Kernels and Derivatives of Kernels

Jungtaek Kim

Pohang University of Science and Technology, Pohang 37673, Republic of Korea

jtkim@postech.ac.kr

June 25, 2021

1 Introduction

In these notes, we describe the popular kernels such as exponentiated quadratic, Matérn 3/2, and Matérn 5/2 kernels [Rasmussen and Williams, 2006], and their derivatives. Before the details of them, we first introduce notations as follows.

Table 1: Notation.

Symbol	Meaning	Dimensionality
$\mathbf{x}$	a data point	$\mathbb{R}^d$
s	a signal scale	$\mathbb{R}$
ℓ	a length scale	$\mathbb{R}$
$\mathbf{L}$	a length-scale matrix	$\mathbb{R}^{d \times d}$
σ_n^2	a noise variance	$\mathbb{R}$
$\mathbf{d}_{ij}$	a difference between $\mathbf{x}_i$ and $\mathbf{x}_j$ (i.e., $\mathbf{x}_i - \mathbf{x}_j$)	$\mathbb{R}^d$
$k(\cdot, \cdot)$	a kernel	$\mathbb{R}$
$\widehat{k}(\cdot, \cdot)$	a kernel with observation noise	$\mathbb{R}$

Note that a length-scale matrix $\mathbf{L}$ for automatic relevance determination [Rasmussen and Williams, 2006] is a diagonal matrix:

$$\mathbf{L} = \begin{bmatrix} \ell_1 & 0 & \cdots & 0 \\ 0 & \ell_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \ell_d \end{bmatrix} \in \mathbb{R}^{d \times d}. \tag{1}$$

In addition, a kernel with observation noise $\widehat{k}(\mathbf{x}_i, \mathbf{x}_j)$ is defined as

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j) + \sigma_n^2 \delta_{ij}, \tag{2}$$

where δ_{ij} is the Dirac delta function, which is 1 if $i = j$, and 0 otherwise.

2 Exponentiated Quadratic Kernel

Exponentiated quadratic kernel (also known as squared exponential kernel) is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = s^2 \exp\left(-\frac{1}{2\ell^2}(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)\right), \tag{3}$$

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = s^2 \exp\left(-\frac{1}{2\ell^2}(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)\right) + \sigma_n^2 \delta_{ij}, \tag{4}$$

for the case with a length scale ℓ, or

$$k(\mathbf{x}_i, \mathbf{x}_j) = s^2 \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{L}^{-2} (\mathbf{x}_i - \mathbf{x}_j)\right), \tag{5}$$

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = s^2 \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{L}^{-2} (\mathbf{x}_i - \mathbf{x}_j)\right) + \sigma_n^2 \delta_{ij}, \tag{6}$$

for the case with a length-scale matrix $\mathbf{L}$.

3 Derivatives of Exponentiated Quadratic Kernel

The derivatives of (4) with respect to hyperparameters, s, ℓ, and σ_n are defined as

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial s} = 2s \exp\left(-\frac{1}{2\ell^2}(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)\right), \tag{7}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell} = s^2 \exp\left(-\frac{1}{2\ell^2}(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)\right) \cdot \frac{1}{\ell^3}(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j), \tag{8}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_n} = 2\sigma_n \delta_{ij}. \tag{9}$$

The derivatives of (6) with respect to hyperparameters, s, $\ell_1, \ldots, \ell_d$, and σ_n are defined as

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial s} = 2s \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{L}^{-2} (\mathbf{x}_i - \mathbf{x}_j)\right), \tag{10}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell_k} = s^2 \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{L}^{-2} (\mathbf{x}_i - \mathbf{x}_j)\right) \cdot \frac{1}{\ell_k^3}(x_{ik} - x_{jk})^2, \tag{11}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_n} = 2\sigma_n \delta_{ij}, \tag{12}$$

for $k = 1, \ldots, d$, where x_{ik} is the k-th element of $\mathbf{x}_i$. The elaborate derivation of (11) is

$$\begin{aligned}
\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell_k} &= \frac{\partial}{\partial \ell_k} \left(s^2 \exp\left(-\frac{1}{2} \mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right) + \sigma_n^2 \delta_{ij} \right) \\
&= s^2 \exp\left(-\frac{1}{2} \mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right) \cdot \frac{\partial}{\partial \ell_k} \left(-\frac{1}{2} \mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right) \\
&= s^2 \exp\left(-\frac{1}{2} \mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right) \cdot -\frac{1}{2} \cdot \frac{\partial}{\partial \ell_k} \left(\frac{(x_{i1} - x_{j1})^2}{\ell_1^2} + \cdots + \frac{(x_{id} - x_{jd})^2}{\ell_d^2} \right) \\
&= s^2 \exp\left(-\frac{1}{2} \mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right) \cdot \frac{1}{\ell_k^3} (x_{ik} - x_{jk})^2. \qquad (13)
\end{aligned}$$

4 Matérn 3/2 Kernel

Maérn 3/2 kernel is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \exp\left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right), \qquad (14)$$

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \exp\left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) + \sigma_n^2 \delta_{ij}, \qquad (15)$$

for the case with a length scale ℓ, or

$$k(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \exp\left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right), \qquad (16)$$

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \exp\left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) + \sigma_n^2 \delta_{ij}, \qquad (17)$$

for the case with a length-scale matrix $\mathbf{L}$.

5 Derivatives of Matérn 3/2 Kernel

The derivatives of (15) with respect to hyperparameters, s, ℓ, and σ_n are defined as

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial s} = 2s \left(1 + \frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \exp\left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right), \qquad (18)$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell} = 3s^2 \exp\left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \cdot \frac{1}{\ell^3} \mathbf{d}_{ij}^\top \mathbf{d}_{ij}, \qquad (19)$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_n} = 2\sigma_n \delta_{ij}. \qquad (20)$$

The elaborate derivation of (19) is

$$\begin{aligned}
\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell} &= \frac{\partial}{\partial \ell} \left(s^2 \left(1 + \frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \exp \left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) + \sigma_n^2 \delta_{ij} \right) \\
&= s^2 \exp \left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \cdot \frac{\sqrt{3}}{\ell^2} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \\
&\quad - s^2 \exp \left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \cdot \frac{\sqrt{3}}{\ell^2} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \\
&\quad + \frac{s^2 \sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \exp \left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right) \cdot \frac{\sqrt{3}}{\ell^2} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \\
&= \frac{3 s^2}{\ell^3} \mathbf{d}_{ij}^\top \mathbf{d}_{ij} \exp \left(-\frac{\sqrt{3}}{\ell} \sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \right). \qquad (21)
\end{aligned}$$

The derivatives of (17) with respect to hyperparameters, s, $\ell_1, \ldots, \ell_d$, and σ_n are defined as

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial s} = 2s \left(1 + \sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right), \qquad (22)$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell_k} = 3 s^2 \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot \frac{1}{\ell_k^3} (x_{ik} - x_{jk})^2, \qquad (23)$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_n} = 2 \sigma_n \delta_{ij}, \qquad (24)$$

for $k = 1, \ldots, d$, where x_{ik} is the k-th element of $\mathbf{x}_i$. The elaborate derivation of (23) is

$$\begin{aligned}
\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell_k} &= \frac{\partial}{\partial \ell_k} \left(s^2 \left(1 + \sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) + \sigma_n^2 \delta_{ij} \right) \\
&= s^2 \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot -\sqrt{3} \cdot \frac{\partial}{\partial \ell_k} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \\
&\quad + s^2 \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot \sqrt{3} \cdot \frac{\partial}{\partial \ell_k} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \\
&\quad + s^2 \sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot -\sqrt{3} \cdot \frac{\partial}{\partial \ell_k} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \\
&= -3 s^2 \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot \frac{\partial}{\partial \ell_k} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \\
&= -3 s^2 \left(\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right)^{\frac{1}{2}} \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot \frac{1}{2} \left(\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right)^{-\frac{1}{2}} \\
&\quad \cdot \frac{\partial}{\partial \ell_k} \left(\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij} \right) \\
&= -3 s^2 \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \cdot \frac{1}{2} \cdot \frac{\partial}{\partial \ell_k} \left(\frac{(x_{i1} - x_{j1})^2}{\ell_1^2} + \cdots + \frac{(x_{id} - x_{jd})^2}{\ell_d^2} \right) \\
&= 3 s^2 \exp \left(-\sqrt{3} \sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} \right) \frac{(x_{ik} - x_{jk})^2}{\ell_k^3}. \qquad (25)
\end{aligned}$$

6 Matérn 5/2 Kernel

Maérn 5/2 kernel is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} + \frac{5}{3\ell^2}\mathbf{d}_{ij}^\top \mathbf{d}_{ij}\right) \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right), \tag{26}$$

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} + \frac{5}{3\ell^2}\mathbf{d}_{ij}^\top \mathbf{d}_{ij}\right) \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) + \sigma_n^2 \delta_{ij}, \tag{27}$$

for the case with a length scale ℓ, or

$$k(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \sqrt{5}\sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} + \frac{5}{3}\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}\right) \exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}}\right), \tag{28}$$

$$\widehat{k}(\mathbf{x}_i, \mathbf{x}_j) = s^2 \left(1 + \sqrt{5}\sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}} + \frac{5}{3}\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}\right) \exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top \mathbf{L}^{-2} \mathbf{d}_{ij}}\right) + \sigma_n^2 \delta_{ij}, \tag{29}$$

for the case with a length-scale matrix $\mathbf{L}$.

7 Derivatives of Matérn 5/2 Kernel

The derivatives of (27) with respect to hyperparameters, s, ℓ, and σ_n are defined as

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial s} = 2s \left(1 + \frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} + \frac{5}{3\ell^2}\mathbf{d}_{ij}^\top \mathbf{d}_{ij}\right) \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right), \tag{30}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell} = \frac{5s^2}{3} \left(1 + \frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) \cdot \frac{1}{\ell^3}\mathbf{d}_{ij}^\top \mathbf{d}_{ij}, \tag{31}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_n} = 2\sigma_n \delta_{ij}. \tag{32}$$

The elaborate derivation of (31) is

$$\begin{aligned}
\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell} &= \frac{\partial}{\partial \ell} \left(s^2 \left(1 + \frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} + \frac{5}{3\ell^2}\mathbf{d}_{ij}^\top \mathbf{d}_{ij}\right) \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) + \sigma_n^2 \delta_{ij}\right) \\
&= s^2 \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) \cdot \frac{\sqrt{5}}{\ell^2}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \\
&\quad - s^2 \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) \cdot \frac{\sqrt{5}}{\ell^2}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \\
&\quad + \frac{s^2\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right) \cdot \frac{\sqrt{5}}{\ell^2}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}} \\
&\quad - \frac{10s^2}{3\ell^3}\mathbf{d}_{ij}^\top \mathbf{d}_{ij} \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top \mathbf{d}_{ij}}\right)
\end{aligned}$$

$$+ \frac{5s^2}{3\ell^2}\mathbf{d}_{ij}^\top\mathbf{d}_{ij} \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}\right) \cdot \frac{\sqrt{5}}{\ell^2}\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}$$

$$= \frac{5s^2}{3\ell^3}\mathbf{d}_{ij}^\top\mathbf{d}_{ij} \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}\right)$$

$$+ \frac{5s^2\sqrt{5}}{3\ell^4}\mathbf{d}_{ij}^\top\mathbf{d}_{ij} \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}\right)\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}$$

$$= \frac{5s^2}{3\ell^3}\mathbf{d}_{ij}^\top\mathbf{d}_{ij} \exp\left(-\frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}\right)\left(1 + \frac{\sqrt{5}}{\ell}\sqrt{\mathbf{d}_{ij}^\top\mathbf{d}_{ij}}\right). \tag{33}$$

The derivatives of (29) with respect to hyperparameters, s, $\ell_1, \ldots, \ell_d$, and σ_n are defined as

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial s} = 2s\left(1 + \sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}} + \frac{5}{3}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\right)\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right), \tag{34}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell_k} = \frac{5s^2}{3}\left(1 + \sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot \frac{1}{\ell_k^3}(x_{ik} - x_{jk})^2, \tag{35}$$

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_n} = 2\sigma_n\delta_{ij}, \tag{36}$$

for $k = 1, \ldots, d$, where x_{ik} is the k-th element of $\mathbf{x}_i$. The elaborate derivation of (35) is

$$\frac{\partial \widehat{k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \ell_k}$$

$$= \frac{\partial}{\partial \ell_k}\left(s^2\left(1 + \sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}} + \frac{5}{3}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\right)\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) + \sigma_n^2\delta_{ij}\right)$$

$$= s^2\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot -\sqrt{5} \cdot \frac{\partial}{\partial \ell_k}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}$$

$$+ s^2\sqrt{5}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot \frac{\partial}{\partial \ell_k}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}$$

$$+ s^2\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot -\sqrt{5} \cdot \frac{\partial}{\partial \ell_k}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}$$

$$+ \frac{5s^2}{3}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot \frac{\partial}{\partial \ell_k}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}$$

$$+ \frac{5s^2}{3}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot -\sqrt{5} \cdot \frac{\partial}{\partial \ell_k}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}$$

$$= -5s^2\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot \frac{\partial}{\partial \ell_k}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}$$

$$+ \frac{5s^2}{3}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot \frac{\partial}{\partial \ell_k}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}$$

$$- \frac{5s^2\sqrt{5}}{3}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right) \cdot \frac{\partial}{\partial \ell_k}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}$$

$$
\begin{aligned}
&= -5s^2\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{\partial}{\partial\ell_k}\left(\frac{(x_{i1}-x_{j1})^2}{\ell_1^2}+\cdots+\frac{(x_{id}-x_{jd})^2}{\ell_d^2}\right)^{\frac{1}{2}}\\
&\quad+\frac{5s^2}{3}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{\partial}{\partial\ell_k}\left(\frac{(x_{i1}-x_{j1})^2}{\ell_1^2}+\cdots+\frac{(x_{id}-x_{jd})^2}{\ell_d^2}\right)\\
&\quad-\frac{5s^2\sqrt{5}}{3}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{\partial}{\partial\ell_k}\left(\frac{(x_{i1}-x_{j1})^2}{\ell_1^2}+\cdots+\frac{(x_{id}-x_{jd})^2}{\ell_d^2}\right)^{\frac{1}{2}}\\
&= -5s^2\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{1}{2}\left(\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\right)^{-\frac{1}{2}}\cdot -2\cdot\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\\
&\quad-\frac{10s^2}{3}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\\
&\quad-\frac{5s^2\sqrt{5}}{3}\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{1}{2}\left(\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}\right)^{-\frac{1}{2}}\cdot -2\cdot\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\\
&= 5s^2\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\\
&\quad-\frac{10s^2}{3}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\\
&\quad+\frac{5s^2\sqrt{5}}{3}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\cdot\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\\
&= \frac{5s^2}{3}\exp\left(-\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right)\frac{(x_{ik}-x_{jk})^2}{\ell_k^3}\left(1+\sqrt{5}\sqrt{\mathbf{d}_{ij}^\top\mathbf{L}^{-2}\mathbf{d}_{ij}}\right). \qquad (37)
\end{aligned}
$$

References

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.