Non-Stationary Covariance Models for Discontinuous Functions as Applied to Aircraft Design Problems

Trent Lukaczyk

December 14, 2012

1 Introduction

1.1 Application

The NASA N+2 Supersonic Aircraft Project is a collaborative effort between Lockheed Martin Corporation and Stanford University that aims to advance the state of supersonic passenger jet technology. Stanford University is contributing to this by developing a generic Multi-Disciplinary Analysis and Optimization (MDAO) framework capable of identifying aircraft designs that demonstrate increased fuel efficiency by reducing drag and reduced sonic boom loudness from a standpoint of external aerodynamics. In order to evaluate the aircraft performance, we perform high-fidelity Computational Fluid Dynamics (CFD) simulations using our unstructured Euler solver SU^2. This can generate visualizations of the airflow around the aircraft as well as estimates of drag and sonic boom loudness.

By changing the design of the aircraft with shape-deforming design variables, we can build a performance response surface. A response surface is an n-dimensional surface that characterizes the relationship of design variable inputs and performance parameter outputs. Traditionally, it is necessary to model a complex problem's response surface with a less expensive "surrogate" model because direct evaluations of the underlying problem are computationally expensive. Two surrogate models commonly used in MDAO are Second Order Least Squares Regression, and Kriging, a form of weighted linear regression. Thus in a broader sense, most machine learning approaches are applicable to this problem.

1.2 Problem

An important result from Chung[1] is the discovery of discontinuities in the response surface for sonic boom loudness. Figure 2 shows this behavior found in CFD simulations of a similar business jet with variations in the radii of two fuselage sections. The existence of these design space discontinuities for supersonic aircraft problems is a major motivation to find approaches capable of modeling them. If a machine learning based model can be applied, significant gains in efficiency can be found when optimizing on these problems.

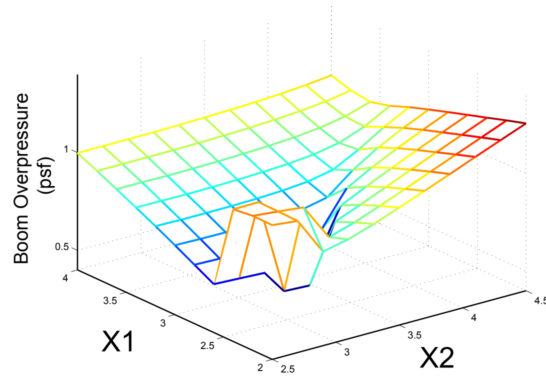

Figure 1: Supersonic Business Jet Response Surface Discontinuities [1]

1.3 Approach

Most machine learning models applied to surrogate modeling assume stationarity in the feature space. For example, a typical approach to weighted least-squares regression assumes that that the length-scale parameter, l, is constant with changes in input features. A non-stationary covariance model might allow l to vary in the feature space, allowing it to shrink near areas of high variation in the data, or grow in areas of low variation. This would allow it to model rapid changes in function value around discontinuities, and maintain smooth features elsewhere.

These models are difficult to manage because the variation of the covariance parameter must be learned as a function of the feature space and even as a function of the target variables. A proposed simplification to this algorithm explored in this report is to iteratively update the local tau parameter as a function of local gradient until the regression model converges. This way the model may be able to learn the location of discontinuities.

2 Present Work

2.1 Non-stationary Weighted Linear Regression

A preliminary form of non-stationary regression was build from the weighted linear regression model.

To generate an approximation to sampled data of an piecewise smooth function, I am using weighted linear regression with a Gaussian weight model.

$$Y = \Theta X, \tag{1}$$

where X is a design matrix with rows of feature vectors. Model parameters Θ are solved with the normal equations,

$$\Theta = (X^T W X)^{-1}(X^T W Y), \tag{2}$$

where W is a diagonal matrix of weights chosen by the function,

$$W_{i,i} = exp\left(-\frac{(x^{(i)} - x)^2}{2l(x)^2}\right). \tag{3}$$

Non-stationarity is easily added by allowing l to be a function of the input features. However this function must now be learned as well.

To estimate the distribution of l, I constructed a second machine learning estimator of the probability of target variable y having a discontinuity at x conditioned by the gradient $\frac{\partial y}{\partial x}$,

$$P\left(w(x) = 1 \,\middle|\, \frac{\partial y}{\partial x}\right) = h(\theta^\top x) = 1 - g\left(\frac{\partial y}{\partial x}\right), \tag{4}$$

where w is the indicator function for the discontinuity, and g is some mapping function which was chosen to be a form of the Gaussian distribution,

$$g(z) = \exp\left(-\sum_{d}^{D} \frac{(\partial y / \partial x_d)^2}{\sigma^2}\right), \tag{5}$$

where σ is a scaling parameter that must be learned and the normalization constant has been dropped to ensure that the distribution's output ranges in $[0, 1]$.

The length scale is then itself scaled by the probability of it *not* being a discontinuity,

$$l(x) = l_o \cdot \tau(x) = l_o \cdot g\left(z(x)\right), \tag{6}$$

where l_o is a baseline length scale, another parameter that must be learned.

2.2 Non-stationary Factor Analysis

Non-stationarity can also be applied in the context of factor analysis, where we build the model,

$$f | x^*, x, f \sim \mathcal{N}\left(f^*, \mathbb{V}[f^*]\right), \tag{7}$$

based on the prior distribution assumptions

$$\begin{gathered} \begin{bmatrix} f_p \\ f_k^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} k(x_p, x_q) & k(x_p, x_j^*) \\ k(x_k^*, x_q) & k(x_k^*, x_j^*) \end{bmatrix}\right), \\ \{\, f_i(x_i) \,|\, i = 1, ..., n \,\} \,,\, \{\, f_t^*(x_t^*) \,|\, t = 1, ..., m \,\}. \end{gathered} \tag{8}$$

and apply a non-stationary kernel $k(\cdot, \cdot)$ derived from the relation from Gibbs provided in Rasmussen[2]:

$$k(x, x^*) = \left(\frac{2l_o(\tau(x)\tau(x^*))}{l_o(\tau^2(x) + \tau^2(x^*))}\right)^{\frac{D}{2}} \exp\left(-\sum_{d}^{D} \frac{(x_d - x_d^*)^2}{\tau^2(x_d) + \tau^2(x_d^*)}\right), \tag{9}$$

where τ has been assumed isotropic but non-stationary in the feature space.

2.3 Iterative Discontinuity Steepening

To steepen the discontinuity, I am iterating the response surface to improve the estimate of τ. The basic process is to generate an initial guess of the response surface, and calculate its gradient. Then use this estimate of the gradient to update the discontinuity indicator τ. Then the response surface is regenerated using the updated tau, and the process repeats until convergence.

3 Results

To test these approaches, I built on top of an existing matlab code for Gaussian process regression, adding the iterative discontinuity steepening algorithm and non-stationary covariance functions.

I ran numerical tests with a toy data set made of with a sinusoidal function plus a small amount of random noise, and a discontinuous shift of the data at $x = 0$. Applying the iterative method described above and using non-stationary weighted least squares regression, the response surface model of the data was able to converge to capture the discontinuity in 100 iterations.

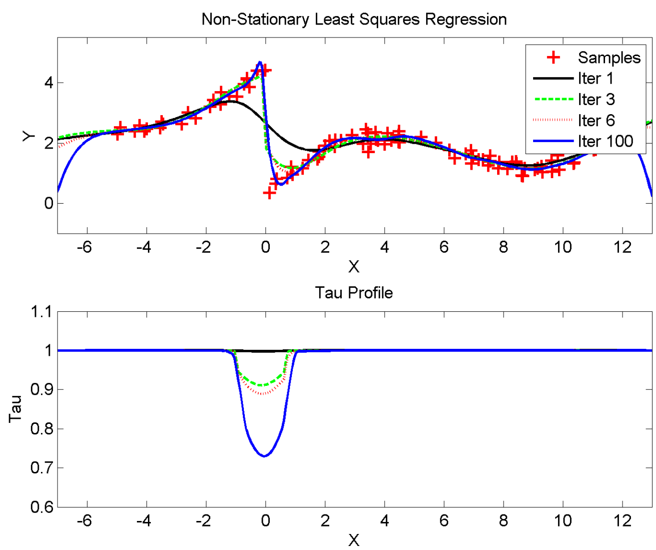

Figure 2: Non-Stationary Least-Squares Weighted Regression Example

Another test case I ran was one with no noise but fewer data, and using the factor analysis approach. An example regression is shown below.

The advantage here was that the scaling parameter b for the discontinuity model could be easily learned as part of the marginal likelihood maximization. Because of the sparsity of the data (which must be acceptable for this method to be useful for response surface modeling), there is more ringing around the discontinuity than that in figure 2. Building an indicator of where more data could be added as part of an adaptive sampling process could be a future improvement to this method.

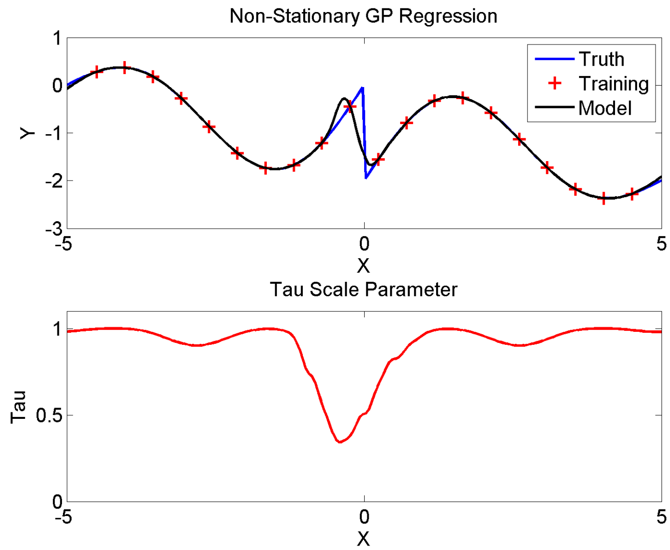

Figure 3: Non-Stationary Least-Squares Weighted Regression Example

4 Conclusions and Future Work

This report has presented the development of a machine learning model which attempts to identify locations of discontinuities in a response surface. A modified Gaussian distribution was chosen to model the distribution of a discontinuity indicator informed by the gradient of the response surface, and an iterative method was used to build an estimate of this indicator. Using two toy examples, this method was shown to capture discontinuities in the training data.

More work is needed to validate this approach. For example, the generalization and estimation error could be evaluated in more thorough sense by generating a random set of toy models with different locations and magnitudes of discontinuities. If the model as built is well suited for this problem, the average generalization and estimation error across this test set would be low.

Additional future work would be to include additional observations of gradients of the training data to better inform the fit. Adding adaptive sampling capabilities to refine the training data round the discontinuity could also be advantageous. Finally, this approach can be applied to the supersonic passenger jet design problem to evaluate it's usefulness in a real engineering application.

References

[1] Chung, H., and J.J Alonso, "Design of a Low-Boom Supersonic Business Jet Using Cokriging Approximation Models," 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA 2002-5598, Atlanta, GA, September 2002.

[2] Rasmussen, C.E. and C.K.I. Williams Gaussian Processes for Machine Learning, MIT Press, Cambridge MA, 2006.

[3] Stephensen, J., K. Gallagher, and C.C. Holmes "Beyond Kriging - Dealing with discontinuous spatial data fields using adaptive prior information and Bayesian Partition Modeling," Geological Society Special Publications, London, January 2004.