Model Reduction for Chemical Kinetics: An Optimization Approach

Linda Petzold
Dept. of Mechanical and Environmental Engineering, University of California, Santa Barbara, CA 93106

Wenjie Zhu
Dept. of Computer Science, University of Minnesota, Minneapolis, MN 55455

The kinetics of a detailed chemically reacting system can potentially be very complex. Although the chemist may be interested in only a few species, the reaction model almost always involves a much larger number of species. Some of those species are radicals, which are very reactive species and can be important intermediaries in the reaction scheme. A large number of elementary reactions can occur among the species; some of these reactions are fast and some are slow. The aim of simplified kinetics modeling is to derive the simplest reaction system which retains the essential features of the full system. An optimization-based method for reduction of the number of species and reactions in chemical kinetics models is described. Numerical results for several reaction mechanisms illustrate the potential of this approach.

Introduction

The kinetics of a detailed chemically reacting system can potentially be very complex. Although the chemist may be interested in only a few species, the reaction model almost always involves a much larger number of species, some of which are radicals. These are very reactive species (usually of low concentration), which can be important intermediaries in the reaction scheme. A large number of elementary reactions can occur among the species; some of these reactions are fast and some are slow. The aim of simplified kinetics modeling is to derive the simplest reaction system which retains the essential features of the full system.

The conventional technique (Williams, 1985; Peters, 1985, 1991) is to systematically apply the so-called steady-state approximation to the appropriate radicals, the partial-equilibrium approximation to the fast reversible reactions, and to ignore the very slow (and therefore unimportant) reactions. The investigator is responsible for identifying the appropriate radicals, the fast elementary reactions and the very slow ones, by making intelligent order of magnitude estimates using information gathered from a detailed examination of available data. A skilled and knowledgeable chemical kineticist is usually needed, and the results obtained are expected to be valid only in some limited domain of initial and operating conditions for a limited interval of time. Usually, the obtained reactions are not elementary reactions and actually represent groups of reactions lumped together. This makes the reduced model more accurate, but the physical meaning of the elementary reactions may be lost.

Sensitivity analysis has often been used for the purpose of obtaining information to develop a reduced-order mechanism. In many cases, this approach has been successful (Seigneur et al., 1982; Gautier and Carr, Jr., 1985; Carr, Jr. et al., 1986). However, for some systems, sensitivity analysis may be misleading. Defining as parameters the rate constants of the chemical reactions, sensitivity analysis determines the change in the species concentration for small perturbations of the rate constants. If a reaction is slow and unimportant, it can be identified in this way. However, sensitivity analysis may also single out fast reactions which are important and, therefore, should not be deleted. To see how this can happen, consider for example, determining the sensitivity with respect to perturbation of the parameters λ_1, λ_2, of the following linear stiff differential equation

$$y' = (\lambda_1 + \lambda_2)y - (\lambda_1 + \lambda_2)\left[f(t) + \frac{g(t)}{\lambda_2}\right] + \frac{g'(t)}{\lambda_2} + f'(t),$$

$$y(0) = \frac{g(0)}{\lambda_2} + f(0) \quad (1)$$

Correspondence concerning this article should be addressed to L. Petzold.

where $|\lambda_1|, |\lambda_2| \gg 0$. This problem has solution $y = f(t) + g(t)/\lambda_2$. Let $s_i = dy/d\lambda_i$, $i = 1, 2$. Then, the sensitivities satisfy

$$s_1' = (\lambda_1 + \lambda_2) s_1 + \left\{ y - \left[f(t) + \frac{g(t)}{\lambda_2} \right] \right\}, \quad s_1(0) = 0$$

$$s_2' = (\lambda_1 + \lambda_2) s_1 + \left\{ y - \left[f(t) + \frac{g(t)}{\lambda_2} \right] \right\} + \frac{(\lambda_1 + \lambda_2) g(t)}{\lambda_2^2} - \frac{g'(t)}{\lambda_2^2}, \quad s_2(0) = -\frac{g(0)}{\lambda_2^2}.$$

Let $\lambda_1 = -1{,}000$, $\lambda_2 = +10$. The solution for s_1 is $s_1 = 0$. If the term (reaction) corresponding to λ_1 is deleted from the original system, the reduced system will be unstable. We note that sensitivity is not giving the wrong answer in this example. Instead, it is giving the correct answer but to the wrong question. Sensitivity tells us the change in the solution corresponding to a small change in the parameter. By deleting the term corresponding to λ_1, we are making a large change in that parameter. Also, sensitivity tells us only the change in the solution to the original system corresponding to a small change in the parameter. It may not yield enough information about the sensitivity of neighboring solutions with respect to the parameter to guarantee stability of the reduced system. See also Allara and Edelson (1975), Edelson (1981), and Brown et al. (1997).

The problem of obtaining a reduced model for the chemical kinetics problem has been considered by a number of authors. In an interesting series of articles (Lam and Goussis, 1993; Lam, 1993), Lam et al. have proposed the computational singular perturbation (CSP) method for automatically determining appropriate simplified kinetics models. The CSP method identifies the fast and slow modes as the system advances in time. The reduced-order model at any given time is solvable by explicit time-stepping methods, because the fast modes have been identified and approximated by algebraic constraints, leaving only the slower modes in the differential system.

Another approach has recently been taken by Maas and Pope (1992): the slow manifold approach. The aim is to find a reduced system with a much smaller number of variables and equations than the original system. If it is possible to describe the chemical system with a very small number of variables, then one can use a table look-up scheme to model the chemistry in complicated processes like turbulent reacting flows. The method employs an analysis of the eigenvalues and eigenvectors of a local linearized system to identify the fast and slow modes and obtain a reduced system. In contrast to other reduced mechanisms, this method does not provide simple or closed-form expressions for the chemical kinetics. Thus, it is primarily useful in the context of table-lookup as described above.

Yet, another method is described by Vajda et al. (1985). This method also employs an eigenvalue-eigenvector analysis of the local linear problem.

A great deal of work has been done on developing reduced-order models for linear differential systems (see Grimme (1994) and Grimme et al. (1994) and the references therein). These methods also proceed by identifying the large modes and constructing a reduced-order model based on the subspace defined by these large modes. Stability of the reduced-order model is an important consideration.

Chemical mechanisms are sufficiently nonlinear that a global approach may be warranted. When nonlinear effects are important, with current methods based on a linear methodology, it is up to the chemist to be alert to the possibility of secondary nonlinear effects in the mechanism leading to parameter interactions, and to uncover their causes on the basis of available chemical knowledge (Vajda et al., 1985). Methods based on local linear analysis can be helpful in pointing out a potential problem, but it still requires a chemist to solve it. This article describes our work to develop a global, nonlinear approach which has the potential to automatically find a reduced-order mechanism which inherits the stability and captures the nonlinear behavior of the original system in the regime of interest. The reduced system of equations will have some physical relevance and interpretation and would usually need to change only a few times during the course of a given problem.

Principles of the New Algorithm

Our objective has been to develop a numerical method for model reduction of chemical kinetics which could produce a reduced model with the following properties:

(1) The reduced-order model accurately models those properties of the original system which are of interest, over the given problem domain.

(2) The reduced-order model inherits the stability properties of the original system.

(3) The reduced-order model is simpler and much cheaper to evaluate than the original system.

(4) Nonlinear behavior should be preserved by the reduced-order model.

(5) The reduced-order model makes physical sense and potentially offers some physical insight.

We expect to be able to use this reduced-order model over a range of operating conditions which are nearby to the conditions for which it is developed. It is not reasonable for this class of problems to seek a reduced-order model which is valid over all possible operating conditions. To see why this is true, consider, for example, chemical kinetics at two very different temperatures (K). Then, it may be that a completely different set of reactions is important in the two systems.

Given n chemical species with mass fractions y_i ($i = 1, \cdots, n$) and N reactions where $F_r(y)$ ($r = 1, \cdots, N$) denotes the reaction rate of the rth physical elementary reaction, the original ODE system for a constant-volume batch reactor is given by

$$y' = g(y) = \sum_{r=1}^{N} S_r F_r(y), \tag{2}$$

where $S_r \in \mathfrak{R}^n$ are the stoichiometric coefficients for reaction F_r. (For example, for the reaction $O_2 + H \rightleftharpoons OH + O$, if the forward and reverse reactions are treated explicitly, then there are two reactions and two reaction terms $F_1 = k_{1,f}[O_2][H]$ and $F_2 = -k_{1,b}[OH][O]$, whereas if the forward

and reverse reactions are coupled, there is one reaction and one reaction term $F_1 = k_{1,f}[O_2][H] - k_{1,b}[OH][O]$. In our implementation, whether the forward and reverse reactions are treated explicitly or coupled is determined by the way in which the reactions have been input to Chemkin (Kee et al., 1989).) Note here S_r are vectors and F_r are scalars, one for each reaction. Typically, $N \gg n$ because it is hard to know in advance which reactions will be important. The objective is to choose some (very small) subset of the reactions to use in the mechanism so that the behavior of the reduced system is as close as possible to that of the original system, given a range of operating conditions.

If S is the matrix whose columns are the stoichiometric vectors $S \in \mathfrak{R}^{n \times N}$, and $F \in \mathfrak{R}$ is the vector of nonlinear reaction terms, then the original system can be written more compactly as

$$y' = SF(y). \tag{3}$$

The reduced system is given by

$$z' = SDF(z), \tag{4}$$

where $D \in \mathfrak{R}^{N \times N}$ is a diagonal matrix whose diagonal elements d_i are either 1 or 0 (depending on whether or not reaction i is selected for the reduced mechanism). Then, the problem of finding the reduced mechanism can be written as a constrained optimization

$$\begin{aligned} \min \quad & \| y - z \| \\ & \text{subject to} \\ & y' = SF(y), \quad y(0) = y_0 \\ & z' = SDF(z), \quad z(0) = y_0, \quad 0 \le t \le b \qquad (5) \\ & \sum_{i=1}^{N} d_i = k, \quad d_i = 0 \text{ or } 1, \qquad (6) \end{aligned}$$

where the minimum is over $d_1, \ldots, d_N$, each d_i can take the value 0 or 1, and $k \ll N$. Currently, in our approach, the number k of reactions for the reduced mechanism is chosen by the user. If it is important to be sure that the reduced model is valid over a range of operating conditions, one should add additional constraints and the corresponding variables which are the differential equations for the original model and for the reduced-order model, but with different initial conditions. By posing this problem directly as an optimization, there is also a natural measure for the error due to the reduced-order model. We note that the norm should be weighted according to user tolerances for relative and absolute errors. In some situations it may be desirable to minimize over some other measure of the error; our methodology can easily accommodate such a modification.

Solving the Nonlinear Integer Programming Problem

It would appear that solving the discrete optimization problem (Eq. 6) directly could be very costly. There are almost no mature methods for nonlinear integer programming problems without convex or polynomial properties, which is the case for Eq. 6. Also, note that it may not be necessary to find the absolute minimum; any reduced mechanism for which $\| y - z \|$ is small enough would meet our needs. Here, we present one method that we have found to be effective.

Consider solving the continuous optimization problem

$$\begin{aligned} \min \quad & \| y - z \| \\ & \text{subject to} \\ & y' = SF(y), \quad y(0) = y_0 \\ & z' = SDF(z), \quad z(0) = y_0, \quad 0 \le t \le b, \\ & k_1 \le \sum_{i=1}^{N} d_i \le k_2, \; 0 \le d_i \le 1 \; (i = 1, \cdots, N) \\ & g(d_1, \cdots, d_N) \le r \qquad (7) \end{aligned}$$

where k_1 and k_2 are lower and upper bounds for the number of reactions in the reduced model, $g(d_1, \cdots, d_N)$ is a nonlinear function which, when equal to 0, forces the d_i to take integer values, and r is a small positive number which relaxes the nonlinear constraint. Different kinds of nonlinear constraints can be used. The one we are currently using is from Gisvold and Moe (1972)

$$g = \sum_{i=1}^{N} \left(d_i - d_i^2 \right)^{\beta} = 0, \tag{8}$$

where $\beta \ge 2$ is a parameter which controls the shape of the normalized, symmetrical Beta-function integrand. The gradients of the constraint are zero at integer points $d_i = 0$ or $d_i = 1$ for $i = 1, \ldots, n$. In our numerical experiments, β is taken to be 2. We solve the continuous optimization problem (Eq. 7) via a sequential quadratic programming (SQP) method (Gill et al., 1985). To obtain the reduced model, solutions $d_1, \ldots, d_N$ to the continuous optimization problem (Eq. 7) are rounded to 0 or 1.

For problems where the initial number of reactions is very large, a grouping approach can be used to drastically reduce the computational work. The reactions in the original mechanism are divided randomly into a number of groups, each containing several reactions. Thus, each optimization parameter d_i is associated with a reaction group and the number of optimization parameters is reduced from the number of reactions in the mechanism to the number of groups. The optimization process is then applied to these reaction groups to obtain a reduced number of groups. The grouping process may be repeated several times to reduce the original mechanism to a desired size. For further details, see Zhu (1998).

The continuous optimization solver we are using can find only local optimal solutions for the nonlinear programming problems. The chances of finding a global optimum are usually increased by choosing a starting point that is "sufficiently close" (Gill et al., 1994). There are several ways to choose the initial guess. One way is to utilize chemical insight to guess a group of important reactions and set their corresponding d_i to 1 and the rest to 0. Another possibility is to use the greedy method to find an initial guess.

In the greedy method, the reactions are deleted from the original model one by one; each time, we drop the reaction which causes the smallest error between the original model and the current reduced model, under the given norm. Then, all the reactions are ranked according to the order that they are deleted from the model. The initial guess is obtained by retaining the set of the most important reactions under the greedy criterion. The number of these reactions can be equal to the number of reactions we want to keep. The greedy method can in fact result in very good initial guesses, as we have observed in our preliminary experiments. The problem with this approach is that when we are dealing with large reaction models, the cost of finding a greedy guess can be very high; it is of order $O(N^2)$.

Another possibility is to use a scanning method. In the scanning method, the error caused by deleting a single reaction from the original model is computed by deleting one reaction at a time. Then, the reactions are ordered according to their errors with the one that causes the smallest error considered to be the most unimportant reaction. If only a small number of reactions are to be deleted from the model, this method can be used to generate the initial guess. However, when a large number of reactions are to be deleted from the model, the initial guess generated by this method is not reliable. On the other hand, this method can be used to identify the most important reactions. Using this method, we preselect those reactions whose absence will cause extremely large errors. Preselecting the most important reactions can avoid some difficulties, as we will explain in detail in the section on post recovery process. Optimization is then done over the remaining reactions. Similarly, those reactions that cause extremely small errors can be predeleted, which works well according to our numerical tests. This reduces the number of variables in the optimization problem. A post-recovery process described later can be used to recover the reactions that have been predeleted, but can actually reduce the error significantly.

The simplest way to generate an initial guess is to set all the parameters to 1. This can be done when the model consists of a large number of reactions and we need to reduce it to a relatively small size. In this case, the two methods mentioned above are not appropriate and, if we do not have any chemical insight into the model, we have no other choice. Starting from all parameters equal to 1 is, in fact, a good choice in some situations as we will see later.

Implementation and Numerical Results

In our implementation, the continuous optimization is done via DASOPT, a code for parameter estimation and optimal control of differential-algebraic systems (Petzold et al., 1997). The software consists of two main parts: the optimizer and the differential-algebraic equation (DAE) solver. The optimizer chooses the parameters, and the DAE solver computes the objective value and its derivatives with respect to the parameters. The optimizer uses this information to generate another set of parameters until the optimal solution is reached. In DASOPT, the optimizer is the SQP-based solver SNOPT (Gill et al., 1994). This software can solve optimization problems with very large, sparse Hessian matrices. The DAEs are solved by DASPKSO (Brenan et al., 1995; Maly and Petzold, 1996), which also performs sensitivity analysis to obtain the derivatives of the objective with respect to the parameters. The residual functions for the sensitivity equations, and the Jacobian matrix required by DASPKSO are supplied by ADIFOR (Bischof et al., 1992), an automatic differentiation tool.

It is interesting to note that the calculation of sensitivities is an integral part of this approach. However, there is an important difference between the proposed method and sensitivity analysis alone. During the optimization, the sensitivities for each iteration are computed by perturbing the rate constants around some value, determined by the optimization method, which could be closer to zero than to the nominal value of the rate constant. Thus, the information is available to decide more reliably whether a reaction could be deleted from the mechanism.

The vector field of the DAE is formed via Chemkin (Kee et al., 1989). An additional equation which finds the norm of the objective function is added to the system formed by Chemkin

$$u' = \sum_{k=1}^{K} (y_k - z_k)^2. \tag{9}$$

The reduced model depends not only on the norm but also on the initial concentrations of the species, and the initial temperature and pressure. Since we use the mass fractions of the chemical species, y_i is between 0 and 1. As the number of species in a chemical mechanism increases, the mass fractions of most species are very small while the temperature is usually very high. In our test mechanisms, temperatures are of the scale 10^3. To balance the effects of the species and temperature in the objective function, we scale up the mass fractions of the species in the objective function. In the numerical results of this article, the scale-up for the species is taken to be 10^3, which makes the total effect of the species comparable to that of the temperature.

The error of the reduced model is compared to the norm of the error which results when all the reactions are deleted. The time interval is important in our problem. Since chemical mechanisms are often highly stiff, they can reach a steady state very fast. Taking a long time interval will give slow reactions more importance. A short time interval will catch the fast reactions, that is, catch the stiffness. Here, we have taken short time intervals. In this situation, the error will usually be small because it is an integral over the time interval taken. We also scale up the norm according to the integration interval. This results in a better scaling for the objective function and its gradients; thus, the optimizer can perform better. In this article, the scale up for the norm is taken as the reciprocal of the time interval.

Numerical results for our method applied to three chemical mechanisms are given below. In our computations, the parameters are associated with the reactions in the input file for Chemkin. If the original file contains reversible reactions, then each parameter is associated with a reversible reaction. If forward and backward reactions are input separately, then each forward and backward reaction is associated with a parameter. It is not difficult to configure our algorithm to consider separately the forward and backward reactions when the input Chemkin file contains only reversible reactions.

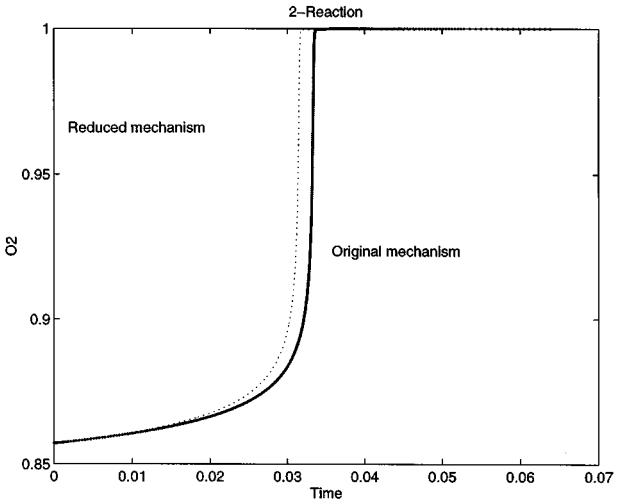

Figure 1. 2-Reaction reduced mechanism for Example 1: O_2.

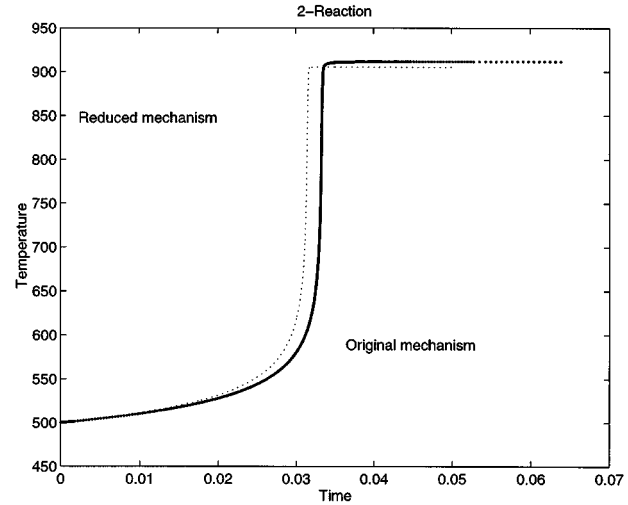

Figure 3. 2-Reaction reduced mechanism for Example 1: temperature.

A detailed description of the corresponding chemical mechanisms can be found on the Web at http://www.engineering.uscb.edu/~cse/, as well as the reduced mechanisms which were obtained.

Example 1: Six-reaction ozone mechanism

Using the methods proposed in this article, we reduced this very small mechanism to three reactions on the time interval [0, 1.0]. The initial mole fractions for species O, O2 and O3 are 0.0, 0.9, and 0.1, respectively, and the initial temperature is 500 K. The pressure is kept at one standard atmosphere. By observing the iteration process of the optimizer, we find that reactions 1 and 3 are always picked by the optimizer, but it encounters difficulty when deciding which of the other ones it should pick: 4 or 6. This suggests that reactions 4 and 6 are similar in their effect on the original system. Picking either one of them results in similar total errors. We reduced further the system to 2 reactions. This time, the optimizer gave the answer as 1 and 3, which we can almost guess from the previous process. The temperature and mass fraction trajectories for major species are shown in Figures 1–3. These results are very similar to those of three reactions, which are not presented here. We can see from these figures that there is a time delay in the ignition for the reduced model. If we preselect reactions 1 and 3 and let the optimizer pick the third important reaction, it will easily find 2. We conjecture that this kind of behavior may be due to the nonlinearity of the system.

Reactions 1–3 yield a very good reduced model: there is almost no error, as shown in Figures 4–6. Reaction 2 seems to be controlling the ignition time. This suggests that a post-recovery process may be necessary to get the secondary important reactions like 2 after the most important reactions (1

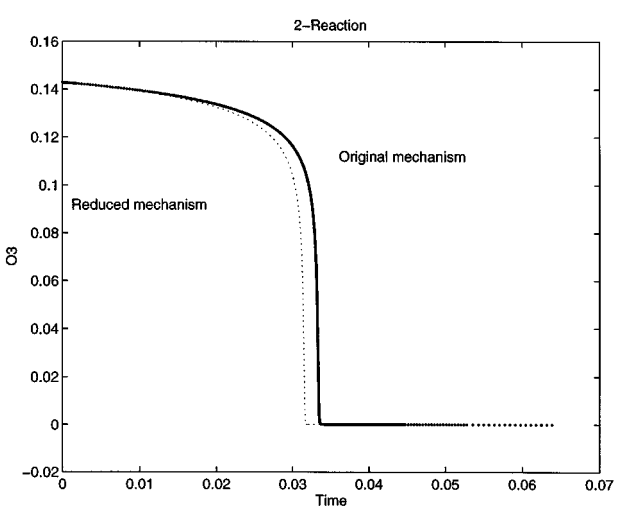

Figure 2. 2-Reaction reduced mechanism for Example 1: O_3.

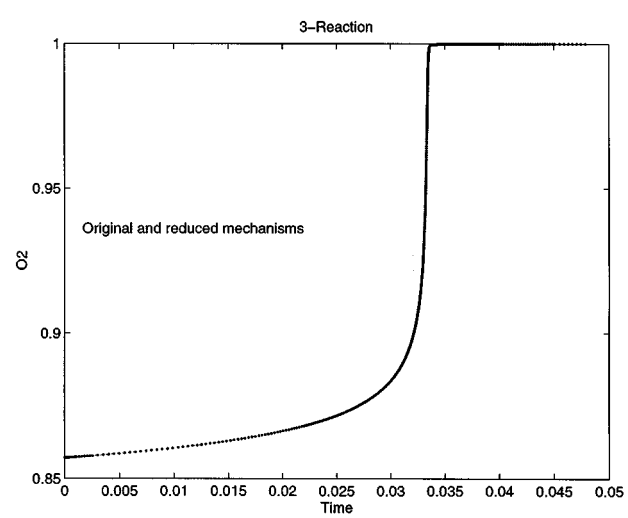

Figure 4. 3-Reaction reduced mechanism for Example 1: O_2.

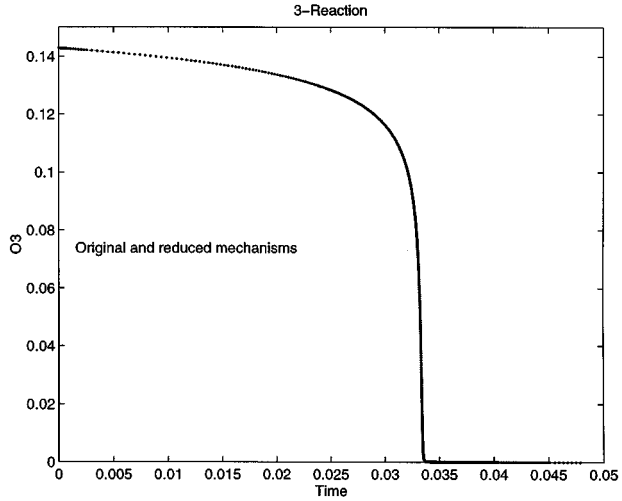

Figure 5. 3-Reaction reduced mechanism for Example 1: O_3.

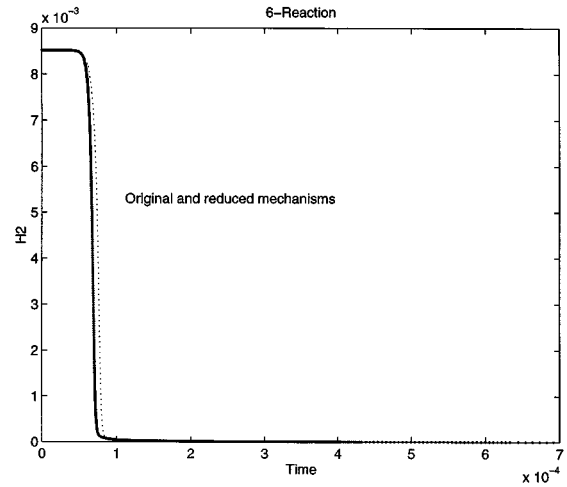

Figure 7. 6-Reaction reduced mechanism for Example 2: H_2.

and 3) are found. This can be accomplished by several approaches which will be described in the section on post recovery process.

Since we are using the scaled L_2 norm of the temperature and species mass fractions, the optimizer captures only the perturbations of species with significant mass fraction. To see the perturbations of species with very small mass fraction, we need to weight their corresponding terms in the L_2 norm, which can be done if one is interested in the change of some specific species. The scale-up can be determined based on the concentration of the species of interest. We just need to make the effect of these species in the objective comparable to that of temperature. For example, if the concentration of a species is around 0.01 during the time interval under consideration, a scale-up of 10^5 is appropriate. Depending on the problem, different scales can be used for different species. If a species of interest has only a very small concentration, a large scale may be used to put more importance on its concentration in the error norm.

Example 2: 20-reaction mechanism

This mechanism is similar to the example mechanism in Kee et al. (1989), except that the element N and all reactions in which it is involved are not included here. It describes combustion for a hydrogen-oxygen mixture. In this example, the initial mole fractions for species H_2 and O_2 are 0.12 and 0.88, respectively, and the initial temperature is 1,000 K. The initial mole fractions for other species are set to 0. The pressure is kept at one standard atmosphere. One reduced model is given by the first six elementary reactions of the original mechanism. The solution of this reduced model matches the original model well on the time interval $[0.0, 10^{-2}]$ on which the optimization is done. Figures 7–10 show temperature and some of the species which change significantly in the time interval considered. As the figures manifest, after a short ig-

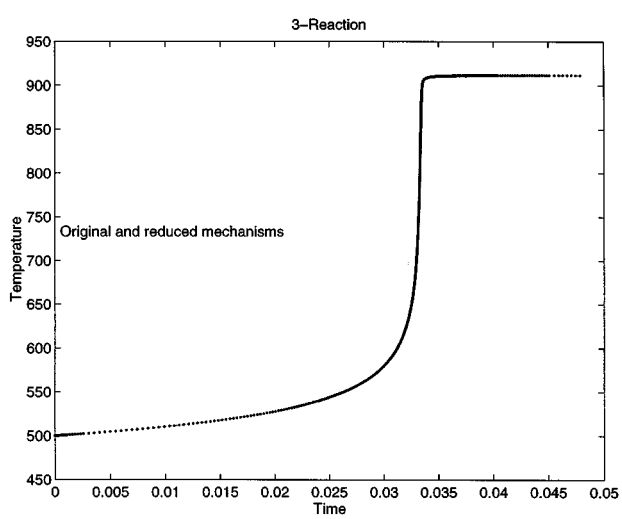

Figure 6. 3-Reaction reduced mechanism for Example 1: temperature.

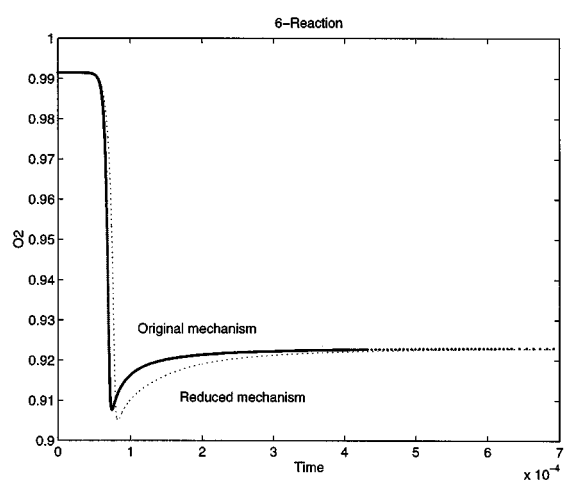

Figure 8. 6-Reaction reduced mechanism for Example 2: O_2.

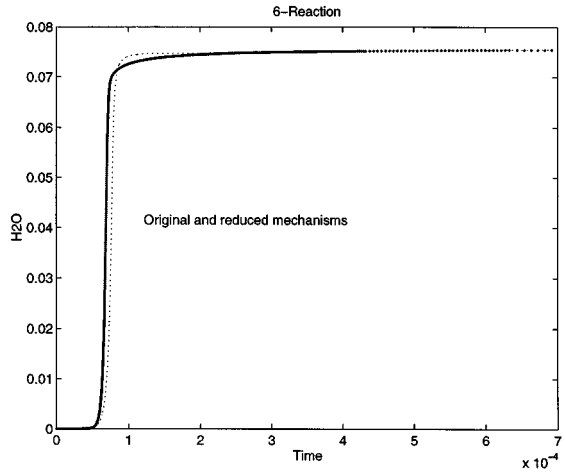

Figure 9. 6-Reaction reduced mechanism for Example 2: H_2O.

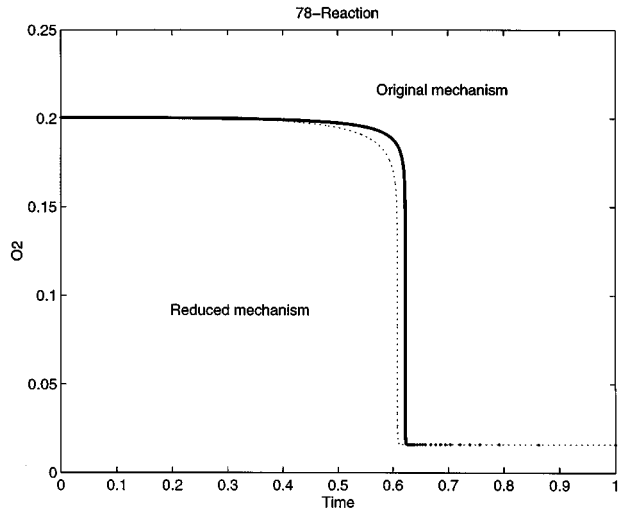

Figure 11. 78-Reaction reduced mechanism for Example 3: O_2.

nition phase, the system reaches steady state and the mass fractions of the species do not change much.

Example 3: GRI mechanism

GRI-Mech 1.2 (Frenklach et al., 1995) is one of the detailed chemical reaction mechanisms capable of the best representation of methane combustion phenomena.

The original GRI mechanism 1.2 contains 177 reversible reactions. We reduced this mechanism to 78, 42, and 32 reactions on the time interval [0,1]. In this example, the initial mole fractions for species O_2, CH_4, N_2 and AR are 0.18, 0.09, 0.64, and 0.09, respectively, and the initial temperature is 1,000 K. The initial mole fractions for other species are set to 0. The pressure is kept at one standard atmosphere.

The 78-reaction reduced mechanism was obtained by dividing the 177 reactions in the original mechanism into 59 groups and then selecting 26 from them using the optimization method only. It models the steady state of the system very well, as shown in Figures 11–15.

A reduced mechanism of 42 reactions was obtained by preselecting the 17 most important reactions using the scanning method described earlier and optimizing over the remaining reactions. The results are satisfactory, as shown in Figures 16–20, although the steady state of the original mechanism and that of the reduced mechanism are slightly different.

The smallest reduced mechanism we obtained so far without using species reduction or the post recovery process (to be described later) is a 32-reaction mechanism. This reduced model is obtained using preselecting and predeleting methods together with the optimization method. This mechanism models the ignition time and the steady state of the original mechanism well, but the shape of the trajectories is not captured perfectly (Figures 21–25). Numerical experiments show

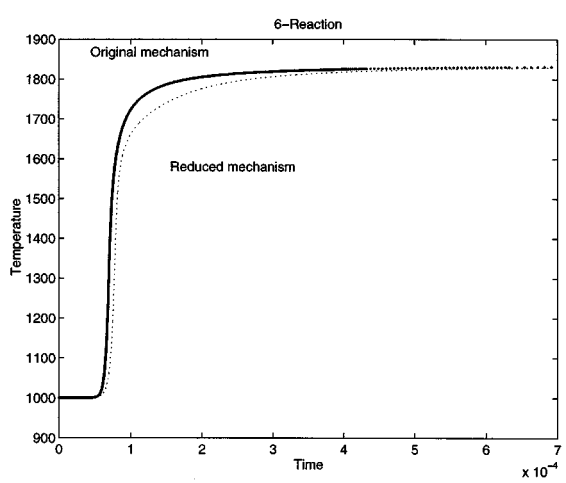

Figure 10. 6-Reaction reduced mechanism for Example 2: temperature.

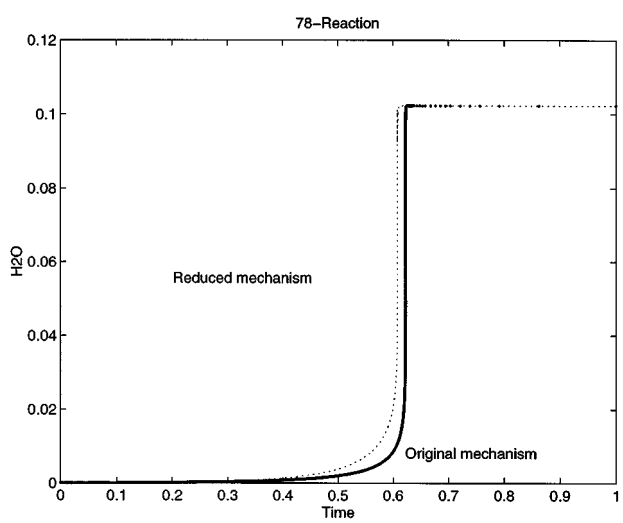

Figure 12. 78-Reaction reduced mechanism for Example 3: H_2O.

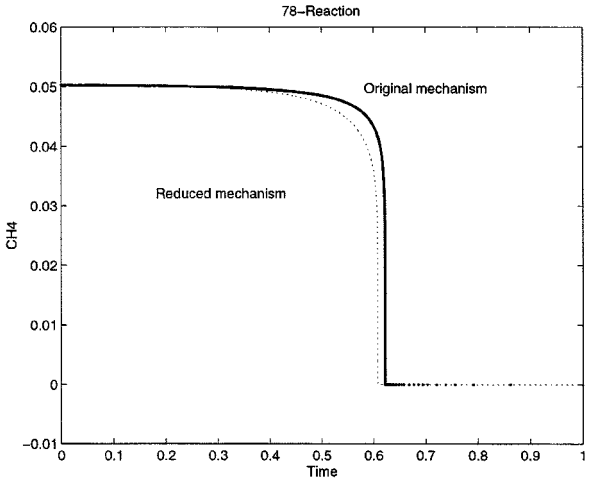

Figure 13. 78-Reaction reduced mechanism for Example 3: CH_4.

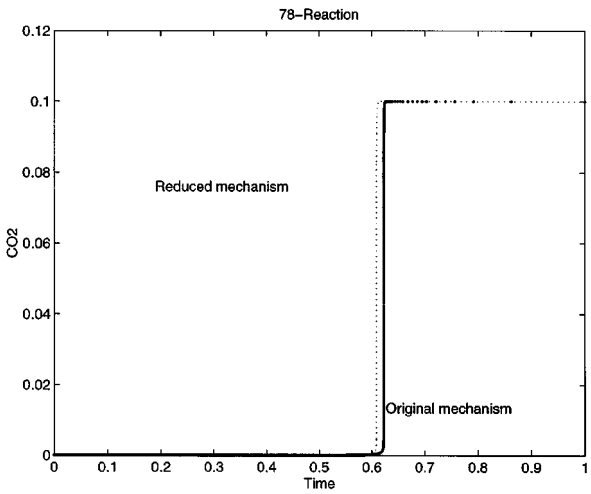

Figure 14. 78-Reaction reduced mechanism for Example 3: CO_2.

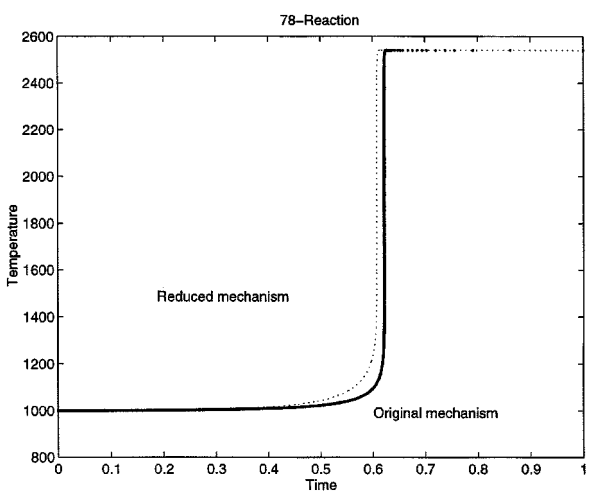

Figure 15. 78-Reaction reduced mechanism for Example 3: temperature.

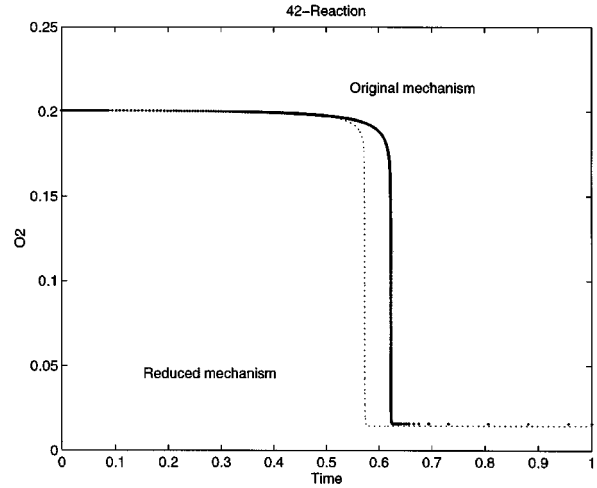

Figure 16. 42-Reaction reduced mechanism for Example 3: O_2.

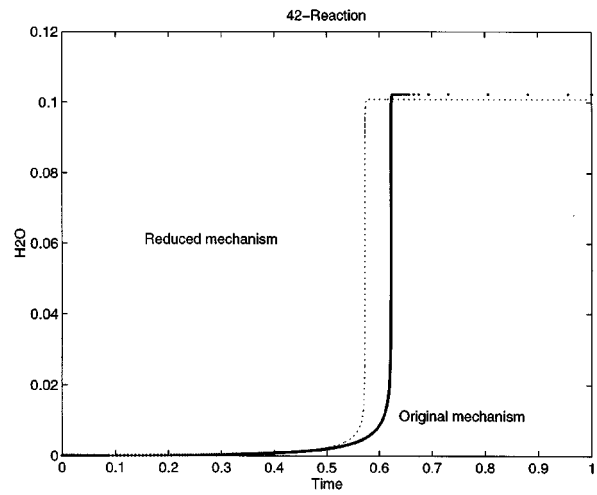

Figure 17. 42-Reaction reduced mechanism for Example 3: H_2O.

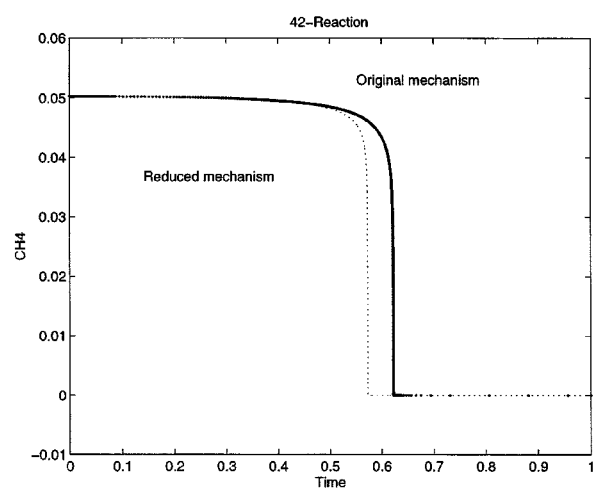

Figure 18. 42-Reaction reduced mechanism for Example 3: CH_4.

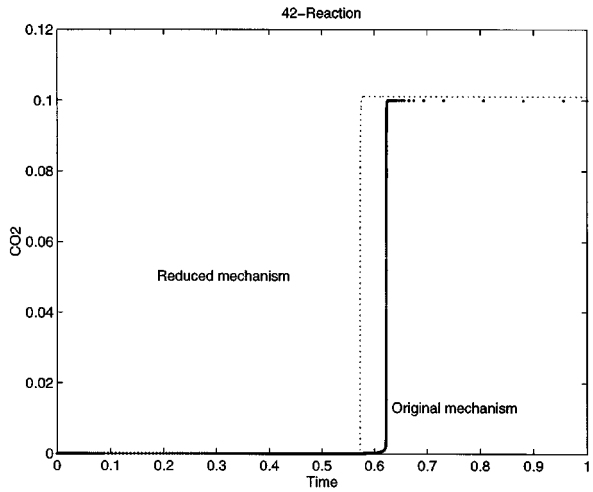

Figure 19. 42-Reaction reduced mechanism for Example 3: CO_2.

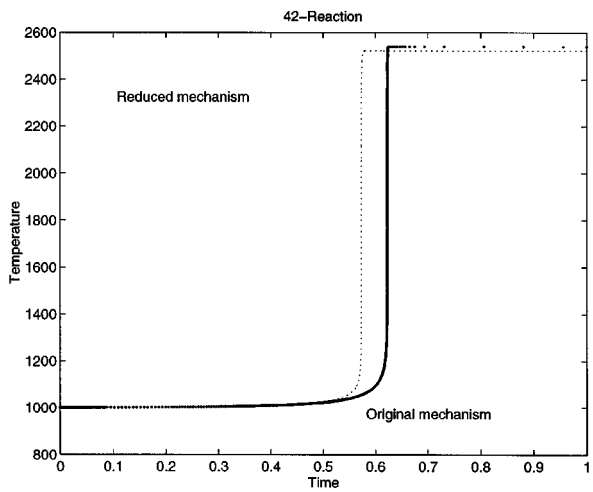

Figure 20. 42-Reaction reduced mechanism for Example 3: temperature.

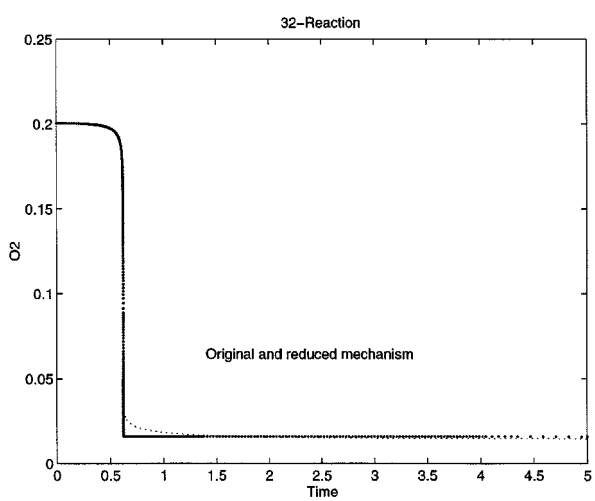

Figure 21. 32-Reaction reduced mechanism for Example 3: O_2.

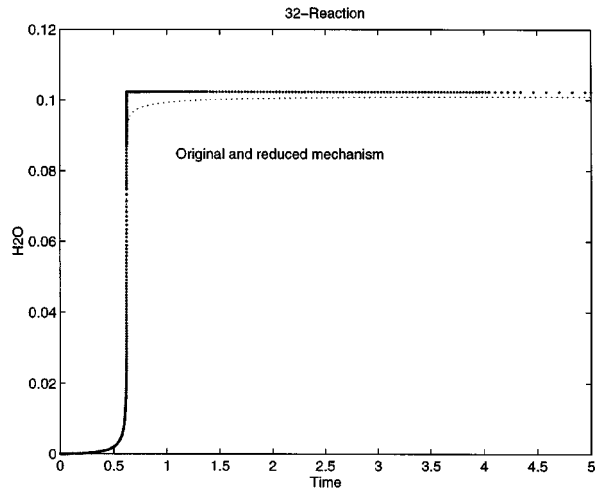

Figure 22. 32-Reaction reduced mechanism for Example 3: H_2O.

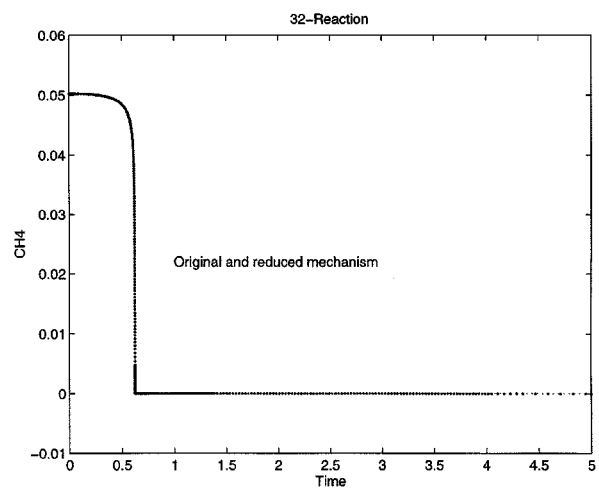

Figure 23. 32-Reaction reduced mechanism for Example 3: CH_4.

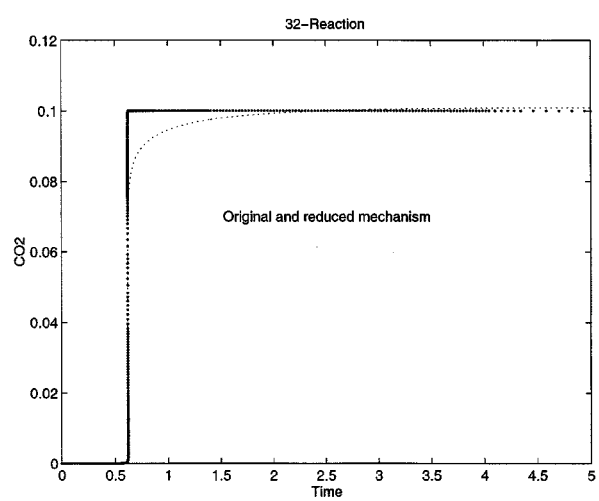

Figure 24. 32-Reaction reduced mechanism for Example 3: CO_2.

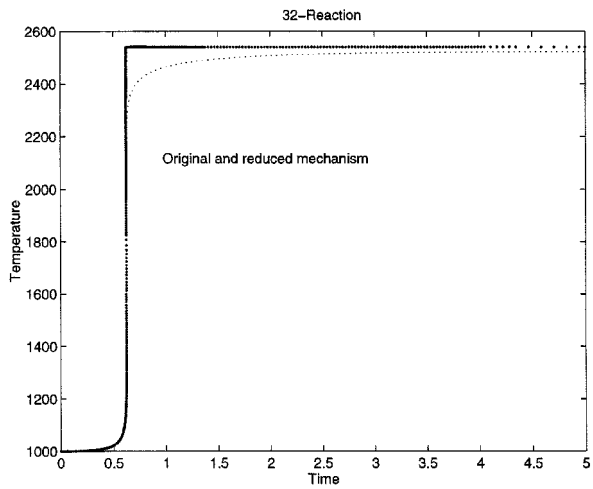

Figure 25. 32-Reaction reduced mechanism for Example 3: temperature.

that it is difficult to reduce the number of reactions to below 30 without using the post recovery process if the steady state of the original mechanism is to be modeled accurately. A smaller reduced model with 26 reactions is obtained using the combination of discrete approach on species (described in the next section), and the post recovery process on reactions which we will outline in the post recovery section.

Reducing the Number of Species

In many problems, reduction of the number of species is also important. Reduction of the number of species can reduce the number of ODEs to be solved and save computation time.

As a consequence of reducing the number of reactions, some species may no longer be explicitly present in any of the remaining reactions. Thus, they will remain unchanged by the reduced model. These species can then be deleted in the sense that their concentration is fixed at the initial value. Note that deleted species may still take part in some reactions if their initial concentrations are not zero. This is because some of the reactions involving species not deleted may be enhanced by the "third body" M. For example, in the reaction $H + O_2 + M = HO_2 + M$, M represents the effect of other species on this reaction and the rate-of-progress for this reaction is the product of $\Sigma_{k=1}^{K}(\alpha_k)[X_k]$ and the rate-of-progress of $H + O_2 = HO_2$. Here, $[X_k]$ are the concentrations of species and $\alpha_k \geq 0$ are weight constants for species. Another consideration is that even if some species do not participate in the chemical reaction explicitly, they need to be considered in the equation-of-state for the gas for a given system pressure.

We might also consider actively seeking out those species that can be deleted in the sense described above. One possibility might be to adapt our general approach to reduce the number of species directly, that is, to associate with each species a parameter and then do the optimization. Because the number of species is much smaller than the number of reactions, the sensitivity analysis for the parameters is less expensive. The optimization problem can be formulated as

$$\begin{aligned} \min \quad & \| y - z \| \\ & \text{subject to} \\ & y' = SF(y), \quad y(0) = y_0 \\ & z' = SDF(z), \quad z(0) = y_0, \quad 0 \leq t \leq b, \\ & k_1 \leq \sum_{i=1}^{N} e_i \leq k_2, \\ & g(e_1, \ldots, e_N) \leq r \end{aligned} \tag{10}$$

where e_i is a parameter associated with species; and $g(e_1, \ldots, e_N)$ is a nonlinear function which, when equal to 0, forces the e_i to take integer values. The difference is that now the elements of the diagonal matrix D are computed by

$$d_i = \Pi_{j \in R_i} e_j \tag{11}$$

where R_i is the set of species which take part in reaction i. Thus, when $e_j = 0$ for some $j \in R_i$, the reaction i is in fact deleted because $d_i = 0$. This optimization problem is well defined when all the parameters take values of 0 or 1. Unfortunately, Eq. 11 is highly nonlinear. Numerical experiments show that the rounding process is much more troublesome than in reducing the number of reactions. Thus an alternative approach must be considered.

Since, as we mentioned before, the number of species is usually much less than the number of reactions, scanning on the species is much less expensive than scanning on the reactions. For example, the GRI model contains 177 reactions but only 32 species. For larger mechanisms, the difference is even more significant. The scanning on the species works as follows. We simply pick a species and delete all the reactions that it is involved in as a reactant or product to arrive at a reduced mechanism. The solution of this mechanism is compared with that of the original mechanism to obtain the error caused by the deletion of this species. Repeating this process on all the species in the mechanism completes one scanning on the species.

The result of scanning on the species for the GRI model is listed in Table 1, where the species are numbered the same way as in the GRI1.2 model. In Table 1, the temperatures are those of the reduced models at the end of the time interval when only the corresponding species are deleted from the model. The errors in Table 1 are given by the scaled L_2 norm of the previous section, which is computed via Eq. 9. The Reactions Remaining column gives the sizes of the reduced models when only the corresponding species is deleted from the model. The last column Kept in Reduced Model indicates whether a species is kept in the two reduced models which we will discuss later; one of them has 14 species and the other has 17 species. Since the starting temperature is 1,000, we can see from the list that deletion of a single species from the set of {4–7, 13–18, 20} will result in a large change in the model, since the temperature does not increase. So, we expect that these 11 species will have to be kept. Deletion of species 1–3, which causes the temperature of the reduced model to be higher than that of the original model, results in the next largest error. Adding them to this set makes the

Table 1. Results of the Scanning on Species for GRI Mechanism

Species No.	Species Name	Temperature (K) at the End	Error Norm	Reactions Remaining	Kept in Reduced Model
1	H_2	2,563.846	327,129.140	147	Both
2	H	2,578.462	145,312.066	93	Both
3	O	2,548.987	566,965.623	140	Both
4	O_2	999.9999	1,673,740.07	148	Both
5	OH	1,000.195	1,673,394.24	119	Both
6	H_2O	1,000.299	1,673,278.93	149	Both
7	HO_2	1,040.797	1,642,273.18	150	Both
8	H_2O_2	2,540.034	5,276.86882	167	None
9	C	2,540.034	0.00010808	172	None
10	CH	2,540.034	0.00365596	161	None
11	CH_2	2,540.034	780.640676	154	None
12	$CH_2(S)$	2,540.034	899.774282	156	None
13	CH_3	1,000.000	1,673,739.91	140	Both
14	CH_4	1,000.000	1,673,740.02	162	Both
15	CO	1,031.249	1,638,600.76	144	Both
16	CO_2	2,205.987	88,755.8553	169	Both
17	HCO	1,009.929	1,660,079.89	153	Both
18	CH_2O	1,000.464	1,673,008.16	149	Both
19	CH_2OH	2,540.034	1,403.02205	164	None
20	CH_3O	1,206.514	1,534,933.98	162	Both
21	CH_3OH	2,540.034	7,138.21136	165	None
22	C_2H	2,540.034	0.00004374	169	None
23	C_2H_2	2,540.034	0.37952949	159	None
24	C_2H_3	2,540.034	4,218.68781	166	None
25	C_2H_4	2,540.034	114,146.134	165	17 Species
26	C_2H_5	2,540.034	75,320.4101	166	17 Species
27	C_2H_6	2,540.034	662,120.500	170	17 Species
28	HCCO	2,540.034	0.01016712	165	None
29	CH_2CO	2,540.034	0.01770426	167	None
30	HCCOH	2,540.034	0.00000011	175	None
31	N_2	2,540.034	0.00000000	177	None
32	AR	2,540.034	0.00000000	177	None

steady state of the reduced system model the original system very well but the ignition time is not modeled very well, as Figure 26 shows. This reduced model contains 14 species and 55 reactions. Note that keeping a species does not mean to keep all the reactions in which it is involved, although deleting a species does mean to delete all the reactions it takes part in. Here, we are not using the steady-state temperature as the sole criteria. The steady-state temperature is used together with the L_2 error norm which is taken on the whole time interval. Using the steady-state temperature only may lead to reduced models with large errors in the ignition time.

An interesting result is that when species 25–27 are also included, the reduced model has almost no error. This reduced model has 17 species and 67 reactions. The figures for this reduced model are not given since they are the same as for the original model.

This result of the scanning on the species suggests that scanning on the species may be a better choice than scanning on the reactions. A modified approach works like this: first the species are scanned and a reduced model which causes almost no error is found. This is much less expensive than scanning the reactions first. Then, the reduced model is scanned a second time on reactions. This reduced model is then fed to the optimizer for further reduction. The work of the optimization is thus reduced considerably and also it can work more efficiently if further reduction is possible. Figure 27 shows the temperature of a 38-reaction reduced model, which is obtained from scanning on the reactions of the 67-reaction model and no optimization. In this case, the reduced

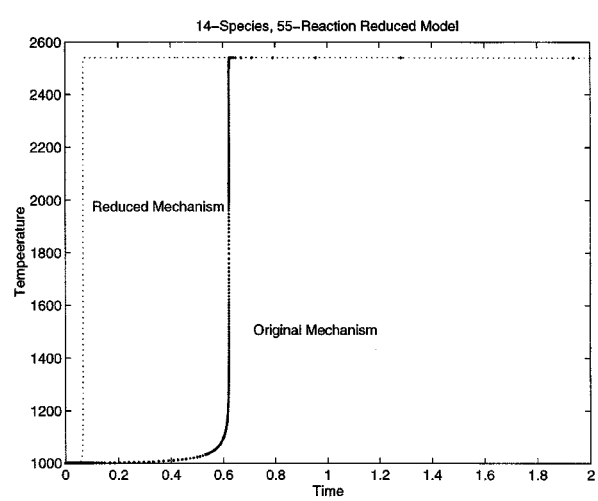

Figure 26. Reduced model of 14 species and 55 reactions for Example 3: temperature.

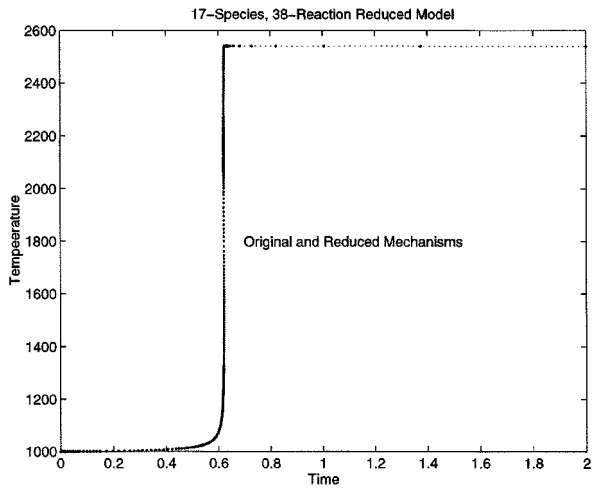

Figure 27. Reduced model of 17 species and 38 reactions for Example 3: temperature.

model results in almost no error and leaves very little space for the optimizer for further reduction, as our numerical experiments demonstrate. The model may be further reduced via optimization or other means, but in our experience this will result in a larger error in the ignition time. Figures for species are not given here, because they are also the same as that of the original mechanism. Note that species are not actually deleted from the model; instead, their concentrations are fixed. So, they still appear in the species list of the input file to Chemkin.

Post Recovery Process and Preselection of Important Reactions

Even though we are taking strict thresholds, the step that rounds to an integer solution can still cause problems. A post recovery process is needed to improve the result, as shown in Example 1. Our numerical experiments show that in some extreme cases, a threshold as small as 0.001 can fail to give a satisfactory integer solution. These failures are due to the high degree of nonlinearity of the optimization problem and the huge difference in the reaction rate constants. The reason for the first difficulty is obvious. The reason for the second difficulty can be illustrated by the following example.

Consider the following scalar ODE

$$y' = -(\lambda_1 + \lambda_2)y, \quad y(0) = 1.0$$

where $\lambda_1 \gg \lambda_2 > 0$. We hope to use the optimization method to select one term which can approximate the original system best. The parameterized system is

$$z' = -(d_1\lambda_1 + d_2\lambda_2)z.$$

Reducing the system to one term requires solving the optimization problem with constraint $d_1 + d_2 = 1$. The correct integer solution should be $d_1 = 1$, $d_2 = 0$.

For simplicity, we take the error norm as

$$\int_0^{t_1} (z(t) - y(t))\,dt.$$

When $(d_1\lambda_1 + d_2\lambda_2)t_1$ and $(\lambda_1 + \lambda_2)t_1$ are large enough, exponential terms in the error can be neglected and the error can be approximated by

$$1/(d_1\lambda_1 + d_2\lambda_2) - 1/(\lambda_1 + \lambda_2).$$

The derivatives of the error norm with respect to d_1 and d_2 are $-\lambda_1(d_1\lambda_1 + d_2\lambda_2)^{-2}$ and $-\lambda_2(d_1\lambda_1 + d_2\lambda_2)^{-2}$. When $d_1 \ll d_2 < 1$ but $d_1^2\lambda_1 \gg 1$, the gradient with respect to d_1 can be small and the optimizer may prematurely declare the current d_1, d_2 to be the optimal solution when the optimality tolerance is not very strict; thus, the rounding process gives a wrong answer as it takes $d_2 = 1$, $d_1 = 0$.

To solve this problem, a post recovery process is needed to improve the result. This can be done in several ways:

- use the optimization method to pick several more reactions;
- perform a sensitivity analysis at 0 of the reactions rounded to 0 to find those reactions that have an extremely large negative derivative value; and
- use the greedy method to find the deleted reactions that improve the objective function most.

The 3-reaction reduced model for Example 1 was obtained via the first approach. We have also used the second approach to obtain a 26-reaction reduced model for Example 3, which is reduced from the 17-species, 38-reaction reduced model in the previous section. We first try to reduce the 38-reaction model to 19 reactions. After a rounding process with threshold 0.03, an integer solution which keeps 19 reactions is obtained, but it yields a poor objective. We then use sensitivity analysis to recover the 6 reactions which have the largest negative derivatives, to obtain this 26-reaction reduced model which is actually better than the larger models we obtained before. The results are shown in Figures 28–32.

Due to the large differences in scale of the reaction rates, some reactions are much more important than others. These most important reactions can be preincluded in the reduced model. This not only saves computation time but can also avoid some numerical difficulties. There are some subsets of reactions of the original mechanism whose deletion results in no ignition, yielding very large objective values. However, the gradients of the objective can be very small when the optimization parameters for these reactions are at 0. The reason is that the remaining subsets of reactions lack more than one key reaction for the ignition. So, the objective is not sensitive to the perturbation in each single parameter. These remaining subsets of reactions are bad local minimums and should be avoided. This is most easily done by preselecting the most important reactions. Thus, preselecting the most important reactions can help the convergence of the optimizer.

Some subsets of reactions that lack one key reaction for the ignition can be very sensitive to small perturbations in species concentrations and reaction rates of the key reaction.

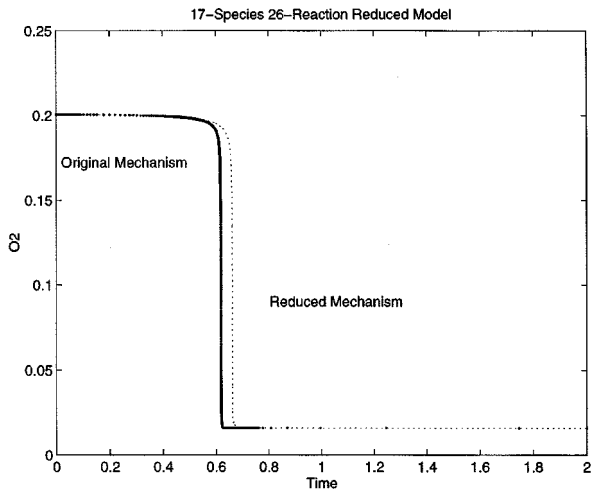

Figure 28. Reduced model of 17 species and 26 reactions for Example 3: O_2.

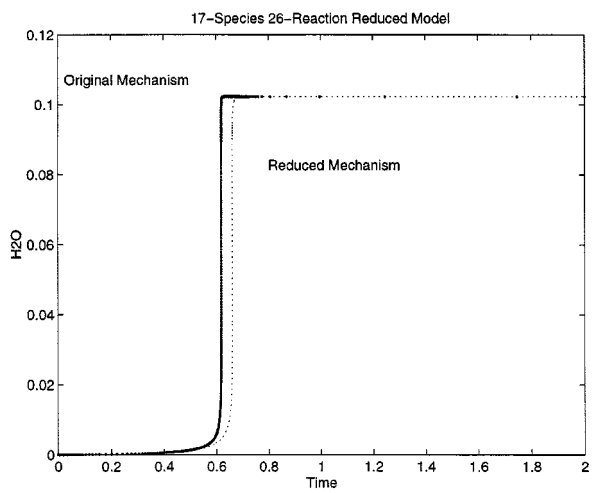

Figure 29. Reduced model of 17 species and 26 reactions for Example 3: H_2O.

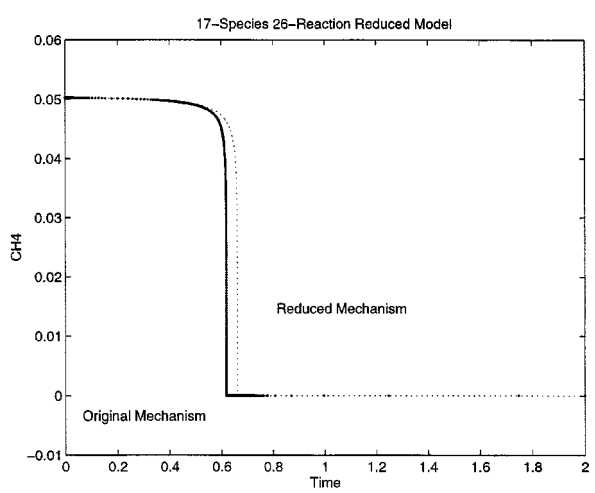

Figure 30. Reduced model of 17 species and 26 reactions for Example 3: CH_4.

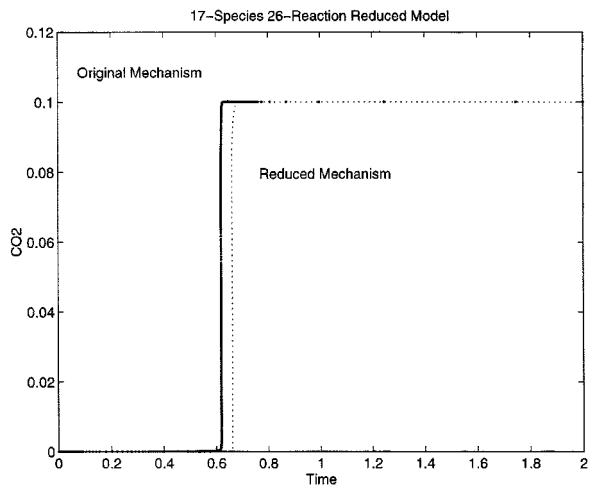

Figure 31. Reduced model of 17 species and 26 reactions for Example 3: CO_2.

The dynamical systems of these subsets of reactions can have very ill-conditioned Jacobian matrices and more seriously, extremely large-scale differences in the sensitivity variables. We have encountered examples with a difference in the scale of the sensitivity variables as large as 10^{-100} and 10^{+50}. Actually, the sensitivity to the key reaction is extremely large, that is, a very small perturbation of the reaction constant from 0 can result in ignition and, thus, reduce the error norm drastically. However, some other reactions may not be so important; a small perturbation to each one of these reaction constants from 0 will not result in ignition; thus, the sensitivities to these reaction constants are almost 0. As we have no prior knowledge of the scaling of the sensitivity variables, scaling can be a very serious problem without a technique to do automatic scaling. Preselecting the most important reactions can alleviate this difficulty.

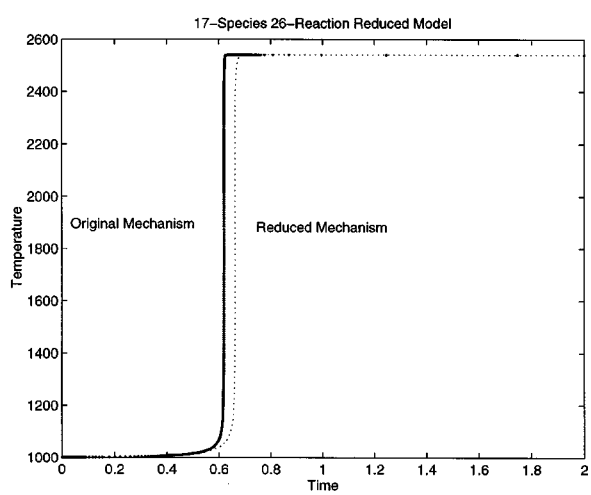

Figure 32. Reduced model of 17 species and 26 reactions for Example 3: temperature.

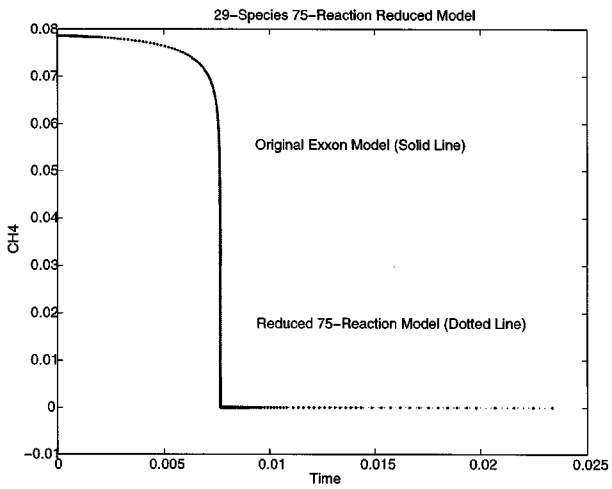

Figure 33. 75-Reaction model vs. Exxon model: CH_4.

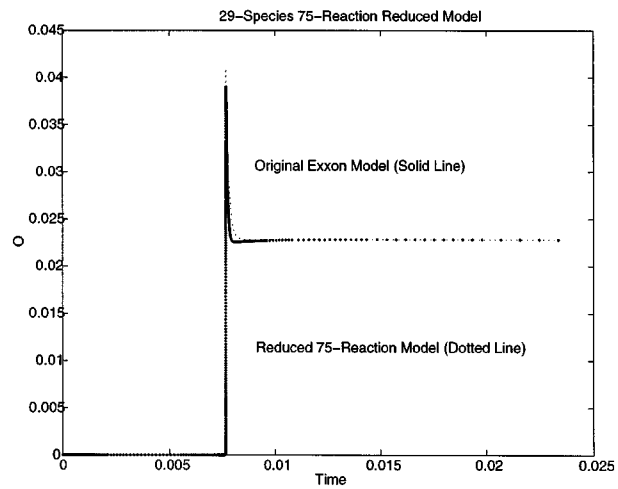

Figure 35. 75-Reaction model vs. Exxon model: O.

Reducing the Exxon Model

In this section, we use the reduction process applied to the Exxon model (Mims et al., 1994) as an example to outline the general algorithm which, in our experience, works best for most cases.

The Exxon model contains an extensive set of gas-phase radical reactions. This homogeneous, one-dimensional (time) model is the largest well-established gas-phase chemical kinetics model we have attempted to reduce. It contains 447 reversible reactions and 116 species. We will present the reduced model under the case of mixture at the lean flammability limit. We take the initial mole fractions of CH_4, O_2, and N_2 as 0.14, 0.56, and 0.03, respectively, and set the initial concentrations of the remaining species to zero. The original model is reduced on the time interval [0.0, 1.0] with initial temperature at 1,273 K. The pressure is kept at one standard atmosphere as before. We use the algorithm introduced in the previous sections. The error in the species concentrations is scaled up in the total error norm by 10^3 as before.

(1) *Scanning Method on Species.* In this step of the reduction process, the species are scanned using the greedy method and we find only 31 species need to be kept to obtain a new model with 142 reactions, which gives an error that can be neglected.

(2) *Scanning Method on Reactions.* In this step, these 142 reactions are scanned and a new reduced model with 75 reactions is obtained. In this step, the error is obtained by comparing the solution of the 142-reaction model and the 75-reaction model. A comparison of species concentrations and temperature between the 75-reaction reduced model and the original Exxon model is given in Figures 33–38.

(3) *Continuous Optimization Method.* In this step, we use our optimization method on the 75-reaction model to reduce it further. To save computation time, we take the 75-reaction model as our original model since it is very close to the original Exxon model.

(a) *Pre-Selection.* We preselect nine of the 75 reactions, which cause the largest errors, using the result of the

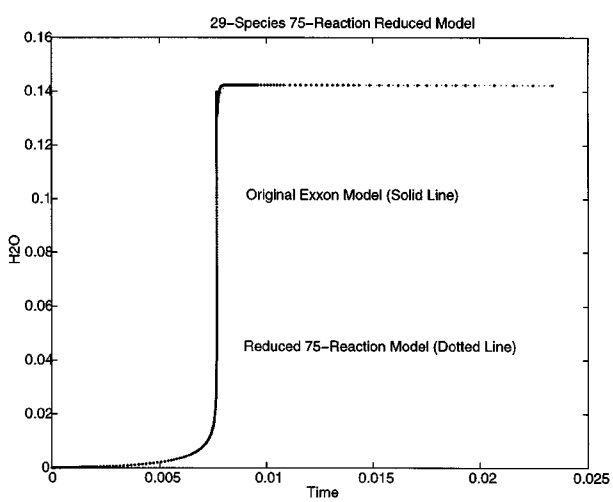

Figure 34. 75-Reaction model vs. Exxon model: H_2O.

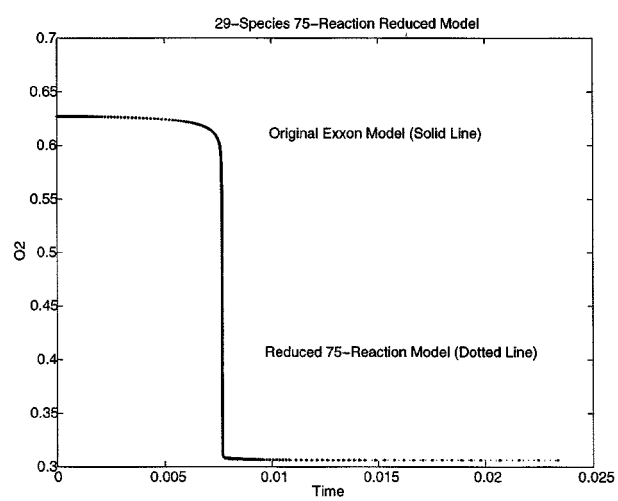

Figure 36. 75-Reaction model vs. Exxon model: O_2.

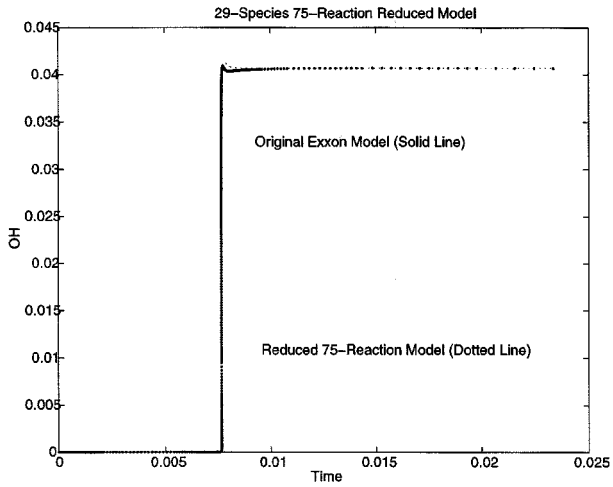

Figure 37. 75-Reaction model vs. Exxon model: OH.

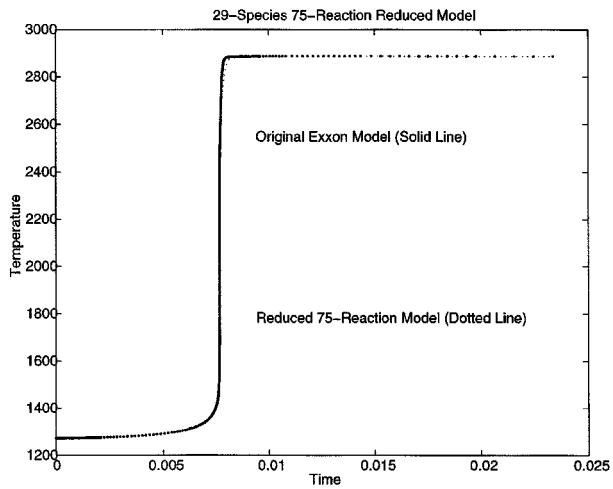

Figure 38. 75-Reaction model vs. Exxon model: temperature.

second step. These reactions will not enter the optimization process. So, the optimization is actually done only on 66 reactions.

(b) *Choosing the Initial Guess.* We choose the initial guess according to the scanning result of the second step. The 15 most important of the remaining 66 reactions are chosen to have initial parameter values of 1, and parameters associated with other reactions are set to 0.

(c) *Optimization.* We use the optimization approach and obtain a solution with 34 reactions.

(d) *Recovery Process.* We use the recovery method to pick up another eight reactions. We choose the reactions to be recovered by the scale of the sensitivity of the error norm to their corresponding parameters. Reactions with the largest negative sensitivities are recovered. The resulting 42-reaction model can approximate the 75-reaction model very well and, thus, the original Exxon model too. Only the first 23 species of this model are actually changing during the physical reaction.

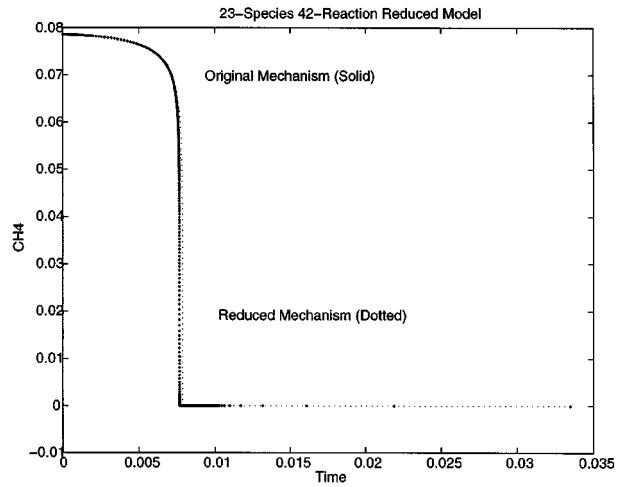

Figure 39. 42-Reaction model vs. 75-reaction model: CH_4.

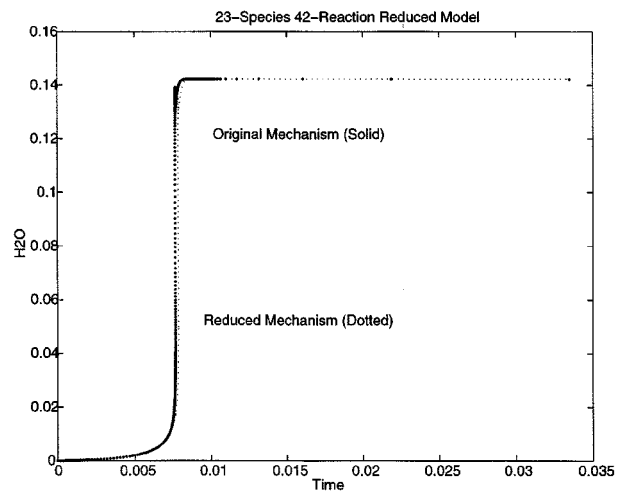

Figure 40. 42-Reaction model vs. 75-reaction model: H_2O.

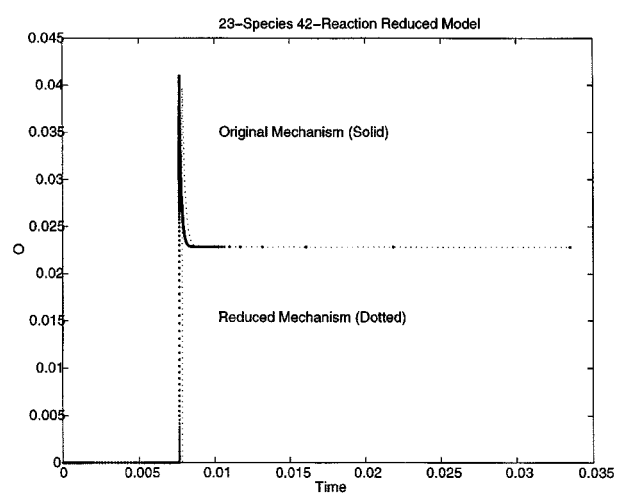

Figure 41. 42-Reaction model vs. 75-reaction model: O.

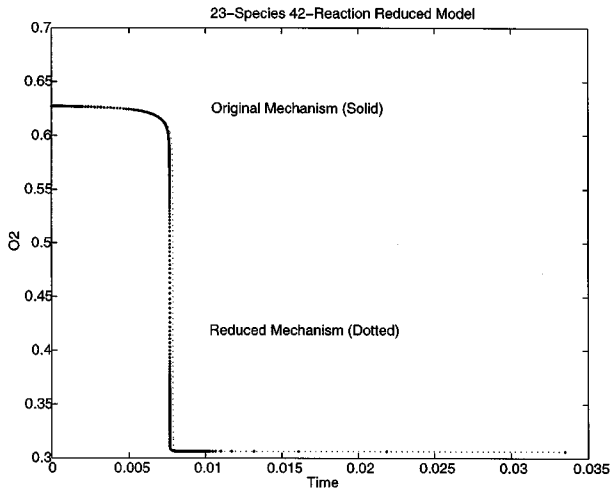

Figure 42. 42-Reaction model vs. 75-reaction model: O_2.

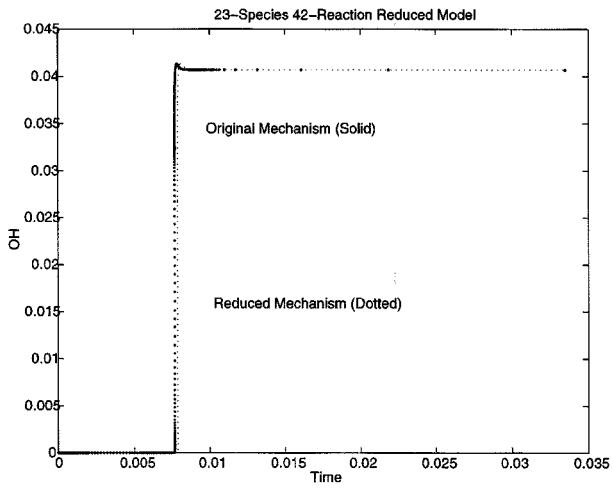

Figure 43. 42-Reaction model vs. 75-reaction model: OH.

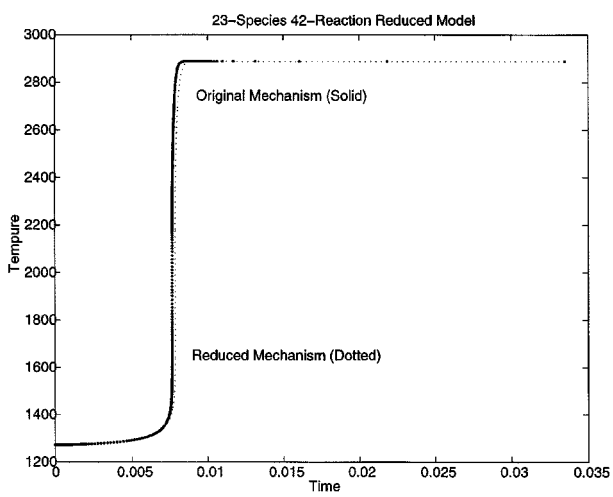

Figure 44. 42-Reaction model vs. 75-reaction model: temperature.

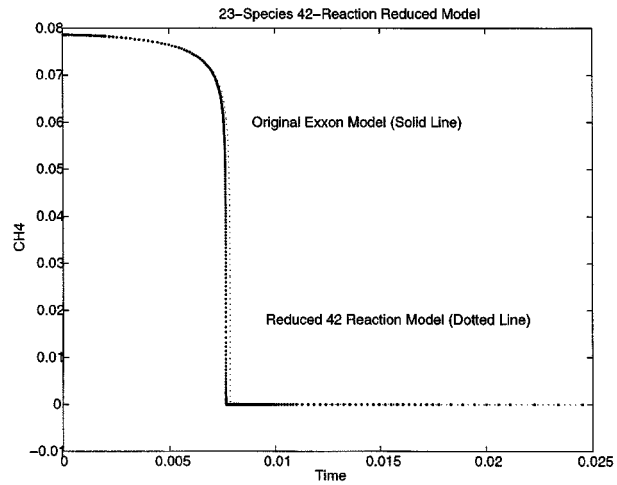

Figure 45. 42-Reaction model vs. Exxon model: CH_4.

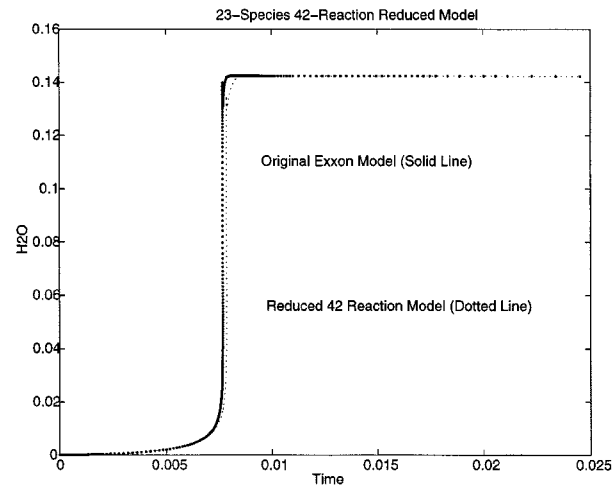

Figure 46. 42-Reaction model vs. Exxon model: H_2O.

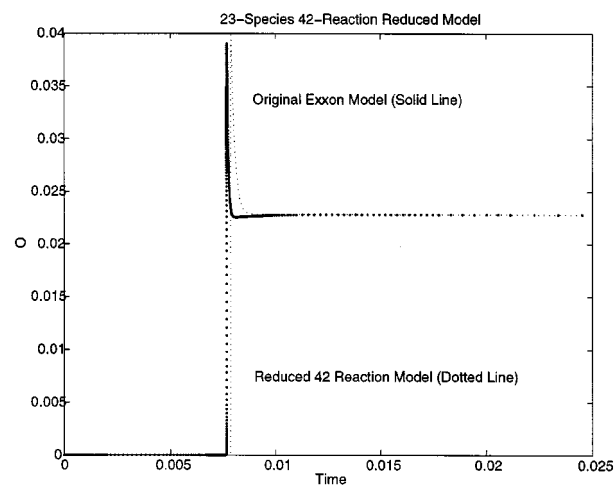

Figure 47. 42-Reaction model vs. Exxon model: O.

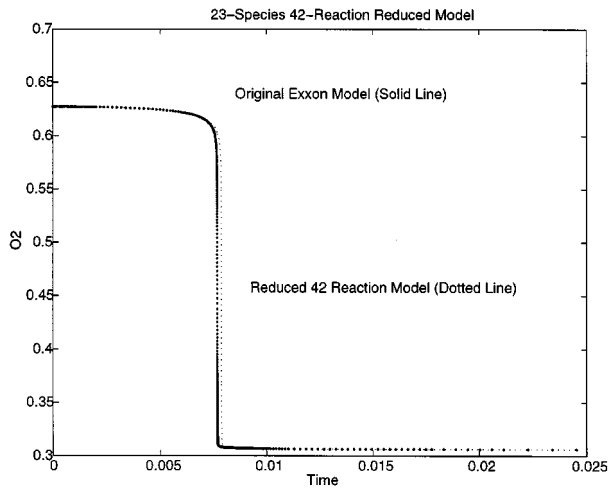

Figure 48. 42-Reaction model vs. Exxon model O_2.

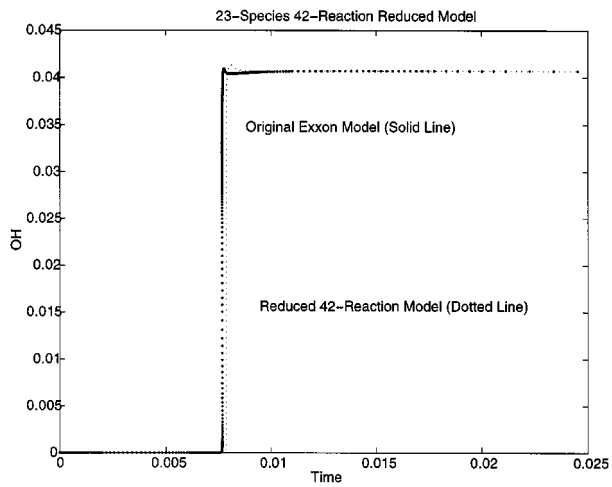

Figure 49. 42-Reaction model vs. Exxon model: OH.

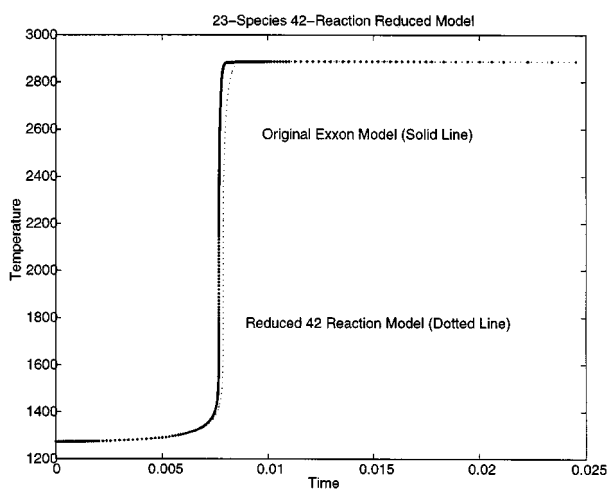

Figure 50. 42-Reaction model vs. Exxon model: temperature.

The species concentrations and temperature of the 75-reaction model and the 42-reaction model are compared in Figures 39–44. In these figures, the original model refers to the 75-reaction model and the reduced model refers to the 42-reaction model. To show that the 42-reaction model is actually very close to the original Exxon model, a comparison between them is given in Figures 45–50.

Acknowledgments

This work was partially supported by DOE contract No. DE-F603-98ER25354 and Exxon Research and Engineering Co., with computing resources from the DOE National Energy Research Scientific Computing Center and the Minnesota Supercomputer Inst.

We would like to thank Dr. Tony Dean of Exxon Research and Engineering Co. and Dr. Bill Green of Massachusetts Institute of Technology for bringing this problem to our attention, and for their interest, advice, and support for this work. Thanks are also due to Prof. J. Ben Rosen of the University of Minnesota for his encouragement and advice in the area of numerical methods for mixed integer nonlinear programming problems.

Literature Cited

Allara, D. L., and D. Edelson, "A Computational Modeling Study of the Low-Temperature Pyrolysis of *n*-Alkanes; Mechanisms of Propane, *n*-Butane, and *n*-Pentane Pyrolyses," *Int. J. Chem. Kinet.*, **7**, 479 (1975).

Bazaraa, M., and C. Shetty, *Nonlinear Programming*, Wiley, New York (1979).

Bischof, C., A. Carle, G. Corliss, A. Griewank, and P. Hovland, "ADIFOR—Generating Derivative Codes from Fortran Programs," *Scientific Programming*, **1**, 11 (1992).

Brenan, K. E., S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, 2nd ed., SIAM, Philadelphia (1995).

Brown, N. J., G. Li, and M. L. Koszykowski, "Mechanism Reduction via Principal Component Analysis," *Int. J. Chem. Kinet.*, **29**, 393 (1997).

Carr, R. W., Jr., D. G. Peterson, and F. K. Smith, "Flash Photolysis of 1,3-Dichlorotetrafluoroacetone in the Presence of Oxygen, Kinetics and Mechanism of the Oxidation of the Chlorodifluoromethyl Radicals," *J. of Physical Chemistry*, **90**, 697 (1986).

Edelson, D., "Computer Simulation in Chemical Kinetics," *Science*, **214**, 981 (1981).

Frenklach, M., H. Wang, M. Goldenberg, G. P. Smith, D. M. Golden, C. T. Bowman, R. K. Hanson, W. C. Gardiner and V. Lissianski, "GRI-Mech—An Optimized Detailed Chemical Reaction Mechanism for Methane Combustion," Gas Research Institute, Chicago, Topical Report, No. GRI-95/0058 (1995).

Gautier, O., and R. W. Carr, Jr., "Variational Sensitivity Analysis of a Photochemical Smog Mechanism," *Int. J. of Chemical Kinetics*, **17**, 1347 (1985).

Gill, P. E., W. Murray, M. A. Saunders, and M. H. Wright, "Model Building and Practical Aspects of Nonlinear Programming," *Computational Mathematical Programming, Series F: Computer and Systems Science*, **15** (1985).

Gill, P. E., W. Murray, and M. A. Saunders, "Large-Scale SQP Methods and Their Application in Trajectory Optimization," *Int. Series of Numerical Mathematics*, **115**, 29, Birkhauser, Switzerland (1994).

Gisvold, K. M., and J. Moe, "A Method for Nonlinear Mixed-Integer Programming and Its Application to Design Problems," J. of Eng. for Industry: Trans. of ASME, **94**, 353 (1972).

Grimme, E., "An Implicitly Restarted Lanczos Method for the Model Reduction of Stable Large-Scale Systems," MS Thesis, University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering, Urbana-Champaign, IL (1994).

Grimme, E., D. Sorensen, and P. VanDooren, "Stable Partial Realizations via an Implicitly Restarted Lanczos Method," preprint (1994).

Kee, R. J., F. M. Rupley, and J. A. Miller, "Chemkin-II: A Fortran Chemical Kinetics Package for the Analysis of Gas Phase Chemical Kinetics," Sandia Report SAND89-8009B, UC-706, Sandia National Laboratories, Livermore, CA (1989).

Lam, S. H., "Using CSP to Understand Complex Chemical Kinetics," *Combust. Sci. and Tech.*, **89**, 375 (1993).

Lam, S. H., and D. A. Goussis, "Conventional Asymptotics and Computational Singular Perturbation for Simplified Kinetics Modelling," preprint, Mechanical and Aerospace Engineering, Princeton Univ., Princeton, NJ (1993).

Maas, U., and S. B. Pope, "Simplifying Chemical Kinetics: Intrinsic Low-Dimensional Manifolds in Composition Space," *Combustion and Flame*, **88**, 239 (1992).

Maly, T., and L. R. Petzold, "Numerical Methods for Sensitivity Analysis of Differential-Algebraic Systems," *Applied Numerical Mathematics*, **20**, 57 (1996).

Mims, C. A., R. Mauti, A. M. Dean, and K. D. Rose, "Radical Chemistry in Methane Oxidative Coupling: Tracing of Ethylene Secondary Reactions with Computer Models and Isotopes," *J. Phys. Chem.*, **98**, 13357 (1994).

Peters, N., "Numerical and Asymptotic Analysis of Systematically Reduced Reaction Schemes for Hydrocarbon Flames," *Lecture Notes in Physics*, **241**, 90, Springer Verlag, Berlin (1985).

Peters, N., *Reducing Mechanisms, Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane-Air Flames*, Springer-Verlag, Berlin-Heidelberg (1991).

Petzold, L., J. B. Rosen, P. E. Gill, L. O. Jay, and K. Park, "Numerical Optimal Control of Parabolic PDEs Using DASOPT," *Large Scale Optimization with Applications, Part II: Optimal Design and Control*, L. Biegler, T. Coleman, A. Conn and F. Santosa, eds., *IMA Volumes in Mathematics and its Applications*, Springer Verlag, New York, **93**, 271 (1997).

Rosen, J. B., Personal communication (1995).

Seigneur, C., G. Stephanopoulos, and R. W. Carr, Jr., "Dynamic Sensitivity Analysis of Chemical Reaction Systems," *Chem. Eng. Sci.*, **17**, 845 (1982).

Vajda, S., P. Valko, and T. Turanyi, "Principal Component Analysis of Kinetic Models," *Int. J. of Chemical Kinetics*, **17**, 55 (1985).

Williams, F. A., *Combustion Theory, The Fundamental Theory of Chemically Reacting Systems*, 2nd ed., Benjamin Cummings Publishing Co., Menlo Park, CA (1985).

Zhu, W., "Model Reduction for Chemical Kinetics: An Optimization Approach," PhD Thesis, Dept. of Computer Science, Univ. of Minnesota (1998).

Manuscript received Mar. 26, 1998, and revision received Jan. 18, 1999.