad \times \left[\left(((\Lambda_a^{-1} + \Lambda_b^{-1}) \tilde{\Sigma}_{t-1} + \mathbf{I})^{-1} (\Lambda_a^{-1} + \Lambda_b^{-1}) \right)^\top \right. \\ & \quad \left. - (\hat{z}_{ij} - \tilde{\mu}_{t-1})^\top \Xi (\hat{z}_{ij} - \tilde{\mu}_{t-1}) \right], \end{aligned} \quad (30)$$

which concludes the computations for the partial derivative in Eq. (18).

B.2.3 Derivative of the Cross-Covariance with Respect to the Input Distribution

For the cross-covariance

$$\begin{aligned} \text{cov}_{f, \tilde{\mathbf{x}}_{t-1}}[\tilde{\mathbf{x}}_{t-1}, \Delta_t^a] &= \tilde{\Sigma}_{t-1} \mathbf{R}^{-1} \sum_{i=1}^n \beta_{a_i} q_{a_i} (\tilde{\mathbf{x}}_i - \tilde{\mu}_{t-1}), \\ \mathbf{R} &:= \tilde{\Sigma}_{t-1} + \Lambda_a, \end{aligned}$$

we obtain

$$\begin{aligned} & \frac{\partial \text{cov}_{f, \tilde{\mathbf{x}}_{t-1}}[\Delta_t, \tilde{\mathbf{x}}_{t-1}]}{\partial \tilde{\mu}_{t-1}} \\ &= \tilde{\Sigma}_{t-1} \mathbf{R}^{-1} \sum_{i=1}^n \beta_i \left((\tilde{\mathbf{x}}_i - \tilde{\mu}_{t-1}) \frac{\partial q_i}{\partial \tilde{\mu}_{t-1}} + q_i \mathbf{I} \right) \end{aligned} \quad (31)$$

$\in \mathbb{R}^{(D+F) \times (D+F)}$ for all target dimensions $a = 1, \dots, E$.

The corresponding derivative with respect to the covariance matrix $\tilde{\Sigma}_{t-1}$ is given as

$$\begin{aligned} & \frac{\partial \text{cov}_{f, \tilde{\mathbf{x}}_{t-1}}[\Delta_t, \tilde{\mathbf{x}}_{t-1}]}{\partial \tilde{\Sigma}_{t-1}} \\ &= \left(\frac{\partial \tilde{\Sigma}_{t-1}}{\partial \tilde{\Sigma}_{t-1}} \mathbf{R}^{-1} + \tilde{\Sigma}_{t-1} \frac{\partial \mathbf{R}^{-1}}{\partial \tilde{\Sigma}_{t-1}} \right) \sum_{i=1}^n \beta_{a_i} q_{a_i} (\tilde{\mathbf{x}}_i - \tilde{\mu}_{t-1}) \\ & \quad + \tilde{\Sigma}_{t-1} \mathbf{R}^{-1} \sum_{i=1}^n \beta_{a_i} (\tilde{\mathbf{x}}_i - \tilde{\mu}_{t-1}) \frac{\partial q_{a_i}}{\partial \tilde{\Sigma}_{t-1}}. \end{aligned} \quad (32)$$

REFERENCES

- [1] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 6th edition, July 2000.
- [2] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*, October 2008. Version 20081110.