A NOVEL FRAMEWORK TO DETERMINE COMPLEX PROCESS FEASIBILITY

Margherita Geremia¹, Fabrizio Bezzo¹, Marianthi G. Ierapetritou²*
¹CAPE_Lab – Computer-Aided Process Engineering Laboratory. Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova (PD), Italy
²Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

Keywords: Feasibility analysis, Design Space, topological data analysis, adaptive sampling, surrogate models.

Feasibility analysis can be exploited to identify the subset of combinations of uncertain input parameters that satisfy all the process and quality constraints, i.e., the design space (DS) of the process. In the presence of disjoint feasibility regions and for computationally expensive nonconvex problems, the use of surrogate-based approaches has been successfully adopted to properly predict the boundaries of the process feasibility space (Wang and Ierapetritou, 2017). It is worth noticing that the choice and prediction accuracy of a suitable surrogate model strongly depends on the specific process of interest, and on the dataset that is available for training. In this context, we aim at investigating how to correctly identify the process feasibility space relying on the available dataset, and determining the minimum number of sampling points that are necessary to uncover the complexity of the original feasibility function. The aim is to compare the performance of different candidate surrogates, while uncovering the complexity of the process feasibility space based on the inclusion of additional sample points up to the attainment of a pre-set level of prediction accuracy.

Methodology and Results

We propose a novel framework to acquire information on the complexity of the feasible space and accurately predict the feasibility boundaries with the minimum number of training data. First, we couple Topological Data Analysis (TDA) (Smith and Zavala, 2021) and data interrogation (Sun and Braatz, 2021) to reconstruct the dataset complexity and restrict the number of candidate surrogate models to the most promising ones that can be further trained. Then, we compute the Bayesian information criterion (BIC) to evaluate quality of fitting and predictive performance of the different models. If none of the trained surrogates guarantees the preset level of accuracy (i.e., stop criterion), new sampling points are needed. The implementation of an adaptive algorithm locates additional points along the boundaries of the feasible region, which are included in the training dataset. The procedure is repeated until the stopping condition is achieved based on the preset level of accuracy. The methodology is tested on pharmaceutical case studies and numerical problems with complex feasible regions, such as the Gomez function (Sasena et al., 2002). Although the initial number of points and sampling technique affect the reconstruction of the problem complexity, the inclusion of adaptive points promotes a fast identification of all disjointed feasibility regions and significant accuracy improvement. Figure 1 shows the surrogate-based feasibility approximation of the Gomez function with 98% accuracy, that is attained after the addition of 154 adaptive points to the initial training dataset (25 Sobol’s samples). The workflow can be efficiently utilized for higher-dimensional case studies; thus, it can be implemented for real manufacturing processes, where the accurate description of the DS is central for process development.

References

* To whom all correspondence should be addressed: mgi@udel.edu