Differentiating Gaussian Processes

Andrew McHutchon

April 17, 2013

1 First Order Derivative of the Posterior Mean

The posterior mean of a GP is given by,

$$\bar{f}_* \;=\; \boldsymbol{k}(\boldsymbol{x}_*, X)\, K(X,X)^{-1}\, \boldsymbol{y} \;\triangleq\; \boldsymbol{k}(\boldsymbol{x}_*, X)\, \boldsymbol{\alpha} \tag{1}$$

Only the $\boldsymbol{k}(\boldsymbol{x}_*, X)$ term depends on the test point $\boldsymbol{x}_*$, therefore to calculate the slope of the posterior mean we just need to differentiate the kernel. For the squared exponential covariance function the derivative of the kernel between $\boldsymbol{x}_*$ and a training point $\boldsymbol{x}_i$ is,

$$\begin{aligned}
\frac{\partial k(\boldsymbol{x}_*, \boldsymbol{x}_i)}{\partial \boldsymbol{x}_*} &= \frac{\partial}{\partial \boldsymbol{x}_*}\left\{\sigma_f^2 \exp\left(-\frac{1}{2}(\boldsymbol{x}_* - \boldsymbol{x}_i)^T \Lambda^{-1} (\boldsymbol{x}_* - \boldsymbol{x}_i)\right)\right\} \\
&= \frac{\partial}{\partial \boldsymbol{x}_*}\left\{-\frac{1}{2}(\boldsymbol{x}_* - \boldsymbol{x}_i)^T \Lambda^{-1} (\boldsymbol{x}_* - \boldsymbol{x}_i)\right\} k(\boldsymbol{x}_*, \boldsymbol{x}_i) \\
&= -\Lambda^{-1}(\boldsymbol{x}_* - \boldsymbol{x}_i)\, k(\boldsymbol{x}_*, \boldsymbol{x}_i)
\end{aligned} \tag{2}$$

which is a $D \times 1$ vector. To compute the derivative of the posterior mean we need to concatenate this derivative for each of the training points. It is helpful to define, $\tilde{X}_* \;=\; [\boldsymbol{x}_* - \boldsymbol{x}_1,\; ...,\; \boldsymbol{x}_* - \boldsymbol{x}_N]^T$, which is an $N \times D$ matrix.

$$\begin{aligned}
\frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*} &= \frac{\partial \boldsymbol{k}(\boldsymbol{x}_*, X)}{\partial \boldsymbol{x}_*}\, \boldsymbol{\alpha} \\
&= -\Lambda^{-1}\, \tilde{X}_*^T \left(\boldsymbol{k}(\boldsymbol{x}_*, X)^T \odot \boldsymbol{\alpha}\right)
\end{aligned} \tag{3}$$

which is a $D \times 1$ vector. $\odot$ represents an element-wise product.

2 Distribution over First Order Derivatives of Posterior Functions

In the previous section we found the derivative of the posterior mean of a GP. However, it is possible to find the distribution over derivatives of functions drawn from the GP posterior. Consider the random GP function values at two test point locations,

$$\begin{aligned}
f(\boldsymbol{x}_*) &= \bar{f}(\boldsymbol{x}_*) \;+\; z_* \\
f(\boldsymbol{x}_* + \boldsymbol{\delta}) &= \bar{f}(\boldsymbol{x}_* + \boldsymbol{\delta}) \;+\; z_\delta
\end{aligned} \tag{4}$$

where,

$$P(z_*, z_\delta) \;=\; \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} k_{**} - \boldsymbol{k}_*^T K^{-1} \boldsymbol{k}_* & k_{*\delta} - \boldsymbol{k}_*^T K^{-1} \boldsymbol{k}_\delta \\ k_{\delta *} - \boldsymbol{k}_\delta^T K^{-1} \boldsymbol{k}_* & k_{\delta\delta} - \boldsymbol{k}_\delta^T K^{-1} \boldsymbol{k}_\delta \end{bmatrix}\right) \tag{5}$$

The derivative is,

$$\begin{aligned}
\frac{\partial f_*}{\partial \boldsymbol{x}_*} &= \lim_{\boldsymbol{\delta}\to 0} \frac{f(\boldsymbol{x}_* + \boldsymbol{\delta}) - f(\boldsymbol{x}_*)}{\boldsymbol{x}_* + \boldsymbol{\delta} - \boldsymbol{x}_*} \\
&= \lim_{\boldsymbol{\delta}\to 0} \frac{\bar{f}(\boldsymbol{x}_* + \boldsymbol{\delta}) \; + \; z_\delta - \bar{f}(\boldsymbol{x}_*) \; - \; z_*}{\boldsymbol{\delta}} \\
&= \lim_{\boldsymbol{\delta}\to 0} \frac{\bar{f}(\boldsymbol{x}_* + \boldsymbol{\delta}) \; - \bar{f}(\boldsymbol{x}_*)}{\boldsymbol{\delta}} \; + \; \lim_{\boldsymbol{\delta}\to 0} \frac{z_\delta \; - \; z_*}{\boldsymbol{\delta}} \\
&= \frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*} \; + \lim_{\boldsymbol{\delta}\to 0} \frac{z_\delta \; - \; z_*}{\boldsymbol{\delta}}
\end{aligned} \tag{6}$$

This is a random variable, the mean of which is given by the first term, and the variance comes from the second. The variance of the second term is found as follows,

$$\begin{aligned}
\mathbb{V}\left[\lim_{\boldsymbol{\delta}\to 0} \frac{z_\delta \; - \; z_*}{\boldsymbol{\delta}}\right] &= \lim_{\boldsymbol{\delta}\to 0} \frac{1}{\boldsymbol{\delta}^2} \left(\mathbb{V}\left[z_\delta\right] \; + \; \mathbb{V}\left[z_*\right] \; - \; \mathbb{C}\left[z_\delta,\, z_*\right] \; - \; \mathbb{C}\left[z_*,\, z_\delta\right]\right) \\
&= \lim_{\boldsymbol{\delta}\to 0} \frac{1}{\boldsymbol{\delta}^2} \left(k_{\delta\delta} - \boldsymbol{k}_\delta^T K^{-1} \boldsymbol{k}_\delta + k_{**} - \boldsymbol{k}_*^T K^{-1} \boldsymbol{k}_* - (k_{*\delta} - \boldsymbol{k}_*^T K^{-1} \boldsymbol{k}_\delta) - (k_{\delta *} - \boldsymbol{k}_\delta^T K^{-1} \boldsymbol{k}_*)\right) \\
&= \lim_{\boldsymbol{\delta}\to 0} \frac{1}{\boldsymbol{\delta}^2} \left(k_{\delta\delta} \; - \; k_{*\delta} \; - \; k_{\delta *} \; + \; k_{**} \; - \; (\boldsymbol{k}_\delta - \boldsymbol{k}_*)^T K^{-1} (\boldsymbol{k}_\delta - \boldsymbol{k}_*)\right) \\
&= \frac{\partial^2 k(\boldsymbol{x}_1^*,\, \boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_1^* \, \partial \boldsymbol{x}_2^*} \; - \; \frac{\partial \boldsymbol{k}(\boldsymbol{x}_1^*,\, X)}{\partial \boldsymbol{x}_1^*} \, K^{-1} \, \frac{\partial \boldsymbol{k}(X,\, \boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_2^*}
\end{aligned} \tag{7}$$

which is a $D \times D$ matrix - the variances and covariances of the derivatives w.r.t. each dimension in $\boldsymbol{x}_*$. Thus,

$$P\left(\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right) \; = \; \mathcal{N}\left(\frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*},\, \frac{\partial^2 k(\boldsymbol{x}_1^*,\, \boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_1^* \, \partial \boldsymbol{x}_2^*} \; - \; \frac{\partial \boldsymbol{k}(\boldsymbol{x}_*,\, X)}{\partial \boldsymbol{x}_*} \, K^{-1} \, \frac{\partial \boldsymbol{k}(X,\, \boldsymbol{x}_*)}{\partial \boldsymbol{x}_*}\right) \tag{8}$$

We see that the mean of the distribution of derivatives is the derivative of the posterior mean. This is to be expected as both differentiation and expectation are linear operations and so are commutative.

3 Expected Squared Derivative

When propagating variances through first order Taylor series models, one uses the square of the first order derivative. We could also take the square inside the expectation which might lead to a better model.

The square of the expected derivative is given by,

$$\mathbb{E}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right]^2 \; = \; \frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*} \, \frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*}^T \tag{9}$$

We can find the expected squared derivative as follows,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right)^2\right] &= \mathbb{V}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right] \; + \; \mathbb{E}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right]^2 \\
&= \frac{\partial^2 k(\boldsymbol{x}_1^*,\, \boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_1^* \, \partial \boldsymbol{x}_2^*} \; - \; \frac{\partial \boldsymbol{k}(\boldsymbol{x}_*,\, X)}{\partial \boldsymbol{x}_*} \, K^{-1} \, \frac{\partial \boldsymbol{k}(X,\, \boldsymbol{x}_*)}{\partial \boldsymbol{x}_*} \; + \; \frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*} \, \frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*}^T
\end{aligned} \tag{10}$$

Compared to equation 9 the expected squared derivative is inflated by the variance of the derivative

4 Derivatives with Uncertain Inputs

We can also ask what the distribution over the derivatives is when the input location is Gaussian distributed, i.e.,

$$\boldsymbol{x}_* \; \sim \; \mathcal{N}\left(\boldsymbol{\mu},\, \Sigma\right) \tag{11}$$

4.1 The mean

We can use the rule of iterated expectations to find the mean derivative,

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right] &= \mathbb{E}_{\boldsymbol{x}_*}\left[\mathbb{E}_{f_*}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right]\right] \\
&= \mathbb{E}_{\boldsymbol{x}_*}\left[\frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*}\right] \\
&= \mathbb{E}_{\boldsymbol{x}_*}\left[-\Lambda^{-1}\,\tilde{X}_*\,\left(\boldsymbol{k}(\boldsymbol{x}_*, X)^T \odot \boldsymbol{\alpha}\right)\right] \\
&= -\Lambda^{-1}\,\mathbb{E}_{\boldsymbol{x}_*}\left[\sum_{i=1}^{N}\alpha_i\,(\boldsymbol{x}_* - \boldsymbol{x}_i)\,k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] \\
&= -\Lambda^{-1}\sum_{i=1}^{N}\alpha_i\,\mathbb{E}_{\boldsymbol{x}_*}\left[(\boldsymbol{x}_* - \boldsymbol{x}_i)\,k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] \\
&= -\Lambda^{-1}\sum_{i=1}^{N}\alpha_i\,\mathbb{E}_{\boldsymbol{x}_*}\left[\boldsymbol{x}_*\,k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] \;-\; \alpha_i\,\boldsymbol{x}_i\mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] \qquad (12)
\end{aligned}$$

To find the two expectations it is useful to note the squared exponential kernel is closely related to the Gaussian p.d.f.,

$$\begin{aligned}
k(\boldsymbol{x}_*, \boldsymbol{x}_i) &= \sigma_f^2\,\exp\left(-\frac{1}{2}(\boldsymbol{x}_* - \boldsymbol{x}_i)^T\,\Lambda^{-1}\,(\boldsymbol{x}_* - \boldsymbol{x}_i)\right) \qquad (13) \\
&= \sigma_f^2\,(2\pi)^{D/2}\,|\Lambda|^{1/2}\,\mathcal{N}\left(\boldsymbol{x}_*;\,\boldsymbol{x}_i,\,\Lambda\right) \qquad (14)
\end{aligned}$$

and also to quote the area under a product of two Gaussians,

$$\begin{aligned}
\int \mathcal{N}(\boldsymbol{x};\,\boldsymbol{\mu}_1, \Sigma_1)\,\mathcal{N}(\boldsymbol{x};\,\boldsymbol{\mu}_2, \Sigma_2)\,d\boldsymbol{x} &= (2\pi)^{-D/2}\,|\Sigma_1 + \Sigma_2|^{-1/2}\,\exp\left(-\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T(\Sigma_1 + \Sigma_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) \qquad (15) \\
&\triangleq Z \qquad (16)
\end{aligned}$$

We start with the simpler of the two expectations,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] &= \int k(\boldsymbol{x}_*, \boldsymbol{x}_i)\,p(\boldsymbol{x}_*)\,d\boldsymbol{x}_* \\
&= \sigma_f^2\,(2\pi)^{D/2}\,|\Lambda|^{1/2}\int \mathcal{N}\left(\boldsymbol{x}_*;\,\boldsymbol{x}_i,\,\Lambda\right)\,\mathcal{N}\left(\boldsymbol{x}_*;\,\boldsymbol{\mu},\,\Sigma\right)\,d\boldsymbol{x}_* \\
&= \sigma_f^2\,(2\pi)^{D/2}\,|\Lambda|^{1/2}\,(2\pi)^{-D/2}\,|\Lambda + \Sigma|^{-1/2}\,\exp\left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^T(\Lambda + \Sigma)^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right) \\
&= \sigma_f^2\,|\Lambda|^{1/2}\,|\Lambda + \Sigma|^{-1/2}\,\exp\left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^T(\Lambda + \Sigma)^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right) \\
&= |\Sigma\Lambda^{-1} + I|^{-1/2}\,k(\boldsymbol{x}_i,\,\boldsymbol{\mu},\,\Lambda + \Sigma) \qquad (17)
\end{aligned}$$

where the third argument to the covariance function specifies the lengthscales.

The second expectation,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_*}\left[\boldsymbol{x}_*\,k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] &= \int \boldsymbol{x}_*\,k(\boldsymbol{x}_*, \boldsymbol{x}_i)\,p(\boldsymbol{x}_*)\,d\boldsymbol{x}_* \\
&= \sigma_f^2\,(2\pi)^{D/2}\,|\Lambda|^{1/2}\int \boldsymbol{x}_*\,\mathcal{N}\left(\boldsymbol{x}_*;\,\boldsymbol{x}_i,\,\Lambda\right)\,\mathcal{N}\left(\boldsymbol{x}_*;\,\boldsymbol{\mu},\,\Sigma\right)\,d\boldsymbol{x}_* \qquad (18)
\end{aligned}$$

which is the mean of the product of two Gaussian distributions (times a constant). Therefore,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}_*}\left[\boldsymbol{x}_*\, k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right] &= \sigma_f^2\,(2\pi)^{D/2}\,|\Lambda|^{1/2}\,Z\,\left(\Lambda^{-1}+\Sigma^{-1}\right)^{-1}\,\left(\Lambda^{-1}\boldsymbol{x}_i\,+\,\Sigma^{-1}\boldsymbol{\mu}\right)\\
&= \sigma_f^2\,|\Lambda|^{1/2}\,|\Sigma+\Lambda|^{-1/2}\,\Lambda\,\exp\left(-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})^T(\Sigma+\Lambda)^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right)\,(\Sigma+\Lambda)^{-1}\,\Sigma\,\left(\Lambda^{-1}\boldsymbol{x}_i\,+\,\Sigma^{-1}\boldsymbol{\mu}\right)\\
&= |\Sigma\Lambda^{-1}+I|^{-1/2}\,k(\boldsymbol{x}_i,\,\boldsymbol{\mu},\,\Lambda+\Sigma)\,\Lambda\,(\Sigma+\Lambda)^{-1}\,\left(\Sigma\,\Lambda^{-1}\boldsymbol{x}_i\,+\,\boldsymbol{\mu}\right)\\
&= \mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*, \boldsymbol{x}_i)\right]\,\Lambda\,(\Sigma+\Lambda)^{-1}\,\left(\Sigma\,\Lambda^{-1}\boldsymbol{x}_i\,+\,\boldsymbol{\mu}\right) \qquad (19)
\end{aligned}$$

Putting equations 17 and 19 into equation 12 gives,

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right] &= -\Lambda^{-1}\sum_{i=1}^{N}\,\alpha_i\,\left(\mathbb{E}_{\boldsymbol{x}_*}\left[\boldsymbol{x}_*\,k(\boldsymbol{x}_*,\boldsymbol{x}_i)\right]\,-\,\boldsymbol{x}_i\,\mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*,\boldsymbol{x}_i)\right]\right)\\
&= -\Lambda^{-1}\sum_{i=1}^{N}\,\left(\Lambda\,(\Sigma+\Lambda)^{-1}\,\left(\Sigma\,\Lambda^{-1}\boldsymbol{x}_i\,+\,\boldsymbol{\mu}\right)\,-\,\boldsymbol{x}_i\right)\,\alpha_i\,\mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*,\boldsymbol{x}_i)\right]\\
&= \sum_{i=1}^{N}\,\left(\left(\Lambda^{-1}\,-\,(\Sigma+\Lambda)^{-1}\,\Sigma\,\Lambda^{-1}\right)\boldsymbol{x}_i\,-\,(\Sigma+\Lambda)^{-1}\,\boldsymbol{\mu}\right)\,\alpha_i\,\mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*,\boldsymbol{x}_i)\right]\\
&= (\Sigma+\Lambda)^{-1}\sum_{i=1}^{N}\,(\boldsymbol{x}_i\,-\,\boldsymbol{\mu})\,\alpha_i\,\mathbb{E}_{\boldsymbol{x}_*}\left[k(\boldsymbol{x}_*,\boldsymbol{x}_i)\right]\\
&= |\Sigma\Lambda^{-1}+I|^{-1/2}\,(\Sigma+\Lambda)^{-1}\sum_{i=1}^{N}\,(\boldsymbol{x}_i\,-\,\boldsymbol{\mu})\,\alpha_i\,k(\boldsymbol{x}_i,\,\boldsymbol{\mu},\,\Lambda+\Sigma) \qquad (20)\\
&= |\Sigma\Lambda^{-1}+I|^{-1/2}\,(\Sigma+\Lambda)^{-1}\,\tilde{X}_*^T\,(\boldsymbol{\alpha}\,\odot\,\boldsymbol{k}(X,\,\boldsymbol{\mu},\,\Lambda+\Sigma)) \qquad (21)
\end{aligned}$$

4.2 The variance

We can use the rule of total variance to find the variance of the derivative,

$$\begin{aligned}
\mathbb{V}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right] &= \mathbb{E}_{\boldsymbol{x}_*}\left[\mathbb{V}_{f_*}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right]\right]\,+\,\mathbb{V}_{\boldsymbol{x}_*}\left[\mathbb{E}_{f_*}\left[\frac{\partial f_*}{\partial \boldsymbol{x}_*}\right]\right]\\
&= \mathbb{E}_{\boldsymbol{x}_*}\left[\frac{\partial^2 k(\boldsymbol{x}_1^*,\,\boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_1^*\,\partial \boldsymbol{x}_2^*}\,-\,\frac{\partial \boldsymbol{k}(\boldsymbol{x}_*,\,X)}{\partial \boldsymbol{x}_*}\,K^{-1}\,\frac{\partial \boldsymbol{k}(X,\,\boldsymbol{x}_*)}{\partial \boldsymbol{x}_*}\right]\,+\,\mathbb{V}_{\boldsymbol{x}_*}\left[\frac{\partial \bar{f}_*}{\partial \boldsymbol{x}_*}\right]\\
&= \mathbb{E}_{\boldsymbol{x}_*}\left[\frac{\partial^2 k(\boldsymbol{x}_1^*,\,\boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_1^*\,\partial \boldsymbol{x}_2^*}\,-\,\frac{\partial \boldsymbol{k}(\boldsymbol{x}_*,\,X)}{\partial \boldsymbol{x}_*}\,K^{-1}\,\frac{\partial \boldsymbol{k}(X,\,\boldsymbol{x}_*)}{\partial \boldsymbol{x}_*}\right]\,+\,\mathbb{V}_{\boldsymbol{x}_*}\left[-\Lambda^{-1}\,\tilde{X}_*^T\,\left(\boldsymbol{k}(\boldsymbol{x}_*,X)^T\odot\boldsymbol{\alpha}\right)\right] \qquad (22)
\end{aligned}$$

We will calculate these expectations separately. Firstly the expectation of the second derivative of the kernel (see section 5 for derivation of the second derivative),

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial^2 k(\boldsymbol{x}_1^*,\,\boldsymbol{x}_2^*)}{\partial \boldsymbol{x}_1^*\,\partial \boldsymbol{x}_2^*}\right] &= \mathbb{E}\left[\Lambda^{-1}\,k\left(\boldsymbol{x}_*,\boldsymbol{x}_*\right)\right]\\
&= \Lambda^{-1}\,\sigma_f^2 \qquad (23)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \boldsymbol{k}(\boldsymbol{x}_*,X)}{\partial x_i^*}\,K^{-1}\,\frac{\partial \boldsymbol{k}(X,\,\boldsymbol{x}_*)}{\partial x_j^*}\right] &= \Lambda_i^{-1}\,\mathbb{E}\left[[\tilde{X}_*^{(i)T}\,\odot\,\boldsymbol{k}\left(\boldsymbol{x}_*\,,X\right)]\,K^{-1}\,\left[\boldsymbol{k}\left(X,\boldsymbol{x}_*\right)\,\odot\,\tilde{X}_*^{(j)}\right]\right]\,\Lambda_j^{-1}\\
&= \Lambda_i^{-1}\,\left(\mathbb{E}[\tilde{X}_*^{(i)T}\odot\boldsymbol{k}\left(\boldsymbol{x}_*\,,X\right)]\,K^{-1}\,\mathbb{E}[\boldsymbol{k}\left(X,\boldsymbol{x}_*\right)\odot\tilde{X}_*^{(j)}]\right.\\
&\qquad\left.+\,\mathrm{tr}\left(K^{-1}\,\mathbb{C}[\tilde{X}_*^{(i)T}\odot\boldsymbol{k}\left(\boldsymbol{x}_*\,,X\right),\,\tilde{X}_*^{(j)T}\odot\boldsymbol{k}\left(\boldsymbol{x}_*\,,X\right)]\right)\right)\,\Lambda_j^{-1} \qquad (24)
\end{aligned}$$

where,

$$\tilde{X}_*^{(i)} \;=\; [x_1^{(i)} - x_*^{(i)},\; \ldots,\; x_N^{(i)} - x_*^{(i)}]^T \tag{25}$$

which is a $N \times 1$ vector.

$$\begin{aligned}
\mathbb{E}_{x_*}[(x_1^{(i)} - x_*^{(i)})\, k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)] \;&=\; x_1^{(i)}\, \mathbb{E}\,[k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)] \;-\; \mathbb{E}[x_*^{(i)}\, k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)] \\
&=\; x_1^{(i)}\, \bar{k}(\boldsymbol{\mu},\, \Sigma,\, \boldsymbol{x}_1,\, \Lambda) \;-\; \bar{k}(\boldsymbol{\mu},\, \Sigma,\, \boldsymbol{x}_1,\, \Lambda)(\Sigma_{ii}\, \Lambda_i^{-1} + 1)^{-1}(\Sigma_{ii}\, \Lambda_i^{-1} x_1^{(i)} + \mu_i) \\
&=\; \left(x_1^{(i)} \;-\; (\Sigma_{ii}\, \Lambda_i^{-1} + 1)^{-1}(\Sigma_{ii}\, \Lambda_i^{-1} x_1^{(i)} + \mu_i)\right) \bar{k}(\boldsymbol{\mu},\, \Sigma,\, \boldsymbol{x}_1,\, \Lambda) \\
&=\; \bar{k}(\boldsymbol{\mu},\, \Sigma,\, \boldsymbol{x}_1,\, \Lambda)\, (x_1^{(i)} - \mu_i)\, (\Sigma_{ii}\, \Lambda_i^{-1} + 1)^{-1}
\end{aligned} \tag{26}$$

Defining $\boldsymbol{U}$ to be a $N \times 1$ vector with elements, $U_k \;=\; x_k^{(i)} - \mu_i$,

$$\mathbb{E}_{x_*}[\tilde{X}_*^{(i)} \;\odot\; k\,(X,\; \boldsymbol{x}_*)] \;=\; \bar{k}(X,\, \boldsymbol{\mu},\, \Sigma,\, \Lambda) \odot \boldsymbol{U}\, (\Sigma_{ii}\, \Lambda_i^{-1} + 1)^{-1} \tag{27}$$

which is a $N \times 1$ vector.

$$\begin{aligned}
&\mathbb{C}_{x_*}[(x_1^{(i)} - x_*^{(i)})\, k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)\,,\; (x_2^{(j)} - x_*^{(j)})\, k\,(\boldsymbol{x}_2,\; \boldsymbol{x}_*)] \\
&\qquad=\; x_1^{(i)}\, x_2^{(j)}\, \mathbb{C}_{x_*}[k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)\,,\; k\,(\boldsymbol{x}_2,\; \boldsymbol{x}_*)] \;-\; x_1^{(i)} \mathbb{C}_{x_*}[\, k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)\,,\; x_*^{(j)}\, k\,(\boldsymbol{x}_2,\; \boldsymbol{x}_*)] \\
&\qquad\qquad -\; x_2^{(j)}\, \mathbb{C}_{x_*}[x_*^{(i)}\, k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)\,,\; k\,(\boldsymbol{x}_2,\; \boldsymbol{x}_*)] \;+\; \mathbb{C}_{x_*}[x_*^{(i)}\, k\,(\boldsymbol{x}_1,\; \boldsymbol{x}_*)\,,\; x_*^{(j)}\, k\,(\boldsymbol{x}_2,\; \boldsymbol{x}_*)] \\
&\qquad=\; x_1^{(i)}\, x_2^{(j)}\, C_{kk}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma) - x_1^{(i)}\, C_{xkk}^{(j)}(\boldsymbol{\mu}, \boldsymbol{x}_2, \boldsymbol{x}_1, \Sigma) - x_2^{(j)}\, C_{xkk}^{(i)}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma) + C_{xkxk}^{(ij)}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma)
\end{aligned} \tag{28}$$

The C terms are derived and defined in section 6. Therefore,

$$\begin{aligned}
&\mathbb{C}[\tilde{X}_*^{(i)T} \odot \boldsymbol{k}\,(\boldsymbol{x}_*\,, X)\,,\; \tilde{X}_*^{(j)T} \odot \boldsymbol{k}\,(\boldsymbol{x}_*\,, X)] \\
&\qquad=\; (X^{(i)}\, X^{(j)T}) \odot C_{kk}(\boldsymbol{\mu}, X, X, \Sigma) - X^{(i)} \odot C_{xkk}^{(j)}(\boldsymbol{\mu}, X, X, \Sigma) - X^{(j)T} \odot C_{xkk}^{(i)}(\boldsymbol{\mu}, X, X, \Sigma) + C_{xkxk}^{(ij)}(\boldsymbol{\mu}, X, X, \Sigma)
\end{aligned} \tag{29}$$

which is a $N \times N$ matrix.

Finally we need, $\mathbb{V}_{\boldsymbol{x}_*}\left[\Lambda^{-1}\, \tilde{X}_*^T\; (\boldsymbol{k}(X, \boldsymbol{x}_*) \odot \boldsymbol{\alpha})\right]$, which we will break up and compute as,

$$\begin{aligned}
&\mathbb{C}_{\boldsymbol{x}_*}\left[\Lambda_i^{-1}\, \tilde{X}_*^{(i)T}\; (\boldsymbol{k}(X, \boldsymbol{x}_*) \odot \boldsymbol{\alpha})\,,\; \Lambda_j^{-1}\, \tilde{X}_*^{(j)T}\; (\boldsymbol{k}(X, \boldsymbol{x}_*) \odot \boldsymbol{\alpha})\right] \\
&\qquad=\; \Lambda_i^{-1}\, \Lambda_j^{-1}\, \mathbb{C}_{\boldsymbol{x}_*}\left[\sum_k (x_k^{(i)} - x_*^{(i)})\, k(\boldsymbol{x}_k, \boldsymbol{x}_*)\, \alpha_k,\; \sum_l (x_l^{(j)} - l_*^{(j)})\, k(\boldsymbol{x}_l, \boldsymbol{x}_*)\, \alpha_l\right] \\
&\qquad=\; \Lambda_i^{-1}\, \Lambda_j^{-1} \sum_k \sum_l \alpha_k\, \alpha_l\, \left(x_k^{(i)}\, x_l^{(j)}\, C_{kk} - x_k^{(i)}\, C_{xkk}^{(j)} - x_l^{(j)}\, C_{xkk}^{(i)} + C_{xkxk}^{(ij)}\right) \\
&\qquad=\; \Lambda_i^{-1}\, \Lambda_j^{-1}\, \boldsymbol{\alpha}_k^T\, \left(X^{(i)}\, X^{(j)T} \odot\, C_{kk} - X^{(i)}\, C_{xkk}^{(j)} - X^{(j)T}\, C_{xkk}^{(i)} + C_{xkxk}^{(ij)}\right)\, \boldsymbol{\alpha}_l
\end{aligned} \tag{30}$$

5 Differentiating the Squared Exponential kernel

The squared exponential kernel is given by,

$$k(\boldsymbol{x}_1,\, \boldsymbol{x}_2) \;=\; \sigma_f^2\, \exp\left(-\frac{1}{2}(\boldsymbol{x}_1 \;-\; \boldsymbol{x}_2)^T\, \Lambda^{-1}\, (\boldsymbol{x}_1 \;-\; \boldsymbol{x}_2)\right) \tag{31}$$

where the hyperparameters are the signal variance σ_f^2 and a characteristic length-scale for each dimension, $\{l_i\}_{i=1}^D$. The squared length-scales are collected into a $D \times D$, diagonal matrix Λ,

$$\Lambda \;=\; \begin{bmatrix} l_1^2 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & l_D^2 \end{bmatrix} \tag{32}$$

The derivative of this kernel with respect to the first argument is,

$$\begin{aligned}
\frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_1} &= \frac{\partial}{\partial \boldsymbol{x}_1}\left\{\sigma_f^2 \exp\left(-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)\right)\right\} \\
&= \frac{\partial}{\partial \boldsymbol{x}_1}\left\{-\frac{1}{2}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)^T \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)\right\} k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{x}_1}\left\{\boldsymbol{x}_1^T\Lambda^{-1}\boldsymbol{x}_1 - \boldsymbol{x}_2^T\Lambda^{-1}\boldsymbol{x}_1 - \boldsymbol{x}_1^T\Lambda^{-1}\boldsymbol{x}_2\right\} k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \\
&= -\frac{1}{2}\left(2\Lambda^{-1}\boldsymbol{x}_1 - 2\Lambda^{-1}\boldsymbol{x}_2\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \\
&= -\Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)
\end{aligned} \tag{33}$$

which is a $D \times 1$ vector. We can also take the derivative w.r.t. the second argument,

$$\begin{aligned}
\frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2} &= \frac{\partial}{\partial \boldsymbol{x}_2}\left\{\sigma_f^2 \exp\left(-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)\right)\right\} \\
&= \frac{\partial}{\partial \boldsymbol{x}_2}\left\{-\frac{1}{2}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)^T \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)\right\} k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{x}_2}\left\{-\boldsymbol{x}_2^T\Lambda^{-1}\boldsymbol{x}_1 - \boldsymbol{x}_1^T\Lambda^{-1}\boldsymbol{x}_2 + \boldsymbol{x}_2^T\Lambda^{-1}\boldsymbol{x}_2\right\} k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \\
&= -\frac{1}{2}\left(-2\Lambda^{-1}\boldsymbol{x}_1 + 2\Lambda^{-1}\boldsymbol{x}_2\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \\
&= \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)
\end{aligned} \tag{34}$$

Note that the only difference between these two derivatives (equations 33 and 34) is the minus sign in equation 33. This comes about because the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is calculated as $\boldsymbol{x}_1 - \boldsymbol{x}_2$ and hence increasing x_1 increases the separation and so decreases the covariance; the opposite is true for $\boldsymbol{x}_2$.

It is trivial to extend these results for the case when one of the inputs is a collection of points, such as a $N \times D$ training matrix X,

$$X = \left[\boldsymbol{x}_1, \ \ldots, \ \boldsymbol{x}_2\right]^T \tag{35}$$

$$\frac{\partial \boldsymbol{k}\left(X, \boldsymbol{x}_*\right)}{\partial \boldsymbol{x}_*} = \boldsymbol{k}\left(X, \ \boldsymbol{x}_*\right) \odot \tilde{X}_* \, \Lambda^{-1} \tag{36}$$

where we define $\tilde{X}_*$ to be the $N \times D$ matrix, $\left[\boldsymbol{x}_1 - \boldsymbol{x}_*, \ \ldots, \ \boldsymbol{x}_N - \boldsymbol{x}_*\right]^T$. Note that $\boldsymbol{k}\left(X, \boldsymbol{x}_*\right)$ is a $N \times 1$ column vector.

Building on equations 33 and 34 we can now find the second derivative (cross term),

$$\begin{aligned}
\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_2} &= \frac{\partial}{\partial \boldsymbol{x}_1}\left\{\Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)\right\} \\
&= \Lambda^{-1}\left(I - \left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right)^T \Lambda^{-1}\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)
\end{aligned} \tag{37}$$

We can see the following relationship,

$$\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_1^2} = \frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2^2} = -\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_1 \partial \boldsymbol{x}_2} = -\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2 \partial \boldsymbol{x}_1} \tag{38}$$

We can summarise the derivatives as follows,

$$\frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2} = \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \tag{39}$$

$$\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2^2} = -\Lambda^{-1} k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) + \Lambda^{-1}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) \frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2}^T \tag{40}$$

To find higher derivatives we need to switch to writing down elements of the derivative. First rephrase the first two derivatives,

$$\frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2^{(i)}} \;=\; \Lambda_{ii}^{-1}\;\left(x_1^{(i)} - x_2^{(i)}\right)\; k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \tag{41}$$

$$\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(j)} \partial x_2^{(i)}} \;=\; -\Lambda_{ij}^{-1}\, k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) \;+\; \Lambda_{ii}^{-1}\;\left(x_1^{(i)} - x_2^{(i)}\right)\; \frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(j)}} \tag{42}$$

$$\frac{\partial^3 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(k)} \partial x_2^{(j)} \partial x_2^{(i)}} \;=\; -\Lambda_{ij}^{-1}\, \frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(k)}} \;-\; \Lambda_{ik}^{-1}\, \frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(j)}} \;+\; \Lambda_{ii}^{-1}\;\left(x_1^{(i)} - x_2^{(i)}\right)\; \frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(k)} \partial x_2^{(j)}} \tag{43}$$

$$\frac{\partial^4 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(l)} \partial x_2^{(k)} \partial x_2^{(j)} \partial x_2^{(i)}} \;=\; -\Lambda_{ij}^{-1}\, \frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(l)} \partial x_2^{(k)}} - \Lambda_{ik}^{-1}\, \frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(l)} \partial x_2^{(j)}} - \Lambda_{il}^{-1}\, \frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(k)} \partial x_2^{(j)}} + \Lambda_{ii}^{-1}\, \left(x_1^{(i)} - x_2^{(i)}\right)\, \frac{\partial^3 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(l)} \partial x_2^{(k)} \partial x_2^{(j)}} \tag{44}$$

We sometimes need to evaluate these derivatives for $\boldsymbol{x}_1 \;=\; \boldsymbol{x}_2$,

$$k((\boldsymbol{x}_1, \boldsymbol{x}_2)|_{\boldsymbol{x}_1 \;=\; \boldsymbol{x}_2} \;=\; \sigma_f^2 \tag{45}$$

$$\left.\frac{\partial k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial \boldsymbol{x}_2^{(i)}}\right|_{\boldsymbol{x}_1 \;=\; \boldsymbol{x}_2} \;=\; 0 \tag{46}$$

$$\left.\frac{\partial^2 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(j)} \partial x_2^{(i)}}\right|_{\boldsymbol{x}_1 \;=\; \boldsymbol{x}_2} \;=\; -\Lambda_{ij}^{-1}\, \sigma_f^2 \tag{47}$$

$$\left.\frac{\partial^3 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(k)} \partial x_2^{(j)} \partial x_2^{(i)}}\right|_{\boldsymbol{x}_1 \;=\; \boldsymbol{x}_2} \;=\; 0 \tag{48}$$

$$\left.\frac{\partial^4 k\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right)}{\partial x_2^{(l)} \partial x_2^{(k)} \partial x_2^{(j)} \partial x_2^{(i)}}\right|_{\boldsymbol{x}_1 \;=\; \boldsymbol{x}_2} \;=\; \Lambda_{ij}^{-1}\, \Lambda_{kl}^{-1}\, \sigma_f^2 \;+\; \Lambda_{ik}^{-1}\, \Lambda_{jl}^{-1}\, \sigma_f^2 \;+\; \Lambda_{il}^{-1}\, \Lambda_{jk}^{-1}\, \sigma_f^2 \tag{49}$$

6 Squared Exponential Kernel Moments

The squared exponential kernel,

$$k(\boldsymbol{x}_1,\, \boldsymbol{x}_2) \;=\; \sigma_f^2\, \exp\left(-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\, \Lambda^{-1}\, (\boldsymbol{x}_1 - \boldsymbol{x}_2)\right) \tag{50}$$

$$\;=\; \sigma_f^2\, (2\pi)^{D/2}\, |\Lambda|^{1/2}\, \mathcal{N}\left(\boldsymbol{x}_1;\, \boldsymbol{x}_2,\, \Lambda\right) \tag{51}$$

Its mean,

$$\mathbb{E}_{\boldsymbol{x}}[k(\boldsymbol{x}, \boldsymbol{x}_1)] \;=\; \sigma_f^2\, |\Lambda|^{1/2}\, |\Lambda + \Sigma|^{-1/2}\; \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{x}_1)^T\, (\Lambda + \Sigma)^{-1}\, (\boldsymbol{\mu} - \boldsymbol{x}_1)\right)$$

$$\;=\; |\Sigma\, \Lambda^{-1} + I|^{-1/2}\, k(\boldsymbol{\mu},\, \boldsymbol{x}_1,\, \Lambda + \Sigma) \tag{52}$$

$$\;\triangleq\; \bar{k}(\boldsymbol{\mu}, \boldsymbol{x}_1, \Sigma, \Lambda) \qquad\qquad (1 \times 1) \tag{53}$$

The product of two squared exponential kernels,

$$k(\boldsymbol{x}, \boldsymbol{x}_1)\, k(\boldsymbol{x}, \boldsymbol{x}_2) \;=\; k\left(\frac{\boldsymbol{x}_1}{2}, \frac{\boldsymbol{x}_2}{2}, \frac{\Lambda}{2}\right) k\left(\boldsymbol{x}, \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{2}, \frac{\Lambda}{2}\right) \tag{54}$$

$$\;=\; \sigma_f^2\, \pi^{D/2}\, |\Lambda|^{1/2}\, k\left(\frac{\boldsymbol{x}_1}{2}, \frac{\boldsymbol{x}_2}{2}, \frac{\Lambda}{2}\right) \mathcal{N}\left(\boldsymbol{x};\, (\boldsymbol{x}_1 + \boldsymbol{x}_2)/2,\, \Lambda/2\right) \tag{55}$$

The mean of a product,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}}[k(\boldsymbol{x}_1, \boldsymbol{x})\, k(\boldsymbol{x}, \boldsymbol{x}_2)] &= k\left(\boldsymbol{x}_1/2, \boldsymbol{x}_2/2, \Lambda/2\right)\, \bar{k}\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)/2,\, \boldsymbol{\mu}_1,\, \Sigma,\, \Lambda/2\right) && (56)\\
&= |2\,\Sigma\,\Lambda^{-1} + I|^{-1/2}\, k\left(\boldsymbol{x}_1/2, \boldsymbol{x}_2/2, \Lambda/2\right)\, k\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)/2,\, \boldsymbol{\mu}_1,\, \Lambda/2 + \Sigma\right) && (57)\\
&\triangleq E_{kk}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma) \qquad\qquad (1 \times 1) && (58)
\end{aligned}$$

Therefore, the covariance of two kernels,

$$\begin{aligned}
\mathbb{C}_{\boldsymbol{x}}[k(\boldsymbol{x}, \boldsymbol{x}_1),\, k(\boldsymbol{x}, \boldsymbol{x}_2)] &= |2\,\Sigma\,\Lambda^{-1} + I|^{-1/2}\, k\left(\boldsymbol{x}_1/2, \boldsymbol{x}_2/2, \Lambda/2\right)\, k\left((\boldsymbol{x}_1 + \boldsymbol{x}_2)/2,\, \boldsymbol{\mu},\, \Lambda/2 + \Sigma\right)\\
&\quad - |\Sigma\,\Lambda^{-1} + I|^{-1}\, k(\boldsymbol{\mu},\, \boldsymbol{x}_1,\, \Lambda + \Sigma)\; k(\boldsymbol{\mu},\, \boldsymbol{x}_2,\, \Lambda + \Sigma) && (59)\\
&\triangleq C_{kk}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma) \qquad\qquad (1 \times 1) && (60)
\end{aligned}$$

The covariance between x times a covariance function with another covariance function,

$$\begin{aligned}
&\mathbb{C}_{\boldsymbol{x}}\left[\boldsymbol{x}\, k(\boldsymbol{x},\, \boldsymbol{x}_1),\; k(\boldsymbol{x},\, \boldsymbol{x}_2)\right]\\
&\quad = k\left(\frac{\boldsymbol{x}_1}{2}, \frac{\boldsymbol{x}_2}{2}, \frac{\Lambda}{2}\right)\, \kappa(\boldsymbol{\mu},\, \Sigma,\, (\boldsymbol{x}_1 + \boldsymbol{x}_2)/2,\, \Lambda/2) \;-\; \boldsymbol{\kappa}\left(\boldsymbol{\mu},\, \Sigma,\, \boldsymbol{x}_1,\, \Lambda\right) \bar{k}\left(\boldsymbol{\mu},\, \boldsymbol{x}_2,\; \Sigma,\, \Lambda\right)\\
&\quad \triangleq C_{xkk}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma) \qquad\qquad (D \times 1) && (61)
\end{aligned}$$

The covariance of x times a covariance function with x times another covariance function,

$$\begin{aligned}
&\mathbb{C}_{\boldsymbol{x}}\left[\boldsymbol{x}\, k(\boldsymbol{x},\, \boldsymbol{x}_1),\; \boldsymbol{x}\, k(\boldsymbol{x},\, \boldsymbol{x}_2)\right]\\
&\quad = k\left(\frac{\boldsymbol{x}_1}{2}, \frac{\boldsymbol{x}_2}{2}, \frac{\Lambda}{2}\right)\, \bar{k}\left(\boldsymbol{\mu},\, \Sigma,\, (\boldsymbol{x}_1 + \boldsymbol{x}_2)/2,\, \Lambda/2\right)\\
&\qquad \left[(2\Sigma\,\Lambda^{-1} + I)^{-1}(\Sigma\,\Lambda^{-1}(\boldsymbol{x}_1 + \boldsymbol{x}_2) + \boldsymbol{\mu})(\Sigma\,\Lambda^{-1}(\boldsymbol{x}_1 + \boldsymbol{x}_2) + \boldsymbol{\mu})^T(2\Sigma\,\Lambda^{-1} + I)^{-1} + \Sigma\,(\Sigma + \Lambda/2)^{-1}\Lambda/2\right]\\
&\qquad - \boldsymbol{\kappa}\left(\boldsymbol{\mu},\, \boldsymbol{x}_1,\, \Sigma,\, \Lambda\right)\, \boldsymbol{\kappa}\left(\boldsymbol{\mu},\, \boldsymbol{x}_2,\, \Sigma,\, \Lambda\right)^T\\
&\quad \triangleq C_{xkxk}(\boldsymbol{\mu}, \boldsymbol{x}_1, \boldsymbol{x}_2, \Sigma) \qquad\qquad (D \times D) && (62)
\end{aligned}$$