Analytical Derivatives of Kernel Functions

A. Ho

August 24, 2017

1 Gaussian Processes

Given a one-dimensional set of data, (X, Y), a kernel function, $k(x_1, x_2, \theta)$, describing the covariance between the data points, and a measurement error function, $r(x)$, a Gaussian process regression attempts to fit the data in a statistically rigourous manner. One advantage of this approach lies in the replacement of pre-defined basis functions with the kernel, which allows for a more generalized fit. However, the disadvantage is that the fitted function cannot be expressed in an analytical form, excluding further analysis of the fits with certain procedures.

A GPR prediction of the fit and its confidence interval, evaluated at the points, X_*, can be made using the following set of equations:

$$
\begin{aligned}
\mathbb{E}[Y_*] &= K(X_*, X) \left[K + R \right]^{-1} Y \\
\mathbb{V}[Y_*] &= K_* + R_* + K(X_*, X) \left[K + R \right]^{-1} K(X, X_*)
\end{aligned}
\tag{1}
$$

where $K(X_1, X_2) = k(X_1, X_2, \theta)$, $R(X_1, X_2) = r(X_1)\, r(X_2)\, \delta_{X_1, X_2}$, and the subscripts 1 and 2 indicate separate instances of the same set of data. Additionally, the shorthand notation $K = K(X, X)$, $R = R(X, X)$, $K_* = K(X_*, X_*)$, and $R_* = R(X_*, X_*)$ was used to improve the readability of the equation.

Within this framework, a prediction of the derivative of the fit and its confidence interval can also be made in a straight-forward manner by using the derivative of the kernel, as such:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial Y_*}{\partial X_*} \right] &= \frac{\partial K(X_*, X)}{\partial X_1} \left[K + R \right]^{-1} Y \\
\mathbb{V}\left[\frac{\partial Y_*}{\partial X_*} \right] &= \frac{\partial^2 K_*}{\partial X_1 \partial X_2} + \frac{\partial^2 R_*}{\partial X_1 \partial X_2} + \frac{\partial K(X_*, X)}{\partial X_1} \left[K + R \right]^{-1} \frac{\partial K(X, X_*)}{\partial X_2}
\end{aligned}
\tag{2}
$$

where X_1, X_2 refer to the first and second instances, respectively, of the same coordinate system, as described in the kernel function.

The free parameters of this fitting procedure are described by θ, known as the *hyperparameters*, which can then be optimized as desired by the user. One method of optimizing these hyperparameters is by maximizing the *log-marginal-likelihood*, expressed as follows:

$$
\log p(Y|X) = -\frac{1}{2} Y^T \left[K + R \right]^{-1} Y - \frac{1}{2} \log |K + R| - \frac{1}{2} N \log 2\pi
\tag{3}
$$

where the vertical brackets indicate the determinant of the enclosed matrix and N represents the number of data points. It should be noted that this optimization method selects the hyperparameters such that the fit has the highest probability to match the input data, but provides no guarantee that the physical process behind the data is modelled correctly.

The maximization process can be accelerated by employing the derivative of the log-marginal-likelihood with respect the hyperparameters, as such:

$$\frac{\partial}{\partial \theta_j} \log p(Y|X) = \frac{1}{2} Y^T \left[K + R\right]^{-1} \frac{\partial K}{\partial \theta_j} \left[K + R\right]^{-1} Y$$
$$- \frac{1}{2} \text{tr} \left(\left[K + R\right]^{-1} \frac{\partial K}{\partial \theta_j} \right) \quad (4)$$

which requires the derivative of the kernel function with respect to the hyperparameters.

The following sections describe the kernels programmed within "GPR1D" and their corresponding derivatives. In order to preserve the functionality of the algorithm, all consecutive derivative order must alternate between the two coordinate variables, x_1 and x_2, which can be denoted mathematically as follows:

$$\frac{\partial^n k(x_1, x_2)}{\partial x^n} \equiv \frac{\partial^a}{\partial x_1^a} \frac{\partial^b}{\partial x_2^b} k(x_1, x_2) \quad (5)$$

where $a + b = n$ and $a - b \in \{-1, 0, 1\}$.

2 Squared Exponential Kernel

This kernel is mathematically expressed as:

$$k(x_1, x_2) = \sigma^2 \exp\left(-\frac{r^2}{2l^2}\right), \qquad r = x_1 - x_2 \quad (6)$$

$$\textbf{Hyperparameters:} \qquad \theta = \{\sigma, l\}$$

A generalized form of the nth-derivative of this kernel can be expressed as follows:

$$\frac{\partial^n k}{\partial x^n} = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \frac{\sigma^2}{l^n} \exp\left(-\frac{r^2}{2l^2}\right) \sum_{i=0}^{n/2} C_{n,i} \left(\frac{r}{l}\right)^{n-2i} \quad (7)$$

where the upper limit of the sum is always rounded down to the nearest integer,

$$\frac{\partial r}{\partial x_1} = 1, \quad \frac{\partial r}{\partial x_2} = -1, \quad (8)$$

and

$$C_{n,i} = (-1)^{n-i} \frac{1}{2^i \, i!} \frac{n!}{(n - 2i)!} \quad (9)$$

2.1 Hyperparameter σ Derivatives

The hyperparameter σ derivative of Equation (6) can be expressed as follows:

$$\frac{\partial}{\partial \sigma} [k(x_1, x_2)] = 2\sigma \exp\left(-\frac{r^2}{2l^2}\right) \tag{10}$$

A generalized form of the hyperparameter σ derivative of Equation (7) can be expressed as follows:

$$\frac{\partial}{\partial \sigma}\left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \frac{2\sigma}{l^n} \exp\left(-\frac{r^2}{2l^2}\right) \sum_{i=0}^{n/2} C_{n,i} \left(\frac{r}{l}\right)^{n-2i} \tag{11}$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $C_{n,i}$ is given by Equation (9) and the upper limit of the sum is always rounded down to the nearest integer.

2.2 Hyperparameter l Derivatives

The hyperparameter l derivative of Equation (6) can be expressed as follows:

$$\frac{\partial}{\partial l} [k(x_1, x_2)] = \frac{\sigma^2 r^2}{l^3} \exp\left(-\frac{r^2}{2l^2}\right) \tag{12}$$

A generalized form of the hyperparameter l derivative of Equation (7) can be expressed as follows:

$$\frac{\partial}{\partial l}\left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \frac{\sigma^2}{l^{n+1}} \exp\left(-\frac{r^2}{2l^2}\right) \sum_{i=0}^{(n+2)/2} L_{n,i} \left(\frac{r}{l}\right)^{n+2-2i} \tag{13}$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), the upper limit of the sum is always rounded down to the nearest integer, and

$$L_{n,i} = (-1)^{n+2-i} \frac{1}{2^i\, i!} \frac{n!}{(n+2-2i)!} \left[\frac{(n+2)!}{n!} - 2i\right], \tag{14}$$

3 Rational Quadratic Kernel

This kernel is mathematically expressed as:

$$k(x_1, x_2) = \sigma^2 Q^{-\alpha}, \qquad r = x_1 - x_2 \tag{15}$$

where

$$Q = 1 + \frac{r^2}{2\alpha l^2} \tag{16}$$

$$\textbf{Hyperparameters:} \qquad \theta = \{\sigma, l, \alpha\}$$

3

A generalized form of the nth-derivative of this kernel can be expressed as follows:

$$\frac{\partial^n k}{\partial x^n} = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \frac{\sigma^2}{l^n} Q^{-\alpha-n} \sum_{i=0}^{n/2} C_{n,i} \frac{\Gamma(\alpha+n-i)}{\alpha^{n-i}\,\Gamma(\alpha)} \left(\frac{r}{l}\right)^{n-2i} Q^i \qquad (17)$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $C_{n,i}$ is given by Equation (9), the upper limit of the sum is always rounded down to the nearest integer, and $\Gamma(z)$ is the *Gamma function*, defined as such:

$$\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x)\,dx \qquad (18)$$

which has the following property, exploited within Equation (17):

$$\Gamma(z+1) = z\,\Gamma(z)\ , \qquad \text{for any } z > 0 \qquad (19)$$

3.1   Hyperparameter σ Derivatives

The hyperparameter σ derivative of Equation (15) can be expressed as follows:

$$\frac{\partial}{\partial \sigma}\left[k(x_1,x_2)\right] = 2\sigma\,Q^{-\alpha} \qquad (20)$$

A generalized form of the hyperparameter σ derivative of Equation (17) can be expressed as follows:

$$\frac{\partial}{\partial \sigma}\left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \frac{2\sigma}{l^n} Q^{-\alpha-n} \sum_{i=0}^{n/2} C_{n,i} \frac{\Gamma(\alpha+n-i)}{\alpha^{n-i}\,\Gamma(\alpha)} \left(\frac{r}{l}\right)^{n-2i} Q^i \qquad (21)$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $C_{n,i}$ is given by Equation (9) and the upper limit of the sum is always rounded down to the nearest integer.

3.2   Hyperparameter l Derivatives

The hyperparameter l derivative of Equation (15) can be expressed as follows:

$$\frac{\partial}{\partial l}\left[k(x_1,x_2)\right] = \frac{\sigma^2 r^2}{l^3} Q^{-\alpha-1} \qquad (22)$$

A generalized form of the hyperparameter l derivative of Equation (17) can be expressed as follows:

$$\frac{\partial}{\partial l}\left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \frac{\sigma^2}{l^{n+1}} Q^{-\alpha-n}$$

$$\sum_{i=0}^{(n+2)/2} L_{n,i} \frac{\Gamma(\alpha+n+1-i)}{\alpha^{n+1-i}\,\Gamma(\alpha)} \left(\frac{r}{l}\right)^{n+2-2i} Q^i \qquad (23)$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $L_{n,i}$ is given by Equation (14) and the upper limit of the sum is always rounded down to the nearest integer.

3.3 Hyperparameter α Derivatives

The hyperparameter α derivative of Equation (15) can be expressed as follows:

$$\frac{\partial}{\partial \alpha}\left[k(x_1, x_2)\right] = \sigma^2 \left(\frac{r^2}{2\alpha l^2} - Q \ln Q\right) Q^{-\alpha-1} \tag{24}$$

A generalized form of the hyperparameter α derivative of Equation (17) can be expressed as follows:

$$\frac{\partial}{\partial l}\left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a}\frac{\partial^b r}{\partial x_2^b}\frac{\sigma^2}{l^n}Q^{-\alpha-n-1}\sum_{i=0}^{n/2}C_{n,i}\frac{\Gamma(\alpha+n-i)}{\alpha^{n-i}\Gamma(\alpha)}\left(\frac{r}{l}\right)^{n-2i}Q^i \times$$
$$\left[\frac{\alpha-2i}{\alpha}\frac{r^2}{2\alpha l^2} - \frac{n-i}{\alpha} - Q\left(\ln Q + \psi(\alpha+n-i) - \psi(\alpha)\right)\right] \tag{25}$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $C_{n,i}$ is given by Equation (9) and the upper limit of the sum is always rounded down to the nearest integer.

4 Matern Half-Integer Kernel

This kernel is mathematically expressed as:

$$k(x_1, x_2) = \sigma^2 \exp\left(-\frac{\sqrt{2\nu}r}{l}\right) S_{p,0} \tag{26}$$

$$r = x_1 - x_2, \qquad p \in \mathbb{N}^+, \qquad \nu = p + \frac{1}{2}$$

where

$$S_{p,i} = \begin{cases} \displaystyle\sum_{z=0}^{p-i}\frac{p!}{(2p)!}\frac{(p+z)!}{z!\,(p-i-z)!}\left(\frac{\sqrt{8\nu}r}{l}\right)^{p-i-z}, & \text{for } i \le p \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

$$\textbf{Hyperparameters:} \qquad \theta = \{\sigma, l, p\}$$

A generalized form of the nth-derivative of this kernel can be expressed as follows:

$$\frac{\partial^n k}{\partial x^n} = \frac{\partial^a r}{\partial x_1^a}\frac{\partial^b r}{\partial x_2^b}\left(\sqrt{2\nu}\right)^n\frac{\sigma^2}{l^n}\exp\left(-\frac{\sqrt{2\nu}r}{l}\right)\sum_{i=0}^{n}M_{n,i}\,S_{p,i} \tag{28}$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $S_{p,i}$ is given by Equation (27) and

$$M_{n,i} = (-1)^{n-i}\frac{2^i\,n!}{i!\,(n-i)!} \tag{29}$$

4.1 Hyperparameter σ Derivatives

The hyperparameter σ derivative of Equation (26) can be expressed as follows:

$$\frac{\partial}{\partial \sigma} [k(x_1, x_2)] = 2\sigma \exp\left(-\frac{\sqrt{2\nu}r}{l}\right) S_{p,0} \tag{30}$$

A generalized form of the hyperparameter σ derivative of Equation (28) can be expressed as follows:

$$\frac{\partial}{\partial \sigma} \left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \left(\sqrt{2\nu}\right)^n \frac{2\sigma}{l^n} \exp\left(-\frac{\sqrt{2\nu}r}{l}\right) \sum_{i=0}^{n} M_{n,i} S_{p,i} \tag{31}$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), $S_{p,i}$ is given by Equation (27) and $M_{n,i}$ is given by Equation (29).

4.2 Hyperparameter l Derivatives

The hyperparameter l derivative of Equation (26) can be expressed as follows:

$$\frac{\partial}{\partial l} [k(x_1, x_2)] = \frac{\sqrt{2\nu}r}{l^2} \sigma^2 \exp\left(-\frac{\sqrt{2\nu}r}{l}\right) (S_0 - 2S_1) \tag{32}$$

A generalized form of the hyperparameter l derivative of Equation (28) can be expressed as follows:

$$\frac{\partial}{\partial l} \left[\frac{\partial^n k}{\partial x^n}\right] = \frac{\partial^a r}{\partial x_1^a} \frac{\partial^b r}{\partial x_2^b} \left(\sqrt{2\nu}\right)^n \frac{\sigma^2}{l^{n+1}} \exp\left(-\frac{\sqrt{2\nu}r}{l}\right)$$
$$\left[-\frac{\sqrt{2\nu}r}{l} \sum_{i=0}^{n+1} M_{n+1,i} S_{p,i} - n \sum_{j=0}^{n} M_{n,j} S_{p,j}\right] \tag{33}$$

where $\partial r/\partial x_{1,2}$ and $\partial r/\partial x_2$ are given by Equation (8), $S_{p,i}$ and $S_{p,j}$ are given by Equation (27) and $M_{n,i}$ and $M_{n+1,i}$ are given by Equation (29).

4.3 Additional Notes

Due to restrictions of the hyperparameter, p, resulting from the differentiability requirements of the kernel, it was decided that it would not be treated as a hyperparameter within the "GPR1D" implementation. Instead, p is treated as a constant, with a user-specified value, which is not altered during any of the optimization procedures.

5   Gibbs Kernel

This kernel is mathematically expressed as:

$$k(x_1, x_2) = \sigma^2 \sqrt{\frac{2P}{S}} \exp\left(-\frac{r^2}{S}\right), \qquad r = x_1 - x_2 \tag{34}$$

where

$$P = l(x_1)\, l(x_2), \quad S = l^2(x_1) + l^2(x_2) \tag{35}$$

and $l(z)$ can be any sufficiently smooth function describing how the length-scale changes with the independent coordinate, denoted here as z for clarity. This length-scale, or *warping*, function may contain its own set of hyperparameters, denoted as θ_l, but the function itself must be at least first-order differentiable in order to apply an optimization scheme to them. More complex schemes may require higher-order differntiability. Examples of such warping functions, $l(z)$, will be provided later in this section.

$$\textbf{Hyperparameters:} \qquad \theta = \{\sigma, \theta_l\}$$

A generalized form of the nth-derivative of this kernel can be expressed as follows:

$$\frac{\partial^n k}{\partial x^n} = \sigma^2 \sqrt{\frac{2P}{S}} \exp\left(-\frac{r^2}{S}\right) F_{1,2}^{(n)} \tag{36}$$

where $F_{1,2}^{(n)}$ represents the multiplicative factor resulting from the formulation of the nth-derivative, with the subscript indicating whether the derivative is done with respect to x_1 or x_2 first. A general form of this factor that is applicable to all possible derivative orders has not yet been determined, but the zeroth-, first- and second-derivative factors are provided in this document.

The zeroth-derivative factor, $F_{1,2}^{(0)}$, can be expressed as follows:

$$F_{1,2}^{(0)} = 1 \tag{37}$$

It should be noted that since the derivatives always alternate between being with respect to x_1 and x_2, all even values of n, including $n = 0$, yield expressions which are invariant to the subscript of $F_{1,2}^{(n)}$.

The first-derivative factor, $F_{1,2}^{(1)}$, can be expressed as follows:

$$F_{1,2}^{(1)} = -\frac{2r}{S}\frac{\partial r}{\partial x_{1,2}} + \frac{1}{2P}\frac{\partial P}{\partial x_{1,2}} + \left(\frac{r^2}{S^2} - \frac{1}{2S}\right)\frac{\partial S}{\partial x_{1,2}} \tag{38}$$

where $\partial r / \partial x_{1,2}$ and $\partial r / \partial x_2$ are given by Equation (8), P and S are given by Equation (35) and

$$\frac{\partial P}{\partial x_{1,2}} = l(x_{2,1})\frac{\partial l(x_{1,2})}{\partial x_{1,2}}, \quad \frac{\partial S}{\partial x_{1,2}} = 2l(x_{1,2})\frac{\partial l(x_{1,2})}{\partial x_{1,2}} \tag{39}$$

The second-order factor, $F_{1,2}^{(2)}$, can be expressed as follows:

$$F_{1,2}^{(2)} = \frac{\partial r}{\partial x_1} \frac{\partial r}{\partial x_2} \left(\frac{4r^2}{S^2} - \frac{2}{S} \right) - \frac{A}{S} \left(\frac{4r^2}{S^2} - \frac{6}{S} \right) - \frac{B}{S} +$$

$$\frac{\partial^2 P}{\partial x_1 \partial x_2} \left(\frac{4P\,r^4}{S^4} - \frac{12P\,r^2}{S^3} + \frac{3P}{S^2} + \frac{r^2}{SP} - \frac{1}{4P} \right) \quad (40)$$

where $\partial r/\partial x_1$ and $\partial r/\partial x_2$ are given by Equation (8), P and S are given by Equation (35),

$$\frac{\partial^2 P}{\partial x_1 \partial x_2} = \frac{1}{P} \frac{\partial P}{\partial x_1} \frac{\partial P}{\partial x_2} = \frac{1}{4P} \frac{\partial S}{\partial x_1} \frac{\partial S}{\partial x_2} = \frac{\partial l(x_1)}{\partial x_1} \frac{\partial l(x_2)}{\partial x_2}$$

$$\frac{\partial^2 S}{\partial x_1 \partial x_2} = 0 \quad (41)$$

and

$$A = \frac{r}{P} \left(\frac{\partial r}{\partial x_1} \frac{\partial P}{\partial x_2} + \frac{\partial r}{\partial x_2} \frac{\partial P}{\partial x_1} \right)$$

$$B = \frac{r}{2} \left(\frac{\partial r}{\partial x_1} \frac{\partial S}{\partial x_2} + \frac{\partial r}{\partial x_2} \frac{\partial S}{\partial x_1} \right) \quad (42)$$

where the derivatives of P and S are given by Equation (39).

5.1 Hyperparameter σ Derivatives

The hyperparameter σ derivative of Equation (34) can be expressed as follows:

$$\frac{\partial}{\partial \sigma} \left[k(x_1, x_2) \right] = 2\sigma \sqrt{\frac{2P}{S}} \exp \left(-\frac{r^2}{S} \right) \quad (43)$$

where P and S are given by Equation (35).

The hyperparameter σ derivative of Equation (36) can be expressed as follows:

$$\frac{\partial}{\partial \sigma} \left[\frac{\partial^n k}{\partial x^n} \right] = 2\sigma \sqrt{\frac{2P}{S}} \exp \left(-\frac{r^2}{S} \right) F_{1,2}^{(n)} \quad (44)$$

where $F_{1,2}^{(n)}$ is given for the zeroth-, first- and second-derivatives by Equations (37), (38) and (40), respectively.

5.2 General Hyperparameter $\theta_{l,i}$ Derivatives

The generalized form of the hyperparameter $\theta_{l,i}$ derivative of Equation (34) can be expressed as follows:

$$\frac{\partial}{\partial \theta_{l,i}} \left[k(x_1, x_2) \right] = \sigma^2 \sqrt{\frac{2P}{S}} \exp \left(-\frac{r^2}{S} \right) G \quad (45)$$

where P and S are given by Equation (35) and

$$G = \frac{1}{2P} \frac{\partial P}{\partial \theta_{l,i}} + \left(\frac{r^2}{S^2} - \frac{1}{2S} \right) \frac{\partial S}{\partial \theta_{l,i}} \tag{46}$$

where

$$
\begin{aligned}
\frac{\partial P}{\partial \theta_{l,i}} &= \frac{\partial l(x_1)}{\partial \theta_{l,i}} l(x_2) + l(x_1) \frac{\partial l(x_2)}{\partial \theta_{l,i}} \\
\frac{\partial S}{\partial \theta_{l,i}} &= 2l(x_1) \frac{\partial l(x_1)}{\partial \theta_{l,i}} + 2l(x_2) \frac{\partial l(x_2)}{\partial \theta_{l,i}}
\end{aligned}
\tag{47}
$$

where the set of all $\partial l(z) / \partial \theta_{l,i}$ are dependent on the details of chosen warping function, $l(z)$.

The generalized form of the hyperparameter $\theta_{l,i}$ derivative of Equation (36) can be expressed as follows:

$$\frac{\partial}{\partial \theta_{l,i}} \left[\frac{\partial^n k}{\partial x^n} \right] = \sigma^2 \sqrt{\frac{2P}{S}} \exp \left(-\frac{r^2}{S} \right) \left(F_{1,2}^{(n)} G + \frac{\partial F_{1,2}^{(n)}}{\partial \theta_{l,i}} \right) \tag{48}$$

where P and S are given by Equation (35), $F_{1,2}^{(n)}$ is given for the zeroth-, first- and second-derivatives by Equations (37), (38) and (40), respectively, and G is given by Equation (46). Similar to $F_{1,2}^{(n)}$, a general formulation for $\partial F_{1,2}^{(n)} / \partial \theta_{l,i}$ that is applicable to all possible derivative orders has not yet been determined, but the zeroth-, first- and second-derivative factors are provided in this document.

The hyperparameter $\theta_{l,i}$ derivative of the zeroth-derivative factor, $F_{1,2}^{(0)}$, can be expressed as follows:

$$\frac{\partial F_{1,2}^{(0)}}{\partial \theta_{l,i}} = 0 \tag{49}$$

The hyperparameter $\theta_{l,i}$ derivative of the first-derivative factor, $F_{1,2}^{(1)}$, can be expressed as follows:

$$
\begin{aligned}
\frac{\partial F_{1,2}^{(1)}}{\partial \theta_{l,i}} = &-\frac{1}{2P^2} \frac{\partial P}{\partial x_{1,2}} \frac{\partial P}{\partial \theta_{l,i}} + \left[\frac{2r}{S^2} \frac{\partial r}{\partial x_{1,2}} - \left(\frac{2r^2}{S^3} - \frac{1}{2S^2} \right) \frac{\partial S}{\partial x_{1,2}} \right] \frac{\partial S}{\partial \theta_{l,i}} + \\
&\frac{1}{2P} \frac{\partial}{\partial \theta_{l,i}} \left(\frac{\partial P}{\partial x_{1,2}} \right) + \left(\frac{r^2}{S^2} - \frac{1}{2S} \right) \frac{\partial}{\partial \theta_{l,i}} \left(\frac{\partial S}{\partial x_{1,2}} \right) \quad (50)
\end{aligned}
$$

where $\partial r / \partial x_{1,2}$ and $\partial r / \partial x_2$ are given by Equation (8), P and S are given by Equation (35), their derivatives with respect to x_1 and x_2 are given by Equation (39), their derivatives with respect to $\theta_{l,i}$ given by Equation (47) and

$$
\begin{aligned}
\frac{\partial}{\partial \theta_{l,i}} \left(\frac{\partial P}{\partial x_{1,2}} \right) &= \frac{\partial l(x_{2,1})}{\partial \theta_{l,i}} \frac{\partial l(x_{1,2})}{\partial x_{1,2}} + l(x_{2,1}) \frac{\partial}{\partial \theta_{l,i}} \left(\frac{\partial l(x_{1,2})}{\partial x_{1,2}} \right) \\
\frac{\partial}{\partial \theta_{l,i}} \left(\frac{\partial S}{\partial x_{1,2}} \right) &= 2 \frac{\partial l(x_{1,2})}{\partial \theta_{l,i}} \frac{\partial l(x_{1,2})}{\partial x_{1,2}} + 2l(x_{1,2}) \frac{\partial}{\partial \theta_{l,i}} \left(\frac{\partial l(x_{1,2})}{\partial x_{1,2}} \right)
\end{aligned}
\tag{51}
$$

The hyperparameter $\theta_{l,i}$ derivative of the second-order factor, $F_{1,2}^{(2)}$, can be expressed as follows:

$$\frac{\partial F_{1,2}^{(2)}}{\partial \theta_{l,i}} = H_1 \frac{\partial P}{\partial \theta_{l,i}} + H_2 \frac{\partial S}{\partial \theta_{l,i}} + H_3 +$$
$$\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial^2 P}{\partial x_1 \partial x_2}\right)\left(\frac{4P\,r^4}{S^4} - \frac{12P\,r^2}{S^3} + \frac{3P}{S^2} + \frac{r^2}{SP} - \frac{1}{4P}\right) \quad (52)$$

where P and S are given by Equation (35), their derivatives with respect to $\theta_{l,i}$ are given by Equation (47),

$$H_1 = -\frac{A}{P}\left(\frac{4r^2}{S^3} - \frac{6}{S^2}\right) + \frac{\partial^2 P}{\partial x_1 \partial x_2}\left(\frac{4r^4}{S^4} - \frac{12r^2}{S^3} + \frac{3}{S^2} - \frac{r^2}{SP^2} + \frac{1}{4P^2}\right) \quad (53)$$

$$H_2 = -\frac{\partial r}{\partial x_1}\frac{\partial r}{\partial x_2}\left(\frac{8r^2}{S^3} - \frac{2}{S^2}\right) + \frac{12A}{S^2}\left(\frac{r^2}{S^2} - \frac{1}{S}\right) + \frac{B}{S^2} -$$
$$\frac{\partial^2 P}{\partial x_1 \partial x_2}\left(\frac{16P\,r^4}{S^5} - \frac{36P\,r^2}{S^4} + \frac{6P}{S^3} + \frac{r^2}{S^2 P}\right) \quad (54)$$

$$H_3 = -\frac{r}{P}\left(\frac{4r^2}{S^3} - \frac{6}{S^2}\right)\left[\frac{\partial r}{\partial x_1}\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial P}{\partial x_2}\right) + \frac{\partial r}{\partial x_2}\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial P}{\partial x_1}\right)\right] -$$
$$\frac{r}{2S}\left[\frac{\partial r}{\partial x_1}\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial S}{\partial x_2}\right) + \frac{\partial r}{\partial x_2}\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial S}{\partial x_1}\right)\right] \quad (55)$$

where $\partial r/\partial x_{1,2}$ and $\partial r/\partial x_2$ are given by Equation (8), the derivatives of P and S with respect to x_1 and x_2 are given by Equation (39), and the derivatives of these derivatives with respect to $\theta_{l,i}$ are given by Equation (51), and

$$\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial^2 P}{\partial x_1 \partial x_2}\right) = \frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial l(x_1)}{\partial x_1}\right)\frac{\partial l(x_2)}{\partial x_2} + \frac{\partial l(x_1)}{\partial x_1}\frac{\partial}{\partial \theta_{l,i}}\left(\frac{\partial l(x_2)}{\partial x_2}\right) \quad (56)$$

5.3 Constant Warping Function

This warping function essentially reduces the Gibbs kernel back into the Squared Exponential kernel, outlined in Section 2, and can be expressed as such:

$$l(z) = l_c \quad (57)$$

$$\textbf{Hyperparameters:} \qquad \theta_l = \{l_c\}$$

A generalized form of the nth-derivative of this warping function can be expressed as follows:

$$\frac{\partial^n l}{\partial z^n} = l_c\,\delta_{n,0} \quad (58)$$

where

$$\delta_{i,j} = \begin{cases} 1\,, & i = j \\ 0\,, & \text{otherwise} \end{cases} \quad (59)$$

10

5.3.1 Hyperparameter l_c Derivatives

The generalized form of the hyperparameter l_c derivative of Equation (58) can be expressed as follows:

$$\frac{\partial}{\partial l_c}\left[\frac{\partial^n l(z)}{\partial z^n}\right] = \delta_{n,0} \tag{60}$$

where $\delta_{n,0}$ is given by Equation (59).

5.4 Inverse Gaussian Warping Function

This warping function is useful for reducing the length scale in a single, localized region of interest within the coordinate space, z, and can be mathematically expressed as such:

$$l(z) = l_b - l_p \, \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \tag{61}$$

$$\textbf{Hyperparameters:} \qquad \theta_l = \{l_b, l_p, \mu, \sigma_l\}$$

A generalized form of the nth-derivative of this warping function can be expressed as follows:

$$\frac{\partial^n l}{\partial z^n} = l_b \, \delta_{n,0} - \frac{l_p}{\sigma_l^n} \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \sum_{i=0}^{n/2} C_{n,i} \left(\frac{z-\mu}{\sigma_l}\right)^{n-2i} \tag{62}$$

where $\delta_{n,0}$ is the Kronecker delta as given in Equation (59), $C_{n,i}$ is given by Equation (9) and the upper limit of the sum is always rounded down to the nearest integer.

5.4.1 Hyperparameter l_b Derivatives

The generalized form of the hyperparameter l_b derivative of Equation (62) can be expressed as follows:

$$\frac{\partial}{\partial l_b}\left[\frac{\partial^n l(z)}{\partial z^n}\right] = \delta_{n,0} \tag{63}$$

where $\delta_{n,0}$ is given by Equation (59).

5.4.2 Hyperparameter l_p Derivatives

The hyperparameter l_p derivative of Equation (61) can be expressed as follows:

$$\frac{\partial}{\partial l_p}[l(z)] = -\exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \tag{64}$$

The generalized form of the hyperparameter l_p derivative of Equation (62) can be expressed as follows:

$$\frac{\partial}{\partial l_p}\left[\frac{\partial^n l}{\partial z^n}\right] = -\frac{1}{\sigma_l^n} \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \sum_{i=0}^{n/2} C_{n,i} \left(\frac{z-\mu}{\sigma_l}\right)^{n-2i} \tag{65}$$

where $C_{n,i}$ is given by Equation (9) and the upper limit of the sum is always rounded down to the nearest integer.

5.4.3 Hyperparameter μ Derivatives

The hyperparameter μ derivative of Equation (61) can be expressed as follows:

$$\frac{\partial}{\partial \mu}[l(z)] = -l_p \frac{z-\mu}{\sigma_l^2} \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \tag{66}$$

The generalized form of the hyperparameter μ derivative of Equation (62) can be expressed as follows:

$$\frac{\partial}{\partial \mu}\left[\frac{\partial^n l}{\partial z^n}\right] = \frac{l_p}{\sigma_l^n} \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \sum_{i=0}^{(n+2)/2} C_{n+1,i} \left(\frac{z-\mu}{\sigma_l}\right)^{n+1-2i} \tag{67}$$

where $C_{n,i}$ is given by Equation (9) and the upper limit of the sum is always rounded down to the nearest integer.

5.4.4 Hyperparameter σ_l Derivatives

The hyperparameter σ_l derivative of Equation (61) can be expressed as follows:

$$\frac{\partial}{\partial \sigma_l}[l(z)] = l_p \frac{(z-\mu)^2}{\sigma_l^3} \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \tag{68}$$

The generalized form of the hyperparameter σ_l derivative of Equation (62) can be expressed as follows:

$$\frac{\partial}{\partial \sigma_l}\left[\frac{\partial^n l}{\partial z^n}\right] = \frac{l_p}{\sigma_l^{n+1}} \exp\left(-\frac{(z-\mu)^2}{2\sigma_l^2}\right) \sum_{i=0}^{(n+2)/2} L_{n,i} \left(\frac{z-\mu}{\sigma_l}\right)^{n+2-2i} \tag{69}$$

where $L_{n,i}$ is given by Equation (14) and the upper limit of the sum is always rounded down to the nearest integer.

5.4.5 Additional Notes

Due to strong correlation between the hyperparameter, μ, and the resulting predicted fit from the algorithm, it was decided that it would not be treated as a hyperparameter within the "GPR1D" implementation. Instead, μ is treated as a constant, with a user-specified value, which is not altered during any of the optimization procedures.