Expert System for Optimization of Model Reduction Techniques
Luigi Fortuna, Antonio Gallo, and Giuseppe Nunnari

ABSTRACT: This paper describes an expert system used for model reduction. This system is a useful tool in the heuristic optimization of approximate model design. A set of rules for selection of the appropriate approximation algorithms has been derived from experience. The approach represents a new way of formalizing rules in order to obtain "optimal" reduced order models. The expert system (EXPRED) has been implemented using the M.1 shell of FRAMENTEC, which is one of the most powerful knowledge-based software tools for a personal computer.

Introduction
The purpose of model reduction is to derive low-order models for complex dynamic systems in order to achieve goals such as more efficient simulation of dynamic behavior or reduced computation in on-line adaptive control. A number of methods have been developed and reported in the literature for reduction of linear systems [1]. Any model reduction approach should include appropriate methods for the selection of key model parameters and, in some sense, should involve optimization criteria. Various authors [2] have pointed out that there can be no universal model reduction algorithm due to the diversity of plant characteristics and applications. Indeed, no general model reduction approach has been developed, however, at least from a theoretical viewpoint, such a possibility should not be excluded.

A unified modular software library of programs devoted to system reduction (UMLLSR) has been proposed by the authors [3]. In the UMMLSR library, some 50 approximation algorithms have been implemented, but the software is updated continuously in accordance with developments in the literature so that the number of algorithms continually increases.

The obvious question that arises is which algorithm should be chosen to perform a suitable approximation? A correct choice will optimize the design. As a heuristic approach to selection of the most convenient algorithm, a number of different approximation procedures may be applied to the complex process, and the resulting low-order models compared by interactive simulation. A theoretical comparison of reduction algorithms, which are often characterized by different approaches, is a difficult problem. It is the authors' opinion that the choice of algorithm requires not only formal knowledge available in the literature, but also heuristic understanding derived from wide experience in practical model reduction. With this approach, a computer can perform an automatic search for the "optimal" solution.

An expert system approach to model reduction is proposed in this paper. An expert system allows one to collect and codify not only basic formal knowledge but also the mass of heuristic notions that represent a precious part of the designer's knowledge. This expert system (EXPRED, for expert reduction) is a useful tool for heuristic optimization of approximate model designs. Because the information is collected in an expert system, the expertise in model reduction is available to nonexperts, and it helps in efficient decision making and problem solving.

Approximation Methods
Model reduction techniques can be grouped into three categories: (1) polynomial reduction methods, (2) optimization approaches, and (3) state-space transformation-based techniques. These three categories are discussed here.

Polynomial reduction methods are generally applied in the frequency domain and usually are not computationally intensive. They are used to find low-order transfer functions whose coefficients are chosen to satisfy various criteria, so that the output of the reduced order model matches, as much as possible, the output of the high-order system. These methods are based on matching moments and Markov parameters between the original and reduced order models. The Padé approximation (coupled with various procedures to preserve the stability of the approximate model) is also popular and efficacious, as is the method of continued fraction expansion. An equivalence can be shown between continued fraction expansion and the Padé approximation. Approaches following the Routh method, which retains stability properties and preserves the contribution in the energy sense of the impulse response in low-order models, have been studied recently [4]. A wide class of such approximation methods is valid for multiple-input/multiple-output (MIMO) systems. By combining various concepts reported in the literature, additional methods can be developed.

Optimization approaches form the second category of model reduction techniques. In general, they are based on sequential parametric optimization procedures aiming at the minimization of some index, which measures an "error" between the original and the reduced order models. If the reduced order model has fixed eigenvalues, the model usually can be obtained analytically. Otherwise, the optimal reduced order models must be obtained numerically by the iterative solution of linear matrix equations, leading to high computational efforts that can become prohibitive if the original model order is high. The optimal projection approach can also be used in solving such problems. These methods, often applied in the time domain, are dependent on and limited by the choice of the error index.

An interesting approach based on the minimization of the Hankel norm \( \| G(s) - G_i(s) \| \), which is the norm of the difference between the transfer-function matrices of the original and reduced order models [5], is included in this class of approximation procedures. In a similar manner, it is also possible to have error bounds in terms of the \( L_\infty \) infinity norm.

The third class of approximation methods includes all the procedures that involve the transformation of the original state-space model representation. The approach is based on the evaluation of the suitability of different state coordinate selections in the full-order model and on the selection of a reduced order model maintaining, as much as possible, the original model properties (time re-
Consider a linear time-invariant state equation with an n-dimensional state vector $x$:

$$\dot{x} = Ax + Bu \quad (1a)$$

$$y = Cx \quad (1b)$$

By using a suitable state transformation $T$, such that $\hat{x} = T^{-1}x$, an equivalent system is obtained with $\hat{x}_1$ containing $r$ elements and $\hat{x}_2$ containing $n - r$ elements.

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad (2a)$$

$$y = [\hat{C}_1 \hat{C}_2] \quad (2b)$$

The new state vector is decomposed in accordance with some selected design philosophy. The “more important” state variables are collected in $\hat{x}_1$, and only these are taken into account in the reduced order model.

A usual approach is based on singular-perturbation methods, wherein the original model is divided into slow and fast modes; the fast subsystem can be neglected under suitable assumptions.

The theory of the aggregation can also be included in this class. The state vector $x$, of the reduced order model is obtained as $\hat{x}_r = Kx$, where $K$ is the aggregation matrix. $K$ is selected to provide certain properties of the closed-loop system containing the low-order aggregation model. The applicability and computational convenience of this method depend on the system order and its eigenvalue type.

It is the authors’ opinion that the theory of balanced realization [2], [6] has been a significant contribution to the field of model approximation. In particular, Moore introduced a set of $n$ nonnegative similarity invariants, defined as second-order modes of the system; they represent the weight of each state variable concerning controllability and observability. The state transformation is such that, in the new representation, the controllability and observability Grammians are equal and diagonal. In this way, we get a low-order model neglecting the state components that contribute little to these structural properties.

A new set of similarity invariants (characteristic values) for linear systems is introduced in [6]: they give a measure of the degree of participation of the state variable in a particular representation, namely a closed-loop configuration with Kalman filtering and linear-quadratic-Gaussian (LQG) control (closed-loop balanced realization). In this case, the transformation matrix $T$ is obtained so that the solutions of the two Riccati equations involved in the LQG control (the Kalman filtering and the control Riccati algebraic equations) are equal and diagonal. Recently, the authors [7] have devoted much attention toward balancing the class of symmetric systems using the so-called Cross-Riccati matrix equation.

### Heuristic Optimality in EXPRED

The optimum criterion for the selection of a suitable model reduction procedure essentially depends on the simplicity-accuracy pair of the low-order model. Simplicity refers to the computational effort in obtaining the low-order model, whereas accuracy concerns reliability in comparison with the original model. In order to establish a correct equilibrium between the two requirements, a number of attributes of both the original model and the approximate model must be taken into account. The Table presents a simplified list of the attributes that influence selection of the approximation algorithm.

Experience in model reduction plays a fundamental role in the choice of the approximation method in order to find the right balance between the requirements of simplicity and precision. The heuristic approach allows one to assign correct weights to various specifications and to select the approximation methods that are appropriate for the particular application. For example, for reduced models meant for open-loop schemes (used in simulation and identification), priority is given to precision. For approximate models used in closed-loop schemes, the first aim is closed-loop stability. In the case of a reduced order compensator, especially if the closed-loop scheme is used in real-time control, simplicity comes before accuracy.

Apparently, the plant characteristics, such as eigenvalue spectrum, pole-zero map, model order, and model structure, are also fundamental in selecting a correct approach to model reduction.

A number of questions arise: Which plant and model attributes are more convenient to define? What is the most appropriate criterion to order and weigh such attributes so that the best model order reduction algorithms may be found? It is evident that a mathematical formulation of the problem becomes very difficult, and a unique solution does not exist. The general problem can be tackled only from a heuristic point of view by using a designer’s experience and intuition. Of course, the term “optimal” in this context has to be considered in a heuristic sense.

### Implementation of EXPRED in an Expert System Shell

The standard expert system [8] architecture shown in Fig. 1 consists of two parts: a knowledge base (K.B.) of facts and rules relating to a particular problem or application, and an inference engine (I.E.) that performs the reasoning process to solve specific problems. The expert system can refer to other conventional software tools such as data bases, graphics packages, and user-implemented routines. The expert system oc-
An example of a fact is the following:

```
fact-5: suggested_description_domain = frequency cf 80.
```

This fact, labeled as "fact-5," provides the VALUE "frequency" for the EXPRESSION "suggested_description_domain." The INTEGER value cf = "80" indicates that the fact is considered true with high certainty.

When a fact is encountered that provides a value for the expression currently being sought, the value is copied into a special memory area, called cache, along with the certainty factor and the label of the knowledge-base entry. Afterward the inference engine continues to search for other knowledge-base entries that could provide additional information about the EXPRESSION. The search for the EXPRESSION ends when all possible paths have been exhausted, or it may terminate early if the fact is found to be completely true (cf = 100) or completely false (cf = -100). An important feature of the shell, used for the present application, is the capability to combine certainty factors for the same EXPRESSION. For example, if cf₁ = 80 and cf₂ = 50 are concluded, the inference engine combines the two values to obtain a single overall certainty factor, cf = 86, which will replace the old certainty factor in the cache area, by using the following formula:

```
cf = cf₁ + cf₂(100 - cf₁)/100
```

Analogous formulas can be given to combine certainty factors when cf₁ is negative and cf₂ is positive or cf₁ and cf₂ are both negative.

**Metafacts** are knowledge-base entries that provide information or directions on how to determine a value rather than directly assigning a value to an expression. The most common metafact is a question. For example, while seeking a value for the EXPRESSION "model_order," the inference engine can meet the following metafact entry:

```
question (model_order) = "Is the model order low, medium, or high?"
```

In that case, the expert system prompts for the value of "model order" by displaying the text provided in the second member of the metafact entry.

A rule is an entry of the following form:

```
if PREMISE then CONCLUSION
```

Where PREMISE is generally a proposition of the following form:

```
EXPRESSION = VALUE
```

and CONCLUSION is made up of a proposition of the form

```
EXPRESSION = VALUE cf INTEGER
```

Examples of rules are reported in the next section.

**Knowledge Base in EXPRED**

Knowledge organization is a central topic in artificial intelligence, but in this section we will focus on expertise rather than organization. We are convinced that the efficiency of an expert system mainly depends on the power of the knowledge formalized in the knowledge base. In this sense, the degree of "optimization" is related strictly to the expertise available in the field of interest. To indicate our expertise, we summarize our experience in model reduction, which is the basis of EXPRED.

We have used model reduction during the last seven years in both theoretical and application work. During that period, over 500 model approximations have been made and 800 scientific papers have been consulted. Twenty doctoral candidates have been involved in model reduction research and 50 approximation algorithms have been widely tested. Twenty papers on the theoretical and application aspects of model reduction have been published by the authors.

This experience has been condensed into about 200 rules in the knowledge base of M.1.

A set of rules is reported as follows:

```
rule_04: if role = control and reduced_model_application = off_line and plant = s150 and model_order = low and dominant_poles = yes then approximation_method = perfect_aggregation
```

```
rule_08: if role = control and reduced_model_application = off_line and plant = s150 and model_order = medium and poles_type = simple
```
Some knowledge types allow us to establish a set of rules so that the expert system can select the method with a 100 percent certainty factor, such as in rule 04.

In fact, the "optimal" method to obtain reduced order models for off-line control design, under the hypothesis of dominant eigenvalues, should be the perfect aggregation approximation method, since under this hypothesis the closed-loop system (with a reduced order LQG regulator) is guaranteed to be stable. This conjecture becomes certain if it is supported by computational simplicity: the type [single input/single output (SISO)] and low order of the plant assure this.

Although some items in the premise of rule 08 are equal to those of rule 04, the conclusions of the two rules are different. In fact, in both rules, the evidence that the reduced order model is to be used in off-line control design (ROLE = CONTROL, REDUCED_MODEL_APPLICATION = OFF_LINE) leads us to consider approximation methods that guarantee closed-loop stability when a controller has been designed using a reduced order model. However, in rule 04, the evidence MODEL_ORDER = LOW (indicating an attribute of the original model) supports the previous conclusion, suggesting a not too heavy computational effort in evaluating the aggregation matrix. Instead, in rule 08, the evidence MODEL_ORDER = MEDIUM and POLES_TYPE = COMPLEX_CONJUGATES leads us to take into consideration the possible computational effort that can arise in evaluating the reduced order aggregate models. Moreover, the role of the reduced model and the evidence PLANT = SISO (with symmetric structure) induce us to select the balancing via the Cross-Riccati equation as the most appropriate approximation algorithm. In rule 08, although with lower certainty, another approach can also be considered successful: the balanced-gain approximation. This choice is due to the fact that ASYMPTOTICAL STABILITY = YES and POLE_ZEROS_MAP = YES (the latter evidence meaning a system with some slow poles near some zeros).

Two additional rules follow:

rule 070: if role = modeling and asymptotical_model_stability = yes and plant = MIMO and description_domain = time and (model_order = medium or model_order = high) and (pole_zeros_map = no or pole_zeros_map is unknown) then approximation_method = open_loop_balanced_approximation cf 40

rule 071: if role = modeling and asymptotical_model_stability = yes and plant = MIMO and description_domain = frequency and (model_order = medium or model_order = high) then approximation_method = mixed_methods cf 40.

The different conclusions follow, of course, from the items contained in the following premises: DESCRIPTION_DOMAIN = TIME (rule 070) and DESCRIPTION_DOMAIN = FREQUENCY (rule 071). In rule 070, where a MIMO system described in the time domain is considered, the conclusion indicates an approximation method based on open-loop balancing. In rule 071, given the matrix transfer function of the plant, the conclusion suggests using the approximation by a "mixed_method" (for example, the procedure that couples the continued fraction expansion and the Routh method to preserve the stability of the reduced order model).

Moreover, as can be seen in the demonstrated rules, certainty factors are used to qualify the approximation methods concluded by each rule. Such certainty factors may be automatically combined by the inference engine if different, nonexclusive rules conclude the same value for an expression.

Obviously, the problem to assign a certainty factor to each conclusion is a key point of our approach. Several factors can influence the choice, such as the reliability of the approximation method concerning the attributes contained in the rule premise, the accuracy of the final solution, the computational load, etc. The solution adopted herein to solve this problem is founded on both statistical and heuristic considerations.

Let us first define (heuristically, by using a fuzzy approach) the order of the approximate model. The nth-order model must assure that some important characteristics of the system, in open- or closed-loop schemes, will be maintained. For example, if the original model is stable, the reduced order model must be stable. Or if the reduced order model is used in a lower-order compensator design, the closed-loop scheme with the original model and the reduced order compensator must be asymptotically stable. This leads to the following definition: The order r adopted for the reduced model is the one that makes the "error" between the original and the approximate models bounded, from below and above, by two fuzzy quantities. The rationale is as follows: We fixed the upper bound of the error by defining the precision of the reduced order model. However, in some cases, that precision can only be reached by a very small reduction in model order, so that the approximation cannot be considered fruitful. This is the reason for fixing also a lower bound for the error.
We define in Fig. 2 a heuristic set of curves (equilateral hyperbolas in a specified region) for which the product between quantified measures of simplicity and accuracy will be constant. A certainty factor is associated with each region delimited by two neighboring curves.

The certainty factors have been assigned to the rules "experimentally." For each rule, a considerable number of high-order systems have been approximated, in accordance with the list of heuristic considerations indicating which method to test. The simplicity and accuracy "parameters" for each approximate model were estimated. The certainty factors were then assigned according to the region where "most" of the considered tests fell.

**EXPRED in the Operative Mode**

EXPRED guides a user through the choice of an approximation algorithm in accordance with some criteria of optimality. It actually works in a front panel frame on a PC in an interactive way, the user can invoke UMLLSR, which runs on a VAX 780, to execute reduction subprograms advised by EXPRED (the conceptual knowledge of the model reduction theory has been incorporated into the knowledge base).

A list of facts regarding the approximation scheme in the knowledge base can be found at the end of each consultation. The user can also query EXPRED about each entry in the knowledge base so that the user can learn and gain experience from every consultation. An example of a consultation is shown below:

```
M.1 >> go
Is the role of the model control or modeling?
  >> control
Is the reduced model application on-line or off-line?
  >> off-line
Is the plant SISO or MIMO?
  >> MIMO
Is the description domain frequency or time?
  >> time
Is the original system model symmetric or not?
  >> yes
```

---

**Conclusions**

An expert system for complex model order reduction is proposed. The user can use EXPRED alone to gain some experience in model reduction techniques or in conjunction with the UMLLSR package. EXPRED's features permit increasing the knowledge base. New rules can be added easily (even though the estimation of the certainty factors may not be immediate).

EXPRED is used in our department by various groups of researchers, not experts in model reduction, in order to select the approximation algorithm for their particular reduction problems.

The efficiency of our expert system increases continually because it is widely tested and one can further extend its knowledge base. Our future aim is to connect EXPRED directly to UMLLSR and to develop another part of the knowledge base so that the expert system can reach its conclusions by using intermediate results in concert with UMLLSR.

---

**Acknowledgment**

The authors are grateful to the reviewers for many suggestions in improving this paper.

**References**


