USING SURROGATE MODELS FOR PROCESS DESIGN AND OPTIMIZATION

A.SHOKRY, A.D. BOJARSKI AND A. ESPUÑA

Department of Chemical Engineering, Universitat Politecnica de Catalunya
Av. Diagonal, 020280, Barcelona, Spain.

In this paper we present a methodology for rigorous optimization of nonlinear programming problems in which the objective function can be represented using black box functions. The specific application is process design and operation in which the process is modeled using modular process simulators. These models consume large CPU time to converge, derivatives are not available, in some cases generate noise, and they are seen as black box models. Different techniques are available for replacing those models by simpler and computational inexpensive models. In this case a kriging metamodel is used to replace them. Kriging has high prediction accuracy and capabilities to estimate the prediction variance. Coupled to kriging the "Expected improvement" technique (EI) [3] is used to find the global solution for the kriging metamodel. The methodology is tested with two mathematical functions, and finally, applied to the optimization of the operation parameters of a gas turbine case study, which is simulated using AspenPlus.

1. Introduction

In chemical engineering, many decisions are taken based on the analysis of a process model results. Detailed process models are complex to be coded, so specialized simulation software packages are used. Most of them have modular architectures (Aspen Plus), and they appear to the user as black boxes. They consume large CPU time to converge. Moreover, in case of optimization major difficulties appear: gradient based optimizers, which try to estimate derivatives, end up using noisy estimates because of errors introduced by these simulators (e.g. by termination criteria); consequently, their accuracy is affected badly [2].

To tackle this problem the original model is used as "physical experiment" for generating data points [5]. These data points are used to construct a simpler but accurate model which is called "metamodel" or "surrogate model". These models are accurate empirical approximation describing the relation between input variable(s) and response value(s) of a process. Jones [3] analyzed the most types of metamodels and concluded that non-interpolating (regression) metamodels are unreliable in optimization, because they do not appropriately capture the function shape; it is usually better to use surfaces that interpolate the

data with linear combinations of basis functions. Jones [3] shows that even when using an interpolating metamodel, exploring the metamodel with an arbitrary optimizer can fail even to find local optima. Consequently there is a need for optimization techniques that not only consider the metamodel prediction but also consider the uncertainty about this prediction [10]. One of such techniques is the "maximum expected improvement" which has been presented by Schonlau, [9] and Sasena et al. [8], and tested by Jones [3, 4]. In this work, an optimization methodology of process operation, in which a complex process is modeled by process simulators, is presented. The methodology is based on fitting a kriging metamodel for the whole process flowsheet and then using the EI improvement technique for optimization. The rest of the paper provides an overview of kriging and maximum expected improvement technique. The methodology is tested on mathematical functions and, finally on a gas turbine case study, modeled using AspenPlus.

2. Methodology

2.1. *Sampling*

The selection of sample points used to construct the metamodel is a key point that significantly affects the metamodel prediction accuracy. There are several techniques are available, but because sampling is not the main goal of the paper we just mention that the Hammersley has been used here because its simplicity.

2.2. *Kriging*

The kriging assumes a stochastic process, in which the error in the predicted values is also a function of the input variables (x). The kriging predictor $\hat{Y}(x)$ is then composed of two parts, a polynomial term $f(x)$ (in most cases $f(x) = \mu$) and a departure term $Z(x)$ from that polynomial. So $\hat{Y}(x) = f(x) + Z(x)$.

where $Z(x)$ is a stochastic Gaussian process that represents the uncertainty about the mean of Y(x) with expected value zero $E(Z(x)) = 0$, and a covariance for two points x_i , x_j calculated as: $cov(Z(x_i),Z(x_j)) = \sigma^2 R(x_i,x_j)$. σ^2 is the process variance, and $R(x_i ,x_j)$ is the spatial correlation function (SCF) which is usually selected exponential [3,4,7]. The final predictor of the kriging method is given by the following equation (1).

$$\hat{Y}\left(x_{new}\right) = \hat{\mu} + r^T R^{-1}\left(Y - 1\hat{\mu}\right) \tag{1}$$

Where: r is the $n \times 1$ vector of correlations $R(x_{new}, x_i)$ between the point to be predicted x_{new} and the sample design points. The mean squared error of the predictor is *given* by:

$$\hat{S}^2(x_{new}) = \sigma^2 \left(1 - r^T R^{-1} r + \frac{\left(1 - r^T R^{-1} r\right)^2}{r^T R^{-1} r} \right) \quad (2)$$

A detailed mathematical derivation can be founded in [3]. Figure 1 shows the accuracy of the kriging prediction when tested with Peaks [3] and Branin [2] functions and compared with thin plate spline radial basis function (RBF) and artificial neural network (ANN) metamodels.

2.3. *Optimization using the method of Maximum Expected Improvement*

The expected improvement (EI) approach uses the expected value of the function at a certain untrained data point x, represented by the random variable $Y(x)$ which is normally distributed, with the mean equal to the kriging prediction at this point $\hat{Y}(x)$, and variance equal to the kriging estimated variance $S^2(x)$. It is assumed that the current best minimum value of the objective is F_{min}, hence, the best current solution is expected to get improvement with by an amount I. The expected improvement is given by equation (5), [3].

$$E\left[I(x)\right] = \left(f_{\min} - \hat{y}\right)\phi\left(\frac{f_{\min} - \hat{y}}{s}\right) + s\,\varphi\left(\frac{f_{\min} - \hat{y}}{s}\right) \quad (3)$$

where Φ and φ are the normal cumulative distribution function and density function. The EI value at a point x in the metamodel domain is a function of the predicted value of the metamodel $\hat{Y}(x)$, and the estimated kriging variance $S(x)$ at this point. Thus the point that maximizes the expected improvement is the point in the metamodel domain that has minimum prediction and maximum variance. The method has been tested by [6] and [3]; both found that it converges to a global optimum.

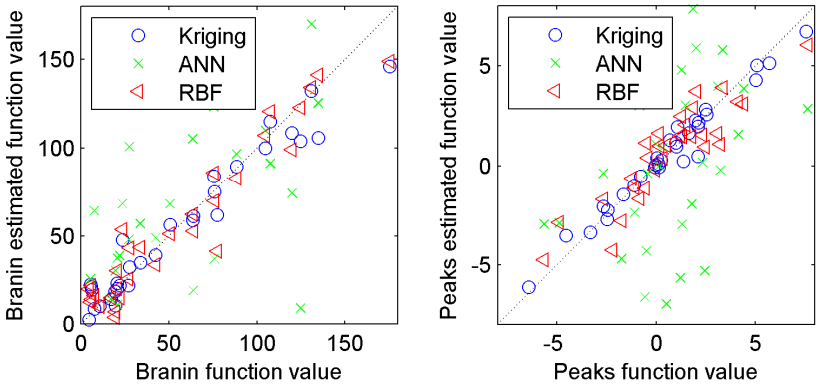

Figure 1. Leave one out cross validation (LOOCV) of kriging, thin plate spline RBF, and ANN metamodels of Branin and Peaks functions. Dotted line shows best possible prediction.

2.4. *Algorithm*

The main steps in the proposed algorithm can be summarized as follows:

Initialization (sampling design and computer experiment)

1. The process flowsheet is explored to identify the independent variables affecting the objective function that will be optimized.
2. Specify the independent variable bounds (metamodel domain).
3. Over the metamodel domain, perform a sampling plan x with a specific design and certain number of sample points.
4. Evaluate the model at these sampling points, and get the corresponding matrix of observations y.

Optimization loop

5. Fit a kriging metamodel by maximizing the likelihood of the observed data *[x y]*, obtaining the metamodel parameters. This optimization is performed using an SQP based algorithm (*fmincon* of Matlab).
6. Maximize the EI and get point x^*. This optimization is done using simulated annealing.
7. Evaluate the real model at x^* and get y^*.
8. Add the new point *[x* y*]* to the original matrix of observations *[X Y]*.
9. Return to step (5) and continue iteration.
    a. Stop if no further improvement in the objective function value y^* (the difference between objective value y^* in two successive iterations is less than specific tolerance) is observed.

3. Applications

3.1. *Mathematical examples*

The methodology has been applied to the Peaks and Branin functions, which are a good test functions due to their nonlinearity and multiple peaks and valleys. A Hammersley design sample is used to generate 31 points used to fit the kriging metamodel. The results are summarized in Table 1, which shows how the proposed algorithm provides accurate results with a few number of function evaluations.

Table 1 . Results of the optimization algorithm using the Peaks and Branin functions.

	Global optimum	*Optimum (proposed algorithm)*	*Function eval. (kriging model)*	*Function eval., (optimization stage)*
Peaks fun.	8.106 (0.009, 1.58)	8.099 (0.004, 1.59)	31	3
Branin fun.	0.397 (-3.14, 12.27)	0.408 (-3.09, 12.15)	31	11

3.2. *Gas turbine*

The methodology has been also applied to the optimization of the operation parameters of a gas turbine (GT). The GT cycle is composed by the system of compressor-combustion chamber-turbine that uses natural gas. A saturation column, before the clean gas combustion, saturates this stream with vapor and nitrogen. GT air cooling has been modeled by taking into consideration four stages GT, thus the air compressor has been modeled as a four steps process as well. The combustion chamber is considered to operate at 15 bars, while air is feed to compressor at atmospheric pressure. All turbines and compressors model units are taken from the AspenPlus model library and are considered to be isentropic. The combustion chamber is modeled as a Gibbs reactor [1]. Each corresponding stage could have the same pressure loss or gain ratio, but in this case the intermediate pressures have been left unspecified and are subject of optimization. The objective function pursued is the maximization of net power (*NP*) obtained, by modifying the cooling air splits fractions and the intermediate stage pressures (of compressors and turbines)

Table 2. Results for the Gas Turbine case study optimization.

	Case 1 (kriging+EI)	**Case 2 (kriging)**	**Case 3 (A+SQP)**
SF_1	0	0	0.0797
SF_2	0	0.1249	0.1011
SF_3	0	0.1507	0.1000
SF_4	0.11329	0	0.0010
PC_2	8.116	9	4
PC_3	14.921	11.0555	10
PT_2	2.0000	2.8225	3
PT_4	1.21334	0.8764	1.5
NP	**239571**	**239490**	**239610**
# metamodel	60	60	0
# optimization	7	21	293

The algorithm has been tested under three different conditions: 1) optimization using kriging and the EI, 2) optimization using kriging, 3) optimization using the original AspenPlus model and Matlab's SQP algorithm. In the 1st and 2nd cases, the algorithm uses a Hammersley design of 60 points. The results show a significant number of function evaluations in the 3rd case. When using the kriging metamodel only (2nd column), the number of function evaluations significantly decreased but the optimizer goes to a local optimum, finally when using the EI coupled with the kriging metamodel the number of function

evaluations is further decreased and also the objective is improved. In case of EI, a better solution can be obtained by decreasing the tolerance in the stopping criteria in the algorithm (step 9), but the number of evaluations will increase, so a trade off should be considered between the optimum accuracy and the computational effort.

4. Conclusions

One important goal of using metamodels to replace first principle models is to reduce the original model function evaluations (experiments). The EI technique shows a great behavior in meeting this goal, in front of other methods currently in use. The results clearly show how the use of the proposed methodology avoids local optima, provides better solutions and reduces the number of required function evaluations.

Acknowledgements

Financial support received from the "Generalitat de Catalunya" with the European Social Fund (FI grants), and the projects TOLERANT (DPI2006-05673) and EHMAN (DPI2009-09386) financed by the Spanish Ministry of Education and Science are fully appreciated.

References

1. Bojarski, A.D; Pérez-Fortes, M.; Velo, E.; Puigjaner, L. (2010 (ECOS2010). Book of Abstracts 1493-1501.
2. Caballero, J.A., Grossmann, I. E. (2008). AICHE J. 54, 2633.
3. Jones, D.R., (2001. J. Glob. Optim. 21(4), 345-383.
4. Jones, D.R.; Schonlau, M.; Welch, W. J. (1998). J. Glob Optim. 13(4), 455-492.
5. Fang, K-T, Li R, Sudjianto A. (2006). Chapman & Hall / CRC.
6. Locatelli M. (1997). Journal of Global Optimization 10, 57–76.
7. Sacks J, Welch WJ, Mitchell WJ. Design and Analysis of Computer Experiments. Stat Sci. 1989; 4:409–435.
8. Sasena, M.J., Papalambros, P.Y. and Goovaerts, P. Proceedings of the 8th AIAA / USAF / NASA / ISMMO Symposium of Multidisciplinary Analysis & Optimization, September 2000. AIAA Paper 2000-4921.
9. Schonlau, M., Welch, W., Jones, D. (1998. New Developments and Applications in Experimental Design, N. Flournoy, W.F. Rosenberger, and W.K. Wong (editors), Institute of Mathematical Statistics, Hayward, California, Vol. 34, 11-25
10. Shinkyu Jeong, M. and Yamamoto, K. (2005). Journal of Aircraft 42, 412-420.