6th European Conference on Computational Mechanics (ECCM 6)
7th European Conference on Computational Fluid Dynamics (ECFD 7)
11 15 June 2018, Glasgow, UK

EFFICIENT CLUSTERING BASED TRAINING SET GENERATION FOR SYSTEMS OF SOLVERS

Francois Sanson[1], Anne W. Eggels[2], Olivier Le Maitre[3], Daan Crommelin[2,4] and Pietro Marco Congedo[1]

[1] Inria, 205 rue de la Vieille Tour, 33405 Talence, France
[2] CWI, P.O. Box 94079 GB Amsterdam, Netherlands
[3] Campus universitaire bt 508, rue John von Neumann, Orsay, France
[4] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Netherland

Key words: *Uncertainty Quantification, Surrogate models, System of Solvers*

Uncertainty propagation in complex industrial solvers demands efficient surrogate model construction methods. The surrogate model substitutes to the computationally expensive solver in order to propagate the input distributions through the solver at minimal computational cost. This is all the more true in the case of Systems of Solvers (SoS). A system of solver is a set of interdependent solvers coupled in order to compute a quantity of interest. The composing solvers are linked through their inputs and outputs such as the outputs of a solver are also inputs of the next one. In this work, we restrict ourselves to directed SoS where the information can only be sent forward: the output of one solver can be used as inputs of a second solver but the outputs of the second solver cannot become inputs of the first one. In this case, it is usually more efficient to build a surrogate model of each solver, and to use their composition to make predictions rather than building directly a surrogate model of the whole system [1]. This accuracy gain comes with an increased complexity from building surrogate models of each solver. The selection of training points for the surrogate model is non obvious because distributions of intermediate variables are in general nonuniform and dependent as they are computed by upstream solvers. One solution used in [1] is to propagate the initial training set obtained from the SoS input distributions (for instance latin hypercube sampling) to the intermediate inputs and use them as training sets. Unfortunately, the propagated samples tend to form clusters in dense regions and to ignore lower density regions, leading to degraded performances of the surrogate model. In this work, we propose an efficient strategy for generating training sets in SoS inspired from [2]. The approach relies on clustering methods for selecting training samples that have a good coverage and at the same time provide a good approximation of the distribution to be propagated. The method is tested with Gaussian Processes (GP) as surrogate models of intermediate solvers. Numerical performances of our method are assessed against more classical techniques on several analytical test functions and on a space object reentry predictor.

REFERENCES

[1] F. Sanson, O. Le Maitre, P.M. Congedo, *Uncertainty Quantification in Systems of Solvers.* UNCECOMP 2017 - 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, Jun 2017, Rhodes, Greece

[2] A.W. Eggels, D.T. Crommelin, J.A.S. Witteveen, *Clustering-based collocation for uncertainty propagation with multivariate correlated inputs*, arXiv preprint arXiv:1703.06112, 2017