The thermochemical state of a single-phase reacting system is $n_s + 1$ dimensional. However, it is widely recognized that low-dimensional manifolds exist in this space. The question then arises as to the best way to approximate these manifolds. Common chemistry reduction approaches for combustion simulation (such as laminar flamelet, equilibrium, and flame-sheet (Burke-Schuman) chemistry) specify a priori the manifold dimensionality and the parameters that form the basis for representing the thermochemical state, which is then prescribed as a unique function of these parameters. If high-fidelity data are available from experiment or DNS, manifolds may be determined via a Principal Component Analysis (PCA), with the principal components (PCs) forming a new basis for describing the thermochemical state. By considering a truncated set of PCs, one may approximate the thermochemical state with rigorous error bounds, and extract a parameterized representation of the thermochemical state, together with transport equations for the principal components.

This paper outlines a methodology for constructing a model for the thermochemical state from DNS data which can be used in a full-scale simulation, with particular focus on the ability to parameterize source terms appearing in the transport equations for the principal components.

1 Introduction

Direct Numerical Simulation (DNS) of practical combustion systems is impossible due to two general complications:
1. **Resolution requirements.** DNS requires that all length and time scales physically realizable by the system of interest be fully resolved. The separation of length scales in turbulent flow implies that resolution requirements scale roughly with the Reynolds number as $\text{Re}^3$ [1].

2. **Number of equations to be solved.** Detailed combustion mechanisms for simple fuels such as methane involve 53 species and 325 reactions [2]. This implies that direct solution of a reacting system involving only methane in air would require solution of 57 strongly coupled partial differential equations (52 species, continuity, 3 momentum, and 1 energy equation). For higher hydrocarbons, the number of equations increases significantly.

Models are commonly introduced to alleviate the cost associated with both of the issues described above. Resolution requirements are commonly addressed by filtering the governing equations spatially and/or temporally to produce the LES and RANS equations. This leads to turbulent closure problems with convective and source terms in particular.

Models to reduce the number of equations to be solved (issue 2 above) vary widely. In general, the thermochemical state of a single-phase reacting system with $n_s$ chemical species is uniquely determined by $n_s + 1$ parameters (e.g. $T$, $p$, and $n_s - 1$ mass fractions, $Y_i$). However, it is widely recognized that in turbulent combustion, lower dimensional manifolds exist in this high-dimensional space [3, 4]. All of the models which address issue 2 above rely, either directly or indirectly, on the fact that in turbulent combustion there exist low-dimensional attractive manifolds to which the thermochemical dynamics quickly relax. These modeling approaches represent attempts to approximate the manifold representing the thermochemical state within some reasonable error bounds.

Among the most common approaches to reduce the number of thermochemical degrees of freedom are:

- Mechanism reduction and solution of a reduced set of species equations [5–8]. The primary difficulty in turbulent closure models for this approach is dealing with closure of the chemical source term, which is highly nonlinear. Common turbulent closure models for source terms when reduced mechanisms are employed span a wide range of complexity from EDC models [9] to PDF transport models [10–12].

- Re-parameterization of the thermochemical state space by a small number of parameters. This is perhaps the most common combustion modeling approach. The most common choice of a parameter for
nonpremixed combustion is the mixture fraction \( f \), which is a particularly convenient choice since if all species diffusivities are equal, the mixture fraction is a conserved scalar \([13–17]\). Because the mixture fraction may be modeled as a conserved scalar, no source terms require closure. However, as with the reduced-mechanism approach, properties such as density, viscosity, etc. must still be calculated from the mixture fraction. Common turbulent closure models for this approach include conditional moment closure \([18]\), and presumed-PDF approaches \([19, 20]\).

This paper will focus on the second of these approaches. Namely, re-parameterizing the thermochemical state by a small number of parameters. Most successful attempts at exploiting these manifolds have been to prescribe \textit{a priori} the parameters that characterize the manifold \([4, 21–24]\). However, such an approach explicitly restricts the subspace that the thermochemistry may access, without providing \textit{a priori} any quantitative error analysis. Indeed, increases in the overlap between timescales associated mixing and those associated with chemical reaction increases the dimensionality of a manifold \([4, 25]\).

Principal Component Analysis (PCA) offers the potential to automate selection of an optimal basis for representing the manifolds which exist in turbulent combustion. The central idea of PCA is to reduce the dimensionality of a data set consisting of a large number of correlated variables while retaining most of the variation present in the original data. The dimension reduction is achieved through an eigenvalue/eigenvector problem, by recasting the data into a lower dimensional space identified by the principal components (PCs). As we will show in §2.1 these are defined as the \( n_\eta \) eigenvectors of the covariance matrix of the data associated with the \( n_\eta \) largest eigenvalues. In many cases (particularly those where manifolds exist) the number of eigenvectors, \( n_\eta \), required to reproduce the original data properly is much smaller than the original dimension of the data.

PCA can provide an optimal representation of the system based on \( n_\eta \ll n_s + 1 \) “optimal” variables, \( \eta \), which are linear combination of the primitive variables \( T, p \) and \( Y_i \). PCA provides a linear mapping from the original variables \( (T, p, Y_i) \) to a set of principal components. Using this mapping, transport equations may be derived for the principal components and the initial/boundary conditions transformed, as discussed in §2.2.1. The entire problem may be then solved using the principal components. For PCA to be effectively exploited as a modeling technique, several conditions must hold. First the thermochemical variables (e.g. \( \rho, \mu, T, \lambda \), etc.) must be parameterized within acceptable error bounds. Second, the source terms for the principal components must be parameterized within acceptable error bounds.
2 Principal Component Analysis

2.1 Formulation

Consider a set of observables, \( X = [x_1, x_2, \ldots, x_m] \), with \( x_i \) the \( i \)th observation comprised of a set of \( n \) variables so that \( X \) is an \( n \times m \) matrix. The basic idea of PCA is to find a basis for representing the data \( X \) such that the data are well represented by a truncated basis [26, 27]. We may define the covariance matrix

\[
R = XX^T
\]

and perform an eigenvector decomposition of \( R \) to obtain

\[
\Lambda = Q^{-1}RQ,
\]

where \( Q \) are the orthonormal eigenvectors of \( R \), which implies that \( Q^{-1} = Q^T \). We may define a new basis to represent the \( i \)th observation in the original data, \( x_i \), as

\[
\eta_i = Q^T x_i.
\]

Equation 3 defines the principal components (PCs) of the data represented by \( X \). The full set of PCs exactly reproduces all observations in the original data, by definition. The PCA analysis recasts the observations in a rotated basis that has some desirable properties:

- The basis is orthonormal.
- The basis optimally represents the variance in the original data.
- The eigenvalues indicate which basis vectors (represented by the orthonormal eigenvectors) are of greatest importance in representing the data.
- PCA is a linear transformation. This is useful since if we have governing equations for the original basis, we may easily obtain governing equations in the new (PC) basis. This will be discussed further in §2.2.1.

\[
\text{Here we have assumed that the data is centered, \textit{i.e.} that its mean is zero. Data may be centered easily by applying a linear transformation if it is not centered.}
\]
The real utility in PCA comes by exploiting the property that PCA maximizes the variance in each PC direction. This means that the rotated coordinate system has the property that the first dimension (corresponding to the largest eigenvalue), is selected to best represent the variance in the data. Subsequent directions each represent the next-largest variance in the data. Therefore, retaining only a truncated basis, \( n_\eta \ll n \) can often approximate the original data remarkably well. We define a transformation matrix \( A \) as a rank-deficient subset of the \( Q \) matrix with \( n_\eta \) rows and \( n \) columns. The rows of \( A \) correspond to the rows of \( Q \) with the \( n_\eta \) largest eigenvalues. We may then approximate \( x \) as

\[
x \approx A \eta.
\]

Likewise, any function of \( x \) may be approximated by \( f(x) \approx f(A\eta) \).

2.2 Application to Combustion

In the context of combustion applications, \( x = [T, p, Y_1, Y_2, \ldots, Y_{n_s-1}]^T \) is the \((n_s+1)\)-dimensional natural basis for describing thermochemical systems. Performing a PCA on this set of variables yields a new \((n_s+1)\)-dimensional basis, \( \eta \), which is rotation of the original basis. Retaining the \( n_\eta \ll (n_s+1) \) PCs with the largest eigenvalues defines a basis for a \( n_\eta \)-dimensional parameterization of the thermochemical state of the system.

2.2.1 Transport Equations for Principal Components

The transport equations for a set of variables \( \Phi = [T, p, Y_1, Y_2, \ldots, Y_{n_s-1}]^T \) may be written as

\[
\frac{D(\Phi)}{Dt} = -\nabla \cdot (j_\Phi) + (s_\Phi),
\]

where \( \frac{D}{Dt} = \rho \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \) is the material derivative operator, \( \mathbf{u} \) is the mass-averaged velocity of the system, \( j_\Phi \) is the mass-diffusive flux of \( \Phi \) relative to the mass-averaged velocity, and \( s_\Phi \) is the volumetric rate of production of \( \Phi \). Since \( A \) is constant in space and time, and PCA represents a linear transformation, we may

\(^2\)Note that the transport equation for temperature has a fairly lengthy source term, involving terms such as \( \frac{\partial p}{\partial t} \), chemical reaction terms, viscous heating terms, etc. Its diffusive term is also lengthy, involving species diffusion, Fourier diffusion, and thermal diffusion.
multiply (5) by $A^T$, with $X = \Phi$, to find transport equations for the PCs,

$$\frac{D}{Dt}(\eta) = -\nabla \cdot [A]^T (j_\Phi) + [A]^T (s_\Phi), \quad (6)$$

$$= -\nabla \cdot (j_\eta) + (s_\eta) \quad (7)$$

or, for the $i^{th}$ PC,

$$\frac{D\eta_i}{Dt} = -\nabla \cdot \sum_{j=1}^{n_s} A_{ji} j_j + \sum_{j=1}^{n_s} A_{ji} s_j, \quad (8)$$

$$= -\nabla \cdot j_{\eta_i} + s_{\eta_i}. \quad (9)$$

While the diffusive term in (6) is not particularly problematic from a modeling perspective, the source term is. We separate this into two issues:

1. Can the PCA represent the source terms of the PCs with reasonable accuracy?

2. How do we account for the effects of turbulent mixing (and the implied subgrid heterogeneity) on $s_\eta$?

In this paper, we consider only the first issue through an $a$ priori analysis of DNS data. This provides an assessment of the potential for this modeling strategy. After demonstrating feasibility, future work will focus on $a$ posteriori assessment of the strategy. Future work will address the second issue.

Considering equation (6), we quickly realize that the source terms of temperature, pressure, and all species contribute to the source term for a PC, (assuming nonzero $A_{ij}$)\(^3\). Therefore, the truncated basis (where $n_\eta \ll n_s + 1$) must not only parameterize the thermochemical state but the source terms as well.

2.3 PCA Modeling Approach

A complete PCA modeling approach requires several ingredients. First, the PCs must be selected using the procedure outlined in §2.1. This requires high-fidelity data, generated via a technique such as DNS or ODT simulation. Once the PCs are selected, transport equations may be derived for each PC as described in §5.

Another key ingredient to a complete PCA modeling approach is defining the initial conditions (ICs) and

\(^3\)In the absence of strong pressure gradients (e.g. shock waves), the coefficient scaling pressure in the PCA will be exceedingly small, as expected since the thermochemical state is relatively insensitive to small pressure changes.
boundary conditions (BCs) on the PCs. The ICs and BCs may be obtained using the transformation matrix \( A \). For Dirichlet conditions on all variables \( \Phi \), we obtain Dirichlet conditions on the PCs, \( \eta \) (ICs are analogously defined).

\[
\Phi = \Phi^* \Rightarrow \eta = A^T \Phi^*
\]

Likewise, Neumann conditions on \( x \) yield Neumann conditions on \( \eta \),

\[
\mathbf{n} \cdot \nabla \Phi = \Theta \Rightarrow \mathbf{n} \cdot \nabla \eta = A^T \Theta.
\]

Mixed conditions on \( x \), for example Neumann conditions on \( x_1 \) and Dirichlet conditions on \( x_2 - x_n \) yield Robin boundary conditions on \( \eta \). Therefore, boundary condition implementation may be slightly more involved, as many codes don’t support Robin conditions.

Diffusion terms in the transport equations for \( \eta \) require evaluation of the diffusive fluxes for each component of \( \Phi \). In turbulent flow calculations, the molecular diffusion term is typically augmented by a “turbulent diffusion” term arising from closure of the convective term. In many cases, and particularly at high Reynolds number, the molecular term is small relative to the turbulent diffusion term and is neglected. However, even when one wishes to retain the full description of molecular diffusion, the treatment with PCA is straightforward. First, \( \Phi \) is obtained from \( \eta \). Next, the diffusive terms for \( \Phi \) are constructed, and the diffusive fluxes for the PCs are calculated as \((j_{\eta}) = [A]^T (j_{\Phi})\).

Source terms for the PCs, \( s_{\eta} \), can be parameterized by \( \eta \) and tabulated \textit{a priori} to avoid run-time calculation. Accurate representation of \( s_{\eta} \) is crucial to this modeling strategy, and will be addressed further in §3.

In turbulent flow, turbulent closure is required for the source terms (convective terms can be treated in a manner consistent with other equations). Depending on the number of PCs, and their correlation, one may consider presumed PDF approaches, quadrature method of moments [12, 28, 29], or transported PDF approaches [10, 11]. This paper will not address specific issues related to turbulent closure.

2.4 Obtaining the Principal Components

As discussed above, a successful application of PCA as a modeling approach requires parameterization of the source terms for each PC. This, in turn, requires that the data that PCA is applied to must have source
terms for all \( \Phi \). This is currently impossible to obtain from experimental data. Therefore, this modeling approach will require availability of computational data generated from reliable chemical mechanisms using methods such as DNS or ODT. Furthermore, the reliability of PCA as a modeling approach will also hinge on the relative invariance of PCs from one dataset to another that is nearby in parameter space.

3 Results

In this section we present results of PCA applied to DNS of CO/H\(_2\) combustion. Details of the DNS dataset, and comparison with several existing modeling approaches may be found elsewhere [4, 25]. In addition to showing results for PCA, results for traditional parameterizations such as mixture fraction and scalar dissipation rate \((f, \chi)\) are shown for reference.

3.1 Parameterization of State Variables

Here we present results for a few state variables to demonstrate typical results of a PCA. We choose to first condition the data on mixture fraction, \( f \), since this is a convenient variable to “force” as the first component. For reference, we compare parameterization by \((f, \chi)\). Figures 1(a) and 1(b) show parameterization of the temperature. The PCA parameterization is slightly better in this case, but not markedly better. Parameterization of minor species tends to be superior by the PCA than by \((f, \chi)\) parameterizations, as shown in figures 2(a) and 2(b), which show results for H radicals.

3.2 Parameterization of Source Terms

As discussed in §2.2.1, the PCs are not conserved variables, and their source terms must also be parameterized. In this section we explore the ability of PCA to parameterize source terms. Figures 3(a) and 3(b) show the scaled chemical source term in the temperature equation,

\[
\frac{1}{\gamma_T \rho c_p} \sum_{i=1}^{n_r} \Delta H_i^r s_i,
\]

where \( \gamma_T \) is the (constant) scaling factor for temperature used in the PCA, \( \Delta H_i^r \) is the enthalpy of reaction for reaction \( i \) and \( s_i \) is the reaction rate of reaction \( i \). Comparing Figures 3(a) and 3(b), it is clear that the
Figure 1: Temperature vs. (a) scalar dissipation rate, $\chi$ and (b) the first principle component, $\eta_1$. Points represent DNS data conditioned on $f_{st}$; red lines represent conditional mean, $T_h|f_{st}$.

Figure 2: H mass fraction vs. (a) scalar dissipation rate, $\chi$ and (b) the first principle component, $\eta_1$. Points represent DNS data conditioned on $f_{st}$; red lines represent conditional mean, $Y_h|f_{st}$, and black lines represent the conditional standard deviation, $\sigma_H|f_{st}$. 
PCA parameterization is dramatically better than the \((f, \chi)\) parameterization, with the PCA parameterizing the temperature source term remarkably well.

Figures 4(a) and 4(b) show results for parameterization of the \(\text{H}_2\text{O}_2\) source term, which is the dominant contribution to \(s_{\eta_1}\). Figure 5(b) shows parameterization of the full source term for \(\eta_1\).

3.3 Global versus Local PCA

The results presented thus far have been at the stoichiometric mixture fraction. One may consider whether a PCA performed at \(f_{st}\) is applicable at other \(f\). If the PCA is highly dependent on mixture fraction, then one of two options must be considered

1. Eliminate the mixture fraction as a parameter and seek a global PCA on the entire dataset.

2. Derive transport equations for the conditional PCs (conditioned on mixture fraction). These equations would have exchange terms representing transport in mixture fraction space.

If the PCA obtained at \(f_{st}\) reasonably represents the data at other \(f\), then the transport equations derived in §2.2.1 may be used directly.

Figures 5(a), 5(b) and 5(c) show parameterizations of the source term for \(\eta_1\) at lean \((f = 0.1)\), stoichiometric \((f = 0.4375)\), and rich \((f = 0.6)\), respectively. It appears that a PCA performed at stoichiometric conditions may be applied over the range of mixture fraction with reasonable accuracy.
Figure 3: Temperature source term vs. (a) scalar dissipation rate, $\chi$ and (b) the first principle component, $\eta_1$.

Figure 4: H$_2$O$_2$ source term at stoichiometric conditions as a function of (a) scalar dissipation rate and (b) the first principal component.
Figure 5: Parameterization of $s_{\eta_1}$ by $\eta_1$. Both plots use the same basis, which was obtained at $f_{st}$. 
4 Conclusions

A modeling approach based on PCA has been proposed and tested using DNS data. This modeling approach is complete, with the exception of a turbulent closure model. The model is based on a rotation of the thermochemical state basis from one based on temperature, pressure, and \( n_s - 1 \) species mass fractions to one which maximizes the representation of the variance in the data. PCA achieves this rotation of basis, and has the convenient property that a truncated basis provides an optimal representation of the data. Implementation of the model requires transport equations for the principle components, which are reacting scalars.

Results from \textit{a priori} analysis of this approach using DNS data show great promise. State variables and source terms both are parameterized reasonably well by a two-parameter model. Results also indicate a uniformly better representation of the DNS data using a PCA-based approach over a mixture fraction and dissipation rate parameterization, as is commonly used in combustion modeling.

On balance, we may summarize the strengths and weaknesses of a PCA modeling approach as follows:
**Strengths** | **Weaknesses**
---|---
Allows *a priori* definition of an error tolerance. & PCs are not conserved scalars. Thus, a valid PCA modeling strategy must accurately parameterize scalar source terms.
The number of required parameters naturally follows from the prescribed error tolerance. & PCA requires high-fidelity datasets with complete thermochemical state information, including source terms. This eliminates all currently available experimental data.
Given a fixed number of parameters, PCA provides an “optimal” representation of the data. &

PCA is a linear transformation, which allows simple definition of transport equations, initial and boundary conditions for the PCs from the original basis: $T_i, p, Y_i$. & Data must be available at “relevant” conditions such that the thermochemical state statistics are representative of those that will exist in the target system.
The PCA model is extracted directly from “relevant” datasets. & Model implementation may require some code modification to facilitate proper boundary and initial condition treatment.
Allows for detailed treatment of diffusion in a rigorous manner, *i.e.* without assuming unity Lewis number. &

We anticipate that one-dimensional turbulence and direct numerical simulation will be the primary tools to generate data for PCA.

The findings presented herein suggest a great potential for a PCA-based modeling approach to be used in practical hydrocarbon combustion modeling. Results indicate that a two-parameter PCA representation of DNS data can represent thermo-chemical state variables with superior accuracy to traditional parameterizations such as mixture fraction and scalar dissipation rate. Future work will focus on examining the feasibility of PCA with various fuels and under varying conditions including extinction. Also, *a posteriori* tests will be conducted to determine the effect of nonlinear propagation of errors in source term parameterization.

**References**


