

Global Sensitivity Analysis with Small Sample Sizes

Michael J. Davis* and Wei Liu*

- Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, IL 60439

Abstract—Methods for doing global sensitivity analysis with small sample sizes are discussed. Two changes are made to our previous algorithm. Rather than an HDMR expansion, we use a full polynomial response surface, but limit the number of interaction terms we include. This reduces the sample size by approximately a factor of five. Further reductions are obtained by using sparse regression techniques to find the regression coefficients, particularly the sparse group lasso.

I. INTRODUCTION

Chemical kinetic models of combustion processes consist of hundreds or thousands of reactions describing the chemistry of dozens to hundreds of chemical species [1]. Even for the most studied chemical reactions the rate coefficients are not known with a great deal of precision, and in many cases most of the rate coefficients are estimated [2]. The well-known uncertainty of rate coefficients has led to extensive use of sensitivity analysis. Traditionally sensitivity analysis has been carried out in a local and linear fashion, generated from information near the estimated rate coefficient [3]. If the rate coefficient is sufficiently different from the nominal value the local sensitivity may be an inaccurate estimation of the role of the uncertainty of the rate coefficient in the uncertainty of the target. Global sensitivity analysis ([4] and [5]) samples the full range of uncertainty of all the reactions at once, so that nonlinearities and correlations may be described.

We present here work [6] on using the sparsity of the sensitivity indices to significantly reduce the sample size needed to get a ranking of the sensitivity of reactions in a chemical model. Using sparse techniques, the sample sizes can be much smaller than the number of reactions and thus can require less computational effort than linear one-at-a-time methods that are often used for ignition delays.

II. RESULTS AND DISCUSSION

In order to use sparse regression techniques, the response surface for the ignition target studied here is written in a form [6] that differs from the HDMR expansion we typically have used (for example, Ref. 7):

$$\tau^{(1)}(\{u_i\}) = \sum_{i=1}^m \sum_{k=0}^n a_{ik} u_i^k \quad (1a)$$

$$\tau^{(2)}(\{u_i\}) = \sum_{i=1}^m \sum_{k=1}^n a_{ik} u_i^k + \sum_{j=1}^p \sum_{s=1}^{n-1} \sum_{\substack{k+r \leq s \\ k,r > 0}} b_{jkr} u_{v(j,1)}^k u_{v(j,2)}^r \quad (1b)$$

where m refers to the number of reactions and n refers to the order of the expansion. It is straightforward to generate sensitivity indices (S_i 's) using expansions like those in Eq. (1a) [6]. Equation (1b) uses the information from the fit in Eq. (1a): p is a number determined from the expansion in Eq. (1a). Only reactions in Eq. (1a) whose sensitivity indices are above a threshold are used in the second set of terms in Eq. (1b). The value of p is generally less than 200 in our calculations. We typically use thresholds of $S_i = 0.005$ or $S_i = 0.01$. The indices “ $v(j,1)$ ” and “ $v(j,2)$ ” in Eq. (1b) refer to a pair of reactions that are used in the interaction term based on their S_i 's. It is straightforward to use Eq. (1a) to extract S_i 's. Extraction of S_i 's from Eq. (1b) requires integration, done analytically for the polynomial expansions.

Results for butanol ignition are presented in the middle panel of Fig. 1, generated using the expansions in Eqs. (1a) and (1b). Comparison of this plot and the one above it demonstrates that good results can be obtained with the expansion in Eqs. (1a) and (1b) with much smaller sample sizes. The sample size of 7,000 used for Eqs. (1a) and (1b) is a significant improvement over the previous sample size of 50,000 (top panel of Fig. 1). We have used an HDMR expansion in many studies and found that convergence was achieved much more rapidly for the large S_i 's than for small S_i 's. The slowness of the convergence is a manifestation of a phenomenon called “overfitting” [8], which is alleviated to a significant extent by the expansion in Eqs. (1a) and (1b). The reduction of effort is achieved using ordinary least squares regression for the fits in Eqs. (1a) and (1b). Even further reductions can be found by starting from Eqs. (1a) and (1b) and taking into account the sparseness of the results - most sensitivity coefficients are very small. The additional reduction is achieved by using sparse regression techniques [8].

The savings from sparse regression results from the fact that typically the uncertainty of only a few reactions contribute significantly to the uncertainty in a given target. The sparse regression techniques take advantage of this without any prior knowledge about which reactions contribute. This is accomplished by adding a “penalty” term to the usual error function for least squares fits:

$$E(\mathbf{c}) = \frac{1}{2} \sum_{k=1}^M \left(t_k - \sum_{j=1}^L c_j g_j(\mathbf{u}_k) \right)^2 + \lambda \sum_{j=1}^L |c_j| \quad (2)$$

where t_k refers to the ignition delay time for the k^{th} sample run. The specific algorithm we use is described in the upcoming paper [6].

The expansion described in Eq. (2) is referred to as the LASSO [8]. The penalty term serves two purposes, it

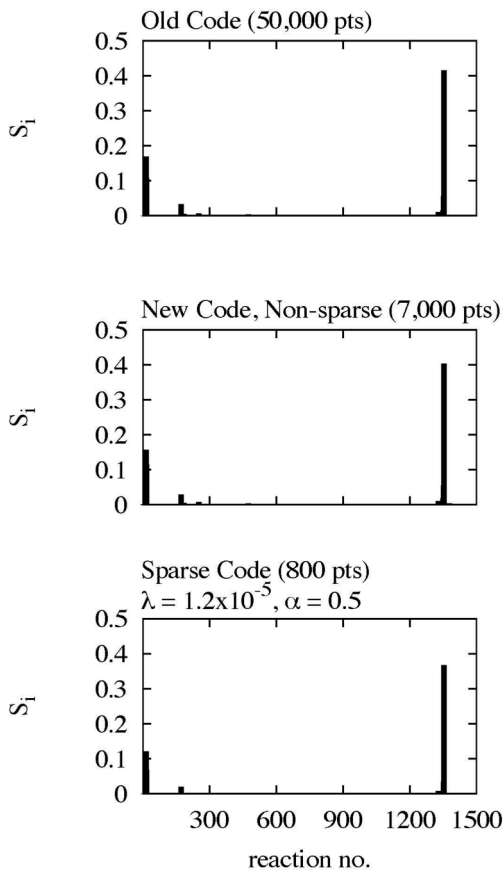


Fig. 1. The top and middle panels show results for butanol ignition with the old algorithm and new algorithm, respectively. The bottom panel shows a calculation using sparse regression.

Butanol Selection (Ignition)

Rxn	No	S_i	Rank	400	800	1200
NC4H9OH+HO2 = C4H8OH-1+H2O2	1353	0.403	1	1	1	1
H2O2+O2 = HO2+HO2 (dupl.)	14	0.158	2	2	2	2
H2O2(+M) = OH+OH(+M)	16	0.116	3	3	3	3
NC4H9OH+HO2 = C4H8OH-3+H2O2	1351	0.055	4	4	4	4
NC4H9OH+HO2 = C4H8OH-2+H2O2	1352	0.032	5	5	5	5
C2H5+HO2 = C2H5O+OH	171	0.0290	6	6	7	7
H2O2+O2 = HO2+HO2 (dupl.)	15	0.0287	7	7	6	6
NC4H9OH+HO2 = C4H8OH-4+H2O2	1350	0.018	8	8	8	8
NC4H9OH+HO2 = C4H8OH-4+H2O2	1329	0.011	9	9	9	9

Columns 3 and 4: OLS, $n = 7000$

Fig. 2. Reaction Selection for n-Butanol ignition is shown.

“regularizes” the solution, reducing overfitting, but more importantly, for our purposes, it selects coefficients, resulting in many coefficients that are zero for specific values of λ . There are a number of ways of picking good values of λ [8], and typically we use cross-validation or an empirical rule we discerned from a number of applications. Equation (2) is a simplification of the procedure used in Ref. [6], the sparse group lasso [9], which recognizes the grouping of expansion coefficients based on individual reactions. There is an additional adjustable parameter, α , set to 0.5 here (it can range from 0.0 to 1.0).

The results of applying the previously described algorithm is presented in the bottom panel of Fig. 1, which demonstrates that we can obtain good agreement between ordinary least squares and sparse techniques, with much smaller sample sizes. Figure 2 has a detailed comparison for all sensitivity coefficients greater than 0.01 (1% of the variance). The comparisons are made for three sample sizes in the sparse case: 400, 800, and 1200. Numbers in red indicate the reactions that are mis-ordered and it can be observed that this only occurs for a single pair which have very close sensitivity coefficients in the much larger ordinary least squares calculation (7000 samples).

ACKNOWLEDGMENT

This work was supported by the Division of Chemical Sciences, Geosciences, and Biosciences, the Office of Basic Energy Sciences, the U.S. Department of Energy, under contract number DE-AC02-06CH11357.

REFERENCES

- [1] C. K. Westbrook, Y. Mizobuchi, T. J. Poinso, P. J. Smith, J. Warnatz, “Computational Combustion”, *Proc. Combust. Inst.* 30, 125-157 (2005).
- [2] H. J. Curran, P. Gaffuri, W. J. Pitz, and C. K. Westbrook, “A Comprehensive Modeling Study of n-Heptane Oxidation”, *Combustion and Flame* 114, 149-177 (1998).
- [3] J. Zador, I. G. Zseley, and T. Turanyi, “Local and global uncertainty analysis of complex chemical kinetic systems”, *Reliab. Eng. Syst. Saf.* 91, 1232-1240 (2006).
- [4] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Sasana, and S. Tarantola, *Global Sensitivity Analysis. The Primer*; John Wiley & Sons: Hoboken NJ (2008).
- [5] A. S. Tomlin and T. Ziehn, “The Use of Global Sensitivity Methods for the Analysis, Evaluation and Improvement of Complex Modelling Systems” in *Coping with Complexity: Model Reduction and Data Analysis*; A. N. Gorban and D. Roose, Eds.; Springer; Berlin, (2010), p. 9.
- [6] M. J. Davis and W. Liu, “Global Sensitivity Analysis with Small Sample Sizes”, to be submitted.
- [7] R. T. Skodje, A. S. Tomlin, S. J. Klippenstein, L. B. Harding, and M. J. Davis, “Theoretical Validation of Chemical Kinetic Mechanisms: Combustion of Methanol”, *J. Phys. Chem. A* 114, 8286-8301 (2010).
- [8] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer (2009).
- [9] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, “A Sparse Group Lasso”, *Journal of Computational and Graphical Statistics*, to be published.