Dimension Reduction for Laminar Methane Combustion

Gábor Janiga*, Klaus Truemper‡
*Lab. of Fluid Dynamics and Technical Flows, University of Magdeburg "Otto von Guericke", Magdeburg, Germany
‡Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA

***Abstract*— In the analysis of combustion processes, simulation is a cost-efficient tool that complements experimental testing. Computational evaluation of detailed physico-chemical models typically requires substantial effort. To avoid the computational bottleneck, reduced chemical schemes, for example, manifold-based or tabulated chemistry methods, have been developed that result in substantially reduced computational effort and memory requirements.**

This work proposes an alternative analysis tool based on the Machine Learning concepts of Subgroup Discovery and Lazy Learning. Its goal is a compact representation of chemical processes using few variables. Efficacy is demonstrated using detailed numerical simulation data of a laminar methane/air combustion process. From these data, the reduction method derives a reduced set of 3 variables from which all other variables are estimated with good accuracy.

I. Introduction

Virtually all transportation systems and the majority of electric power plants rely directly or indirectly on the combustion of fossil fuels. Numerical simulation has become increasingly important for improving the efficiency of these combustion processes. Indeed, simulation constitutes a highly cost-efficient complement of experimental testing of design prototypes.

Analysis of quantitative problems like predictions of intermediate radicals, pollutant emissions, or ignition/extinction limits requires complete chemical models [6]. Until recently, most corresponding computations have been limited to one or two spatial dimensions due to the huge computational effort required for three-dimensional simulations; see [9], [11].

Classical reduction approaches (sensitivity, quasi steady-state, partial equilibrium) are based on an analysis of chemical pathways that identifies the most important ones. The other reactions are then modeled using the main selected variables. Further reduction techniques, for example, ILDM-based or tabulated chemistry methods lead to a considerable speed-up of the computations when compared with a complete reaction mechanism, with small loss of accuracy. They drastically reduce required computational memory as well.

This work proposes an additional analysis tool called REDSUB that is based on Machine Learning methods, specifically Subgroup Discovery [12], [16] and the concept of Lazy Learner [14]. It analyses simulation data of chemical processes using a complete reaction scheme and searches for a compact representation that uses fewer variables while keeping errors within specified limits.

To demonstrate efficacy, a laminar methane/air burner is considered here, as found in many practical applications. The numerical simulation considers a two-dimensional domain with detailed physico-chemical models [10]. For each grid point, the simulation produces values for 29 chemical species, 3 thermodynamic properties (pressure, temperature, enthalpy), and 2 velocity components. Given this information, REDSUB selects 3 variables from which the values of the remaining chemical and thermodynamic variables are estimated with good accuracy.

We begin with a general definition of the reduction problem. Let X be a finite collection of vectors in R^n, the n-dimensional real space. In the present setting, X contains the results of a simulation model for some process. For efficient use of X, for example, for optimization purposes, we want to select a small subset J of $\{1, 2, \ldots, n\}$ and construct a black box so that the following holds. For any vector of $x \in X$, the black box accepts as input the x_j, $j \in J$, and outputs values that are close estimates of the x_j, $j \notin J$.

For a simplified treatment of numerical accuracy and tolerances, we translate, scale, and round values of vectors $x \in X$ so that the entries become uniformly comparable in magnitude. Thus, we select translation constants, scaling, and rounding so that all variables finally have integer values ranging from 0 to 10^6, with both bounds achieved on X. From now on we assume that the vectors of X themselves are of this form.

As a final condition that is part of the problem definition, the index set J selected in the reduction process must be a subset of a specified nonempty set $I \subseteq \{1, 2, \ldots, n\}$ and must contain a possibly empty subset $M \subseteq I$. For example, the x_j, $j \in I$, may represent values that are easily measured in the laboratory and thus support a simplified partial verification of the full model via the reduced model. The subset M contains the indices of variables that we require in the reduced model for any number of reasons.

In the example combustion process discussed in Section III, each vector $x \in X$ contains $n = 31$ values for various chemical species, temperature and pressure. The function $F(x)$ is the enthalpy. The associated vector $y \in R^m$ has $m = 2$ and defines the point in the plane to which the values of x and $F(x)$ apply. Problem of that size are easily handled by REDSUB. Indeed, in principle

the method can handle cases where the vectors x of X have up to several hundred entries. The size of m is irrelevant.

Once REDSUB has identified the index set J, the Lazy Learner of REDSUB can use J and the set X to estimate for any vector where just the values x_j, $j \in J$, are given, the values for all x_j, $j \notin J$, and $F(x)$. This feature allows application of the results in settings similar to that producing X.

II. Algorithm REDSUB

In a traditional approach for the problem at hand, we could (1) estimate the function values with ridge regression [7], [8]; (2) derive the index set J via the estimating coefficients of the ridge regression equations of (1) and covariance considerations; and (3) use ridge regression once more to compute estimating equations that estimate values for the variables $j \notin J$, and $F(x)$ from those for the variables x_j, $j \in J$. The approach typically requires that the user specifies nonlinear transformations of the variables for the regression steps (1) and (3). Here, we propose a new approach that is based on two Machine Learning concepts and does not require user-specified nonlinear transformations. Specifically, we use *Subgroup Discovery* [12], [16] to select *interesting* indices j from the set I; from these indices, the set J is defined. Then we use the concept of *Lazy Learner* to predict values for the x_j, $j \notin J$, and $F(x)$.

A. Subgroup Discovery

The Subgroup Discovery method, called SUBARP, constructs polyhedra that represent important partial characterizations of *level sets* $\{x \in X \mid F(x) \geq c\}$ of $F(\cdot)$ for various values c. SUBARP has been constructed from the classification method Lsquare [2], [3], [4], [5] and the extension of [15], using the general construction scheme of [13] for deriving subgroup discovery methods from rule-based classification methods.

B. Lazy Learner

A Lazy Learner is a method that, well, never learns anything. Instead, it directly derives from a given training set an estimate for the case at hand and then discards any insight that could be gained from those computations [14]. In the present situation, given index set J and training set S, the Lazy Learner estimates for a given vector x the entries x_j, $j \notin J$, and $F(x)$ from the x_j, $j \in J$. The employed estimation process is a particular Locally Weighted Learning algorithm; see [1] for an excellent survey.

III. Application

In [10], a simulation is described for a partially-premixed laminar methane/air combustion process in a 2-dimensional space. Thus, $m = 2$. We skip a review of the model and only mention that the simulation has $n = 32$ variables measuring important characteristics such as mixture composition, temperature, pressure and enthalpy. The mixture composition covers 29 species. The

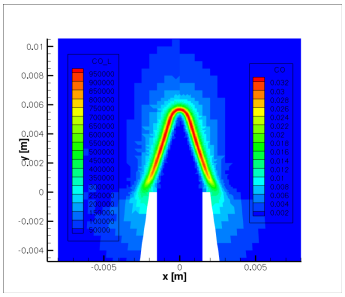

Fig. 1. Demonstration of the model reduction for CO

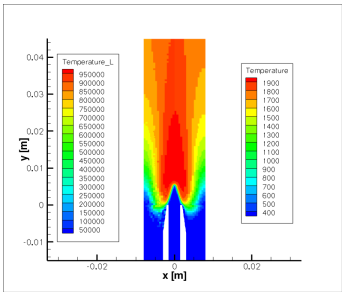

Fig. 2. Demonstration of the model reduction for Temperature

29 species define the set I from which the set J is to be selected. The set M of mandated selections is empty.

For the $7,310$ grid points, the simulation supplies vectors x, each with values for the 31 variables. Thus, the set X consists of $7,310$ vectors. The function $F(x)$ is the enthalpy. We split the data 50/50 into a learning set X from which the reduced model is derived via SUBARP and the Lazy Learner, and a *verification set* V. The split is so done that the y vectors corresponding to the vectors of X constitute a coarser subgrid of the original grid produced by the simulation.

Using an error bound $e = 0.5\%$, REDSUB selects the three variables H_2, H_2O, and N_2. To evaluate the effectiveness of the solution variables, we employ the entire set X in the Lazy Learner and use it as black box to estimate values for the verification set V. The overall error turns out to be 0.26%. Indeed, the estimated values are quite precise. For a demonstration, Figs. 1–2 show sample pictures for two variables, specifically, for CO and the temperature. In these figures, the area of the flame is shown. On the right-hand side of the center vertical axis, the colors encode the values determined by the detailed simulation. On the left-hand side, colors encoding the estimated (and hence scaled) values are displayed. Due to the symmetry, consistency of the color patterns of the two sides implies accurate estimation, and conversely.

Table I shows the correlation between estimated and simulated values. Since H_2, H_2O and N_2 are used to produce the estimates, their correlation values are trivially equal to 1.0. Clearly, estimated and simulated values

TABLE I
CORRELATION OF ACTUAL AND ESTIMATED VALUES USING H_2, H_2O AND N_2

Variable	Correlation	Variable	Correlation
Pressure	0.9923	CH_3	0.9998
Temperature	0.9989	CH_3O	0.9996
H	0.9999	CH_2OH	0.9998
OH	0.9998	CH_4	0.9999
O	0.9998	C_2H	0.9996
HO_2	0.9996	HCCO	0.9996
H_2	1.0000	C_2H_2	0.9998
H_2O	1.0000	CH_2CO	0.9998
O_2	0.9999	C_2H_3	0.9997
CO	0.9999	C_2H_4	0.9997
CO_2	0.9999	C_2H_5	0.9997
CH	0.9997	C_2H_6	0.9997
HCO	0.9998	C	0.9996
CH_2S	0.9998	C_2	0.9996
CH_2	0.9997	N_2	1.0000
CH_2O	0.9998	Enthalpy	0.9929

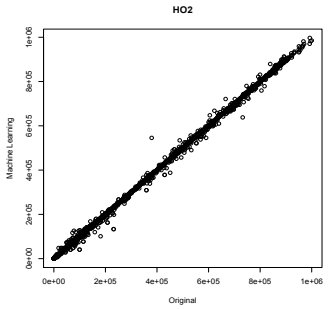

Fig. 3. Correlation between HO_2: simulated vs. estimated values

are highly correlated for all variables. As an example, a scatter-plot is included for the variable HO_2 that shows the relationship between simulated and estimated values. Considering that almost all of the $3,655$ points occur on or very close to the diagonal, estimation errors are clearly low.

The above evaluation can be repeated except that we use a slightly larger error bound of $e = 1\%$. REDSUB selects the two variables H_2 and H_2O for the reduced model. The overall error for the verification set V is now 0.33%. The results (not shown here due to lack of space) demonstrate that model reduction to the two variables H_2 and H_2O still produces useful results, but at a lesser level of accuracy.

IV. CONCLUSION

This work introduces a method called REDSUB for model reduction of reacting flows. REDSUB is based on the Machine Learning techniques of Subgroup Discovery and Lazy Learner. Use of the method does not require any user-specified transformations of simulation data or other manual effort. The effectiveness of REDSUB is demonstrated with simulation data of a methane/air combustion process, where 32 variables representing 29 chemical species of the combustion mixture and 3 thermodynamic properties are reduced to 3 variables. For the remaining variables, the values are estimated with good accuracy over the entire range of possible values. An even smaller model using 2 variables still has reasonable accuracy.

ACKNOWLEDGMENT

Interesting discussions with D. Thévenin are gratefully acknowledged.

REFERENCES

[1] C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. *Artificial Intelligence Review*, 11:11–73, 1997.
[2] S. Bartnikowski, M. Granberry, J. Mugan, and K. Truemper. Transformation of rational and set data to logic data. In *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques*. Springer, 2006.
[3] G. Felici, F. Sun, and K. Truemper. Learning logic formulas and related error distributions. In *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques*. Springer, 2006.
[4] G. Felici and K. Truemper. A MINSAT approach for learning in logic domain. *INFORMS Journal of Computing*, 14:20–36, 2002.
[5] G. Felici and K. Truemper. The lsquare system for mining logic data. In *Encyclopedia of Data Warehousing and Mining*, pages 693–697. Idea Group Publishing, 2005.
[6] R. Hilbert, F. Tap, H. El-Rabii, and D. Thévenin. Impact of detailed chemistry and transport models on turbulent combustion simulations. *Progress in Energy and Combustion Science*, 30:61–117, 2004.
[7] A. E. Hoerl. Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58:54–59, 1962.
[8] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthgonal problems. *Technometrics*, 12:55–67, 1970.
[9] G. Janiga, O. Gicquel, and D. Thévenin. High-resolution simulation of three-dimensional laminar burners using tabulated chemistry on parallel computers. In *2nd ECCOMAS Thematic Conference on Computational Combustion*, pages 1–15, Delft, The Netherlands, 2007.
[10] G. Janiga, A. Gordner, H. Shalaby, and D. Thévenin. Simulation of laminar burners using detailed chemistry on parallel computers. In P. Wesseling, E. Onate, and J. Périaux, editors, *European Conference on Computational Fluid Dynamics (ECCOMAS CFD 2006)*, pages 210/1–210/14, Egmond aan Zee, The Netherlands, 2006.
[11] G. Janiga and D. Thévenin. Three-dimensional detailed simulation of laminar burners on parallel computers. In *Proceedings of the European Combustion Meeting ECM2009*, pages 810175/1–810175/7, Vienna, Austria, 2009.
[12] Willi Klösgen. EXPLORA: A multipattern and multistrategy discovery assistant. In *Advances in Knowledge Discovery and Data Mining*. AAAI Press, 1996.
[13] Nada Lavrač, Bojan Cestnik, Dragan Gamberger, and Peter Flach. Decision support through subgroup discovery: Three case studies and the lessons learned. *Machine Learning*, 57:115–143, 2004.
[14] T. Mitchell. *Machine Learning*. McGraw Hill, 1997.
[15] K. Truemper. Improved comprehensibility and reliability of explanations via restricted halfspace discretization. In *Proceedings of International Conference on Machine Learning and Data Mining (MLDM 2009)*, 2009.
[16] Stefan Wrobel. An algorithm for multi-relational discovery of subgroups. In *Proceedings of First European Conference on Principles of Data Mining and Knowledge Discovery*, 1997.