Dependent Gaussian Processes
for Multivariate Regression

Phillip Boyle and Marcus Frean

School of Mathematics, Statistics and Computer Science
Victoria University of Wellington
New Zealand

Abstract

Gaussian processes are usually parameterised in terms of their covariance functions. However, this makes it difficult to deal with multiple outputs, because ensuring that the covariance matrix is positive definite is problematic. An alternative formulation is to treat Gaussian processes as white noise sources convolved with smoothing kernels, and to parameterise the kernel instead. Using this, we demonstrate Gaussian Process regression over multiple, coupled outputs.

Introduction

- An alternative to directly parameterising a Gaussian Process (GP) covariance function is to treat GPs as the outputs of stable linear filters and parameterise the filter instead.
- If we stimulate a linear filter with Gaussian white noise, $x(t)$, then the output, $y(t)$, is necessarily a Gaussian Process.
- The output is defined by $y(t) = h(t) \star x(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau$, where $h(t)$ is the impulse response of the filter and $\star$ denotes convolution (figure 1a).
- We can construct multiple dependent GPs by stimulating a multiple output linear filter with Gaussian white noise sources (figure 1b).

Strongly Dependent Outputs

- Consider the situation in figure 2 where we observe two strongly dependent outputs. Here, output 1 is uniformly sampled, but output 2 is sparsely sampled.
- If we build a dependent GP model of both the outputs assuming that they are coupled (figure 1b), then the model does a good job at predicting output 2 in its data sparse region.
- Observe that the dependent model has learned the coupling and translation between the outputs, and has filled in output 2 where samples are missing. The control model cannot achieve such in-filling as it is consists of two independent Gaussian processes.

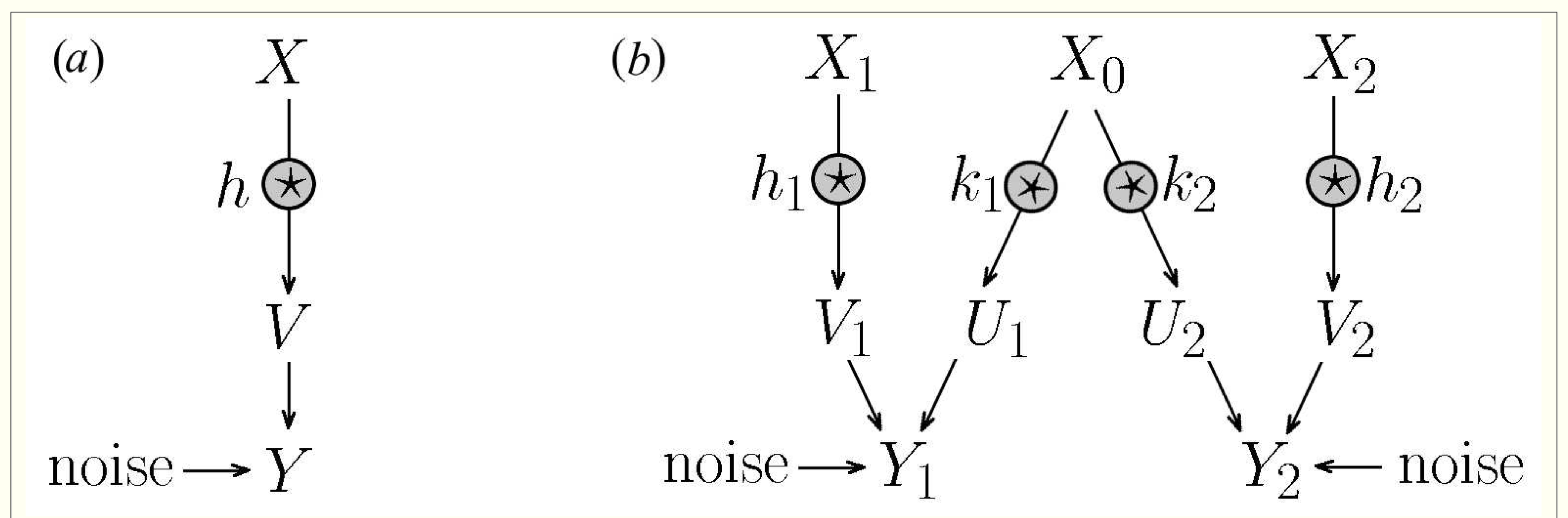

Figure 1
(a) Gaussian process prior for a single output. The output Y is the sum of two Gaussian white noise processes, one of which has been convolved ($\star$) with a kernel (h).
(b) The model for two dependent outputs Y_1 and Y_2. All of X_0, X_1, X_2 and the "noise" contributions are independent Gaussian white noise sources. Notice that if X_0 is forced to zero Y_1 and Y_2 become independent processes as in (a) - we use this as a control model.

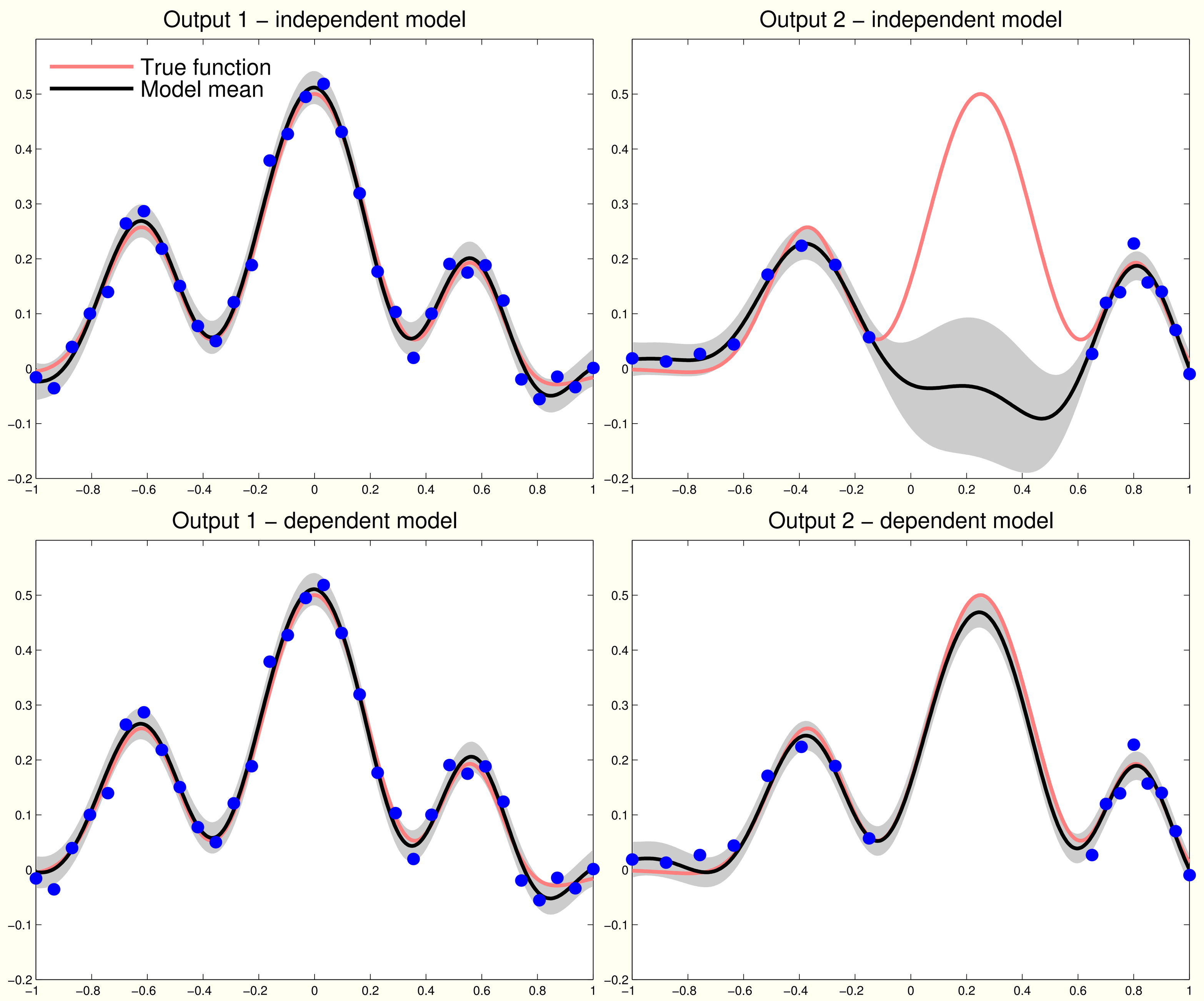

Figure 2 - Example 1
Strongly dependent outputs where output 2 is simply a translated version of output 1, with independent Gaussian noise. The black lines represent the model, the red lines are the true function, and the dots are samples. The shaded regions represent 1σ error in the model prediction. *(Top)* Independent model of the two outputs. *(Bottom)* Dependent model.

Deriving the Covariance Functions

In figure 1, k_1, k_2, h_1, h_2 are parameterised Gaussian kernels where $k_1(s) = v_1 \exp\left(-\frac{1}{2}s^T A_1 s\right)$, $k_2(s) = v_2 \exp\left(-\frac{1}{2}(s-\mu)^T A_2 (s-\mu)\right)$, and $h_i(s) = w_i \exp\left(-\frac{1}{2}s^T B_i s\right)$.

We can derive the set of functions $C_{ij}^Y(d)$ that define the autocovariance ($i = j$) and cross-covariance ($i \neq j$) between outputs i and j, for a given separation d between arbitrary inputs s_a and s_b. By solving a convolution integral, $C_{ij}^Y(d)$ can be expressed in a closed form and is related to the parameters of the Gaussian kernels and the noise variances σ_1^2 and σ_2^2 as follows:

$$C_{11}^Y(d) = C_{11}^U(d) + C_{11}^V(d) + \delta_{ab}\sigma_1^2 \qquad C_{12}^Y(d) = C_{12}^U(d)$$
$$C_{22}^Y(d) = C_{22}^U(d) + C_{22}^V(d) + \delta_{ab}\sigma_2^2 \qquad C_{21}^Y(d) = C_{21}^U(d)$$

where

$$C_{ii}^U(d) = \frac{\pi^{\frac{p}{2}} v_i^2}{\sqrt{|A_i|}} \exp\left(-\frac{1}{4}d^T A_i d\right)$$
$$C_{12}^U(d) = \frac{(2\pi)^{\frac{p}{2}} v_1 v_2}{\sqrt{|A_1 + A_2|}} \exp\left(-\frac{1}{2}(d-\mu)^T \Sigma (d-\mu)\right)$$
$$C_{21}^U(d) = \frac{(2\pi)^{\frac{p}{2}} v_1 v_2}{\sqrt{|A_1 + A_2|}} \exp\left(-\frac{1}{2}(d+\mu)^T \Sigma (d+\mu)\right) = C_{12}^U(-d)$$
$$C_{ii}^V(d) = \frac{\pi^{\frac{p}{2}} w_i^2}{\sqrt{|B_i|}} \exp\left(-\frac{1}{4}d^T B_i d\right)$$

with $\Sigma = A_1(A_1 + A_2)^{-1}A_2 = A_2(A_1 + A_2)^{-1}A_1$.

Having $C_{ij}^Y(d)$, we can construct the covariance matrices $\mathbf{C}_{11}, \mathbf{C}_{12}, \mathbf{C}_{21}$, and $\mathbf{C}_{22}$ and hence define the positive definite symmetric covariance matrix $\mathbf{C}$ for the *combined* output data $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2\}$:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

where $\mathcal{D}_i$ is the observed data from output i.

Coupled Time Series Forecasts

- Consider the observation of multiple time series, where some of the series lead or predict the others.
- We simulated a set of three time series (figure 3) where series 3 is positively coupled to a lagged version of series 1 ($lag = 0.5$) and negatively coupled to a lagged version of series 2 ($lag = 0.6$).
- We built dependent GP models of the three time series and compared them with independent GP models. The dependent GP model incorporated a prior belief that series 3 was coupled to series 1 and 2. The independent GP model had no coupling between its outputs.
- Clearly, the dependent GP model does a far better job at forecasting series 3.

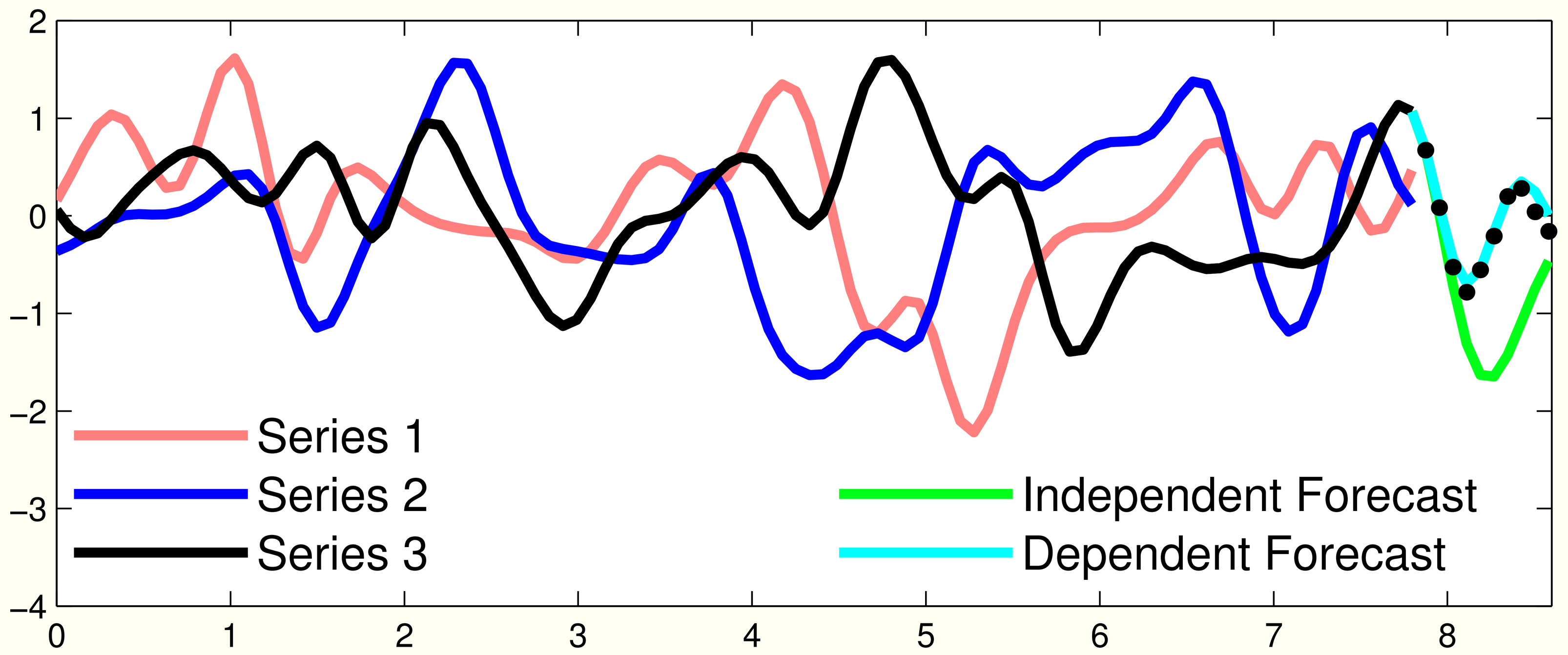

Figure 3 - Example 2
Three coupled time series, where series 1 and series 2 predict series 3. Forecasting begins after 100 steps $t = 7.8$, with the dependent model forecast coloured cyan, and the independent (control) forecast coloured green. The dependent model does a far better job at forecasting the next 10 steps of series 3 (black dots).

Multiple Outputs and Non-stationary Kernels

- We can also model N-dependent-outputs, each defined over a p-dimensional input space by assuming M-independent Gaussian white noise processes $X_1(s) \ldots X_M(s)$, and $M \times N$ kernels.
- The kernels used in need not be Gaussian, and need not be spatially invariant, or stationary. All that we require kernels is that are absolutely integrable.

Conclusions

- Multivariate regression using Gaussian Processes can be achieved by inferring convolution kernels instead of covariance functions. This makes it easy to construct the required positive definite covariance matrices for covarying outputs.
- One application of this work is to learn the spatial or temporal translations between outputs.
- Another application is in the forecasting of multiple time series that are not independent.
- In general, if we wish to use Gaussian Processes to model multiple sets of data that are not independent then we can use dependent GPs.