Fuel 106 (2013) 344–356

Contents lists available at SciVerse ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

High-dimensional, unsupervised cell clustering for computationally efficient engine simulations with detailed combustion chemistry

Federico Perini *

Dipartimento di Ingegneria Meccanica e Civile, Università di Modena e Reggio Emilia, I-41125 Modena, Italy

HIGHLIGHTS

- Reduction of the computational efforts due to chemistry in CFD modeling.
- A new bounding-box-constrained *k*-means clustering algorithm is presented.
- Unsupervised cluster center initialisation based on grid-like domain decomposition.
- Local clustering accuracy is guaranteed by the bounding-box constraint.
- 3–4 Times reduction in overall CPU times is achieved.

ARTICLE INFO

Article history:
Received 22 March 2012
Received in revised form 5 November 2012
Accepted 5 November 2012
Available online 28 November 2012

Keywords:
Cell clustering
Detailed chemistry
High-dimensional space
k-Means
Internal combustion engines

ABSTRACT

A novel approach for computationally efficient clustering of chemically reacting environments with similar reactive conditions is presented, and applied to internal combustion engine simulations. The methodology relies on a high-dimensional representation of the chemical state space, where the independent variables (i.e. temperature and species mass fractions) are normalized over the whole dataset space. An efficient bounding-box-constrained *k*-means algorithm has been developed and used for obtaining optimal clustering of the dataset points in the high-dimensional domain box with maximum computational accuracy, and with no need to iterate the algorithm in order to identify the desired number of clusters. The procedure has been applied to diesel engine simulations carried out with a custom version the KIVA4 code, provided with detailed chemistry capability. Here, the cells of the computational grid are clustered at each time step, in order to reduce the computational time needed by the integration of the chemistry ODE system. After the integration, the changes in species mass fractions of the clusters are redistributed to the cells accordingly. The numerical results, tested over a variety of engine conditions featuring both single- and multiple-pulse injection operation with fuel being injected at 50° BTDC allowed significant computational time savings of the order of 3–4 times, showing the accuracy of the high-dimensional clustering approach in catching the variety of reactive conditions within the combustion chamber.

1. Introduction

The advancements in computational resources have allowed, in recent years, combustion research to achieve quantitative predictive capabilities thanks to the adoption of chemical kinetics models in conjunction with multidimensional simulations [1]; the study of the interactions of physical and chemical processes, especially at the smallest scales [2], as well as the development of novel combustion concepts, which are able to exploit the variety of reactive conditions a fuel–oxidizer system can account for [3–6], is however urging the need for using comprehensive, detailed reaction models [7], which can be made of thousands species and more than 10,000 reactions [8–13]. In order to cope with such mechanism dimensions, a number of approaches have been developed with the aim of either achieving suitable computational time scaling with the mechanism dimension, or of avoiding unnecessary computations. For example, a fast and practical approach is to provide a reduced mechanism subset, by identifying the active species as the ones among which strong reaction exchanges occur [14–18]; other approaches simplify the computational effort by dividing reactions into fast and slow groups [19–22], or implement storage–retrieval techniques for reducing the number of chemistry ODE system integrations by adopting suitable approximations to high-dimensional functions [23–26]. Research is also active in identifying proper ODE integration techniques computationally

* Tel.: +39 0592056100; fax: +39 0592056126.
E-mail address: federico.perini@unimore.it

0016-2361/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fuel.2012.11.015

suitable for stiff chemistry problems: as acknowledged [7], the dominating computational cost when integrating large reaction mechanism is due to factorization of the Jacobian matrix, if implicit or semi-implicit integration methods are adopted; for this reason most approaches aim at either adopting matrix-free integration methods or at reducing the Jacobian-related cost by either expressing it in a sparse format, or by introducing simplified, approximate Jacobian formulations [27–34].

The approach developed in the present work belongs to a number of studies which aim at reducing the overall computational cost due to detailed chemistry computations in multidimensional computational fluid dynamics (CFD) simulations of practical combustion systems – such as internal combustion engines –, by actually reducing the number of reacting environments the detailed chemistry ODE system needs to be integrated into. As Babajimopoulos et al. firstly showed [35], significant computational time savings can be achieved when the instantaneous chemical composition within a multidimensional domain fulfills a pattern-like structure, for instance a fuel–oxidizer charge stratification, such as that occurring in typical homogeneous-charge compression ignition (HCCI) engine combustion. The multidimensional domain can thus be represented as a multi-zone environment, where each zone owns defined temperature and mixture equivalence ratio, so that only one detailed chemistry ODE system needs to be solved per zone in each global advancement time-step, and then the results of the integration can be backward remapped to the original cells, proportionally to their initial compositions. This concept has been generalized by Liang et al. [36], where the recognition of the homogeneous zones has been set up as an evolutionary clustering problem, where the independent variables are cell's temperature and local mixture equivalence ratio. Barths et al. [37] applied a similar approach as a two-way coupling between the actual CFD simulation and a zero-dimensional multi-zone environment made up of a limited number of homogeneous, chemically reacting zones. Shi et al. [38] have shown that significant computational time savings can be achieved also when the clustering procedure is done according to a cell proximity criterion. Another approach has been chosen by Goldin et al. [39], where the whole species and energy composition space has been adopted for describing each computational cell, and applied to 1D and 2D laminar flame computations.

The main idea underlying the present work is that the robustness and the potential of the partitioning approach mainly rely on the smartness of the cell clustering algorithm, intended as its ability to:

- catch the variety of reactive conditions in the multidimensional domain, that cannot be simplified into a unique parameter, and that are usually ruled by species associated to fast timescales;
- automatically identify, at each timestep, the optimal number of clusters;
- minimize the inner inhomogeneity of the agglomerates by covering the whole domain of compositions in a sparse way;
- do not introduce significant computational overhead and thus be suitable for large-scale parallel computations.

The paper thus describes the study of a complete, unsupervised high-dimensional clustering approach, validated and applied for internal combustion engine simulations with detailed chemistry; its presentation is structured as follows. In Section 2, all the aspects of the approach developed for chemistry-based cell clustering are presented: first of all, the clustering problem is defined by introducing the dataset representation, the relationships between chemically-reacting CFD cells and their images in the clustering space, and suitable distance metrics; then, an unsupervised initialisation procedure is reported, which sets the initial cluster partition up as a grid-like structure. A full description and validation of a novel crisp clustering algorithm of the *k*-means class is then presented; the algorithm, named 'bounding-box-constrained' (BBC) *k*-means, is tailored for clustering chemistry-based datasets which typically model thermodynamic systems, whose behavior is strongly non-linear with respect to the variables' values. Finally, in Section 3 the implementation of the proposed procedure into a customized version of the KIVA-4 code [40], provided with detailed chemistry capability, is presented and discussed focusing on the accuracy and on the computational time savings allowed by adoption of this procedure with respect to a standard solution where a chemistry ODE system is integrated in each cell of the domain. The analysis shows that the procedure proved to be robust on a variety of diesel engine cases involving different combustion modes, and that overall speed-ups of at least three times have been achieved for all the tested cases, with almost negligible overhead introduced by the clustering and remapping procedure.

2. Unsupervised high-dimensional clustering (UHDC)

As acknowledged [36,39], the issue of reducing the overall computational demands due to solving chemical kinetics in multidimensional simulations can be addressed as a three-step procedure: (1) optimal clustering of the chemically reactive cell in the computational domain into a number of chemically homogeneous environments; (2) solution of the chemistry ODE system in each cluster, yielding internal energy and species mass fractions source terms at the cluster level; (3) mass-conservating redistribution of the cluster-level time-integrated quantities to the each single cell.

As far as the clustering problem is concerned, two main approaches are possible: crisp clustering algorithms such as the *k*-means [41] consider cluster centers as the average values among their own points; this approach is particularly suitable for grouping CFD domain cells on a chemistry basis thanks to its limited computational demands, where the most time-consuming task is that required by evaluation of point-to-center distances; its intrinsic averaging however tends to deteriorate the diversity of the initial integration conditions of the cells, and needs to be restored by a specific backward remapping procedures. On the other hand, in fuzzy clustering algorithms such as the common fuzzy *c*-means [42] each point belongs to each cluster center to a certain degree of membership; thus, cluster centers usually follow a more disperse distribution in the domain, and every single point can be viewed as the weighted average of all the cluster centers. From the chemical kinetics point of view, this latest aspect could be beneficial, as each cell in the computational grid may be represented as the weighted average of a smaller number of sparse and far-away-from-each-other reactive conditions. However, the strong non-linearity and anisotropic behavior of chemical kinetics in combustion systems make this approach less suitable: the membership exponent approach used to quantify membership values to cluster centers would also require that species mass fraction ranges would be properly scaled so that each problem variable would have the same degree of importance in contributing to the membership function. Furthermore, fuzzy clustering algorithms typically require significantly higher computational efforts, that would render their adoption useful only in presence of huge computational domains.

2.1. High-dimensional clustering problem formulation

The chemical kinetics initial value problems treated in this work are used in order to compute species mass fractions and internal energy source terms as part of the operator-splitting context adopted in the KIVA family of codes [40,43]. The dimensions of the chemistry integration problem are given by the number of the