

**Keywords** — Gaussian Processes, Statistics, Velocity Fields

## I. Overview of Gaussian Processes

Gaussian processes generalize the multivariate Gaussian distribution and can be used to describe a probability distribution over families of functions.

### i. Multivariate Gaussian Distribution

The multivariate Gaussian distribution is used to model *random vectors* (vectors of jointly distributed random variables).

$$\mathbf{x} \in \mathbb{R}^N \sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu} \in \mathbb{R}^N = (\mu_1, \mu_2, \dots, \mu_N)^\top = (\mathbb{E}(x_1), \mathbb{E}(x_2), \dots, \mathbb{E}(x_N))^\top$$

$$\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N} = \mathbb{E}((\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top) = [\text{cov}(x_i, x_j)]_{i,j=1}^N \quad (1)$$

$$x_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$$

### ii. Gaussian Processes (GPs)

- *Gaussian process* (GP): an uncountably infinite collection of random variables; any finite sample is a draw from a MV Gaussian distribution.
- GPs are fully specified by a *mean function*  $m$  and *covariance* (kernel) function  $k$ .
- The kernel function must produce a positive semi-definite matrix when evaluated on a set of input points (or vectors).

We focus on the *squared exponential kernel*  $k: \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ , defined as:

$$k(\mathbf{x}, \mathbf{x}') = \alpha^2 \exp\left(-\frac{1}{2\rho^2} \|\mathbf{x} - \mathbf{x}'\|^2\right). \quad (2)$$

- $\|\cdot\|$  is the Euclidean Norm:  $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_N^2}$
- $\alpha$  and  $\rho$  are *hyperparameters* (chosen, or estimated from data)

## II. Gaussian Process Regression – Univariate $y$

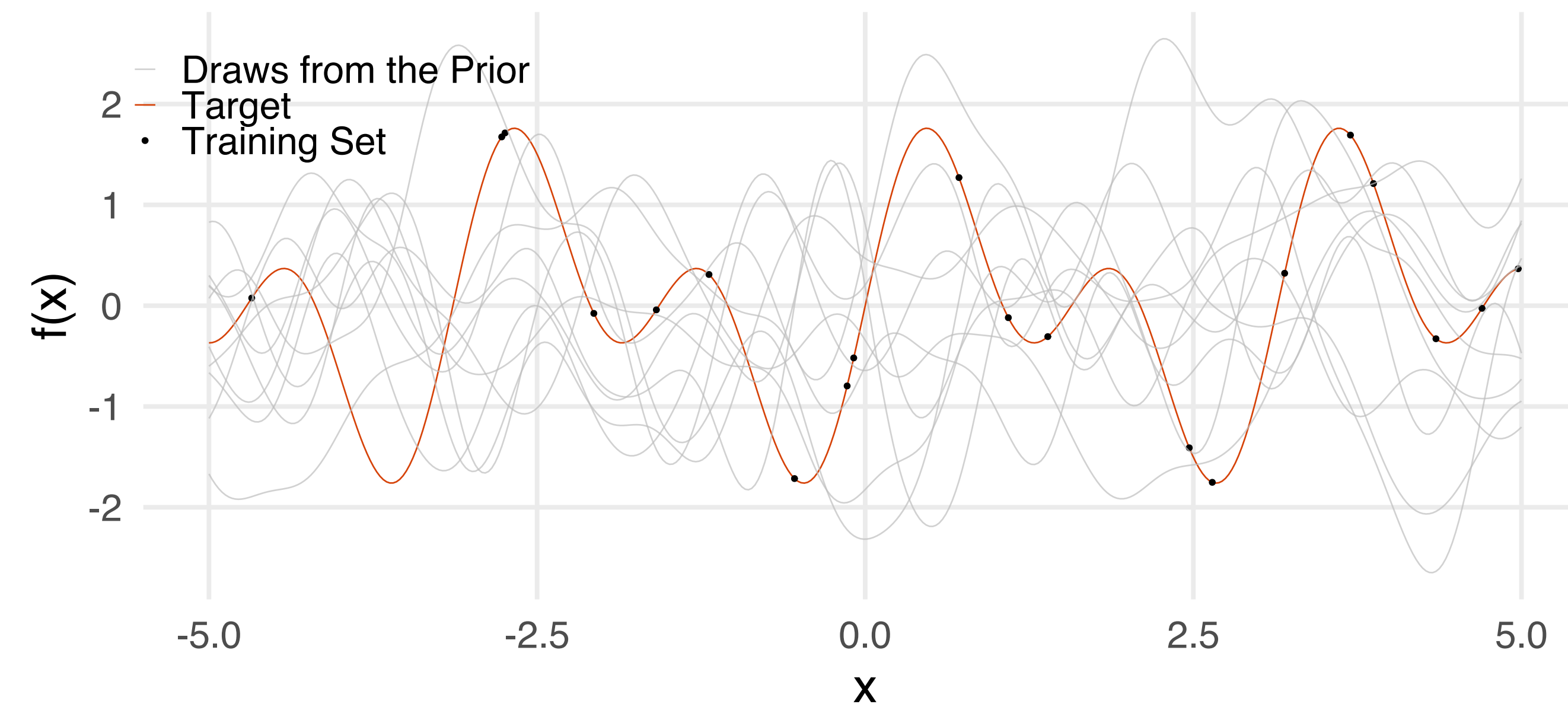
Let  $S = (\mathbf{x}, \mathbf{y}) = \{(\mathbf{x}_i, y_i) : \mathbf{x}_i, y_i \in \mathbb{R}, i \in 1, 2, \dots, N\}$  be a researcher's dataset, and let  $N = 20$  and  $M = 400 - N$ . We wish to use  $S$  to find an unknown function  $f$  that satisfies  $\mathbf{y} = f(\mathbf{x})$ , possibly subject to additive noise  $\varepsilon$ . We can draw samples from the prior distribution:

$$f \sim \mathcal{GP}(\mathbf{0}, k)$$

$$\mathbf{y}_* \sim \mathcal{N}_M(\mathbf{0}, k(\mathbf{x}_*, \mathbf{x}_*)) \quad (3)$$

$$k(\mathbf{x}_*, \mathbf{x}_*) \in \mathbb{R}^{M \times M} = [k(\mathbf{x}_{*i}, \mathbf{x}_{*j})]_{i,j=1}^M.$$

- Draws from the prior distribution (shown in grey) don't necessarily agree with the data points.
- Kernel choice determines properties of  $f$  (e.g., smoothness)



Our prior model for  $f$  and the observed data  $S$  can be combined to form a *posterior* distribution:

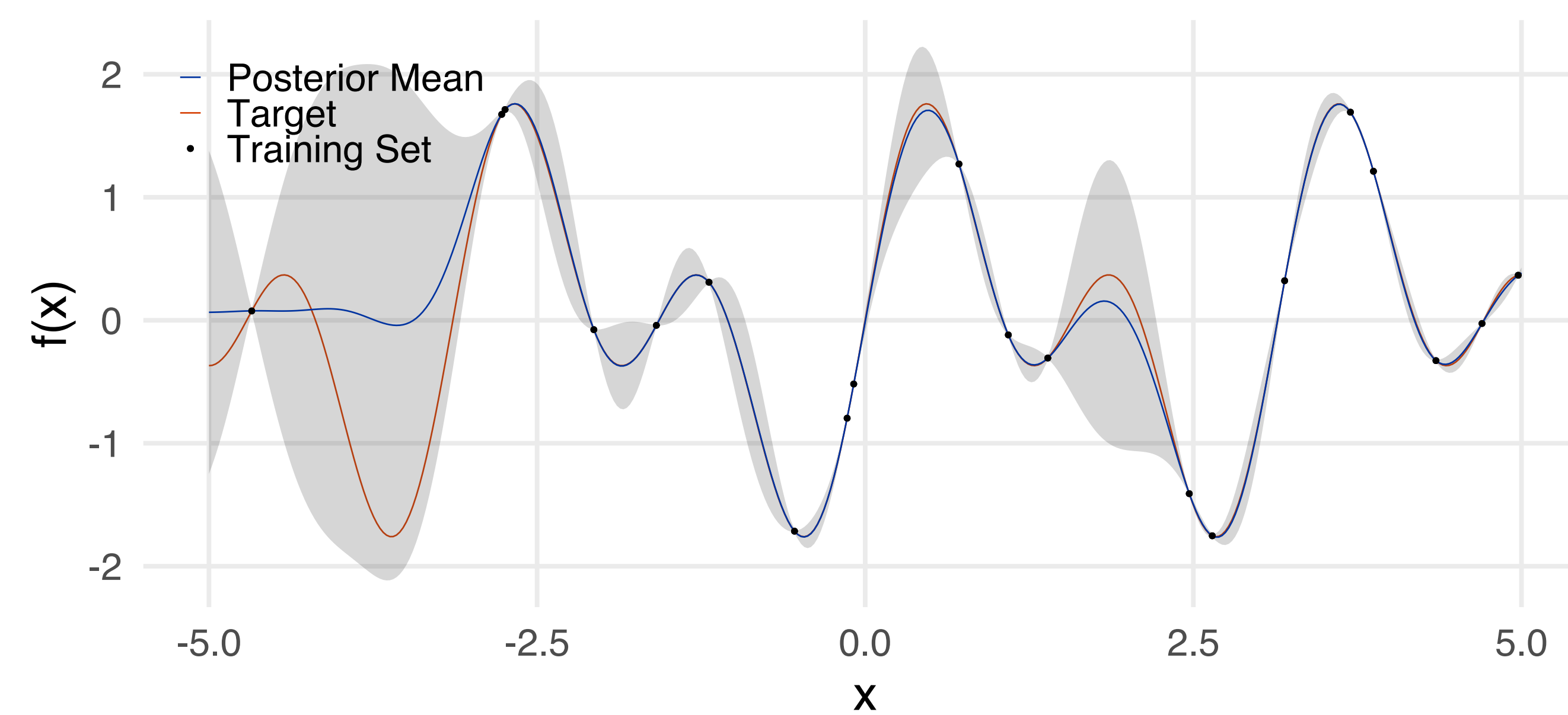
$$\mathbf{y}_* | \mathbf{x}, \mathbf{y}, \mathbf{x}_* \sim \mathcal{N}_M(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$$

$$\hat{\boldsymbol{\mu}} \in \mathbb{R}^M = k(\mathbf{x}_*, \mathbf{x}) k(\mathbf{x}, \mathbf{x})^{-1} \mathbf{y}$$

$$\hat{\boldsymbol{\Sigma}} \in \mathbb{R}^{M \times M} = k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{x}) k(\mathbf{x}, \mathbf{x})^{-1} k(\mathbf{x}, \mathbf{x}_*)^\top \quad (4)$$

$$k(\mathbf{x}, \mathbf{x}) \in \mathbb{R}^{N \times N} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^N$$

$$k(\mathbf{x}_*, \mathbf{x}) \in \mathbb{R}^{M \times N} = [k(\mathbf{x}_{*i}, \mathbf{x}_j)]_{i,j=1}^{M,N}.$$



## III. Multioutput GPR – Vector-valued $\mathbf{y}$

- GPR can be extended to targets with  $>1$  dimensions.
- Velocity fields:  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times 2}$
- Idea: columns of  $\mathbf{Y}$  might not be independent
- Intrinsic Coregionalization Model (ICM): combines the kernel matrix with a similarity matrix  $B$

$$\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times 2}, \mathbf{y} \in \mathbb{R}^{2N} = \text{vec}(\mathbf{Y}), \mathbf{X}_* \in \mathbb{R}^{M \times 2}$$

$$\mathbf{y}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim \mathcal{N}_{2M}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$$

$$\hat{\boldsymbol{\mu}} \in \mathbb{R}^{2M} = K_{\mathbf{X}_* \mathbf{X}} K_{\mathbf{X} \mathbf{X}}^{-1} \mathbf{y}$$

$$\hat{\boldsymbol{\Sigma}} \in \mathbb{R}^{2M \times 2M} = K_{\mathbf{X}_* \mathbf{X}_*} - K_{\mathbf{X}_* \mathbf{X}} K_{\mathbf{X} \mathbf{X}}^{-1} K_{\mathbf{X} \mathbf{X}_*}^\top$$

$$K_{\mathbf{X} \mathbf{X}} \in \mathbb{R}^{2N \times 2N} = B \otimes k(\mathbf{X}, \mathbf{X}) \quad (5)$$

$$K_{\mathbf{X}_* \mathbf{X}} \in \mathbb{R}^{2M \times 2N} = B \otimes k(\mathbf{X}_*, \mathbf{X})$$

$$K_{\mathbf{X}_* \mathbf{X}_*} \in \mathbb{R}^{2M \times 2M} = B \otimes k(\mathbf{X}_*, \mathbf{X}_*)$$

$$B \in \mathbb{R}^{2 \times 2} = \text{corr}(\mathbf{Y}) = \left( \frac{\langle \mathbf{y}_i - \bar{\mathbf{y}}, \mathbf{y}_j - \bar{\mathbf{y}} \rangle}{\|\mathbf{y}_i - \bar{\mathbf{y}}\| \|\mathbf{y}_j - \bar{\mathbf{y}}\|} \right)_{i,j=1}^2$$

