Vector-valued Gaussian Processes

BOISE STATE UNIVERSITY

Andrew Moore, Grady Wright (Advisor) (Department of Mathematics)

Keywords — Gaussian Processes, Statistics, Velocity Fields

I. Overview of Gaussian Processes

Gaussian processes generalize the multivariate Gaussian dist. and can describe probability distributions over functions. [1]

i. Multivariate Gaussian Distribution

$$
\begin{aligned}
\boldsymbol{z} \in \mathbb{R}^N &\sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \;\; z_i \sim \mathcal{N}(\mu_i, \boldsymbol{\Sigma}_{ii}) \\
\boldsymbol{\mu} \in \mathbb{R}^N &= (\mu_1, \mu_2, \ldots, \mu_N)^\top = (\mathbb{E}(z_1), \mathbb{E}(z_2), \ldots, \mathbb{E}(z_N))^\top \\
\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N} &= \mathbb{E}\left((\boldsymbol{z} - \mu)(\boldsymbol{z} - \mu)^\top\right) = \left[\operatorname{cov}(z_i, z_j)\right]_{i,j=1}^N
\end{aligned} \tag{1}
$$

ii. Gaussian Processes (GPs)

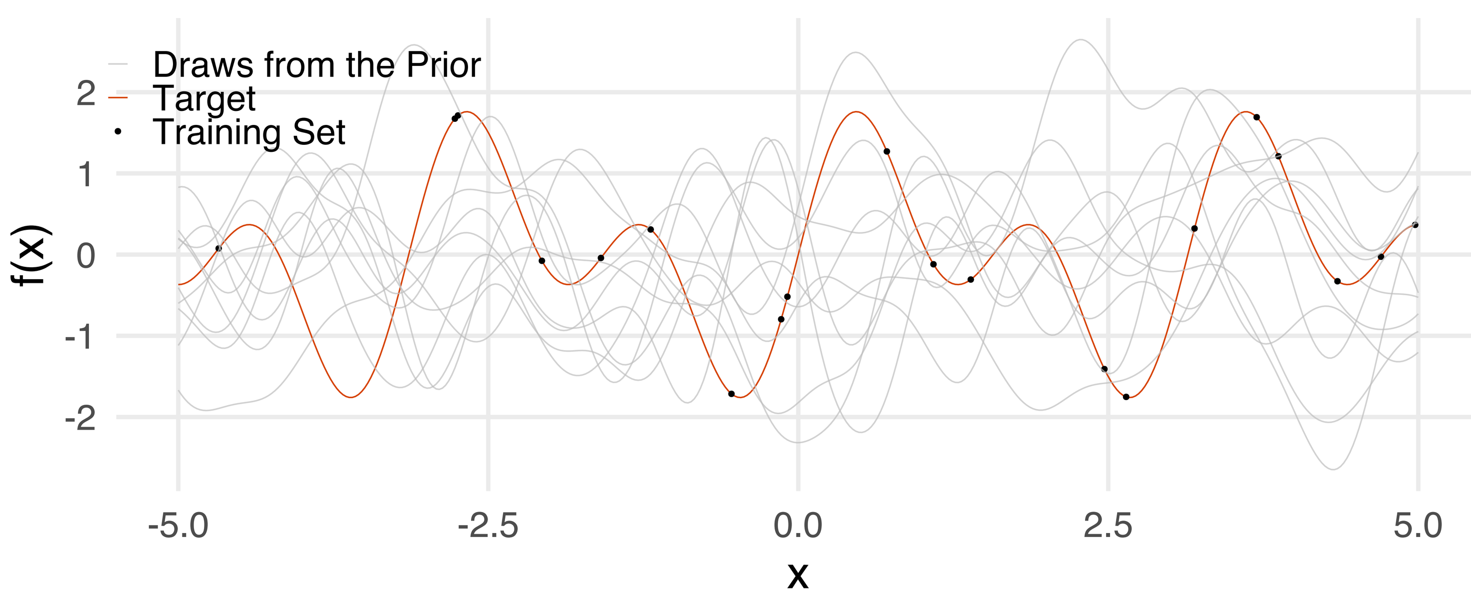
- GP: an uncountably infinite collection of random variables; any finite sample is a draw from a MV Gaussian distribution.
- GPs fully specified by *mean* and *covariance (kernel)* functions.
- The covariance function k must produce a positive semi-definite matrix.
- Squared exponential kernel, $k : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$:

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

- α and ρ are *hyperparameters* (chosen, or estimated from data)

II. Gaussian Process Regression – Univariate $\mathbf{y}$

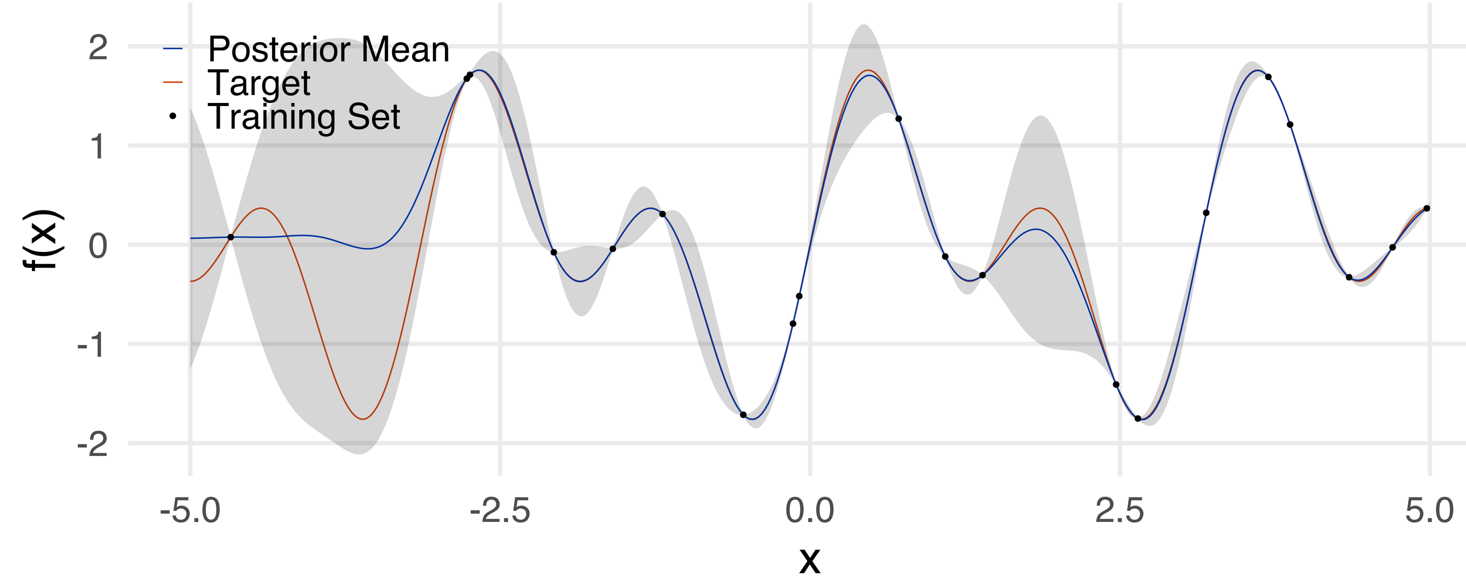
$$
\begin{aligned}
&\text{Have: } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N, \boldsymbol{x}_* \in \mathbb{R}^M \quad \text{Want: } \boldsymbol{y} = f(\boldsymbol{x}) \\
&\text{Prior: } f \sim \mathcal{GP}(\boldsymbol{0}, k) \quad \boldsymbol{y}_* \sim \mathcal{N}_M(\boldsymbol{0}, k(\boldsymbol{x}_*, \boldsymbol{x}_*)) \\
&k(\boldsymbol{x}_*, \boldsymbol{x}_*) \in \mathbb{R}^{M \times M} = \left[k(x_{*i}, x_{*j})\right]_{i,j=1}^M.
\end{aligned} \tag{3}
$$



$$
\text{Posterior: } \boldsymbol{y}_* \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}_* \sim \mathcal{N}_M\left(\hat{\mu}, \hat{\Sigma}\right)
$$

$$
\begin{aligned}
\hat{\mu} \in \mathbb{R}^M &= k(\boldsymbol{x}_*, \boldsymbol{x}) k(\boldsymbol{x}, \boldsymbol{x})^{-1} \boldsymbol{y} \\
\hat{\Sigma} \in \mathbb{R}^{M \times M} &= k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, \boldsymbol{x}) k(\boldsymbol{x}, \boldsymbol{x})^{-1} k(\boldsymbol{x}_*, \boldsymbol{x})^\top
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
k(\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}^{N \times N} &= \left[k(x_i, x_j)\right]_{i,j=1}^N \\
k(\boldsymbol{x}_*, \boldsymbol{x}) \in \mathbb{R}^{M \times N} &= \left[k(x_{*i}, x_j)\right]_{i,j=1}^{M,N}.
\end{aligned} \tag{5}
$$

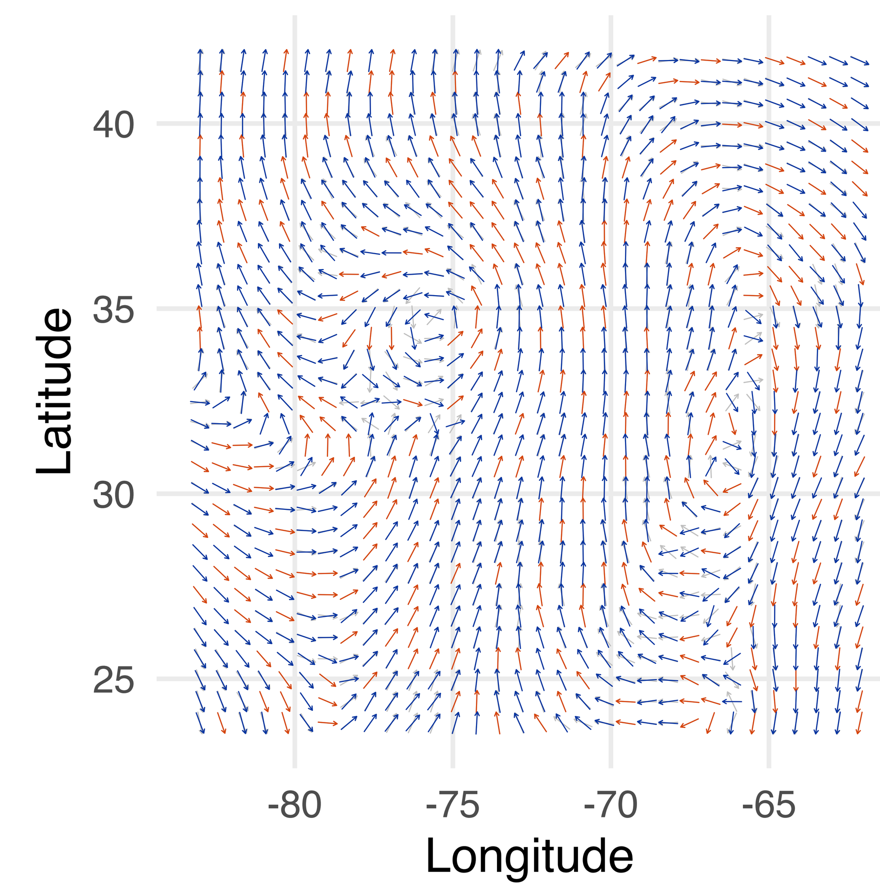


III. Multioutput GPR – Vector-valued $\mathbf{y}$

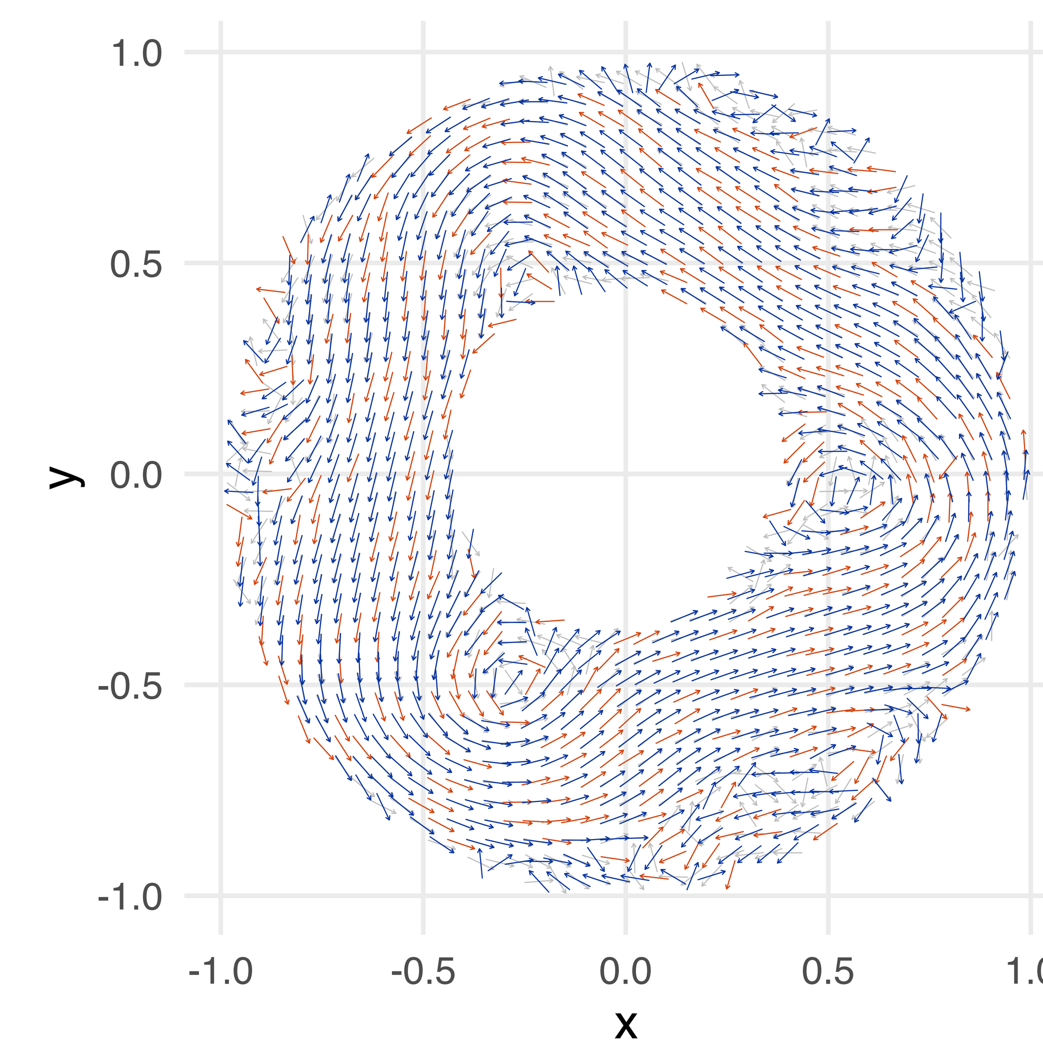
- Velocity fields: $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{N \times 2}$, $\boldsymbol{y} \in \mathbb{R}^{2N} = \operatorname{vec}(Y)$, $\boldsymbol{X}_* \in \mathbb{R}^{M \times 2}$
- Intrinsic Coregionalization Model (ICM) [2], [3]

$$
\begin{aligned}
\boldsymbol{y}_* \mid \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_* &\sim \mathcal{N}_{2M}\left(\hat{\mu}, \hat{\Sigma}\right) \\
\hat{\mu} \in \mathbb{R}^{2M} &= K_{\boldsymbol{X}_* \boldsymbol{X}} K_{\boldsymbol{X}\boldsymbol{X}}^{-1} \boldsymbol{y} \\
\hat{\Sigma} \in \mathbb{R}^{2M \times 2M} &= K_{\boldsymbol{X}_* \boldsymbol{X}_*} - K_{\boldsymbol{X}_* \boldsymbol{X}} K_{\boldsymbol{X}\boldsymbol{X}}^{-1} K_{\boldsymbol{X}_* \boldsymbol{X}}^\top \\
K_{\boldsymbol{X}\boldsymbol{X}} \in \mathbb{R}^{2N \times 2N} &= B \otimes k(\boldsymbol{X}, \boldsymbol{X}) \\
K_{\boldsymbol{X}_* \boldsymbol{X}} \in \mathbb{R}^{2M \times 2N} &= B \otimes k(\boldsymbol{X}_*, \boldsymbol{X}) \\
K_{\boldsymbol{X}_* \boldsymbol{X}_*} \in \mathbb{R}^{2M \times 2M} &= B \otimes k(\boldsymbol{X}_*, \boldsymbol{X}_*) \\
B \in \mathbb{R}^{2 \times 2} = \operatorname{corr}(\boldsymbol{Y}) &= \left(\frac{\langle \boldsymbol{y}_i - \overline{\boldsymbol{y}}_i, \boldsymbol{y}_j - \overline{\boldsymbol{y}}_j \rangle}{\|\boldsymbol{y}_i - \overline{\boldsymbol{y}}_i\| \, \|\boldsymbol{y}_j - \overline{\boldsymbol{y}}_j\|}\right)_{i,j=1}^2
\end{aligned} \tag{6}
$$

Case study: Hurricane Isabel Simulation [4]



Case study: Particle Image Velocimetry [5]



Note: colors reflect sample data, posterior mean, and test points.

Bibliography

[1] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2, no. 3. MIT press Cambridge, MA, 2006.

[2] M. A. Alvarez, L. Rosasco, N. D. Lawrence, and others, “Kernels for vector-valued functions: A review,” *Foundations and Trends® in Machine Learning*, vol. 4, no. 3, pp. 195–266, 2012.

[3] E. V. Bonilla, K. Chai, and C. Williams, “Multi-task Gaussian process prediction,” *Advances in neural information processing systems*, vol. 20, 2007.

[4] W. Chang, *R graphics cookbook: practical recipes for visualizing data*. O'Reilly Media, 2018.

[5] U. Harlander, G. B. Wright, and C. Egbers, “Reconstruction of the 3D flow field in a differentially heated rotating annulus by synchronized particle image velocimetry and infrared thermography measurements,” in *16th International symposium on applied laser techniques to fluid mechanics, Lisbon, Portugal*, 2012.