INVESTIGATION INTO
'Dependent Gaussian Processes [Phillip Boyle and Marcus Frean]' AS A POTENTIAL TOOL FOR THE SUPERVISED MAPPING PROBLEM

Prepared by Tung Nguyen
Prepared for Prof. James MacLean

SUMMARY
Gaussian Process (GP) Regression is usually done in the context of multi-input-single-output. In the presence of multiple outputs, the method of choice is to consider each output as independent. In other words, for each output, an independent regression will be done performed without considering the relationship, or dependency, between that output and other outputs of the system. In this paper, Boyle and Frean present a new technique to incorporate the dependency between outputs into the GP regression. Basically, instead of directly parameterizing the GP covariance function, GPs are treated as the outputs of stable linear filters. And the impulse responses will be parameterized.

The paper provided a full derivation of the new set of covariance functions that are obtained by solving a convolution integral. These covariance functions are then assembled into a new covariance function.

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

In the matrix above, the indices 1 and 2 are the output indices.
It can be observed that this derivation actually arrives at a covariance matrix that can be used in the same manner as the original covariance matrix formulated in [Rasmussen & Williams]. The only difference is that this matrix is much larger in size, due to the dependency information that it carries.

POTENTIAL AREA OF APPLICATION
In this paper, the author shows three numerical examples of how this technique could be applied. The third example is about time-series prediction which is not in our interest at the moment. We shall only focus on the first two examples, which are spatial Gaussian Process Regressions (1-d and 2-d input spaces).

A similar feature between these examples is the fact that the two dependent outputs under consideration are simply shifted, or translated, relative to each other. Looking at the output of example one (**Figure 1**), it can be seen that if the dependency between these outputs is not considered, the data in the sparse region of output 2 is poorly predicted.

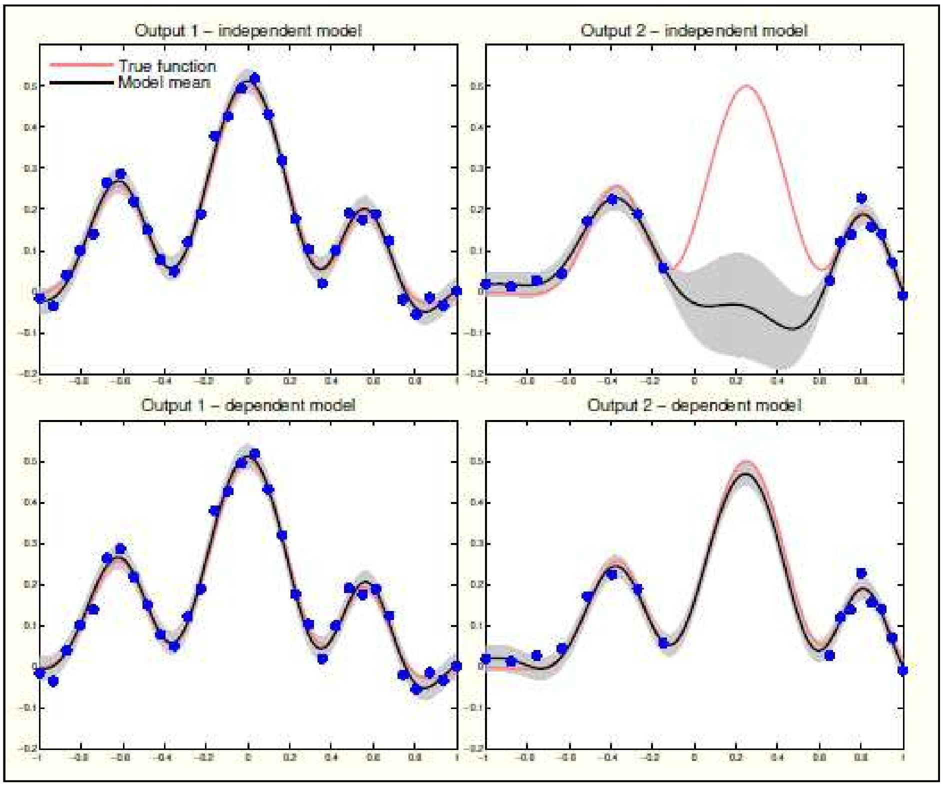

Figure 1 - Top - Independent model of the two outputs
Bottom - Dependent model

The result above indicates that this technique can be useful to our mapping problem. If we consider the four range readings to be our outputs, each of them can make use of the training information that is available to only the others.

In the figure below, we can see that output 2 is just output 1 translated by 90 degrees. The physical mounting positions of these sensors create a dependency between them that cannot be ignored. By incorporating this dependency into our prediction, we essentially enrich our data sets because each separate training set for each range reading (the output) may now augment the others. (**Figure 2**)

Due to the symmetrical mounting configuration of the sensors, the hyperparameters that are used to construct the Gaussian kernels will be the same for all of them. The paper lists the following hyperparameters:

v_i = signal variance, or σ_f in [Rasmussen&Williams]
w_i = signal variance, or σ_f in [Rasmussen&Williams]
A_i = length-scale, or l in [Rasmussen&Williams]
B_i = length-scale, or l in [Rasmussen&Williams]
σ_i = noise variance, or σ_n in [Rasmussen&Williams]
μ = translation between inputs to model coupling between outputs

The paper does not mention how to find the appropriate μ. The authors seem to suggest that users perform an optimization to maximize the likelihood in order to find the optimal set of hyperparameters.

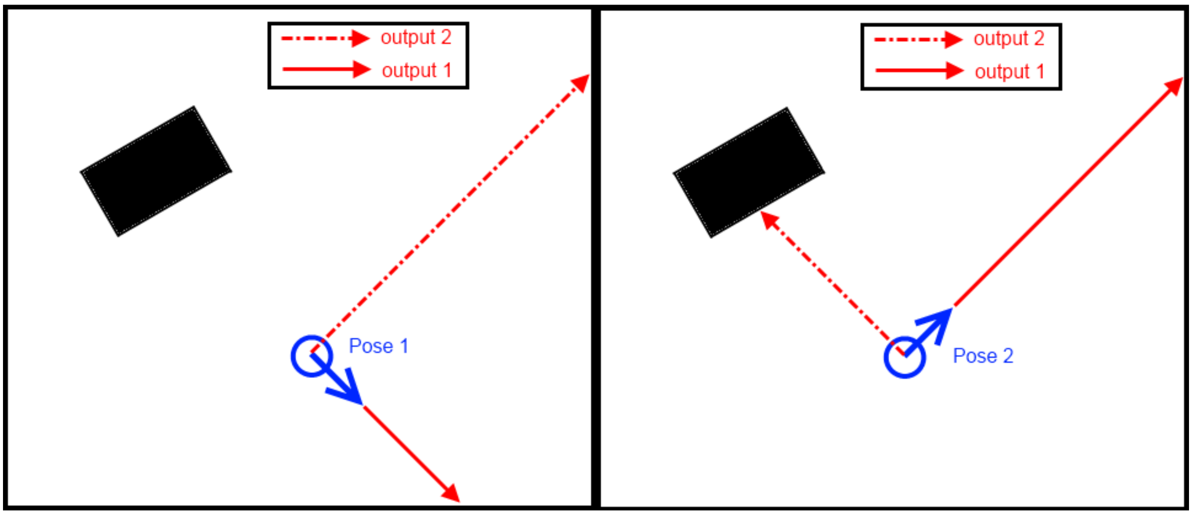

Figure 2 - Dependency between range readings