Aerospace Science and Technology 52 (2016) 115–128

Contents lists available at ScienceDirect

Aerospace Science and Technology

www.elsevier.com/locate/aescte

Aerodynamic database reconstruction via gappy high order singular value decomposition

Ana I. Moreno [1], Artur A. Jarzabek [1], José M. Perales [2], José M. Vega [*,3]

E.T.S.I. Aeronáutica y del Espacio, Universidad Politécnica de Madrid, Plaza Cardenal Cisneros, 3, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 16 October 2015
Received in revised form 23 December 2015
Accepted 5 February 2016
Available online 11 February 2016

Keywords:
Multidimensional databases
Aerodynamic databases
CFD databases
Gappy reconstruction
High order singular value decomposition
Linear algebra

ABSTRACT

A method based on an iterative application of high order singular value decomposition is derived for the reconstruction of missing data in multidimensional databases. The method is inspired by a seminal gappy reconstruction method for two-dimensional databases invented by Everson and Sirovich (1995) [20] and improved by Beckers and Rixen (2003) [21] and Venturi and Karniadakis (2004) [22]. In addition, the method is adapted to treat both noisy and structured-but-non-rectangular databases.

The method is calibrated and illustrated using a three-dimensional toy model database that is obtained by discretizing a transcendental function. The performance of the method is tested on three aerodynamic databases for the flow past a wing, one obtained by a semi-analytical method, and two resulting from computational fluid dynamics.

1. Introduction

Multidimensional databases are used in the aerospace industry and laboratories for many purposes related to, e.g., design, certification, flight control, and flight simulators. Increasing the quality of these databases involves various tasks, such as database compression, reconstruction of lost data, and error filtering. The various dimensions in these databases are usually associated with spatial and temporal coordinates and/or parameters. For instance, both computational fluid dynamics (CFD) simulations and wind tunnel campaigns produce huge databases. Such databases are *strongly correlated* both in the physical space and time dimensions and in the parameter space, due to *redundancies* produced by physical laws, such as those associated with the underlying Navier–Stokes equations. Because of these redundancies, the *actual information* contained in the database is much smaller than the *database size*, which could become huge due to the curse of dimensionality [1] as the number of database dimensions increases. Identifying the actual information is the main ingredient to repair and/or complete the database.

Multidimensional databases can always be considered as families of vectors, isolating one of the dimensions and folding the remaining dimensions together into a single dimension; this means that the number of dimensions can be reduced to two. For instance, three-dimensional databases resulting from time-dependent 2D flows can be treated as collections of vectors, each giving a snapshot at a particular value of time (see, e.g., [2]). Truncated *proper orthogonal decomposition* (POD), which is a variant of *singular value decomposition* (SVD), takes advantage of the redundancies among the vectors and produces a lower dimensional approximation. This is useful in a variety of contexts, such as database generation [3,4], derivation of reduced order models [5–7], compression [8,9], and image processing [10].

However, the two-dimensional reduction looses the individual redundancies along the dimensions that have been folded together. Such redundancies can be accounted for using *high order singular value decomposition* (HOSVD). This is an extension to tensors [11, 12] of standard SVD [13], which only applies to matrices. The advantage of using HOSVD, instead of SVD, is clearly appreciated in both database generation [14] and compression [15]. In the latter case, the compression factor increases exponentially with the number of the database dimensions whose redundancies are accounted for. Thus, identifying as many redundancies as possible in the database alleviates the curse of dimensionality. In addition, following a former idea [3,4] in the context of standard POD (for two-dimensional databases), HOSVD can be combined with interpolation [15] to calculate the database elements at intermediate values of the physical quantities accounted for in the various dimensions. Such combination has been successfully used to speed

* Corresponding author.
E-mail address: josemanuel.vega@upm.es (J.M. Vega).

1 Aeronautical Engineer and PhD. Student, Department of Aircraft and Spacecraft.
2 Professor of Aerospace Engineering, Department of Aircraft and Spacecraft.
3 Professor of Applied Mathematics, Department of Applied Mathematics.

http://dx.doi.org/10.1016/j.ast.2016.02.004

up calculations in both the real time control of reciprocating engines [16] and the conceptual design of aircraft components [17].

Gappy reconstruction consists in reconstructing missing data, which is needed in, e.g., experimental wind tunnel databases due to complete obstruction of sensors or the impossibility of locating sensors in a part of the physical domain. This task has been addressed in the literature in various contexts, since the pioneering work by Yates [18], who applied least squares to fill in missing data. Gappy reconstruction methods are either statistical [19] or deterministic [2–4,20–23]. Concentrating on the latter, they aim at reconstructing one or more gappy vectors. It may happen that:

- A low dimensional approximation of the exact database is known beforehand. The gappy vectors are reconstructed one-by-one, imposing that they be close to the low dimensional approximation.
- No a priori knowledge about the database is available.

Everson and Sirovich [20] (hereafter ES) dealt with both situations, since they developed a two-step method, by:

(i) First reconstructing some of the gappy vectors and calculating a low dimensional approximation of the database. To perform this step, they developed a genuine gappy reconstruction method.
(ii) Then using the low dimensional approximation calculated in step 1 to reconstruct any new gappy vector. The method to perform this step could be seen as an inverse design method.

Bui-Thanh et al. [4] considered the methods introduced in ES to perform steps (i) and (ii) as independent methods, and applied them to some databases resulting from the inviscid aerodynamic flow around an airfoil.

Here, we concentrate in the ES *gappy-reconstruction method* to perform the step (i), which is based on an iterative algorithm. At each iteration step, POD is applied to the database in which the result of the former step is used as a new guess for gappy data, while the original data are maintained at non-gappy positions. This method was seminal, but exhibits the difficulty that the final error strongly depends on both the initial guess and the number of retained POD-modes. The difficulty was solved independently by Beckers and Rixen [21] and Venturi and Karniadakis [22] introducing a second iteration on the number of retained modes. The modified method will be referred to as the ES-BR-VK method and yields fairly robust approximations in cases in which the standard ES gappy-reconstruction method fails. The improved method was extensively tested by Gunes et al. [2] in families of flow snapshots calculated using a CFD solver for specific values of time in the 2D unsteady flow around a cylinder; a comparison was also made with Kriging interpolation [24]. Among other interesting conclusions, Gunes et al. [2] note that the ES-BR-VK method generally works better than Kriging for small gappyness/large temporal resolution, while Kriging is preferred when either the gappyness level is high or the temporal resolution is small. In particular, the ES-BR-VK method does not produce any results (and Kriging interpolation was proposed instead) when gappy data include a fixed spatial region (*black box*) for all values of time, which is needed to reconstruct experimental data in spatial regions that are not accessible to measurements. See also [25] for further applications of the ES-BR-VK method.

As formulated, the ES-BR-VK method applies to (one-dimensional) families of vectors, which can also be seen as matrices. The first goal of the present paper is to extend the method to multidimensional databases using HOSVD. The resulting method (which will be referred to as *gappy-HOSVD* method) works fairly well, both when gappy elements are randomly located along the database and when they are concentrated, and also when gappyness is located black-box-like. In addition, the method will also be adapted to treat noisy databases, which are frequent in industry and, as already noticed in [22], produce a non-monotone behavior of the ES-BR-VK method. In other words, the performance of the gappy reconstruction method worsens as the number of retained modes increases beyond some point, which requires to select the appropriate number of modes (not an easy task). A method will be presented in this paper that is synergic with HOSVD and selects well the 'optimal' numbers of modes along the various database dimensions. The resulting method will be referred to as the *gappy-noisy-HOSVD* method.

On the other hand, the standard application of POD, SVD, and HOSVD can only be made to structured, rectangular/hyper-rectangular databases. Unfortunately, these databases may either be not possible or not convenient because some of the database positions needed to complete the hyper-rectangle are either non-physical (as in, e.g., the flow around an obstacle when Cartesian coordinates are used) or of no interest (e.g., some combinations of the Mach number and the angle of attack in the flow around an aircraft wing). The idea to obtain a HOSVD description in a non-rectangular mesh is to apply the gappy-HOSVD method in an augmented hyper-rectangular database considering as gappy elements those elements that have been added to the original (non-rectangular) database. In other words, the method will fill in the spurious added data in such a way that they are consistent (along the various dimensions) with the original database elements.

Summarizing, the object of this paper is two-fold, since we shall (i) extend the ES-BR-VK to treat multidimensional databases and (ii) adapt the method to cope with noisy and non-rectangular databases. We note that efficient 'universal methods' to repair/complete industrial databases are not to be expected. This is because these tasks most likely depend both on the structure of the underlying redundancies and on the magnitude of background errors (e.g., CFD-generated errors), which are unavoidable in practice. Thus, any method (in particular, the methods developed below) that is intended to work reasonably well may involve a few parameters that should be calibrated for each specific application (e.g., aerodynamics). The methods developed below will be robust, namely the sensibility of the results on the tunable parameters will be small. Also, we note that no assumptions will be made on the nature of the redundancies the methods are based upon. The only requirement is that such redundancies be present. This generality is inherited from the original ES method and makes a difference with gappy reconstruction/interpolation methods based on, e.g., polynomial or radial basis functions, and also with traditional error filtering methods that assume specific error properties, such as Gaussian distributions or zero means.

The methods developed in this paper will be first tested using a *toy model database*, which is obtained discretizing a transcendental function (see Fig. 1). This database is 'clean', which helps to separate difficulties, allowing for both identifying the main requirements of the methods and testing their performance. The methods will also be applied to three aerodynamic databases containing the pressure distribution around a wing for various angle of attack values. The first database is obtained by a semi-empirical method and the remaining two, by using CFD, which inheritly contains some level of noise. All applications will be performed using MATLAB on a desktop PC, with a 3.40 GHz processor, which gives results in a few CPU seconds.

The remaining of the paper is organized as follows. The above mentioned toy model database is described in §2. The basic ability of POD/SVD and HOSVD to compress databases and filtering errors out will be recalled in §3, where notation will be established. The iterative ES method and the doubly iterative BR-VK extension will be revisited in §4, where the extended gappy-HOSVD and gappy-