The Evolutionary Algorithm SAMOA with Use of Design of Experiments

Susanne Zaglauer
BMW AG
80809 München, Germany
susanne.zaglauer@bmw.de

ABSTRACT

In the automotive industry, especially in engine calibration, many technical optimization tasks cannot be solved by common evolutionary algorithms. The algorithms must work with many difficult boundary conditions, like multi-objective optimization and a priori unknown constraints. They must also deal with the growing complexity of the optimization tasks and the huge experimental effort and they must find the global optimum in a reasonable time.

Therefore, the principle target of this contribution is the presentation of an intelligent Design of Experiments (DoE) strategy of the recently developed algorithm $SAMOA$ to reduce significantly the optimization time of $SAMOA$, which is a combination of a genetic algorithm and an evolutionary strategy. This algorithm can handle the mentioned problems and solves multi-objective problems with a priori unknown constraints. It can operate parallel and solve the technical optimization problems in a reasonable time, because of an intelligent DoE strategy.

Categories and Subject Descriptors

G.4 [**MATHEMATICAL SOFTWARE**]: Algorithm design and analysis;
G.3 [**PROBABILITY AND STATISTICS**]: Experimental design

General Terms

Design

Keywords

Evolutionary Algorithm, SAMOA, Design of Experiments

1. INTRODUCTION

The automotive industry is, based on the sales, the most important industry section in Germany and in other countries and a powerful factor of innovation, growth and employment. But with many various framework conditions, like the CO_2 and fuel reduction at equal engine power, the application of the control units becomes more challenging and important for the hole system automotive. Traditional application methods fail because of the growing complexity of the optimization tasks, the huge experimental effort and the contradictory optimization aims. They also cannot deal with the dynamic constraints, which are a priori unknown and change with every measurement, and they cannot find the global optimum in a reasonable time. Therefore, the evolutionary algorithm $SAMOA$ is developed, which can handle with a priori unknown constraints, with multi-criteria problems and can solve complex technical design problems in a reasonable time, because of an intelligent Design of Experiments strategy.

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.
ACM 978-1-4503-1178-6/12/07.

2. SAMOA

In the following the recently developed evolutionary algorithm $SAMOA$ (**S**elf-**A**daptive **M**ulti-criteria **O**nline optimization **A**lgorithm) is presented. At the beginning of the algorithm the individuals must be created. Many common evolutionary algorithms use mostly a random initialization of the individuals. In $SAMOA$ an intelligent Design of Experiments strategy is used, which is further discussed in the next section. The fitness assignment should then define how many offsprings every individual produces. A robust solution, which is also implemented in $SAMOA$, is the nonlinear ranking [1] and in use of multi-criteria problems the S-metric or hypervolume measure [3]. At the selection the individuals, who serve as parents for the next generation, are chosen according to their fitness. The selection probability of an individual is calculated by the fitness value of the individual normalized by the whole fitness of the selection pool. In $SAMOA$ the truncation selection is implemented, because it is faster as other selection methods and is advantageous in parameter optimizations [4]. After the selection follows the intermediary recombination, in which the information of the parents is combined in a special way and the offsprings are created of the parents. After the recombination the mutation of the offsprings is performed. Therefore, the variables of the offsprings are changed by little disturbances, which occur with a low probability. In $SAMOA$ a mutation of real variables with adaption of the covariance matrix is realized [2]. After the production of the offsprings, they must be reinserted in the population. In $SAMOA$ a reinsertion with offspring selection is implemented, in which only a part of the offsprings is reinserted. The algorithm runs as long as a termination criterion is reached. $SAMOA$ uses a combination of the direct termination criterion "Maximum number of generations respectively function evaluations" and the indirect termination criterion "Running mean" to limit the maximum calculation time and to prevent that the optimization runs much longer although no better objective value is found.

To deal with a priori unknown constraints a method is implemented in $SAMOA$, at which the dynamic constraints are modeled by regression models and by a hull model with confidence terms [5]. These models are then added to the objective function with the help of penalty methods. The term dynamic shows, that the information over the limit functions raises dynamically during the optimization.

3. DESIGN OF EXPERIMENTS

An important question is how the measuring points, the initial individuals, should be distributed efficiently in the experimental space. The answer to this question is provided by the Design of Experiments (DoE). The goal is to identify the connections between target and influence factors systematically with as few experiments as possible.

Model-free Design of Experiments

In the practice it is not possible or it takes to many time to calculate any derivatives or covariance matrices from technical design problems and so model-free DoE are very important in this context. They do not care about the used problem structure and try to cover the input space as equally as possible (space filling designs). The most popular space filling designs are the random, the Latin Hypercube distribution and the distance-based criteria, especially the S-optimal design. The random DoE involves randomly assigning experimental conditions. Each member of the population has an equal chance of being included in the sample. The Latin Hypercube design is a statistical sampling method and in this schema only one sampling point is in every column and row of a grid. For this purpose one sampling position is placed randomly in every cell along the grid diagonal. After this the rows of the grid are changed in that way, that a chosen criteria, like the maximizing of the minimal distance of the points, is fulfilled:

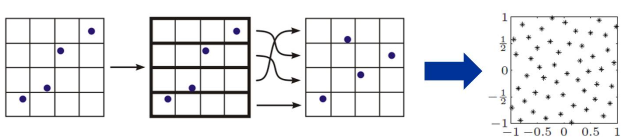

Figure 1: ***The maxmin Latin-Hypercube-Distribution.***

The distance-based criteria are based on the distance $d(x, \mathcal{A})$ from a point x in the p-dimensional Euclidean space $\mathbb{R}^p$ to a set $\mathcal{A} \subset \mathbb{R}^p$. The S-optimality seeks then to maximize the harmonic mean distance from each design point to all the other points in the design:

$$\max_{y\in\mathcal{C}} \frac{N_{\mathcal{D}}}{\sum_{y\in\mathcal{D}} \frac{1}{d(y,\mathcal{D}-y)}},$$

where $\mathcal{C}$ is the set of candidate points and $\mathcal{D}$ is the set of design points. An interesting approach is also the combination between the Latin-Hypercube and the S-optimal design. Therefore, the rows of the Latin-Hypercube grid are changed in that way, that the harmonic mean distance from each design point to all the other points in the design is maximized.

Model-based Design of Experiments

If the derivatives and the covariance matrix of a given problem can be calculated, it is more useful to use model-based DoE. These try to distribute the design points in the input space in a way that the estimation of the parameters is as insensitive as possible to the measurement noise. Two examples are the D-optimal and the Bayesian design. The D-optimal design is the most popular model-based DoE. To estimate the variance of a parameter estimation of any parametric model the Fischer Information Matrix (FIM) can be used. The FIM describes the information content of a random variable $x \sim \mathcal{N}(\mu(\theta), \sigma I)$ at the parameter θ from which the Likelihood-function $L(\theta) = p(x, \theta)$ is dependent

$$FIM(\theta) = \frac{1}{\sigma^2} \frac{\partial \ln p(x,\theta)^\top}{\partial \theta} \frac{\partial \ln p(x,\theta)}{\partial \theta}.$$

The D-optimal design is then defined by the maximization of the determinant of the FIM. But this design is often criticized because of its dependency to the statistical model and the tendency to weight the boundary area of the experimental space significantly. In order to defuse this critique a Bayesian modification of the D-optimal design is used, which adds control points to the D-optimal design to decide, if the model has uncertainties. Control points in a design are useful mainly for the protection against higher order effects. These terms are called potential terms. The assumed model contains only primary terms. Typically the sample size is not large enough to estimate all primary and potential terms simultaneously. The issue is to develop an approach, which estimates the p primary terms precisely during a general traceability is provided for the q potential terms. The Bayesian design maximizes then the determinant of the posterior covariance matrix

$$\max \ det(FIM(\theta) + K),$$

whereby K is a diagonal matrix whose first p diagonal elements are equal to 0 and whose remaining q diagonal elements are equal to 1. For further details see [6].

4. REFERENCES

[1] B. Bagot and H. Pohlheim. Complementary selection and variation for an efficient multiobjective optimization of complex systems. In *GECCO 2004 - Proceedings of the Genetic and Evolutionary Computation Conference*, 2004.

[2] N. H. Christian Igel and S. Roth. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15:1–28, 2007.

[3] C. C. Coello, D. V. Veldhuizen, and G. Lamont. Evolutionary algorithms for solving multi-objective problems. *Kluwer Academic Publishers, New York*, 2002.

[4] H. Mühlenbein and D. Schlierkamp-Voosen. *Evolutionary Computation 1*, chapter : Predictive Models for the Breeder Genetic Algorithm: I. Continous Parameter Optimization., pages 25–49. 1993.

[5] S. Zaglauer. The evolutionary algorithm samoa with use of dynamic constraints. In *GECCO 2012, Genetic and Evolutionary Computation Conference*, 2012.

[6] S. Zaglauer and M. Deflorian. Bayesian D-optimal Design. In *Design of Experiments (DoE) in Engine Development*, 2011.