You will find the figures mentioned in this article in the German issue of MTZ 5/2003 beginning on page 432.

Modellbasierte Online-Optimierung moderner Verbrennungsmotoren
Teil 1: Aktives Lernen

Model-Based Online Optimisation

of Modern Internal Combustion Engines
Part 1: Active Learning

This two-part article presents the model-based optimisation algorithm “mbminimize”. It was developed in a corporate project of the University of Tübingen and the BMW Group for the purpose of optimising internal combustion engines online on the engine test bed. The first part concentrates on the basic algorithmic design, as well as on modelling, experimental design and active learning. The second part will discuss strategies for dealing with limits such as knocking.

By Jan Poland,
Kosmas Knödler,
Thomas Fleischhauer,
Alexander Mitterer,
Stephan Ullmann and
Andreas Zell

1 Introduction

Manufacturers of modern internal combustion engines are confronted with increasing challenges: The legal restrictions concerning exhaust emissions are getting stricter, fuel is becoming more expensive and customers are demanding powerful and comfortable engines with low fuel consumption. As a result, engine complexity has greatly increased, as has the number of adjustable parameters. Tuning the engine thus results in a complex optimisation problem, whose solution requires the aid of computers.

The application of computer-aided offline optimisation has become common. Such approaches are presented, for example, in [1], [2] and [3]. They rely on the following principle: On the basis of an experimental plan, a sufficiently large number of measurements are performed at the test bed. This data, possibly combined with previous knowledge, is used to calculate a computer model that simulates the engine. This model can then be optimised by appropriate algorithms. The optima are again verified at the test bed and finally used for tuning the engine.

In the online optimisation, the optimisation system directly interacts with the test bed, which means that, in the ideal case, the process is fully automatic. This approach carries the further advantage that information gained by measurement can be evaluated immediately. This point will be discussed in detail in sections 3 and 4. Offline optimisation, on the other hand, has the advantage that the computation time can be completely separated from the test bed, i.e. the test bed can be used manually during the computations. The automatic test bed control with appropriate limit handling poses new challenges to the algorithms. This issue will be discussed in the second part of this article.

Some online optimisation systems are currently available, e.g. Cameo [4] and Vega [5]. This report introduces a new online optimisation algorithm, mbminimize. It differs from the known algorithms in several points:

- Modelling: A global model is used instead of distinct models for each operating point. Thus, the information gained by one measurement has an influence on several operating points, thus reducing measurement costs.

- Model types: A heterogeneous committee containing neural networks and cubic regression models is used.
- Experimental design: The model itself determines the experimental design (active learning), and the influence of the objective of the optimisation is increased over time.
- Limit handling: New strategies are developed. They not only allow the approximation of infeasible areas in an approximate way, but also ensure that the engine is reset to safe parameters when a limit occurs (see the second part of this article).

2 Requirements

This article concentrates on the following parameterised optimisation problem that we call "full factorial tuning". For each operating point, one or more parameter combinations for the actuators are requested in such a way that a given objective achieves an optimum, e.g. minimum fuel consumption. For this purpose, some particularities of internal combustion engines have to be considered:

- Measurements at the test bed are expensive and time-consuming. Therefore, a minimum number of measurements are desirable.
- Measurement results are corrupted by noise. Thus, noise handling is important.
- Changing actuators may require a stabilising interval of a considerable duration. Therefore, changes in the operating point in particular should occur as rarely as possible.
- There are limits, i.e. constraints, for the optimisation.
- Safe optima are of particular interest, i.e. optima that possess a large neighbourhood of good objective values.

Because of these points, the new algorithm described in this report is designed as a model-based optimisation algorithm. The global model is defined on the whole search space that is spanned by the operating plane and the actuator dimensions. Thus, the number of measurements is kept to a minimum, and the model allows noise handling. Two model types are used: linear parameterised and non-linear regression models. For linear parameterised models, the model output is a linear combination of (non-linear) basis functions. For example, polynomials of a maximum third degree define the well-known and widely used cubical model. Here, the model output always depends linearly on the vector of free parameters. For a given data set of measurements, the optimal parameter vector is easily estimable if the square error is to be minimised.

Non-linear models on the other hand are non-linear in the parameter vector, too. This results in considerably higher flexibility. The optimum parameter vector now has to be determined by means of non-linear optimisation, so called training. Moreover, the optimum parameter vector is no longer unique, and the model can easily be fitted to the noise rather than to the true function. This is called overfitting. Therefore, appropriate regularisation techniques are indispensable (see for example [6]). A particular type of non-linear model is the feedforward network, a subclass of artificial neural networks (see [7]). This model type is used for the algorithm presented here.

3 Active Learning

Considering the high search space dimension and the expense of each single measurement, the central problem of engine optimisation is clearly the experimental design. This means performing a minimum number of measurements in order to gain maximum information. The information gain is normally quantified by means of statistics, and there are many different methods and a broad range of literature available in this field, for example [8].

An important class of experimental design methods makes use of the model itself. This is done by evaluating the uncertainty of the model on the search space. Different design criteria already arise at this point. One could, for example, try to minimize the average uncertainty or, alternatively, the maximum uncertainty of the model.

The model uncertainty at a given point in the search space depends on the model. For example, the uncertainty of a linear model always attains its maximum at the search space boundary, **Figure 1**. On the other hand, an optimum experimental design contains only boundary points in this case. These considerations can be extended to linear parameterised models, such as cubical models. The model uncertainty at a given point can be expressed by a term that depends on the model and on the actual measurement points (see [8]). An important and easy to evaluate optimality criterion is satisfied if the determinant of the so-called information matrix is at the maximum. Such experimental designs are also called D-optimal, and under certain conditions this is equivalent to minimizing the maximum uncertainty on the search space (see [8]).

It is striking that the measurement results do not occur anywhere in this representation of the model uncertainty. This means that an optimum design can already be calculated at the beginning, and cannot be improved by the information gained by measuring. In other words, this theory favours offline optimisation. This situation is fundamentally changed if non-linear models are considered. In this case, the uncertainty can be expressed by means of an information matrix as well, which now depends on the training results, i.e. it depends on the results of the measurements (see [9]).

If non-linear models are used for approximating the objective function, an online experimental design is appropriate for exploiting the information gained by measuring for the purpose of optimally placing new measurements. We therefore consider the following situation: Given some measurement data and a model, we look for a new point in the search space that is to be determined in order to maximise the expected information gain by measurement. This scenario is referred to as active learning or query, since it is the model itself that decides which point is to be the next one to be trained. A possible criterion is maximising the model uncertainty, which is given by the information matrix and the point. This is again closely linked to D-optimality. **Figure 2** shows an example. The uncertainty of the model (here an artificial neural network) is large in regions with a low data density.

There are several variants of this criterion (see e.g. [9], [10]) that are obtained under slightly different theoretical assumptions. They behave in a very similar way in practice. However, these methods tend to behave unfavourably if the normality assumption is not fulfilled. The accuracy of the uncertainty estimate is increased by taking large samples of model parameters according to their true probability distribution. These Markov Chain Monte Carlo (MCMC, see e.g. [11]) methods require a large amount of computation time.

A further possibility is to use a committee of models. This has several benefits. The expected error is smaller than for single models, different model types can be considered simultaneously (heterogeneous committees) and the model disagreement (i.e. the variance) of the output of the committee members can be used as an estimate for the model uncertainty, i.e. as a query criterion. This query by committee (QBC) criterion (see e.g. [12]) was superior to other criteria in our empirical studies (see [13]) and can be evaluated very quickly. Therefore, it is used in the new algorithm. For homogeneous committees, the QBC criterion can be interpreted as a variant of an MCMC method. Each committee member is a sample of the parameter vector, and the probability distribution is biased toward its peaks because of the training. Moreover, the use of a heterogeneous committee reduces the probability of a completely incorrect model output in spite of a good fit due to an inappropriate model type. **Figure 3** shows a committee consisting of a linear model and a neural network. **Figure 4** illustrates the output of such a committee and the disagreement of the committee members, i.e. the QBC criterion.

In the course of the optimisation, not only the model uncertainty but also the approximated objective should be incorporated into the query criterion to an increasing degree. In this way, the search space is explored more thoroughly in regions with good objective values, and the optima are localised with greater precision. This is implemented in the presented algorithm by excluding parts of the search space with an estimated objective value below a threshold.

The mbminimize algorithm is started with an initial experimental design that is computed according to space-filling or D-optimal criteria. Its order is optimised by the methods described in [14]. After that, further measurement points are determined via active learning, then arranged optimally if necessary and used for measuring. Finally, the approximated optima are computed and verified.

4 Conclusions

The mbminimize algorithm has been tested thoroughly by means of benchmark functions and simulations and has proven to be quite powerful and efficient. For example, all three optima of the Branin function (see [15]) are correctly localised with less than 40 function evaluations. Initial experiments at the engine test bed yielded promising results.

By using a global model and sophisticated online experimental design techniques, in particular active learning, the mbminimize algorithm presented in this report is able to drastically reduce the number of function evaluations compared to common optimisation algorithms. The approach allows a large variety of extensions and combinations with other methods. For example, other model types can be incorporated. Another possibility is enhancement by multi-criterion techniques for optimising several objectives. In the second part of this article, we will present the handling of limits and constraints.

This research project was supported by the BMBF (grant no. 01 IB 805 A/1).

References

[1] Mitterer, A.: Optimierung vielparametriger Systeme in der Antriebsentwicklung, Statistische Versuchsplanung und Künstliche Neuronale Netze in der Steuergeräteauslegung zur Motorabstimmung. Dissertation, Lehrstuhl für Meßsystem- und Sensortechnik, TU München, 2000

[2] Schüler, M.; Hafner, M.; Isermann, R.: Einsatz schneller neuronaler Netze zur modellbasierten Optimierung von Verbrennungsmotoren. Teil 1: Modellbildung des Motor- und Abgasverhaltens. In: MTZ 61 (2000) Nr. 10, S. 704-710

[3] Hafner, M.; Schüler, M.; Isermann, R.. Einsatz schneller neuronaler Netze zur modellbasierten Optimierung von Verbrennungsmotoren. Teil 2: Stationäre und dynamische Optimierung von Verbrauch und Emissionen. In: MTZ 61 (2000) Nr. 11, S. 2-8

[4] Gschweitl, K.; Pfluegl, H.; Fortuna, T.; Leithgoeb, R.: Steigerung der Effizienz in der modellbasierten Motorenapplikation durch die neue CAMEO Online DoE-Toolbox. In: ATZ 103 (2001)

[5] Bredenbeck, J.: Statistische Versuchsplanung für die Online-Optimierung von Verbrennungsmotoren. In: MTZ 60 (1999) Nr. 11, S. 740-744

[6] MacKay, D. J. C.: A practical Bayesian framework for backpropagation networks. In: Neural Computation 4 (1992) S. 448-472

[7] Zell, A.: Simulation Neuronaler Netze. Addison-Wesley, Bonn, 1994

[8] Bandemer, H.; Bellmann, A.: Statistische Versuchsplanung. Teubner, 4. Auflage 1994

[9] Cohn, D. A.: Neural network exploration using optimal experiment design. In: Jack D. Cowan, Gerald Tesauro, and Joshua Alspector, (Editoren), Advances in Neural Information Processing Systems 6 (1994), S. 679-686

[10] Paass, G.: Query sampling for prediction and model selection. In: IPSJ Magazine 38 (1997) Nr. 7, S. 562-568

[11] Neal, R. M.: Probabilistic inference using markov chain monte carlo methods. Technical report, Dept. of Computer Science, University of Toronto, Sep 1993. CRG-TR-93-1

[12] Seung, H.; Opper, M.; Sompolinski, H.: Query by committee. In: Fifth Annual Workshop on Computational Learning Theory, 1992 (ACM, New York), S. 287-294

[13] Poland, J.; Zell, A.: Different criteria for active learning in neural networks: A comparative study. In: M. Verleysen (Editor), Proceedings of the 10th European Symposium on Artificial Neural Networks (2002), S. 119-124

[14] Knödler, K.; Poland, J.; Mitterer, A.; Zell, A.: Optimizing data measurements at test beds using multi-step genetic algorithms. In: N. Mastorakis (Editor), Advances In Fuzzy Systems and Evolutionary Computation (2001), S. 277-282

[15] Dixon, L. C. W.; Szego, G. P.: The optimization problem: An introduction. In: Towards Global Optimization II. North Holland, 1978