Proceedings of the 25th Nordic Seminar on Computational Mechanics
K. Persson, J. Revstedt, G. Sandberg, M. Wallin (Eds.)

Stochastic model reduction applied to inverse problems

Tomasz Garbowski[1] and Maciej Putowski[1]

[(1)]Poznan University of Technology, Institute of Structural Engineering, Piotrowo 5, 60-695 Poznan, tomasz.garbowski@put.poznan.pl, maciej.m.putowski@doctorate.put.poznan.pl

Summary. This article describes the use of Gaussian Processes in model reduction techniques with application to inverse problems. Mainly, the work is focused on the proper construction of the model approximation, namely on training process based on minimal number of learning samples, by making use of automatic samples selection through computed standard deviation of model prediction. An example of application of stochastic surrogate model for the paperboard characterization through biaxial tensile test and DIC measurements is also presented.

Key words: Gaussian processes, inverse analysis, model approximation

Introduction

The inverse analysis often uses a numerical model as a counterpart to experiment in order to build the discrepancy function between experimentally measured and numerically computed quantities, such as displacements, reaction forces, strains, accelerations, etc. If the numerical model is complex and/or has to be computed many times, the iterative minimization procedure becomes very expensive, therefore, not attractive from practical point of view especially when the test has to be performed 'in situ' (i.e. without a computer which can handle heavy computations). The alternative is to use a surrogate which approximates the behavior of the numerical model but is much simpler, thus less expensive. The surrogate is usually constructed as a 'black box' where for the approximation the following methods, among others, are commonly used: Radial Basis Functions (RBFs), Polynomials, Proper Orthogonal Decomposition (POD) combined with RBFs, Artificial Neural Networks (ANNs) or Gaussian Processes (GP) [1, 2].

All listed here approximation techniques require the numerically computed responses (i.e. training samples) in order to build a smooth and accurate analytical approximation of the sought solution. It would be best to use a method which needs the smallest possible number of 'training' points and in the same time is precise and robust. The approximation method based on GP satisfies all above-mentioned requirements: it gives very good results when the number of training examples is limited. Another important feature of GP is that it gives not only the approximation of the mean value of sought solution but also its standard deviation. This feature gives a possibility of automatic and systematic improvement of the solution, because the computed standard deviation of the model prediction provides a localization where the approximation is weak (and therefore, it points out where, in the parameter space, the additional experimental or numerical data are necessary to improve the approximation).

The presented stochastic algorithm is formulated within Bayesian framework, thus provides additional information about the magnitude of correlation between state variables (i.e. measurable quantities) and control variables (i.e. parameters). It is very important to know the relevance of input-output correlation because based on it one can exclude from the model the parameters which do not influence the measurable quantities. Such reduction saves the experimental efforts of finding parameters which appear to be irrelevant in particular simulation.

Model reduction through Gaussian Processes

In order to explain how to construct a model approximation by Gaussian Processes, first, a linear regression model should be considered. It is a linear function of model parameters $\mathbf{w}$ and nonlinear function of the input vector $\mathbf{x}$, and usually is defined as:

$$y\left(\mathbf{x}, \mathbf{w}\right) = \sum_{j=1}^{M} w_j \phi_j\left(\mathbf{x}\right), \tag{1}$$

which simply is a linear combination of fixed, nonlinear basis functions $\phi_j(\mathbf{x})$ of the input variables (e.g. polynomial basis functions).

If we now take N given training patterns $(\mathbf{x}_n, t_n)$, $\mathbf{x}_n$ being the input vector, t_n the response for $n = 1 \ldots N$, then the parameters $\mathbf{w}$ of the linear model can be computed by, for example, penalized least squares method:

$$\mathbf{w} = \left(\mathbf{\Phi}^T\mathbf{\Phi} + \lambda\mathbf{I}\right)^{-1}\mathbf{\Phi}^T\mathbf{t}, \tag{2}$$

where $\mathbf{\Phi}$ is $N \times M$ design matrix with elements defined as $\phi_m(\mathbf{x}_n)$. The regularization parameter λ is called hyperparameter and can be estimated using validation set or by applying Bayesian inference and maximizing evidence of dataset $p(\mathbf{t}|\lambda)$ w.r.t. λ (details are given in [1]).

Gaussian Process

Gaussian process model can be obtained by reformulation of the linear model in terms of dual representation. In this approach, linear model is trained by minimizing a regularized error, which is defined using $N \times N$ symmetric Gram matrix:

$$\mathbf{K} = \mathbf{\Phi}\mathbf{\Phi}^T = \phi(\mathbf{x})^T\phi(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}'), \tag{3}$$

where $k(\mathbf{x}, \mathbf{x}')$ is a kernel function. The vector $\mathbf{k}_n = k(\mathbf{x}_n, \mathbf{x})$ represents n-th row or column of $\mathbf{K}$ matrix.

The prediction for a new input $\mathbf{x}^*$ can be computed by the formula:

$$y(\mathbf{x}^*) = k(\mathbf{x}, \mathbf{x}^*)^T(\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{t}, \tag{4}$$

where $k(\mathbf{x}, \mathbf{x}^*)$ is a covariance between a new input $\mathbf{x}^*$ and the other inputs, $\mathbf{t} = (t_1 \ldots t_N)^T$ is a vector of training target values.

From the Bayesian point of view the dual representation of linear model leads to the Gaussian process, where the kernel function is interpreted as a covariance function of the GP. Application of such regression model for prediction allows to compute the predictive distribution of the target variable $y(\mathbf{x}^*)$ for a new input vector $\mathbf{x}^*$. This requires evaluation of conditional distribution $p(y|\mathbf{t})$, which for the Gaussian processes is a Gaussian distribution with mean and covariance respectively given by:

$$m(\mathbf{x}^*) = \mathbf{k}^T\mathbf{C}^{-1}\mathbf{t}, \tag{5}$$

$$\sigma^2(\mathbf{x}^*) = c - \mathbf{k}^T\mathbf{C}^{-1}\mathbf{k}, \tag{6}$$

where $\mathbf{C}$ is the $N \times N$ covariance matrix given by:

$$C(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') + \beta^{-1}\mathbf{I}, \tag{7}$$

where β is the variance of the target distribution. The covariance function $C(\mathbf{x}, \mathbf{x}')$ defines the property that vectors $\mathbf{x}_n$ and $\mathbf{x}_m$, which are close in input space, should give rise to highly correlated outputs $y(\mathbf{x}_n)$ and $y(\mathbf{x}_m)$.

Covariance function

The covariance function can be any function that will generate a non-negative definite covariance matrix for any ordered set of (input) vectors $(\mathbf{x}_1, \ldots \mathbf{x}_N)$. A stationary, non-isotropic squared exponential covariance function $k(\mathbf{x}, \mathbf{x}')$ is chosen here, and given by:

$$k(\mathbf{x}_n, \mathbf{x}_m) = \nu \exp\left(-\frac{1}{2}\sum_{i=1}^{d} w_i \left(x_n^i - x_m^i\right)^2\right) + b, \tag{8}$$

where the term b represents a bias that controls the vertical offset of the Gaussian process, while ν controls the vertical scale of the process. The w_i parameters allow a different distance measure for each dimension. If w_i is small then the i-th input is downweighted and have little effect on the input.

After defining the covariance function we can make predictions of the new input vectors but it is necessary to learn the hyperparameters. In order to find those parameters one can search for the most probable ones by maximizing the log likelihood function given by:

$$\ln p(\mathbf{t}|\theta) = \frac{1}{2}\ln|\mathbf{C}| - \frac{1}{2}\mathbf{t}\mathbf{C}^{-1}\mathbf{t} - \frac{N}{2}\ln 2\pi, \tag{9}$$

using gradient-based optimization algorithms, such as a first-order batch Levenberg-Marquardt Algorithm (LMA), which provides fast and robust convergence.

Active learning

In order to optimize the training procedure of GP based surrogate model, the active learning criterion, that improve the global model fit, is also utilized here. A new training points are sequentially added in the zones where the model predictions are poor and/or in the vicinity of the minimum of the approximated solution, meaning that the iterative re-training of the surrogate model is performed by adding a new training patterns, sampled in the new locations in the parameter space (selected by the algorithm itself). This approach is very efficient when the number of training samples is limited, so the algorithm starts building approximation with small selection of patterns (randomly or uniformly distributed) and improves the approximation by sampling the parameter space in a clever way (based on its confidence about the quality of the approximation). It stops when the maximum number of training samples is reached.

Application

In the present communication an example of a model characterization is used to show the application of above described model reduction techniques. The examples show the use of GP as numerical model surrogate for characterization of paperboard parameters [3] through biaxial test and DIC measurements techniques combined with inverse analysis. From such test we can identify most of the in-plane parameters of paperboard (for the details see [4]).

In the above examples the GP based approximation serves as a surrogate of numerical model, which in combination with iterative minimization routine (e.g. trust-region algorithm) or evolutionary-based techniques (e.g. particle swarm algorithm) gives very fast and accurate results of model identification. The model reduction approach has the following important advantages w.r.t. the procedure based on evaluation of full numerical model: (1) it's fast (the computational burden is moved to training phase) and (2) it doesn't require to use powerful computers for heavy and repetitive computations of numerical model, so once the model is constructed it can be used 'in situ' on the portable computer.

Summary

The GP approximation model which serves as a numerical model reduction is used here in combination with inverse analysis to solve practical engineering identification problems. The work is mainly focused on the proper construction of the GP model, namely on: (1) training process based on minimal number of training data, by making use of automatic samples selection through computed standard deviation of model prediction and (2) control parameters reduction based on input-output correlation. The successful application of stochastic model reduction techniques for the material model characterization problems is also shown here.

References

[1] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.

[2] C.E. Rasmussen, C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Pres, 2006.

[3] Q.S. Xia, M.C. Boyce, D.M. Parks. A constitutive model for the anisotropic elastic-plastic deformation of paper and paperboard. *International Journal of Solids and Structures*, 61:1–48, 2000.

[4] T. Garbowski, G. Maier, G. Novati. On calibration of orthotropic elastic-plastic constitutive models for paper foils by biaxial tests and inverse analyses. *Structural and Multidisciplinary Optimization*, 46:111–128, 2012.