1. DEFINITIONS

All definitions given are for 2-dimensional inputs.

1.1. Assume.

$$\left[y^*, y_1^\delta, y_2^\delta, y \,\middle|\, \sigma^2, l\right] \sim N\left(0, K\right) \tag{1}$$

$$K = \begin{bmatrix} K^{00}\left(x^*, x^*\right) & K^{10}\left(x_1^\delta, x^*\right)^T & K^{20}\left(x_2^\delta, x^*\right)^T & K^{00}(x, x^*)^T \\ K^{10}\left(x_1^\delta, x^*\right) & K^{11}\left(x_1^\delta, x_1^\delta\right) & K^{21}\left(x_2^\delta, x_1^\delta\right)^T & K^{01}\left(x, x_1^\delta\right)^T \\ K^{20}\left(x_2^\delta, x^*\right) & K^{21}\left(x_2^\delta, x_1^\delta\right) & K^{22}\left(x_2^\delta, x_2^\delta\right) & K^{02}\left(x, x_2^\delta\right)^T \\ K^{00}\left(x, x^*\right) & K^{01}\left(x, x_1^\delta\right) & K^{02}\left(x, x_2^\delta\right) & K^{00}\left(x, x\right) \end{bmatrix} \tag{2}$$

where

- $y = y(x)$: observed target function values (training set)
- $y^* = y(x^*)$: unobserved function values
- $y_k^\delta = \frac{\partial}{\partial x_k} y\left(x^\delta\right)$: vectors of partial derivatives in the kth input dimension
- σ^2: constant variance parameter
- l: length-scale parameter
- $K^{d_1 d_2}(x_1, x_2)$: covariance functions. The superscripts indicate the argument where the derivative is taken (explicit formulas are in the following section).
  - Let $\dim(x) = ncol(x)$.
  - Then $d_1, d_2 \in \{0, 1, 2, ..., \dim(x)\}$, where d_i indicates the dimension to which the derivative is taken for the i-th argument of $K(\cdot, \cdot)$

1.2. **Covariance functions.** Let

- $d = \dim(x)$
- $d_1, d_2 \in \{0, 1, 2, ..., \dim(x)\}$, where d_i indicates the dimension to which the derivative is taken for the i-th argument of $K(\cdot, \cdot)$
- x_i is a row vector. That is, $x_i = \begin{bmatrix} x_{(i,1)} & x_{(i,2)} & \cdots & x_{(i,d)} \end{bmatrix}$
- $g(\cdot)$ is a correlation function.

1.2.1. $d_1 = d_2 = 0 \Rightarrow$ `"matern"/"sqexp"`.

$$
\begin{aligned}
Cov\left(Y\left(x_1\right), Y\left(x_2\right)\right) &= K^{00}\left(x_1, x_2\right) \\
&= \sigma^2 \prod_{k=1}^{d} g\left(x_{(1,k)}, x_{(2,k)}; l_k\right)
\end{aligned}
$$

1.2.2. $d_1 > 0, d_2 = 0$ *or* $d_1 = 0, d_2 > 0 \Rightarrow$ `"matern1"/"sqexp1"`.

$$
\begin{aligned}
\frac{\partial}{\partial x_{(1,i)}} Cov\left(Y\left(x_1\right), Y\left(x_2\right)\right) &= K^{i0}\left(x_1, x_2\right) \\
&= \sigma^2 \frac{\partial}{\partial x_{(1,i)}} \prod_{k=1}^{d} g\left(x_{(1,k)}, x_{(2,k)}; l_k\right) \\
&= \sigma^2 \left[\prod_{\substack{k=1 \\ k \neq i}}^{d} g\left(x_{(1,k)}, x_{(2,k)}; l_k\right)\right] \left[\frac{\partial}{\partial x_{(1,i)}} g\left(x_{(1,i)}, x_{(2,i)}; l_i\right)\right] \\
\frac{\partial}{\partial x_{(2,i)}} Cov\left(Y\left(x_1\right), Y\left(x_2\right)\right) &= K^{0i}\left(x_1, x_2\right) \\
&= -K^{i0}\left(x_1, x_2\right)
\end{aligned}
$$

1.2.3. $d_1, d_2 > 0 \Rightarrow$ `"matern2"/"sqexp2"`.

$$
\frac{\partial^2}{\partial x_{(2,j)} \partial x_{(1,i)}} Cov\left(Y\left(x_1\right), Y\left(x_2\right)\right) = K^{ij}\left(x_1, x_2\right)
$$

$$
= \begin{cases}
\sigma^2 \left[\prod\limits_{\substack{k=1 \\ k \neq i,j}}^{d} g\left(x_k, x'_k; l_k\right)\right] \left[\frac{\partial^2}{\partial x_{(2,i)} \partial x_{(1,i)}} g\left(x_{(1,i)}, x_{(2,i)}; l_i\right)\right], i = j \\
\sigma^2 \left[\prod\limits_{\substack{k=1 \\ k \neq i,j}}^{d} g\left(x_k, x'_k; l_k\right)\right] \left[\frac{\partial}{\partial x_{(2,j)}} g\left(x_{(1,j)}, x_{(2,j)}; l_j\right)\right] \left[\frac{\partial}{\partial x_{(1,i)}} g\left(x_{(1,i)}, x_{(2,i)}; l_i\right)\right], i \neq j
\end{cases}
$$

1.2.4. *Matern correlation function and its derivatives.* If $g(\cdot,\cdot)$ is from the Matern class of covariance functions with parameter 5/2 there is a more simple form. The subscripts for the inputs in $g(\cdot,\cdot)$ are dropped for legibility.
Let $\theta = \sqrt{5}/l$:

$$g\left(x, x'\right) = \left(1 + \theta\left|x - x'\right| + \frac{1}{3}\theta^2\left|x - x'\right|^2\right)\exp\left\{-\theta\left|x - x'\right|\right\} \tag{3}$$

$$\frac{\partial}{\partial x} g\left(x, x'\right) = -\frac{1}{3}\theta^2\left(x - x'\right)\left[1 + \theta\left|x - x'\right|\right]\exp\left\{-\theta\left|x - x'\right|\right\} \tag{4}$$

$$\frac{\partial^2}{\partial x'\partial x} g\left(x, x'\right) = \frac{1}{3}\theta^2\left[1 + \theta\left|x - x'\right| - \theta^2\left(x - x'\right)^2\right]\exp\left\{-\theta\left|x - x'\right|\right\} \tag{5}$$

See calculations of derivatives of $g(\cdot,\cdot)$ in section 3.

2. Setup

2.1. **The posterior density.**

$$\left[l, \sigma^2, y^*, y_1^\delta, y_2^\delta \left| y\right.\right] \propto \left[y, y^*, y_1^\delta, y_2^\delta \left|\sigma^2, l\right.\right]\left[\sigma^2, l\right] \tag{6}$$

$$\propto \left[y^*, y_1^\delta, y_2^\delta \left| y, \sigma^2, l\right.\right]\left[\left[y\left|\sigma^2, l\right.\right]\right]\left[\sigma^2, l\right] \tag{7}$$

The goal is to evaluate densities on the RHS. Specifically,

$$\left[y^*, y_1^\delta, y_2^\delta \left| y, \sigma^2, l\right.\right] \sim N\left(m, S\right)$$

due to the properties of the multivariate normal distribution (https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Conditional_distributions). That is, the formulas from Kriging follow immediately.

To calculate m, S, start by dividing the covariance matrix K into the appropriate blocks:

$$K = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \text{ where} \tag{8}$$

$$\Sigma_{11} = \begin{bmatrix} K^{00}\left(x^*, x^*\right) & K^{10}\left(x_1^\delta, x^*\right)^T & K^{10}\left(x_2^\delta, x^*\right)^T \\ K^{10}\left(x_1^\delta, x^*\right) & K^{11}\left(x_1^\delta, x_1^\delta\right) & K^{11}\left(x_2^\delta, x_1^\delta\right)^T \\ K^{10}\left(x_2^\delta, x^*\right) & K^{11}\left(x_2^\delta, x_1^\delta\right) & K^{11}\left(x_2^\delta, x_2^\delta\right) \end{bmatrix} \tag{9}$$

$$\Sigma_{12} = \Sigma_{21}^T = \begin{bmatrix} K^{00}\left(x, x^*\right)^T \\ K^{01}\left(x, x_1^\delta\right)^T \\ K^{01}\left(x, x_2^\delta\right)^T \end{bmatrix} = \begin{bmatrix} K^{00}\left(x^*, x\right) \\ K^{10}\left(x_1^\delta, x\right) \\ K^{10}\left(x_2^\delta, x\right) \end{bmatrix} \tag{10}$$

$$\Sigma_{22} = K^{00}\left(x, x\right). \tag{11}$$

Then,

$$m = 0 + \Sigma_{12}\Sigma_{22}^{-1}\left(y - 0\right) = \begin{bmatrix} K^{00}\left(x^*, x\right) \\ K^{10}\left(x_1^\delta, x\right) \\ K^{10}\left(x_2^\delta, x\right) \end{bmatrix} K^{00}(x,x)^{-1}y \tag{12}$$

$$S = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{13}$$

2.2. **Example.** Given

$$y\left(x\right) = 11x_1^{10} + 9x_1^8 + 7x_1^6 + 10x_2^9 + 8x_2^7 \tag{14}$$

$$\frac{\partial}{\partial x_1} y\left(x\right) = 110x_1^9 + 72x_1^7 + 42x_1^5 \tag{15}$$

$$\frac{\partial}{\partial x_2} y\left(x\right) = 90x_2^8 + 56x_2^6 \tag{16}$$

where

- y: 10×1
- x^*: 2×2
- x_1^δ: 8×2
- x_2^δ: 8×2

Then including the GP parameters σ^2 and l, there are 20 unknowns. Fixing GP parameters, we can calculate the covariance matrix S required for $\left[y^*, y_1^\delta, y_2^\delta \,|y, \sigma^2, l\right]$.

2.3. **Problem.** Assuming reasonable values for σ^2 and l, following the formula for S given above, the resulting matrix is negative definite. Adding small nuggets does not fix this.

It seems that if I use rather small values of l (all entries less than approximately 0.3), then this problem disappears.

3. Derivatives for the Matern

Recall: $g\left(x, x'\right) = \left[1 + \theta\left|x - x'\right| + \frac{1}{3}\theta^2|x - x'|^2\right] \exp\left\{-\theta\left|x - x'\right|\right\}$

3.1. First derivative.

$$\frac{\partial}{\partial x} g(x,x') = \left[\theta sign(x-x') + \frac{2}{3}\theta^2 (x-x')\right] \exp\{-\theta |x-x'|\} \tag{17}$$

$$+ \left[1 + \theta |x-x'| + \frac{1}{3}\theta^2 (x-x')^2\right] \exp\{-\theta |x-x'|\} (-\theta sign(x-x')) \tag{18}$$

$$= \exp\{-\theta |x-x'|\} \left[-\frac{1}{3}\theta^2 (x-x') - \frac{1}{3}\theta^3 (x-x')^2 sign(x-x')\right] \tag{19}$$

$$= \exp\{-\theta |x-x'|\} \left(-\frac{1}{3}\theta^2 (x-x')\right) [1 + \theta (x-x') sign(x-x')] \tag{20}$$

$$= -\frac{1}{3}\theta^2 (x-x') [1 + \theta |x-x'|] \exp\{-\theta |x-x'|\} \tag{21}$$

3.2. Second derivative.

$$\frac{\partial^2}{\partial x' \partial x} g(x,x') = \frac{\partial}{\partial x'} \left[-\frac{1}{3}\theta^2 (x-x') [1 + \theta |x-x'|] \exp\{-\theta |x-x'|\}\right] \tag{22}$$

$$= \frac{\partial}{\partial x'} \left[\left[-\frac{1}{3}\theta^2 (x-x') - \frac{1}{3}\theta^3 (x-x')^2 sign(x-x')\right] \exp\{-\theta |x-x'|\}\right] \tag{23}$$

$$= \left[\frac{1}{3}\theta^2 - \frac{2}{3}\theta^3 (x-x') (-1) sign(x-x')\right] \exp\{-\theta |x-x'|\} \tag{24}$$

$$+ \left[-\frac{1}{3}\theta^2 (x-x') - \frac{1}{3}\theta^3 (x-x')^2 sign(x-x')\right] \exp\{-\theta |x-x'|\} (\theta sign(x-x')) \tag{25}$$

$$= \exp\{-\theta |x-x'|\} \left[\frac{1}{3}\theta^2 + \frac{2}{3}\theta^3 |x-x'| - \frac{1}{3}\theta^3 |x-x'| - \frac{1}{3}\theta^4 (x-x')^2\right] \tag{26}$$

$$= \exp\{-\theta |x-x'|\} \left[\frac{1}{3}\theta^2 + \frac{1}{3}\theta^3 |x-x'| - \frac{1}{3}\theta^4 (x-x')^2\right] \tag{27}$$

$$= \frac{1}{3}\theta^2 \left[1 + \theta |x-x'| - \theta^2 (x-x')^2\right] \exp\{-\theta |x-x'|\} \tag{28}$$