Reducing the Tabulation Dimension in the In Situ Adaptive Tabulation (ISAT) Method

by

S.B. Pope

FDA 96-04 August 1996

Fluid Dynamics and Aerodynamics Program

Sibley School of Mechanical and Aerospace Engineering

Cornell University Ithaca, New York 14853

Introduction

In the ISAT method for the efficient implementation of combustion chemistry (Pope 1996), the retrieval work and the table storage scale as D^2, where D is the number of degrees of freedom in the thermochemistry. For the GRI 2.11 mechanism for methane, D is of order 50. There is great advantage in being able to reduce the dimensionality used in the table to $D_r \approx \frac{1}{4}D$, say, thus increasing the asymptotic speed and decreasing the storage by a factor of 16 (in this example).

This note describes how the required reduction can be achieved. It uses the notation of the ISAT paper (Pope 1996); and for simplicity takes the scaling matrix to be $\mathbf{B} = \mathbf{I}$, and assumes that $\boldsymbol{\phi} = 0$ is in the accessed region. The reduction is constructed from a table generated using the D degrees of freedom in the thermochemistry.

Retained and Neglected Subspaces

The "retained subspace" $\mathcal{R}$ is a D_r-dimensional subspace of the D-dimensional composition space. The success and accuracy of the method described here depends entirely on the appropriate choice of $\mathcal{R}$.

The "neglected subspace" $\mathcal{N}$ is the $D_n = D - D_r$ orthogonal complement of $\mathcal{R}$.

Let $\{\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^{D_r}\}$ be an orthonormal basis for $\mathcal{R}$, and let these vectors be the columns of a matrix $\mathbf{E}$. Similarly let $\{\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^{D_n}\}$ be an orthonormal basis for $\mathcal{N}$, and let these vectors form the columns of a matrix $\mathbf{F}$. (Note that $\mathbf{e}^i$ is orthogonal to $\mathbf{f}^j$.)

In the e-basis, the retained components of $\boldsymbol{\phi}$ are

$$\tilde{\boldsymbol{\phi}}^r = \mathbf{E}^T \boldsymbol{\phi}, \tag{1}$$

while in the original basis they are

$$\boldsymbol{\phi}^r = \mathbf{E}\tilde{\boldsymbol{\phi}}^r = \mathbf{P}^r \boldsymbol{\phi}, \tag{2}$$

where $\mathbf{P}^r \equiv \mathbf{E}\mathbf{E}^T$ is the perpendicular projection onto $\mathcal{R}$. Similarly, in an obvious notation, the neglected components are

$$\tilde{\boldsymbol{\phi}}^n = \mathbf{F}^T \boldsymbol{\phi}, \tag{3}$$

and

$$\boldsymbol{\phi}^n = \mathbf{F}\tilde{\boldsymbol{\phi}}^n = \mathbf{P}^n\boldsymbol{\phi}, \tag{4}$$

with $\mathbf{P}^n = \mathbf{F}\mathbf{F}^T$.

As implied by the terminology, the reduction is achieved by approximating the D-vector $\boldsymbol{\phi}$ by its projection onto the retained subspace $\boldsymbol{\phi}^r$. This can be represented by the D_r-vector $\tilde{\boldsymbol{\phi}}^r$. Similarly, the $D \times D$ mapping gradient matrix $\mathbf{A}$ is replaced by the $D_r \times D_r$ matrix $\tilde{\mathbf{A}}^r$, defined below.

Directions of Maximum Variation (DMV)

Starting with $D_r = 0$, the retained subspace can be progressively built up by incrementing D_r and suitably defining the additional basis vector $\mathbf{e}^{D_r}$. A possible choice of $\mathbf{e}^{D_r}$ is the *direction of maximum variation* (DMV). Out of all the tabulation points $\boldsymbol{\phi}^0$, all the mappings $\mathbf{R}(\boldsymbol{\phi}^0)$, and the origin, let $\boldsymbol{\phi}^{(a)}$ and $\boldsymbol{\phi}^{(b)}$ be the pair of points that are furthest apart in the neglected subspace. That is $|\mathbf{F}^T(\boldsymbol{\phi}^{(a)} - \boldsymbol{\phi}^{(b)})|$ is maximized by this choice of a and b. We define the *maximum variation* by

$$\zeta \equiv |\mathbf{F}^T(\boldsymbol{\phi}^{(a)} - \boldsymbol{\phi}^{(b)})|, \tag{5}$$

and the DMV by

$$\mathbf{u} \equiv [\boldsymbol{\phi}^{(a)} - \boldsymbol{\phi}^{(b)}]/\zeta. \tag{6}$$

If ζ is zero, then $\mathbf{u}$ is taken to be any vector in $\mathcal{N}$. Note that, correctly, $\mathbf{u}$ is a unit vector, orthogonal to $\mathcal{R}$.

Neglected Mapping

For a given specification of $\mathcal{R}$, let ζ be the maximum variation in the neglected subspace. From any point $\boldsymbol{\phi}$, the mapping $\mathbf{R}(\boldsymbol{\phi})$ can be decomposed into a retained and neglected part:

$$\mathbf{R} = \mathbf{R}^r + \mathbf{R}^n. \tag{7}$$

Let $\mathcal{H}$ be the convex hull formed by all tabulation points $\boldsymbol{\phi}^0$ and mappings $\mathbf{R}(\boldsymbol{\phi}^0)$. It is clear from these definitions, that for all points $\boldsymbol{\phi}$ in $\mathcal{H}$, we have

$$|\mathbf{R} - \mathbf{R}^r| = |\mathbf{R}^n| \leq \zeta. \tag{8}$$

Thus the error in the approximation $\mathbf{R} \approx \mathbf{R}^r$ is bounded by ζ.

There are points inside the ellipsoids of accuracy (EOA) that are outside $\mathcal{H}$. Hence, Eq. (8) is not a tight upper bound on the error incurred: but we assume that it is adequate.

Retained Mapping

For points $\boldsymbol{\phi}$ inside the EOA at $\boldsymbol{\phi}^0$, the linearized mapping is

$$\mathbf{R}^\ell = \mathbf{R}(\boldsymbol{\phi}^0) + \mathbf{A}\,\delta\boldsymbol{\phi}, \tag{9}$$

where $\delta\boldsymbol{\phi} \equiv \boldsymbol{\phi} - \boldsymbol{\phi}^0$. In the $\mathbf{e}-\mathbf{f}$-basis, the retained and neglected components are

$$\begin{aligned} \begin{bmatrix} \tilde{\mathbf{R}}^{\ell,r} \\ \tilde{\mathbf{R}}^{\ell,n} \end{bmatrix} &= \begin{bmatrix} \mathbf{E}^T \\ \mathbf{F}^T \end{bmatrix} \begin{bmatrix} \mathbf{R}(\boldsymbol{\phi}^0) \end{bmatrix} + \begin{bmatrix} \mathbf{E}^T \\ \mathbf{F}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{\phi} \end{bmatrix} \\ &= \begin{bmatrix} \tilde{\mathbf{R}}^r(\boldsymbol{\phi}^0) \\ \tilde{\mathbf{R}}^n(\boldsymbol{\phi}^0) \end{bmatrix} + \begin{bmatrix} \mathbf{E}^T \\ \mathbf{F}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{E}\,\mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{E}^T \\ \mathbf{F}^T \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{\phi} \end{bmatrix} \\ &= \begin{bmatrix} \tilde{\mathbf{R}}^r(\boldsymbol{\phi}^0) \\ \tilde{\mathbf{R}}^n(\boldsymbol{\phi}^0) \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{A}}^r & \mathbf{G} \\ \mathbf{H} & \mathbf{K} \end{bmatrix} \begin{bmatrix} \delta\tilde{\boldsymbol{\phi}}^r \\ \delta\tilde{\boldsymbol{\phi}}^n \end{bmatrix}, \end{aligned} \tag{10}$$

where the last matrix is a partitioning of the transformed mapping gradient matrix $[\mathbf{EF}]^T\mathbf{A}[\mathbf{EF}]$. (Note that $[\mathbf{EF}]$ is a unitary matrix.)

The neglected component $\tilde{\mathbf{R}}^{\ell,n}$ has already been considered. The retained component is

$$\tilde{\mathbf{R}}^{\ell,r} = \tilde{\mathbf{R}}^r(\boldsymbol{\phi}^0) + \tilde{\mathbf{A}}^r\delta\tilde{\boldsymbol{\phi}}^r + \mathbf{G}\delta\tilde{\boldsymbol{\phi}}^n. \tag{11}$$

In the reduced method, $\tilde{\mathbf{R}}^r(\boldsymbol{\phi}^0)$ and $\tilde{\mathbf{A}}^r$ are tabulated, and $\delta\tilde{\boldsymbol{\phi}}^r = \tilde{\boldsymbol{\phi}}^r - \tilde{\boldsymbol{\phi}}^{0,r}$ are known. But the final term

$$\gamma \equiv \mathbf{G}\delta\tilde{\boldsymbol{\phi}}^n, \tag{12}$$

is neglected: it represents the contribution to the retained mapping from the neglected components of the composition. Clearly, the accuracy of the method depends on $|\gamma|$ being small.

Let the SVD of $\mathbf{G}$ be $\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$, let $\{\tilde{\mathbf{v}}^1, \tilde{\mathbf{v}}^2, \ldots, \tilde{\mathbf{v}}^{D_n}\}$ be the columns of $\tilde{\mathbf{V}}$, and let $\tilde{\sigma}_i$ ($i = 1, 2, \ldots, D_n$) be the singular values (i.e., the diagonal elements of $\tilde{\mathbf{\Sigma}}$). Then Eq. (12) can be rewritten

$$\tilde{\mathbf{U}}^T \gamma = \tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T \delta\tilde{\boldsymbol{\phi}}^n. \tag{13}$$

For a given magnitude $|\delta\tilde{\boldsymbol{\phi}}^n|$, the maximum value of $|\gamma|$ that can occur is $\tilde{\sigma}_1|\delta\tilde{\boldsymbol{\phi}}^n|$, and this occurs when $\delta\tilde{\boldsymbol{\phi}}^n$ is aligned with the first singular vector $\tilde{\mathbf{v}}^1$. We define σ_{max} to be the maximum of $\tilde{\sigma}_1$ over all tabulation points, and we define the *direction of maximum sensitivity* (DMS) $\tilde{\mathbf{v}}$ to be $\tilde{\mathbf{v}}^1$ at that point. The neglected term γ is bounded by

$$|\gamma| \leq \sigma_{\mathrm{max}}|\delta\tilde{\boldsymbol{\phi}}^n|. \tag{14}$$

In the original basis, the direction of maximum sensitivity is

$$\mathbf{v} \equiv \mathbf{F}\tilde{\mathbf{v}}. \tag{15}$$

Acceptable, Minimal and Optimal Decompositions

The reduced representation introduces two errors : one due to the neglect of $\tilde{\mathbf{R}}^{\ell,n}$, and one due to the neglect of the influence of $\delta\tilde{\boldsymbol{\phi}}^n$ on $\tilde{\mathbf{R}}^{\ell,r}$ (i.e., γ). For points $\boldsymbol{\phi}$ in $\mathcal{H}$, the sum of these errors is bounded by

$$\varepsilon_n = \zeta + \sigma_{\mathrm{max}}\zeta, \tag{16}$$

(see Eqs. 8 and 14). Given an error tolerance $\varepsilon_{\mathrm{tol}}$, an *acceptable decomposition* is one for which ε_n does not exceed $\varepsilon_{\mathrm{tol}}$.

There is a minimum dimension $D_{r,\mathrm{min}}$ of the retained subspace for which acceptable decompostions exist. An acceptable decomposition with $D_r = D_{r,\mathrm{min}}$ is a *minimal decomposition.* Of the minimal decompositions, one that minimizes ε_n is an *optimal decomposition.*

Algorithm

It is not evident how to construct the optimal or even a minimal decomposition. But the following algorithm produces an acceptable decomposition. It takes into consideration the fact that it is much more expensive to determine the DMS $\mathbf{v}$ than the DMV $\mathbf{u}$. (To determine $\mathbf{v}$, the SVD must be performed for $\mathbf{G}$ at every tabulation point.)

1. Specify a parameter α, e.g. $\alpha = 10$.
2. Evaluate the DMV, $\mathbf{u}$, and the maximum variation, ζ.
3. If $\zeta \geq \alpha \varepsilon_{\text{tol}}$, then increment D_r taking $\mathbf{u}$ as the additional basis vector; go to 2.
4. ($\zeta < \alpha\varepsilon_{\text{tol}}$.) Evaluate the DMS, $\mathbf{v}, \sigma_{\max}$ and ε_n.
5. If $\varepsilon_n \leq \varepsilon_{\text{tol}}$—all done.
6. Increment D_r: if $\sigma_{\max} > 1$ use $\mathbf{v}$, otherwise use $\mathbf{u}$; go to 2.

References

S.B. Pope (1996) "Computationally Efficient Implementation of Combustion Chemistry using In Situ Adaptive Tabulation", Cornell Report FDA 96–02 (submitted to Combustion Theory and Modelling).