

Requirements and new approaches of probabilistic optimal design from a practical point of view considering steam turbines

K.Cremanns*, D. Roos*, R. Voß** *Institute of Modelling and High-Performance Computing, Niederrhein University of Applied Sciences **Siemens AG

Outline

- 1. Introduction
 - Motivation
 - Objective
 - Planned application
- 2. Previous results
 - · Optimized latin hypercube sampling
 - Advanced moving least square approximation
- 3. Outlook
 - Integration into optiSLang
 - Further developments

Outline

1. Introduction

- Motivation
- Objective
- Planned application
- 2. Previous results
 - Optimized latin hypercube sampling
 - Advanced moving least square approximation
- 3. Outlook
 - Integration into optiSLang
 - Further developments

1. Introduction - Motivation



1. Introduction - Objective



1. Introduction – Planned application



1. Introduction – Planned application

"External" constraints:

- Steam temperature (THD)
- Start times
- Required lifetime
- Number of starts
- Number of load cycles
- ...

"Internal" constraints :

- Fulfillment mechanical integrity rotor + blading
- Manufacturability
- Compliance criteria rotor dynamics
- ...

Objectives:

- Long lifetime (h)
 - High number of starts (N)
 - High start speed (MW / min)
 - High number of load cycles (N)
- High performance (MW)
- High efficiency (mu)
- Eigenfrequencies outside critical regions (50Hz, 60 Hz, 100Hz, 120Hz)



Outline

- 1. Introduction
 - Motivation
 - Objective
 - Planned application

2. Previous results

- · Optimized latin hypercube sampling
- Advanced moving least square approximation
- 3. Outlook
 - Integration into optiSLang
 - Further developments

Base : standard latin hypercube sampling

 Parameter space is divided into N (samples or classes) with the same probability of 1 / N, and in each of these classes a random point is selected.

Advantages:

- Low computation time to generate.
- It has a lower variance compared to standard Monte Carlo sampling.
- > The value range of each variable is covered.
- ≻Non collapsing.

Cons:

>It can cause unwanted input parameter correlations.

➢ It does not guarantee a "filling" coverage of the parameter space.



Optimized latin hypercube sampling (OHLS):

- The standard latin hypercube sampling gets improved by optimization.
- Optimization criteria are:

> Mean value of the linear correlation coefficients ρ_{ij} (modified criterion of Owen [1]): $\rho_{avg} = \frac{\sum_{i=2}^{n} \sum_{j=1}^{i-1} |\rho_{ij}|}{n(n-1)/2}; 0 \le \rho_{avg} \le 1$

Euclidean Maximin [2][3] design coefficient (scaled to [0,1]):

$$l_2 = d(x_i, x_j) = \min(-\sqrt{\sum_{k=1}^n |x_{ik} - x_{jk}|^2}; 0 \le l_2 \le 1)$$

- Overall criterion: $f(\rho_{avg}, l_2) = w_1 \rho_{avg} + w_2 l_2 = \Psi$; with $w_1 + w_2 = 1; w_1 = 0.5, w_2 = 0.5$
- Optimization method: Simulated Annealing (SA) [4]

Comparison for N=50 n=2:







Comparison for Matlab Ihsdesign N=100 n=10: criterion maximin

Matlab lhsdesign criterion: maximin 1.000.000 iterations

Own OLHS 10.000 iterations



X2

K. Cremanns, D. Roos, R. Voß

Comparison for Matlab Ihsdesign N=100 n=10: criterion: orthogonality

Own OLHS 10.000 iterations

Matlab Ihsdesign criterion: orthogonality 1.000.000 iterations



Comparison for optiSLang N=100 n=10: Advanced latin hyper cube sampling (ALHS)

optiSLang ALHS (optimized correlation)

Own OLHS 10.000 iterations



Comparison for optiSLang N=100 n=10: Advanced latin hyper cube sampling (ALHS)

x10 0.90 0.90 x9 0.75 x8 0.75 x7 0.60 0.60 x6 x5 0.45 0.45 x4 0.30 0.30 x3 0.15 x2 0.15 x1 0.00 0.00 30

optiSLang ALHS (optimized correlation)



$$\rho_{av,g} = 0.72E-2$$

 $\rho_{avg} = 0.694 \text{E-}3$

Own OLHS 10.000 iterations

Uniformity

Own OLHS 10.000 iterations



• Sequentiell OLHS (SOLHS):

- Start with a OLHS with presented method with e.g. 2 samples.
- Add sample points in that way, that each step the number of samples double (necessary to keep **uniformity**).
- The new points are also optimized with a combination of random picks and SA regarding ψ of the whole new design matrix (existing points + additional points).
- Continue until metamodel is convergent (further steps are adding 4, 8, 16... samples).
- Advantages:
 - This method can result in a lower number of needed calculated samples, but still in a oversampling, because the fixed number of additional sets.
 - Over the iterations it's possible to check the convergence of the metamodel.
 - Every new created LHS is still a OLHS, with optimal settings for maximin and correlation.

• Example for 64 samples with SOLHS (steps: adding 2, 4, 8, 16, 32 samples)



- × Start 2
- add 2
- add 4
- * add 8
- add 16
- add 32

Comparison for Matlab N=64 n=2: maximin and SOHLS 64 samples





 $l_2 = 0.035$





 $l_2 = 0.041$

Comparison for Matlab N=64 n=2: orthogonality and layer 64 samples

Matlab OLHS criterion: orthogonality MCS (100.000) Iterations



Own OLHS 10.000 iterations layer 64 samples





Standard moving least square approximation (MLS) [6] :

$$\begin{split} \widetilde{y}(x) &= p^T(x)a(x) \quad ; \text{Approximation of the results of the test points} \\ p(x) &= [1, x_1, x_2, ..., x_n, x_1^2, x_2^2, ..., x_n^2, x_1 x_2, ... x_n x_m]^T \quad ; \text{Polynomial basisfunction} \\ a(x) &= (P^T W(x) P)^{-1} P^T W(x) y(x) \quad ; \text{Moving coefficients depending on testpoint x} \\ P &= [p^T(x_1) p^T(x_2) ... p^T(x_N)] \quad ; \text{Contains all polynomial basisfunction of the support points} \\ y(x) &= [y(x_1) y(x_2) ... y(x_N)] ; \text{Contains the results of the objective function for all support points} \\ W(x) &= diag[w_1(x - x_1) w_2(x - x_2) ... w_N(x - x_N)] \quad ; \text{Overall weighting matrix (diagonal matrix) contains for each testpoint a seperate weighting \\ (1) &= \int_{-\infty}^{\infty} e^{-(\frac{\|x - x_N\|}{D\alpha}\|^2} \|x - x_N\| \le D \quad ; \text{Gaussian weighting function} \end{split}$$

 $w_{N}(\|x-x_{N}\|) = \begin{cases} e^{-(\frac{\|x-x_{N}\|}{D\alpha})^{2}} & \|x-x_{N}\| \leq D \\ 0 & \|x-x_{N}\| > D \end{cases}$; Gaussian weighting function $\alpha = \frac{1}{\sqrt{-\log_{10}(0.001)}}$; Constant

D = Self- selected parameter affecting model accuracy

Advanced moving least square approximation (AMLS):

- A new concept of weighting, which not only includes a weighting for each test point, but also per variable. This means there exist not only one weighting matrix and D for the whole approximation but a weighting matrix and D per variable and if present for the crossterms.
- The D's are chosen through optimization with a particle swarm optimization algorithm.
- The optimization objective is the generalized coefficient of determination:

$$R^{2} = \left(\frac{\sum_{N=1}^{k} (y^{k} - \mu_{y})(\hat{y}^{k} - \mu_{\hat{y}})}{(N-1)\sigma_{y}\sigma_{\hat{y}}}\right)^{2}; 0 \le R^{2} \le 1$$

- Whereby the sample points are divided in equal subsets (0.2*N), so that every sample point is support- and testpoint, then the average R^2 is calculated (cross validation).
- α is not longer a constant, but a further optimization variable, in order to better fit the problem, rather than to change the weighting function. Therefore is also used a new formulation of the Gaussian weighting function:

$$w_N(\|x - x_N\|) = \begin{cases} \frac{e^{-(\alpha \|x - x_N\|/D)^2} - e^{-\alpha^2}}{(1 - e^{-\alpha^2})} \\ 0, \end{cases}$$



Shape of the Gaussian weight functions with different control parameter α with $R = ||x - x_N|| / D$

Benchmark of the AMLS:

- Used Metamodels [9]: MLS (D is through optimization), AMLS, super vector machine [10], Gaussian-process-regression [11], optiSLang MLS (gaussaian weighting function, same D like MLS, all parameters have a quadratic basis polynom)
- Evaluation criterion: $RMSE = \sqrt{\frac{1}{N_{test}}\sum_{k=1}^{N_{test}}(y(x_k) \hat{y}_e(x_k))^2}$
- $N_{Support} = 120$; $N_{test} = 500$; n = 2 4 (testpoints are not used, to create the metamodel)
- Testfunctions:
 - Rosenbrock: $y = 100(x_1^2 x_2)^2 + (x_1 1)^2; -2 \le x_i \le 2$
 - Normal PDF shape: $y = \frac{1}{1 + x_1^4 + 5x_2^4 + 2x_2^2}; -3 \le x_i \le 3$
 - Sixhump Camelback: $y = 4x_1^2 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 4x_2^2 + 4x_2^4; -2 \le x_i \le 2$
 - Own testfunction 1 $y = x_1 2 + x_2^2 + \sin(x_3) + \cos(x_4) + x_5^3; -5 \le x_i \le 5$
 - Own testfunction 2 $y = x_1 + x_1^2 x_2^2 + \sin(x_3)^2 + \cos(x_4 x_2) + x_5^3 x_1; -5 \le x_i \le 5$

Rosenbrock: $y = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2; -2 \le x_i \le 2$



Rosenbrock: $y = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2; -2 \le x_i \le 2$

Response surface AMLS Response surface osl_MLS MLS approximation of Output Output -500 $\begin{array}{c} -2.0 \\ -1.5 \\ -0.5 \\ x2 \end{array} \\ 0.0 \\ x2 \end{array} \\ 0.5 \\ 1.0 \\ 1.5 \\ 2.0 \\ 2.0 \\ 1.5 \end{array} \\ 1.5 \\ 1.0 \\ 1.0 \\ 1.5 \\ 1.0$ -0 0.5 0 -0.5 -1 -1.5 -2 0.5 x1 $-1.5 -1 -0.5 0 \\ -1.5 -1 -0.5 0 \\ -1.5 -1 -0.5 0 \\ -1.5 -1 \\ -1.$ -16 1.5

Rosenbrock:
$$y = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2; -2 \le x_i \le 2$$

Rosenbrock response surface for x1



red = Approximation green = Calculation

Normal PDF shape:
$$y = \frac{1}{1+x_1^4+5x_2^4+2x_2^2}; -3 \le x_i \le 3$$



Normal PDF shape:
$$y = \frac{1}{1+x_1^4+5x_2^4+2x_2^2}; -3 \le x_i \le 3$$

Response surface AMLS

Response surface osl_MLS

0.704 0.675 0.650 0.625 1.0 0.600 0.575 0.550 0.525 0.8 0.500 0.8 0.475 0.450 0.425 0.6 0.400 y 0.375 0.6 0.350 Output 0.325 0.4 0.300 0.275 0.250 0.4 0.2 0.225 0.200 0.175 0.150 0.0 0.2 0.125 0.100 3 0.075 2 0.050 -3 -2 1 0.025 -1 C -0.011 0 x2 0 x1 x1 3

K. Cremanns, D. Roos, R. Voß

Sixhump Camelback: $y = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4; -2 \le x_i \le 2$



Sixhump Camelback: $y = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4; -2 \le x_i \le 2$

Response surface AMLS

Response surface osl_MLS

MLS approximation of Output



Sixhump Camelback: $y = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4; -2 \le x_i \le 2$



red = Approximation green = Calculation

Own testfunction 1: $y = x_1 2 + x_2^2 + sin(x_3) + cos(x_4) + x_5^3$; $-5 \le x_i \le 5$



Own testfunction 1: $y = x_1 2 + x_2^2 + sin(x_3) + cos(x_4) + x_5^3$; $-5 \le x_i \le 5$

Approximationspoints AMLS

Approximationspoints osl_MLS (x1, x3, x4 constant)



MLS approximation of Output

Own testfunction 1: $y = x_1 2 + x_2^2 + sin(x_3) + cos(x_4) + x_5^3$; $-5 \le x_i \le 5$



red = Approximation green = Calculation

Own testfunction 2: $y = x_1 + x_1^2 x_2^2 + \sin(x_3)^2 + \cos(x_4 x_2) + x_5^3 x_1; -5 \le x_i \le 5$



Own testfunction 2: $y = x_1 + x_1^2 x_2^2 + \sin(x_3)^2 + \cos(x_4 x_2) + x_5^3 x_1; -5 \le x_i \le 5$

Response surface AMLS

Response surface osl_MLS (x3, x4, x2 constant)



Conclusion:

- For problems without variables, which need to be filtered / constant, the AMLS showed better results as optiSLangs MoP with Gaussian Weighting and optimized D = MLS.
- For Problems with more than 2 variables, it occurs a advantage through filtered or constant parameters.
- The AMLS is for every testfunction better than the standard MLS with optimized D.

Plans:

- Our objective is to use metamodells for robust design optimization, therefore very accurate metamodels are necessary.
- It is planned to use optiSLangs variable reduction methods in combination with our metamodells to further improve the metamodels.

Outline

- 1. Introduction
 - Motivation
 - Objective
 - Planned application
- 2. Previous results
 - Optimized latin hypercube sampling
 - Advanced moving least square approximation
- 3. Outlook
 - Integration into optiSLang
 - Further developments

3. Outlook - Integration into optiSLang

• Own developments in python can easily be integrated into optiSLang v 4.

Advantages:

- Usage of existings methods / post processing and great process integration possibilities for other softwares, like ANSYS, ABAQUS, Excel.
- Extract data of different kinds of formats to work together. (ETK)
- No need to programm interfaces.
- Variables in Python can be decleared as parameters for optiSLang -> a lot of opportunities to work with own developments and methods / post processing of optiSLang.



Developement of:

- AMLS (tests of different optimization algorithms for a high number of variables).
- Changing polynom coefficients regarding the variables (thanks to T. Most).
- Combination of sampling and estimination of the prognosis quality of the metamodell -> convergence analysis of the prognosis quality to sample the minimum number of designs.
- Chance-constraints stochastic multiobjective optimization on metamodels (optimization taking into account the probability of failure of the constraints).
- Simulationmodel for the different fractions of the hp/ip rotor developement.

Literature

- [1] Owen, A. (1994). Controlling correlations in Latin hypercube samples. J. Amer. Statist. Assoc. 89, 1517-1522.
- [2] Johnson, M., Moore, L. and Ylvisaker, D. (1990) Minimax and maximin distance designs. Journal of Statistical Planning and Inference 26 (2): 131-148
- [3] Morris, M. D., Mitchell, T. J. (1995), Exploratory designs for computer experiments. Journal of Statistical Planning and Inference 43 (3): 381-402
- [4] Kirkpatrick, S.; Gelatt Jr, C. D.; Vecchi, M. P. (1983). "Optimization by Simulated Annealing". Science 220 (4598): 671–680
- [5] web.stanford.edu
- [6] Lancaster P. and Salkauskas K. (1981). Surfaces Generated by Moving Least Squares Methods. Mathematics of Computation, Vol. 37, 141-158.
- [7] Eberhart, R. C. and J. Kennedy (1995). A new optimizer using particle swarm theory. pp. 39–43. Nagoya, Japan. IEEE Service Center Piscataway NJ: Proceedings of the Sixth International Symposium on Micromachine and Human Science.
- [8] www.quantcom.de/diemorphogenetischefelder.html

Literature

- [9] Pedregosa et al (2011)., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830.
- [10] C. Cortes, V. Vapnik (1995), Support-vector networks, Machine Learning, 20, 273-297
- [11] C. E. Rasmussen, C. K. I. Williams (Ed. T Diettrich), (2006), Gaussian Processes for Machine Learning, MIT Press.

Appendix

The advantages:

- Failure probabilities instead of safety factors
- Better understanding of parameters through sensitivity analysis
- Increased flexibility to perform changes within the parameter space
- The possibility to use extended application limits in contrast to the deterministic design
- Avoid interface conflicts and no need for expert knowledge in all areas / tools
- Optimal compromise solutions for the requirements

Simulated Annealing (SA):

- Start is a random LHS with a design matrix D, where each column stands for a design point.
- By interchanging p (p < n) elements of two randomly selected columns -> $D_{\rm try}\,$ is a new design matrix.
- If $\psi(D_{try})$ is better than $\psi(D)$ D gets D_{try} . If $\psi(D_{try})$ is worse than $\psi(D)$ a random decision is made, if still D gets D_{try} or whether D_{try} is discarded. That D gets D_{try} will happen with the probability:

$$\pi = exp(-[\Psi(D_t ry) - \Psi(D)]/t)$$
 , t=self-selected parameter.

• This random decision prevents that only a local minimum is found.

- The result of the optimization is D_{best}.
- Start design matrix D is the best sampling regarding the overall ciriterion ψ out of 500 generated standard latin hypercube samplings



Comparison for Matlab N=32 n=2: maximin and SOHLS 32 samples









 $l_2 = 0.065$

Comparison for Matlab N=16 n=2: maximin and SOHLS 16 samples



Own SOLHS 10.000 iterations



Comparison for Matlab N=16 n=2: orthogonality and layer 16 samples

Matlab OLHS criterion: orthogonality 100.000 iterations



Own SOLHS 10.000 iterations



07.11.2014 WOST 2014

Comparison for Matlab N=32 n=2: orthogonality and layer 32 samples

Matlab OLHS criterion: orthogonality 100.000 iterations





07.11.2014 WOST 2014



07.11.2014 WOST 2014

Particle Swarm algorithm [7]:

- An initial population ("swarm") of possible candidates ("particle") move through the parameter space .
- The direction of the "particle" is guided by the knowledge of its best position (local optimum), and the best known position of the "swarm leader" (global optimum).
- If new better positions are discovered, they are used to steer the "swarm".
- This process is repeated a certain number of iterations until an optimal solution is found.

Advantages:

- Suitable for multiobjective optimization.
- High number of input variables possible.

Disadvantage:

• Suitable coefficients for the "swarm" to determine behavior (8 coefficients).



Rosenbrock:
$$y = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2; -2 \le x_i \le 2$$



Red = Calculation

Own testfunction 1: $y = x_1 2 + x_2^2 + sin(x_3) + cos(x_4) + x_5^3$; $-5 \le x_i \le 5$



Blue = Approximation Red = Calculation

Sixhump Camelback: $y = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4; -2 \le x_i \le 2$



Red = Calculation