J. Fluid Mech. (2004), *vol.* 519, *pp.* 315–336. © 2004 Cambridge University Press
DOI: 10.1017/S0022112004001338 Printed in the United Kingdom

Gappy data and reconstruction procedures for flow past a cylinder

By DANIELE VENTURI[1] AND GEORGE EM KARNIADAKIS[2]†

[1]DIENCA, Department of Energy, Nuclear and Environmental Engineering, University of Bologna, 40136, Bologna, Italy

[2]Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

(Received 27 October 2003 and in revised form 14 July 2004)

We investigate the possibility of using proper orthogonal decomposition (POD) in reconstructing complete flow fields from gappy data. The incomplete fields are created from DNS snapshots of flow past a circular cylinder by randomly ommiting data points. We first examine the effectiveness of an existing method and subsequently introduce modifications that make the method robust and lead to the maximum possible resolution at a certain level of spatio-temporal gappiness. We simulate three levels of gappiness at approximately 20%, 50% and 80% in order to investigate the limits of applicability of the new procedure. We find that for the two lower levels of gappiness both the temporal and spatial POD modes can be recovered accurately leading to a very accurate representation of the velocity field. The resulting resolution is improved by more than five times compared to the existing method. However, for 80% gappiness only a few temporal modes are captured accurately while the corresponding spatial modes are noisy. We explain this breakdown of the method in terms of a simple perturbation analysis. This new methodology can be a building block in an effort to develop effective data assimilation techniques in fluid mechanics applications.

1. Introduction

We address in this work the issue of data assimilation in fluid mechanics, using the two-dimensional flow past a circular cylinder as a prototype problem. While data assimilation is routinely used in climate and ocean modelling, this is not the case with more classical fluid mechanics applications. However, the recent rapid developments in quantitative imaging techniques, e.g. particle image velocimetry (PIV) and magnetic resonance imaging (MRI), and the simultaneous advances in large-scale simulation open the possibility for integrating seamlessly simulation and experiment. In the new computational paradigm, simulation and experiment will become a symbiotic feedback system for diverse flow-based applications, e.g. controlled flow–structure interactions, smart combustion systems, active microfluidic networks, etc.

The type of data assimilation most appropriate for fluid mechanics will depend on the specific application. In a typical case, the initial flow simulation will be based on parameters and spatio-temporal modes extracted from a single state or an ensemble of states of the experiment with subsequent continuous injection of new data into

† Author to whom correspondence should be addressed: gk@dam.brown.edu

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

the ongoing simulation. In some cases, the output of the simulation may be used to directly influence the experimental input, e.g. to steer the measurements to the proper location. In other cases, there may be only a feed-forward input from the experiment to the simulation for predicting accurately the evolution of a given experimental state.

At the heart of this integration is the ability to reconstruct flow fields from a finite number of experimental observations at a controlled level of accuracy. Even for the relatively simple experiment on flow past a cylinder, information due to 'shadowing' (i.e. obstructed view) and proximity to the boundaries may be missing or the frequency of the measurements may be below the required temporal resolution of the corresponding simulation. Therefore, we have to work with gappy data where the spatio-temporal regions of 'missingness' are known in advance or where missingness occurs at random (MAR). This problem is of course not new and researchers have been working on it for many decades. One of the first effective approaches was developed by Yates (1933) who proposed filling-in missing data with least-squares estimates. This may at first seem to be circular and of little practical use but it turns out to be quite effective in exploiting redundancy in the available measurements. Several statistical approaches for data imput building on the original ideas of Yates are presented in Little & Rubin (2002). Among them, local Kriging is an effective statistical method combined with least squares and has been used with success in geology and other fields to interpolate spatial data. Unlike other estimation procedures, Kriging provides a measure of the error and associated confidence in the estimates (Stein 1999). In the current paper, we will compare local Kriging against the proposed new method for reconstructing gappy fields.

In this work we follow a deterministic approach based on proper orthogonal decomposition (POD) combined with the least-squares approach. This method was first proposed by Everson & Sirovich (1995) for image reconstruction and it has been used with success in Tan, Willcox & Damodaran (2003) for steady flow past an airfoil. The method breakes down if there are any time instants for which the data are missing everywhere (i.e. absence of a snapshot) or if there are any spatial subregions for which the data are missing at all times. Here we apply this method to unsteady flow past a circular cylinder and investigate ways of improving its performance. The fundamental question we address is: What is the maximum possible resolution that can be achieved given a certain degree of gappiness in the data.

This is a rather complex issue as it could also depend on the type of missingness, for example close to the cylinder boundaries the relative importance of missing data is higher. However, POD provides the best representation (in the average sense) for a given field and produces a hierarchical set of spatio-temporal scales. This, in turn, helps in establishing accuracy criteria and thus it addresses the question of maximum possible resolution. Indeed, in the current work we have extended the Everson–Sirovich approach in a way that leads to the best possible reconstruction independent of the initial guess for filling in the missing data. In addition, we have developed and verified a robust criterion for selecting the optimum number of modes for reconstruction.

We have selected the flow past a cylinder as a test problem due to the previous experience with this flow that shows that a low-dimensional representation indeed exists (Ma, Karamanos & Karniadakis 2000; Ma *et al.* 2003; Cao & Aubry 1993; Deane *et al.* 1991). This is true both in two and three dimensions but in this paper we concentrate on reconstructing two-dimensional gappy fields. We have investigated two states at Reynolds number 100 and 500 (based on the cylinder diameter) and three degrees of gappiness of approximately 20%, 50% and at 80%.

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

The paper is organized as follows. In § 2 we review the method proposed in Everson & Sirovich (1995) and we present a modification that makes the method robust and significantly more accurate. We also address the question of temporal versus spatial resolution, which becomes particularly important in gappy fields. In § 3 we present the main results, and finally in § 4 we discuss some of the open issues of the new methodology in the context of data assimilation in fluid mechanics.

2. POD theory and gappy data

In this section we first present the algorithm of Everson & Sirovich and subsequently we present an extended procedure that eliminates dependence on the initial condition. It also leads to a significant increase in resolution given a specific set of gappy data. We then address the question of *resolution equivalence* in space–time in the case of incomplete data sets.

2.1. *Initial-condition-independent method*

Let us consider a (d-dimensional) vector flow field $\boldsymbol{u}(\boldsymbol{x}, t) \in (L^2(\Omega \times T))^d$ defined in the spatio-temporal domain $(\Omega \times T)$. We assume that we have available a finite number N of snapshots of the flow field. We can then look for a *biorthogonal representation* of $\boldsymbol{u}(\boldsymbol{x}, t)$ in the form (Aubry, Guyonnet & Stone 1991; Aubry 1991)

$$\boldsymbol{u}(\boldsymbol{x}, t) = \sum_{k=1}^{N} \lambda_k \boldsymbol{\Phi}_k(\boldsymbol{x})\, \psi_k(t), \tag{2.1}$$

where $\boldsymbol{\Phi}_k(\boldsymbol{x})$ and $\psi_k(t)$ are the orthonormal spatial and temporal modes, respectively.

We would like to determine the unknown functions $\boldsymbol{\Phi}_k(\boldsymbol{x})$ and $\psi_k(t)$ by minimizing the functional (in the energy norm)

$$\begin{aligned} F[\boldsymbol{\Phi}_k, \psi_k] &= \left\| \boldsymbol{u}(\boldsymbol{x}, t) - \sum_{k=1}^{N} \lambda_k \boldsymbol{\Phi}_k(\boldsymbol{x}) \psi_k(t) \right\|_{L^2}^2 \\ &= \int_T \int_\Omega \left(\boldsymbol{u}(\boldsymbol{x}, t) - \sum_{k=1}^{N} \lambda_k \boldsymbol{\Phi}_k(\boldsymbol{x}) \psi_k(t) \right) \cdot \left(\boldsymbol{u}(\boldsymbol{x}, t) - \sum_{k=1}^{N} \lambda_k \boldsymbol{\Phi}_k(\boldsymbol{x}) \psi_k(t) \right) \mathrm{d}\boldsymbol{x}\, \mathrm{d}t \end{aligned} \tag{2.2}$$

with respect to an arbitrary variation of $\boldsymbol{\Phi}_k(\boldsymbol{x})$ and $\psi_k(t)$. This minimization leads to the Euler–Lagrange equations

$$\int_\Omega \boldsymbol{u}(\boldsymbol{x}, t) \cdot \boldsymbol{\Phi}_j(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \lambda_j \psi_j(t), \tag{2.3}$$

$$\int_T \boldsymbol{u}(\boldsymbol{x}, t) \psi_j(t)\, \mathrm{d}t = \lambda_j \boldsymbol{\Phi}_j(\boldsymbol{x}). \tag{2.4}$$

Following Aubry & Lima (1995*a*,*b*), Aubry *et al.* (1991) and Aubry, Guyonnet & Lima (1995) we define the linear integral operators U and U^*:

$$(U\boldsymbol{\Phi}_j)(t) = \int_\Omega \boldsymbol{u}(\boldsymbol{x}, t) \cdot \boldsymbol{\Phi}_j(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}, \tag{2.5}$$

$$(U^* \psi_j)(\boldsymbol{x}) = \int_T \boldsymbol{u}(\boldsymbol{x}, t) \psi_j(t)\, \mathrm{d}t, \tag{2.6}$$

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

where U^* is the adjoint of U. By inserting the biorthogonal decomposition of $\boldsymbol{u}(\boldsymbol{x}, t)$ from equation (2.1) in (2.3) and (2.4) we find

$$U\boldsymbol{\Phi}_j = \lambda_j \psi_j, \tag{2.7}$$

$$U^*\psi_j = \lambda_j \boldsymbol{\Phi}_j. \tag{2.8}$$

These are called *dispersion relations* and they provide the link between the spatial and the temporal evolution of the system. If we apply the operator U^* to (2.7) and the operator U to (2.8) we obtain the eigenvalue problems

$$U^*U\boldsymbol{\Phi}_j = \int_\Omega S(\boldsymbol{x}, \boldsymbol{x}')\boldsymbol{\Phi}_j(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' = \lambda_j^2 \boldsymbol{\Phi}_j, \tag{2.9}$$

$$UU^*\psi_j = \int_T T(t, t')\psi_j(t')\,\mathrm{d}t' = \lambda_j^2 \psi_j. \tag{2.10}$$

The operators U^*U and UU^* are symmetric and positive. The kernel of U^*U is the two-point *spatial correlation* while the kernel of UU^* is the two-point *temporal correlation*, given respectively by

$$S(\boldsymbol{x}, \boldsymbol{x}') = \int_T \boldsymbol{u}(\boldsymbol{x}, t)\cdot\boldsymbol{u}(\boldsymbol{x}', t)\,\mathrm{d}t \tag{2.11}$$

$$T(t, t') = \int_\Omega \boldsymbol{u}(\boldsymbol{x}, t)\cdot\boldsymbol{u}(\boldsymbol{x}, t')\,\mathrm{d}\boldsymbol{x}. \tag{2.12}$$

The symmetric version (see Aubry 1991) of the proper orthogonal decomposition (2.1) corresponds to the spectral analysis of the operator U and consists of square roots of proper values (eigenvalues) and proper functions (eigenfunctions) of the correlation operators U^*U and UU^*. We order the eigenvalues as

$$\lambda_1^2 \geqslant \lambda_2^2 \geqslant \cdots \lambda_N^2 \geqslant 0.$$

The above formulation assumes the completeness of the flow field in the spatio-temporal domain. However, modifications are required if there exists a spatio-temporal region in which the field $\boldsymbol{u}(\boldsymbol{x}, t)$ is missing or is corrupted, thus leading to incomplete dynamics. We would like to reconstuct this gappy field using the POD orthogonal modes. This is the problem first considered in Everson & Sirovich (1995) for an image reconstruction static problem. The gappy field can be written in the following way:

$$\boldsymbol{u}_g(\boldsymbol{x}, t) = (u(\boldsymbol{x}, t)m_1(\boldsymbol{x}, t),\, v(\boldsymbol{x}, t)m_2(\boldsymbol{x}, t),\, w(\boldsymbol{x}, t)m_3(\boldsymbol{x}, t)) \tag{2.13}$$

where $m_j(\boldsymbol{x}, t)$ tracks the spatio-temporal missingness. It is defined by

$$m_j(\boldsymbol{x}, t) = \begin{cases} 1 & \text{if the } j\text{th component of the field is known in } (\boldsymbol{x}, t) \\ 0 & \text{if the } j\text{th component of the field is missing in } (\boldsymbol{x}, t). \end{cases} \tag{2.14}$$

A similar spatio-temporal modulation of travelling wave solutions was studied in Aubry, Lima & Rahibe (2003); see also Aubry & Lima (1995*a*). The procedure proposed by Everson & Sirovich completes the missing spatio-temporal dynamics starting from a certain initial guess for the unknowns and proceeds iteratively. At the heart of the method is again the minimization of a functional but in a new norm, defined in the space–time domain where the field is known. Let us denote by $\widetilde{\boldsymbol{u}}(\boldsymbol{x}, t)$ a completed field obtained based on some initial guess. The standard Everson–Sirovich method employs the time-average value at that location as initial guess. Subsequently, we perform POD of $\widetilde{\boldsymbol{u}}(\boldsymbol{x}, t)$ to obtain the guessed spatial and temporal modes. This

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

decomposition has the form

$$\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)=\sum_{k=1}^{N}\widetilde{\lambda}_k\widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x})\,\widetilde{\psi}_k(t), \tag{2.15}$$

where $\widetilde{\psi}_k(t)$ is the kth guessed temporal mode and $\widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x})$ is the kth guessed spatial mode. The new functional for minimization is then

$$\begin{aligned} F_g[\widetilde{\xi}_k] &= \left\| \widetilde{\boldsymbol{u}}(\boldsymbol{x},t)-\sum_{k=1}^{M}\widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x})\widetilde{\xi}_k(t)\right\|^2_{\text{Gappy}} \\ &= \left(\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)-\sum_{k=1}^{M}\widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x})\widetilde{\xi}_k(t),\ \widetilde{\boldsymbol{u}}(\boldsymbol{x},t)-\sum_{k=1}^{M}\widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x})\widetilde{\xi}_k(t)\right)_{\text{Gappy}}, \end{aligned} \tag{2.16}$$

where the 'Gappy' norm is defined on the support of $\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)$, i.e. the spatio-temporal domain on which the values of $\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)$ are known with certainty. M is the number of modes that we use in the reconstruction process, which is different from the number of snapshots N ($M\leqslant N$). Minimization of the new functional (2.16) leads to the linear system of algebraic equations

$$\sum_{j=1}^{M}(\widetilde{\boldsymbol{\Phi}}_i(\boldsymbol{x}),\ \widetilde{\boldsymbol{\Phi}}_j(\boldsymbol{x}))_{\Omega_g(t)}\widetilde{\xi}_j(t)=(\widetilde{\boldsymbol{u}}(\boldsymbol{x},t),\ \widetilde{\boldsymbol{\Phi}}_i(\boldsymbol{x}))_{\Omega_g(t)},\quad i=1,\ldots,M. \tag{2.17}$$

The unknowns are the new (non-normalized) temporal modes $\{\widetilde{\xi}_k(t)\}$, and $\Omega_g(t)$ is the gappy spatial domain at time t. Note that the $M\times M$ matrix

$$[\widetilde{\boldsymbol{K}}]_{ij}=(\widetilde{\boldsymbol{\Phi}}_i(\boldsymbol{x}),\widetilde{\boldsymbol{\Phi}}_j(\boldsymbol{x}))_{\Omega_g(t)} \tag{2.18}$$

has time-dependent coefficients.

The Everson–Sirovich method is based on solving the system (2.17) for each guess and consists of the following steps:

(*a*) use time-average values as initial guesses at the locations $m_j(\boldsymbol{x},t)=0$ to obtain N snapshots of an initial complete field $\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)$;

(*b*) perform POD of $\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)$ to obtain N guessed spatial modes $\{\widetilde{\boldsymbol{\Phi}}_i(\boldsymbol{x})\}$;

(*c*) select the number of modes M to be employed in the reconstruction;

(*d*) construct the matrix $[\widetilde{\boldsymbol{K}}]_{ij}=(\widetilde{\boldsymbol{\Phi}}_i(\boldsymbol{x}),\widetilde{\boldsymbol{\Phi}}_j(\boldsymbol{x}))_{\Omega_g(t)}$ and the vector $[\boldsymbol{f}]_i=(\widetilde{\boldsymbol{u}}(\boldsymbol{x},t),\widetilde{\boldsymbol{\Phi}}_i(\boldsymbol{x}))_{\Omega_g(t)}$;

(*e*) solve the $M\times M$ linear system: $\widetilde{\boldsymbol{K}}\widetilde{\xi}=\widetilde{\boldsymbol{f}}$ for the unknowns $\{\widetilde{\xi}_k(t)\}$;

(*f*) construct a new vector field as follows:

$$\widetilde{\boldsymbol{w}}(\boldsymbol{x},t)=\sum_{k=1}^{M}\widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x})\,\widetilde{\xi}_k(t) \tag{2.19}$$

and overwrite the previous guess, i.e. set

$$\widetilde{\boldsymbol{u}}(\boldsymbol{x},t)=\widetilde{\boldsymbol{w}}(\boldsymbol{x},t)\quad\text{only if}\quad m_j(\boldsymbol{x},t)=0;$$

(*g*) upon convergence stop, otherwise go to (*b*).

Details on the convergence will be given in the flow examples in the next section. The method breaks down when the matrix $\widetilde{\boldsymbol{K}}$ is singular. This includes, for example, the case in which a snapshot is missing (i.e. $\Omega_g(t^*)=0$). Also, the method breaks down

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

if there are any spatial subregions for which the data is missing at all times. With the above procedure the optimum number of modes M_o in the reconstruction depends on the initial guess. Specifically, by *optimum number* of modes we mean the number M_o for which the error is the smallest among all possible converged reconstructions.

To this end, we have developed the following extension of the Everson–Sirovich procedure that does not depend on the initial guess and, in addition, enhances accuracy significantly. This will be demonstrated in the next section. We summarize here the main steps of the extended procedure.

Extended procedure

(*a*) Perform the standard Everson–Sirovich procedure but employ only $M = 2$ modes in the reconstruction.

(*b*) Use the converged result from the previous step as a new initial guess and apply the Everson–Sirovich procedure but now employ $M = 3$ modes in the reconstruction.

(*c*) Proceed similarly for the nth iteration until the eigenspectrum obtained does not change anymore.

Although more costly, our results suggest that the iterative procedure leads to the maximum possible resolution of the true eigenspectrum and thus of possible accuracy in reconstructing the flow field. The final solution will only depend on the degree of gappiness and not on the initial guesses in the gappy subregions. Alternatively, one could stop at an earlier iteration if accurate resolution only of the first few modes is desired, that is accept incomplete convergence in the eigenspectrum.

2.2. *Temporal versus spatial resolution*

For complete data snapshots a certain spatial mode is resolved accurately if the corresponding temporal mode is resolved accurately (Aubry 1991; Sirovich 1987). In this section we will show that this is not necessarily true for perturbed data – the perturbation being the converged error for the reconstructed field. To this end, we start with equation (2.4)

$$\int_T \widetilde{\boldsymbol{u}}\widetilde{\psi}_k \,\mathrm{d}t = \widetilde{\lambda}_k \widetilde{\boldsymbol{\Phi}}_k. \tag{2.20}$$

Here the tilded quantities are the perturbed ones, that is the ones obtained from a converged guess for the missing data. We also define

$$\widetilde{\boldsymbol{u}} = \boldsymbol{u} + \delta\boldsymbol{u}, \quad \widetilde{\psi}_k = \psi_k + \delta\psi_k, \quad \widetilde{\lambda}_k = \lambda_k + \delta\lambda_k, \quad \widetilde{\boldsymbol{\Phi}}_k = \boldsymbol{\Phi}_k + \delta\boldsymbol{\Phi}_k \tag{2.21}$$

where the deltas denote perturbations with respect to the 'true' quantities.

By inserting equations (2.21) into (2.20) we obtain

$$\int_T (\boldsymbol{u}\psi_k + \delta\boldsymbol{u}\psi_k + \boldsymbol{u}\delta\psi_k + \delta\boldsymbol{u}\delta\psi_k)\,\mathrm{d}t = \lambda_k\boldsymbol{\Phi}_k + \lambda_k\delta\boldsymbol{\Phi}_k + \delta\lambda_k\boldsymbol{\Phi}_k + \delta\lambda_k\delta\boldsymbol{\Phi}_k. \tag{2.22}$$

In adddition, for the 'true' field and the 'true' spatial and temporal modes we have the relationship (2.4) and inserting it into (2.22) gives

$$\int_T (\delta\boldsymbol{u}\psi_k + \boldsymbol{u}\delta\psi_k + \delta\boldsymbol{u}\delta\psi_k)\,\mathrm{d}t = \lambda_k\delta\boldsymbol{\Phi}_k + \delta\lambda_k\boldsymbol{\Phi}_k + \delta\lambda_k\delta\boldsymbol{\Phi}_k. \tag{2.23}$$

Let us now assume that the observed kth temporal mode $\widetilde{\psi}_k$ is perfectly resolved, that is

$$\delta\psi_k = 0 \quad (\widetilde{\psi}_k = \psi_k), \qquad \delta\lambda_k = 0 \quad (\widetilde{\lambda}_k = \lambda_k). \tag{2.24}$$

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

By inserting the conditions (2.24) in (2.23) we obtain

$$\int_T \delta \boldsymbol{u} \psi_k \, \mathrm{d}t = \lambda_k \delta \boldsymbol{\Phi}_k = \int_T \delta \boldsymbol{u} \widetilde{\psi}_k \, \mathrm{d}t$$

and therefore

$$\delta \boldsymbol{\Phi}_k(\boldsymbol{x}) = \frac{1}{\widetilde{\lambda}_k} \int_T \delta \boldsymbol{u}(\boldsymbol{x}, t) \widetilde{\psi}_k(t) \, \mathrm{d}t. \tag{2.25}$$

The corresponding formula for the perturbation in the kth temporal mode, given a perfectly resolved spatial mode (i.e. $\delta \boldsymbol{\Phi}_k = 0$, $\delta \lambda_k = 0$) is

$$\delta \psi_k(t) = \frac{1}{\widetilde{\lambda}_k} \int_\Omega \delta \boldsymbol{u}(\boldsymbol{x}, t) \cdot \widetilde{\boldsymbol{\Phi}}_k(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{2.26}$$

Equation (2.25) shows that almost certainly we have a perturbation in the kth spatial mode even though the corresponding temporal mode may be perfectly resolved. Similarly, equation (2.26) shows that almost certainly we have a perturbation in the kth temporal mode even though the corresponding spatial mode may be perfectly resolved. As we will see in the flow examples in the next section, this perturbation is negligible for a small or modest degree of gappiness but for very large gappiness we find that the temporal resolution is different from the spatial resolution. Given this discussion, the reconstruction procedures that we presented in the previous section optimize the temporal resolution.

3. Flow reconstruction

We consider here two-dimensional incompressible flow past a circular cylinder at Reynolds number $Re = 100$ and $Re = 500$. We have simulated the time-dependent flow using the spectral element/hp method (Karniadakis & Sherwin 1999) on a mesh consisting of 412 spectral triangular elements of order $P = 8$. After we had established fully periodic states, we extrated from the direct numerical simulation (DNS) 40 and 50 equidistant snapshots, for $Re = 100$ and $Re = 500$, respectively. Subsequently, we generated a spatio-temporal gappiness by randomly discarding the values of the solution on some nodes in every snapshot. The spatio-temporal 'gappiness percentage' here is defined to be the number of nodal data points that are missing with respect to the total number of nodal data points. Since the spectral elements are not equal and they also have different shape, this gappiness percentage is not directly related to the spatial gappineness at every time. In other words a 20% spatio-temporal gappiness in a region where the grid is very coarse is not the same as 20% gappiness where the grid is very fine. Moreover, in order to simulate an 'experimental-like' dataset, we also implemented the possibility of a different spatio-temporal gappiness close to the cylinder. We also note that disregarding information at some grid points affects greatly the Lagrangian interpolation process involved in the spectral element method. This, in turn, implies that the solution obtained from the gappy field, without any treatment, will be greatly oscillatory. Figure 1 shows the computational domain, and in table 1 we provide specific details on the four different cases that we have investigated.

3.1. *Reconstruction based on the Everson–Sirovich procedure*

We have investigated the reconstruction of velocity fields at both $Re = 100$ and $Re = 500$ but the results are very similar, so here we will present primarily the results for $Re = 100$ (case A). A typical result of the convergence history from the application

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

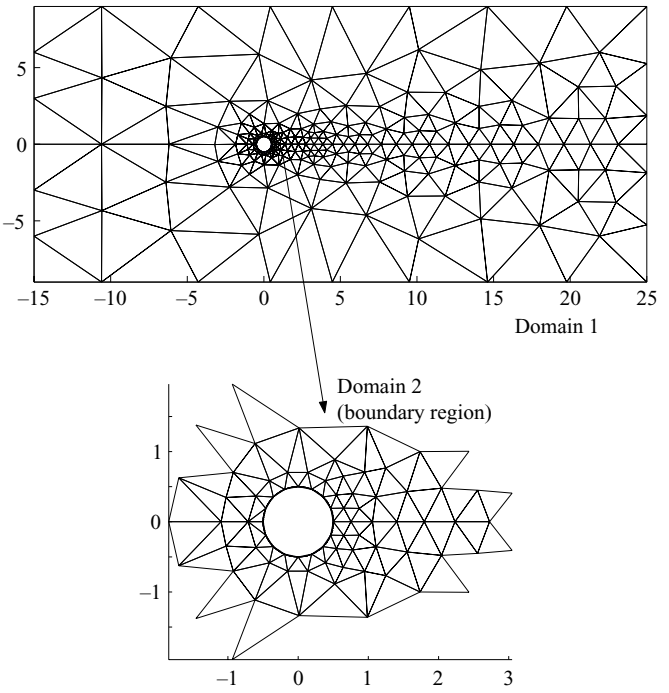


FIGURE 1. Computational domain and a close-up of the boundary region.

Case	Reynolds number	%Gappiness (Domain 1)		%Gappiness (Domain 2)	
A	100	18.9 U	19.5 V	23.4 U	22.9 V
B	500	21.2 U	21.2 V	21.2 U	21.1 V
C	500	48.4 U	50.0 V	46.7 U	44.7 V
D	500	75.4 U	76.3 V	79.6 U	76.2 V

TABLE 1. Definition of the gappy data fields.

of the Everson–Sirovich method, using a time-average initial guess, is shown in figure 2; specifically, the relative error in the two velocity components is plotted versus iteration number. The relative error for the streamwise component is defined as

$$\frac{\left\|u_M^{(i)} - u\right\|_2^2}{\|u\|_2^2} \equiv \frac{\displaystyle\int_T \int_\Omega \left(\widetilde{u}_M^{(i)}(\boldsymbol{x}, t) - u(\boldsymbol{x}, t)\right)^2 \mathrm{d}\boldsymbol{x}\,\mathrm{d}t}{\displaystyle\int_T \int_\Omega (u(\boldsymbol{x}, t))^2 \mathrm{d}\boldsymbol{x}\,\mathrm{d}t}, \tag{3.1}$$

where $\widetilde{u}_M^{(i)}(\boldsymbol{x}, t)$ denotes the M-modes reconstructed streamwise component of the velocity field at the current ith iteration, while $u(\boldsymbol{x}, t)$ denotes the 'true' component obtained from the DNS before the the random data-discarding process. The relative error for the crossflow velocity component is defined similarly to equation (3.1). The relative error is plotted for several values of M coresponding to different reconstructions. The best case here is obtained for $M_o = 8$ as it corresponds to the lowest error in both the u and the v components. The convergence history for

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

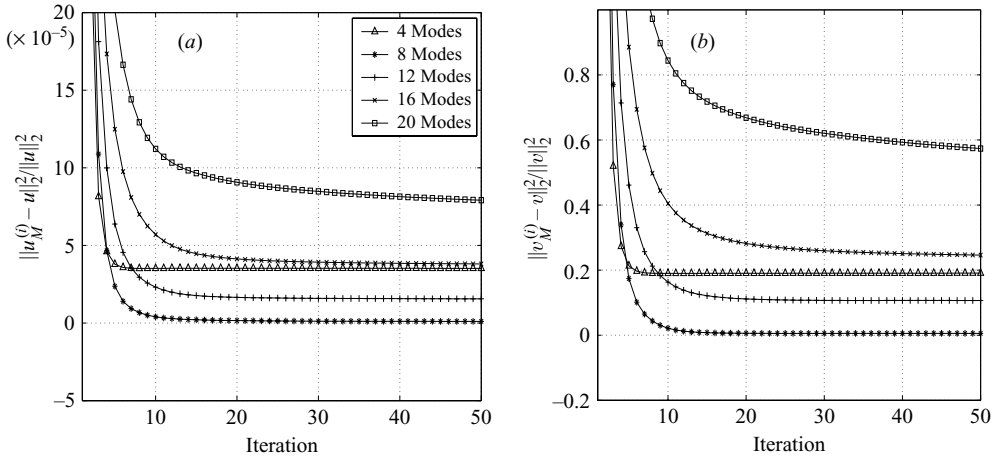


FIGURE 2. Case A: convergence history of the relative error for the steamwise (*a*) and crossflow (*b*) velocity components. The initial guess is taken as the time average of the available data.

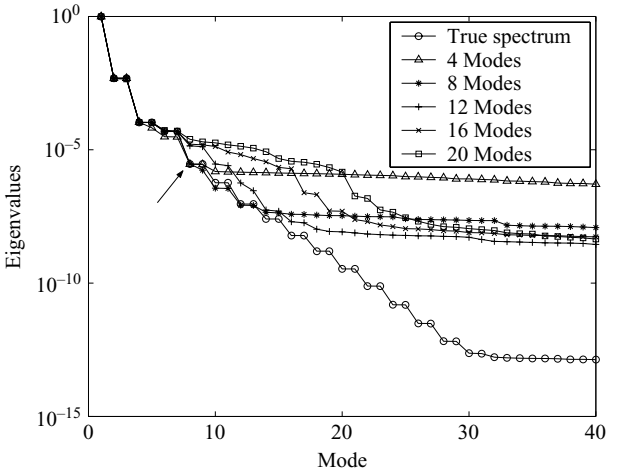


FIGURE 3. Case A: comparison of reconstructed with 'true' spectra. The arrow indicates the deviation of the $M_o = 8$ eigenspectrum from the 'true' one. The initial guess is taken as the time average of the available data.

$Re = 500$ at approximately the same gappiness level (case B) is similar except that the best reconstruction is achieved with $M_o = 10$ modes. The main result for both cases is that the error in the reconstruction does not decrease monotonically with the number of modes used in the reconstruction. This is not too surprising as including more higher (not resolved) modes may introduce noise that may degrade the quality of the reconstructed solution. For example, in case A reconstruction with $M = 12, 16$ or 20 leads to reconstructed fields of lower quality than for $M_o = 8$. The following few plots will demonstrate this result more systematically.

First, we compare the eigenspectra from the reconstructed fields to the 'true' spectrum from the full DNS in figure 3. We see that the case with $M_o = 8$ resolves the spectrum more accurately than all other reconstructions but it misses the upper part (high wavenumbers) of the 'true' spectrum. However, it is still an open question if

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

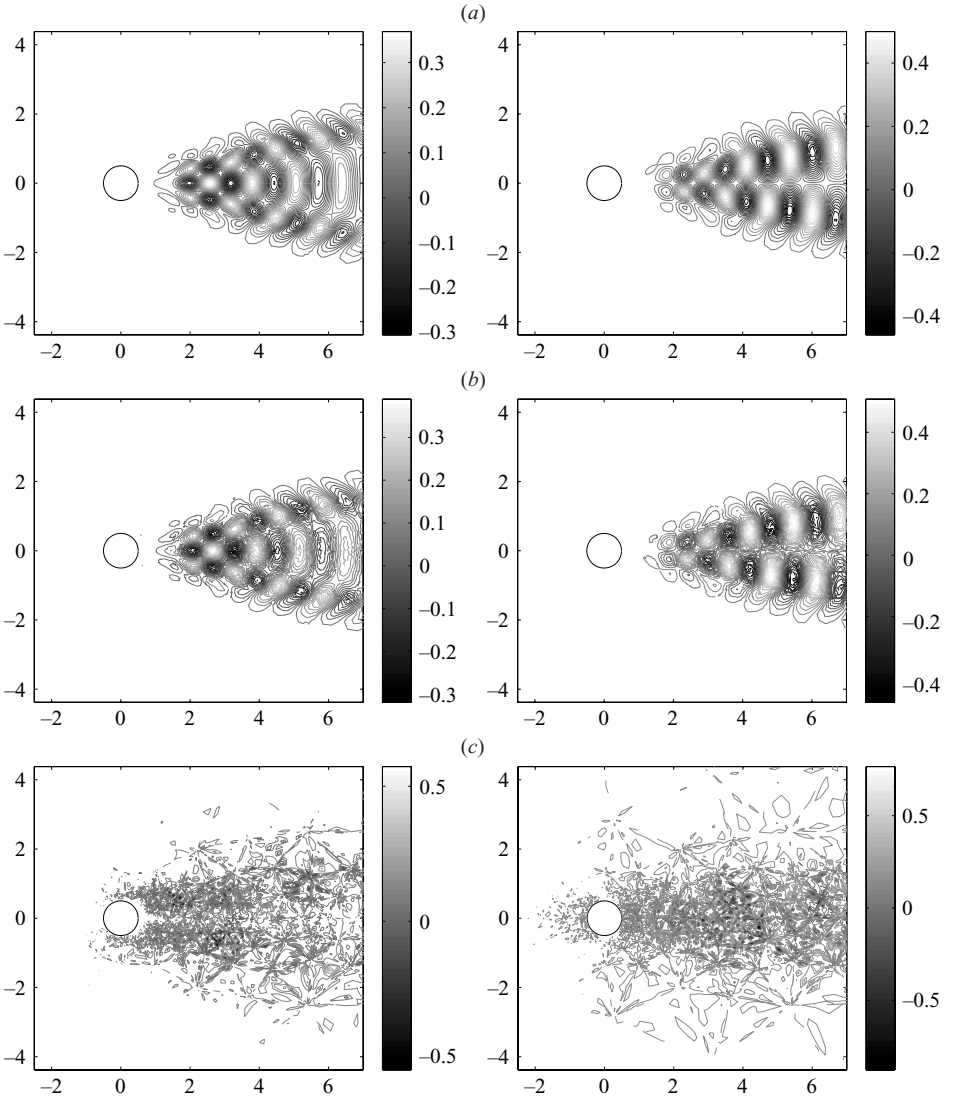


FIGURE 4. Case A: comparison of the 8th spatial POD mode using (*b*) $M_o = 8$ and (*c*) $M = 16$ with (*a*) the DNS ('true') mode. Left: streamwise component; right: crossflow component.

this is the maximum number of (temporal) modes that can be resolved for the degree of gappiness that we consider here (case A). We will revisit this issue in the next section but first let us document that indeed with $M_o = 8$ we resolve accurately all the corresponding spatial modes up to the deviation points obtained from figure 3; see the arrow for the case $M_o = 8$. We note that the number of resolved modes corresponding to the deviation point in this plot is very close to the optimum number of modes M_o in this case. This is a typical result for low to modest levels of gappiness.

In figure 4 we plot the two components of the 8th spatial POD mode from the reconstructed fields as well as the original DNS field. We see that for $M_o = 8$ we can resolve this mode but for $M = 16$ the reconstruction is not as accurate, in accord with

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

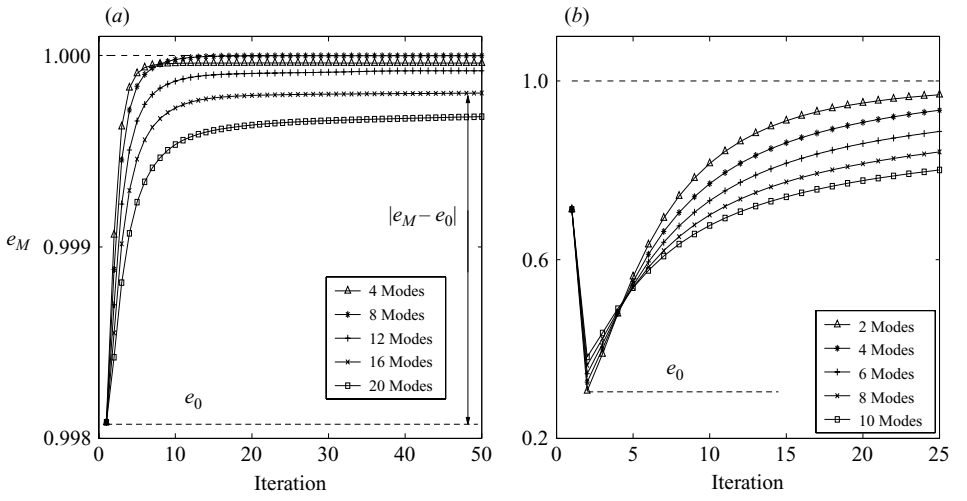


FIGURE 5. Convergence history of the total relative energy using different numbers of modes for reconstruction. (*a*) Case A, time-average initial guess. (*b*) Case D, random initial guess.

our findings from the eigenspectrum comparison. Comparison of lower modes, e.g. the 4th mode, shows that both reconstructed fields resolve this mode accurately. We find that in this case with relatively small gappiness a certain temporal mode is resolved accurately if the corresponding spatial mode is resolved accurately. This result has been obtained before (e.g. see Sirovich 1987; Aubry *et al.* 1993, 1991) for complete data sets. However, as we indicated in § 2.2 this may not be true in general for gappy fields. We will return to this point later when we study case D corresponding to the highest level of gappiness.

3.1.1. *Criterion for optimum number of modes*

In the discussion so far we have assumed that we know the 'true' eigenspectrum in order to find the optimum number of modes. Next, we propose a criterion based on the total energy that leads to the same result and does not require knowledge of the 'true' solution. To this end, we start from the observation that the convergence history of the total energy, after a possible initial oscillatory path, typically becomes monotonic. This is demonstrated in the two plots in figure 5.

Specifically, in figure 5 we plot the convergence history of the total energy normalized with the 'true' energy of the system, defined as

$$e_M \equiv \frac{\displaystyle\int_T \int_\Omega \tilde{\boldsymbol{u}}_M^{(i)}(\boldsymbol{x},t)\cdot\tilde{\boldsymbol{u}}_M^{(i)}(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t}{\displaystyle\int_T \int_\Omega \boldsymbol{u}(\boldsymbol{x},t)\cdot\boldsymbol{u}(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t} = \frac{\displaystyle\sum_{j=1}^{N} \tilde{\lambda}^2_{j(M)}(i)}{\displaystyle\sum_{j=1}^{N} \lambda_j^2}. \tag{3.2}$$

Here $\tilde{\lambda}^2_{j(M)}(i)$ is the jth eigenvalue of the M-modes reconstructed correlation matrix at the ith iteration, and λ_j^2 is the jth eigenvalue of the 'true' correlation matrix. Also, $\tilde{\boldsymbol{u}}_M^{(i)}(\boldsymbol{x},t)$ is the M-modes reconstructed velocity field at the ith iteration while $\boldsymbol{u}(\boldsymbol{x},t)$ is the 'true' field. We see that in both cases the reconstructed energy approaches the 'true' energy from below. If we assume that the converged relative energy does

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

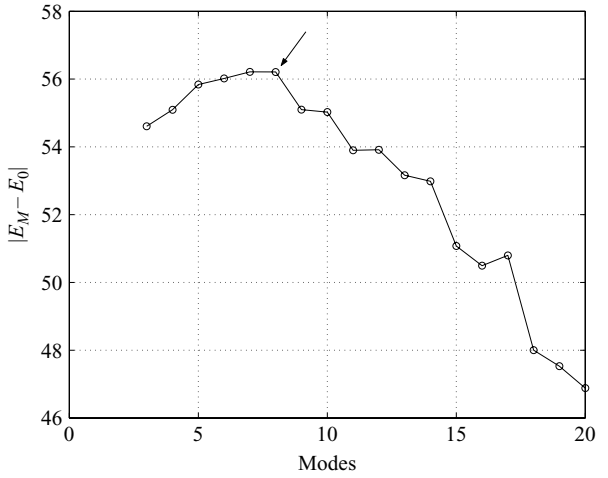


FIGURE 6. Case A: verification of the selection criterion for the optimum number of modes $M_o = 8$.

not overshoot the line $e_M = 1$ then it is possible to identify the reconstruction which minimizes the difference between the 'true' and the reconstructed energy, i.e. the curve which is closest to $e_M = 1$. Such identification has been made possible by a choice of a reference level of energy e_0 (see figure 5).

Note that in case D shown in figure 5(*b*), convergence has not been achieved within 25 iterations; in fact, full convergence reconstruction with $M = 3$ modes leads to the least error. The aforementioned assumption of absence of overshoot (or undershoot, if the reconstructed energy approaches the 'true' energy from above) is good if the degree of gapiness is not very high. Based on such observations, we can now formulate the following criterion for the selection of the optimum number of modes M_o:

$$\text{Choose } M \text{ such that: } |E_M - E_0| \text{ is maximum.}$$

Here, E_M is the total energy for the converged reconstuction using M modes (i.e. the numerator in equation (3.2)), and E_0 is the total energy associated with the fixed reference value. Application of this criterion for the flow of case A leads to the results of figure 6 that confirm the validity of our selection criterion. Clearly, this criterion is independent of the knowledge of the true solution. We have produced similar results for all cases listed in table 1 that confirm this optimization criterion. For example, the optimum number modes for case B is $M_o = 10$, for case C is 7 and for case D is 5; in all cases the time average was used as initial guess.

In contrast, the criterion based on

$$\pi_M \equiv \sqrt{\sum_{k=1}^{N} \left(\widetilde{\lambda}_{k(M)}(i) - \widetilde{\lambda}_{k(M)}(i-1)\right)^2} \tag{3.3}$$

used for example in Tan *et al.* (2003) may not lead to the optimum number of modes as defined here. In order to illustrate this point we plot in figure 7 π_M versus the iteration number for different reconstructions. The results indicate that with $M = 4$ modes we obtain the fastest convergence; however we have seen that the optimum number of modes for reconstruction for case A is $M_o = 8$.

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

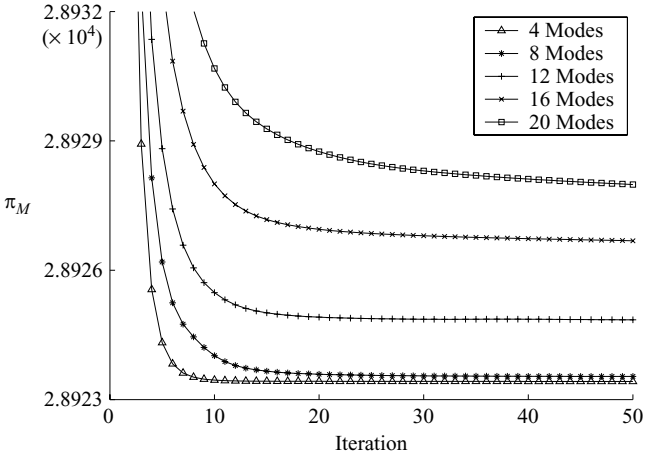


FIGURE 7. Case A: convergence history using the ratio π_M.

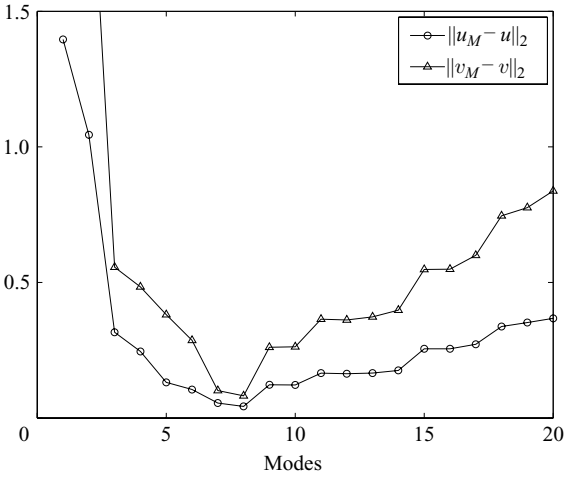


FIGURE 8. Case A: error in the energy norm of the reconstructed field versus the number of modes M.

3.1.2. *Reconstructed velocity field*

So far we have compared the eigenspectrum and the spatial POD modes of the reconstructed fields with the 'true' field. We now turn our attention to the flow field. First, we demonstrate that indeed the reconstructed field is most accurate for $M_o = 8$ in the mean-square sense. In figure 8 we plot the error (in the energy norm) of the two components of the velocity field as a function of the number of modes employed in the reconstruction. We see that we achieve the minimum value for $M_o = 8$, and the same is true for the maximum pointwise error (not shown here). Because of the non-monotonic convergence of the Everson–Sirovich procedure with the number modes M, we expect that for $M > M_o$ we may add noise to the reconstructed field as indicated by the loss of accuracy in figure 8. Indeed, this is the case as demonstrated in the flow fields shown in figure 9. We compare here the 8th snapshot but similar results are valid for other time instants. For a more quantitative comparison we plot the crossflow velocity profiles in the near wake at $x/d = 2$ and compare reconstructions

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

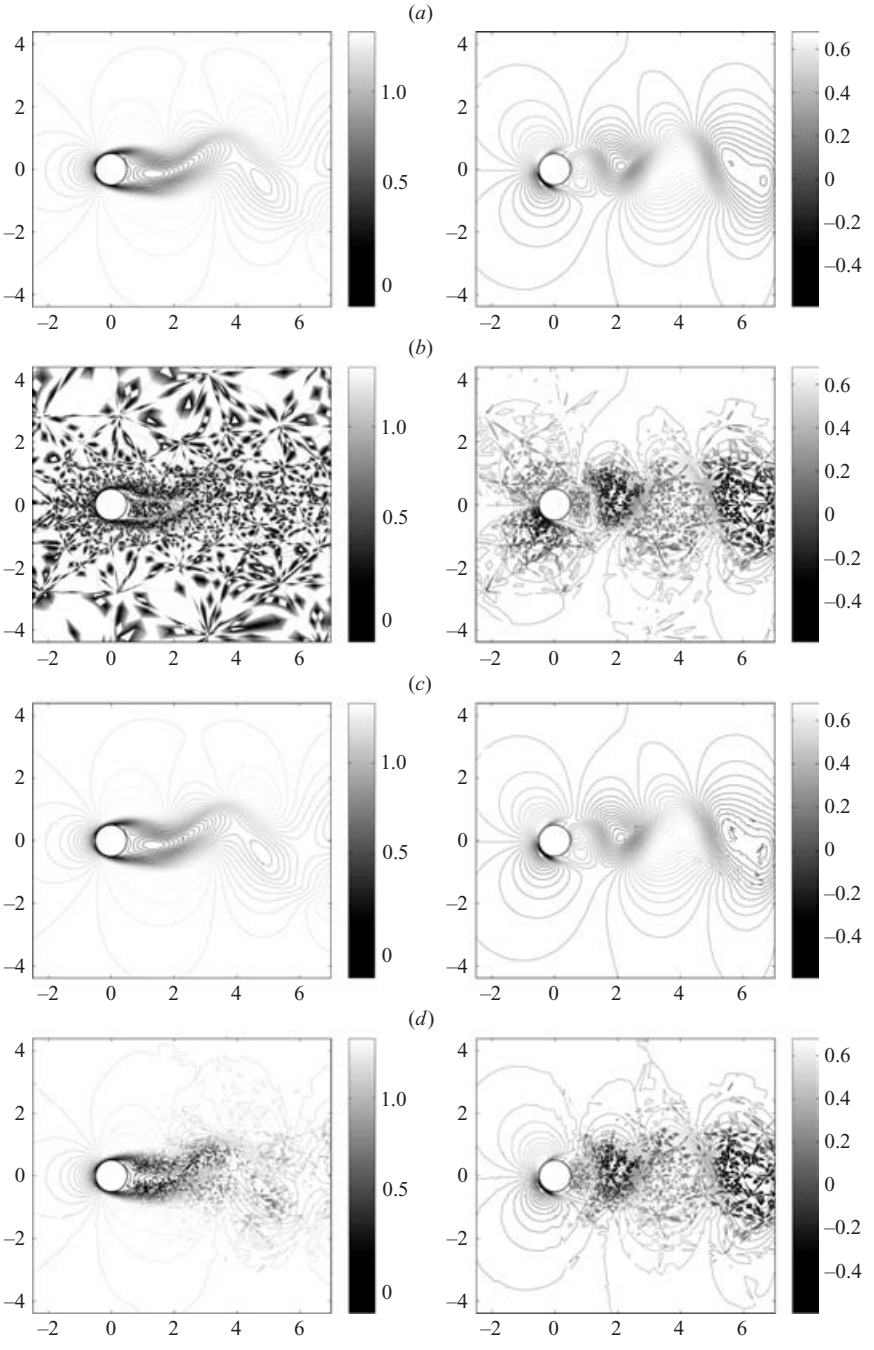


FIGURE 9. Case A: comparison between the 'true' field (*a*) and the reconstructed field for the eighth snapshot using $M_o = 8$ modes (*c*) and $M = 20$ (*d*). The figure also includes the gappy field (*b*), which is the starting point of the iterative procedure. The initial guess is the time average of the available data. Left: streamwise component; right: crossflow component.

corresponding to $M_o = 8$ and $M = 20$, see figure 10. In particular, what we plot in this figure are the interpolated data at $x/d = 2$ using the 'true' DNS field, the gappy field, and the reconstructed field. This interpolation introduces wiggles in the profile

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

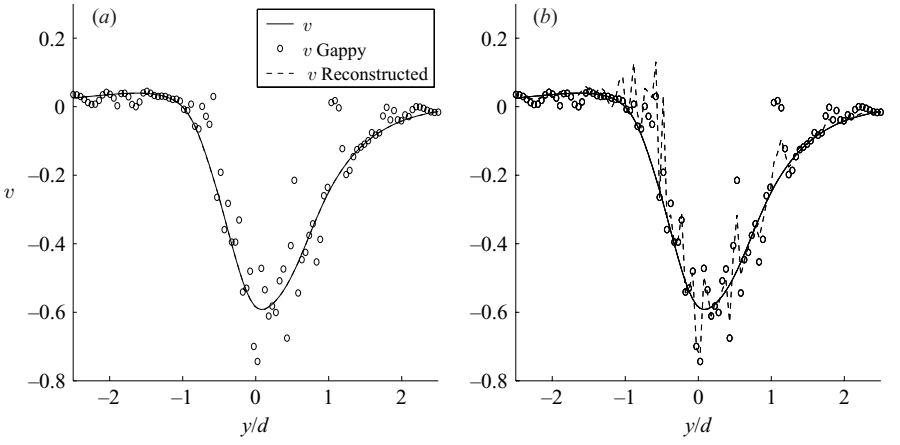


FIGURE 10. Case A: crossflow velocity profile at $x/d = 2$. Solid line denotes the 'true' velocity profile, symbols denote the interpolated profile from the gappy field, and dashed line denotes the interpolated profile from the recostructed field using (*a*) $M_o = 8$ modes and (*b*) $M = 20$. The initial guess is the time average of the available data.

for the gappy field. The results corresponding to the optimum number of modes are clearly superior.

3.2. *Reconstruction based on the extended procedure*

The straightforward application of the Everson–Sirovich procedure starting from different initial guesses for the gappiness leads to different numbers of resolved modes for the same flow field. To demonstrate this we compare the reconstructed results corresponding to two different initial guesses for case A, which was analysed in the previous section. Specifically, we use

(*a*) the time average of the available data (as before)

$$U(\boldsymbol{x}) = \frac{1}{T}\int_T u_g(\boldsymbol{x}, t)\,\mathrm{d}t, \quad V(\boldsymbol{x}) = \frac{1}{T}\int_T v_g(\boldsymbol{x}, t)\,\mathrm{d}t, \tag{3.4}$$

where $u_g(\boldsymbol{x}, t)$ and $v_g(\boldsymbol{x}, t)$ are the two components of the gappy velocity field;

(*b*) a random fluctuation defined by

$$u_r(\boldsymbol{x}, t) = U(\boldsymbol{x}) f_u(\boldsymbol{x}, t), \quad v_r(\boldsymbol{x}, t) = V(\boldsymbol{x}) f_v(\boldsymbol{x}, t), \tag{3.5}$$

where $f_u(\boldsymbol{x}, t)$ and $f_v(\boldsymbol{x}, t)$ are random functions following a *normal* probability distribution with values in $[-1, 1]$.

In figure 11 we plot the eigenspectra obtained starting from the random initial guess. We note that the 'deviation point' (denoted by the arrow), which defines the number of resolved temporal modes, is different for the random initial condition case. In particular, the number of resolved modes is decreased from 8 (time-average initial guess) to 3 (random initial guess). We found similar results for all cases listed in table 1. For example, we found that for case D (greatest gappiness) the number of resolved temporal modes with a time-average initial guess is 5 while with a random initial guess it is 3. Correspondingly, for case C the number of resolved temporal modes drops from 7 to 3.

We now apply the new procedure described in § 2. To this end, we construct a sequence of guesses by simply increasing iteratively the number of modes used in the reconstruction process and taking as an initial guess for the next run the converged

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

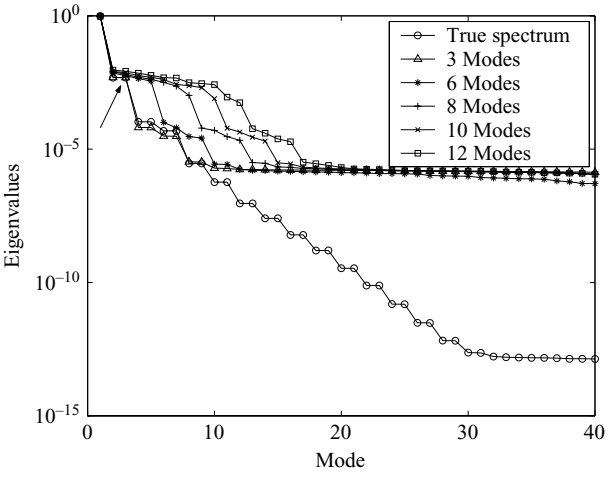


FIGURE 11. Case A: comparison of eigenspectra obtained using a random initial guess.

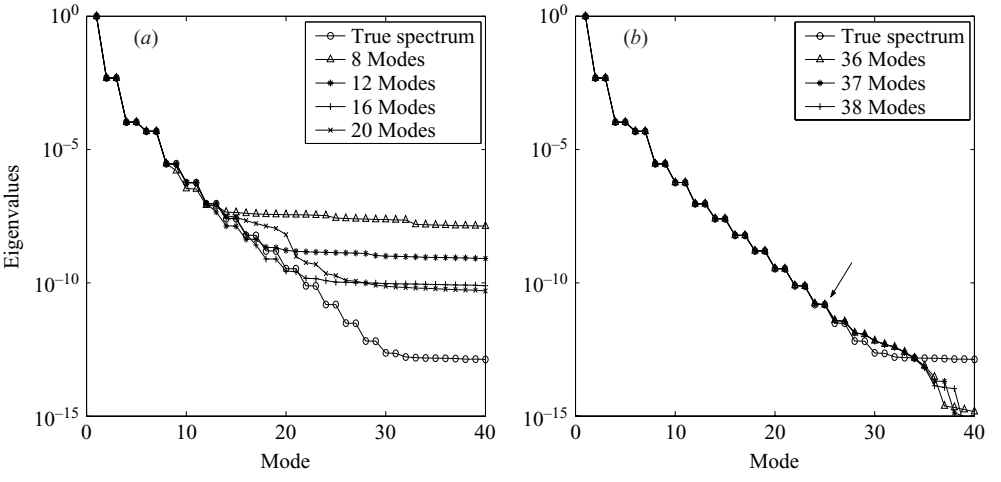


FIGURE 12. Case A: eigenspectra obtained with the new method iterating up to mode 8 (*a*) and up to mode 35 (*b*). The initial condition is a random guess.

guess from the previous run. In figure 12 we plot the eigenspectra obtained with the modified procedure by iterating up to the 8th mode and also up to the 35th mode. The results shown in this figure were obtained with the random initial values but the same results are obtained starting from the time-average initial guess. This and other tests we have perfomed show that the number of resolved modes obtained with the new iterative procedure is independent of the initial guess. Such results, in turn, suggest that the new procedure produces the *maximum* possible number of resolved modes given a gappy data field. For case A this value is 27 compared to the value 3 obtained with the standard Everson–Sirovich procedure for the random initial condition (see figure 11). Therefore, we have increased the number of resolved temporal modes with the extended procedure by nine times for the random initial condition and more than three times for the time-average initial condition. Similarly,

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

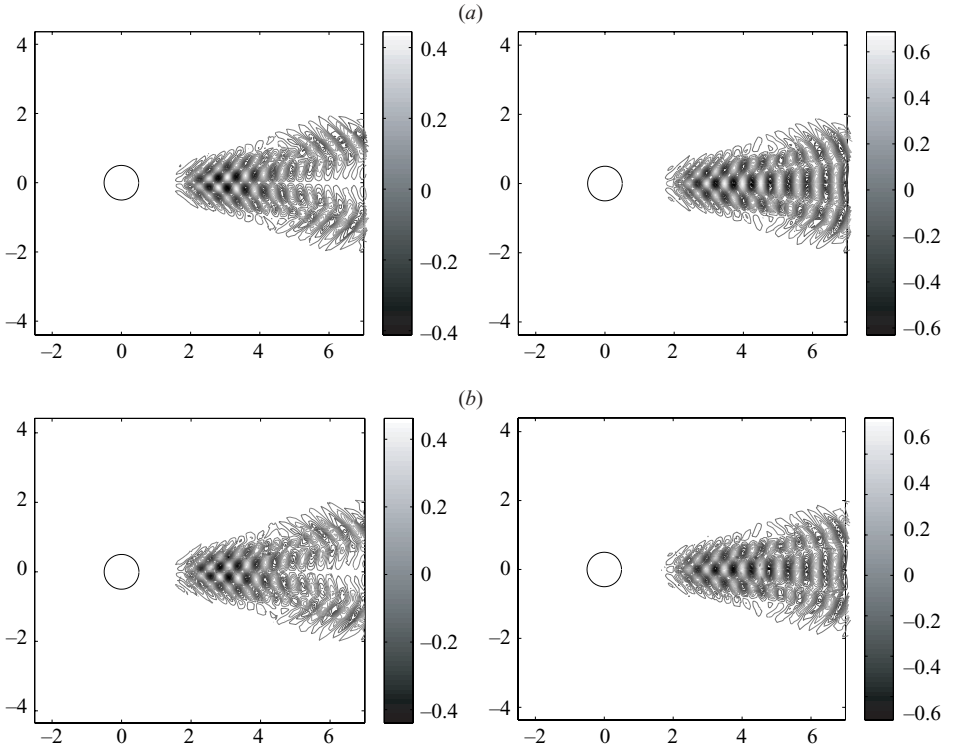


FIGURE 13. Case A: comparison of reconstructed 19th spatial mode (using the new method) (*b*) with the corresponding 'true' DNS mode (*a*). Shown are isocontours of the two velocity components. Left: streamwise component; right: crossflow component.

for case C we have increased the resolution by five times with respect to the random initial guess and have more than doubled it with the time-average initial guess.

In order to verify that indeed the increased number of resolved modes leads to a better representation of the velocity field we examined the accuracy of modes above the eighth mode resolved accurately with the standard Everson–Sirovich procedure. In figure 13 we plot isocontours of the 19th mode and compare them with the exact mode obtained from the full DNS. We observe that the agreement is very good using the new extended method; in fact, we have verified that this good agreement is valid up to the 27th mode as suggested by the resolved eigenspectrum.

3.2.1. *Very large gappiness*

In the discussion so far we have focused mostly on case A since cases B and C in table 1 behave similarly. We now analyse in some more detail case D that corresponds to very large gappiness of about 78% at $Re = 500$. We consider again the two aforementioned initial guesses which lead to different spectra of the reconstructed fields. For brevity we do not show the spectra here but we have found that the resolved number of modes is $M_o = 3$ for the random initial guess and $M_o = 5$ for the time-average initial guess using the standard Everson–Sirovich procedure. Using the new method and iterating up to mode 20 we obtain the same resolved eigenspectrum (independently of the initial guess), as shown in figure 14. In particular, the number of resolved modes for this flow field is increased to 7, i.e. we have doubled the number of resolved temporal modes.

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

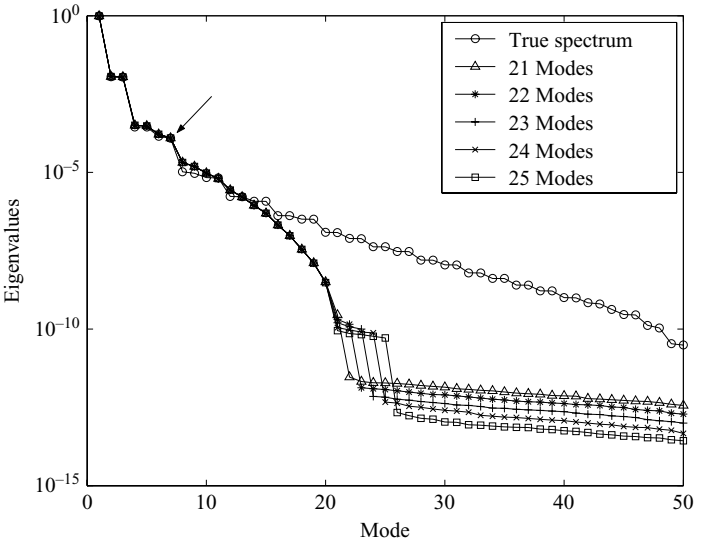


FIGURE 14. Case D: eigenspectrum obtained using the new iterative procedure.

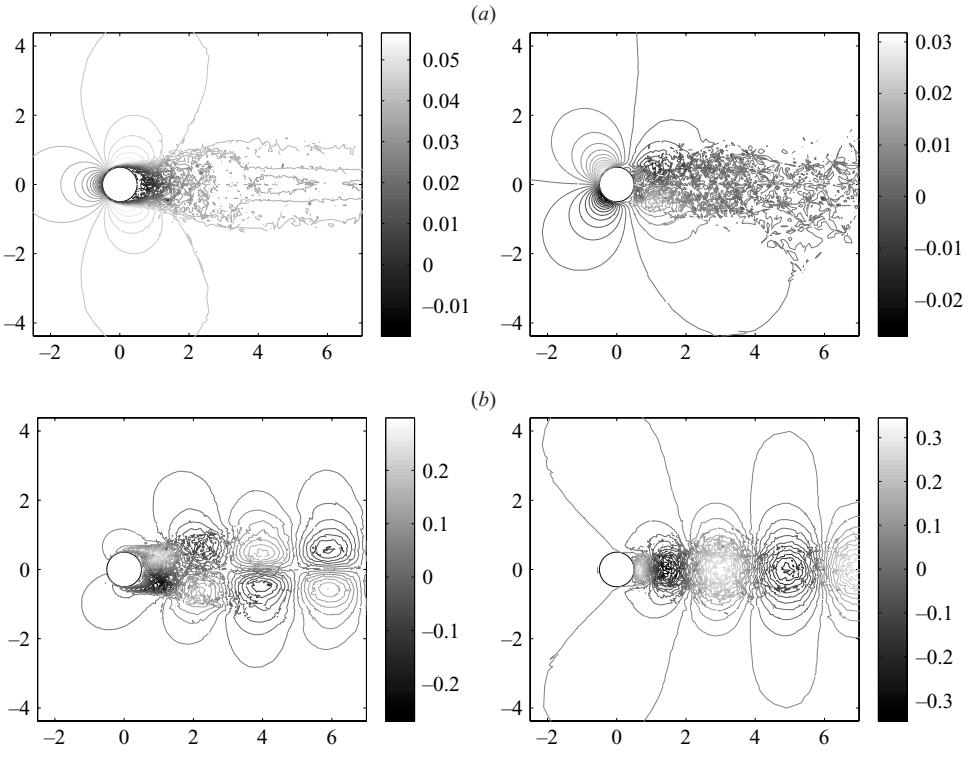


FIGURE 15. Case D: reconstruction of (*a*) the first and (*b*) the second spatial modes. Left: streamwise component; right: crossflow component.

However, unlike the other three cases, in case D we have realized the effects of noise on the spatial modes as analysed in § 2.2 and directly suggested by equation (2.25). We plot in figure 15 the two components of the first and the second spatial

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

	Local linear interp.	Everson–Sirovich	Local Kriging	Extended proc.
$\|\|\widetilde{u}-u\|\|_\infty$	0.2664	0.0581	0.0722	0.0538
$\|\|\widetilde{v}-v\|\|_\infty$	0.2443	0.09	0.0674	0.0665
$\|\|\widetilde{u}-u\|\|_2$	0.1630	0.0427	0.0070	0.0068
$\|\|\widetilde{v}-v\|\|_2$	0.1168	0.0817	0.0051	0.0049

TABLE 2. Case A: maximum pointwise error (L_∞) and average error (L_2) for the two velocity components using different reconstruction procedures.

	Local Kriging	Extended proc.
$\|\|\widetilde{u}-u\|\|_\infty$	1.0138	0.5628
$\|\|\widetilde{v}-v\|\|_\infty$	0.4990	0.4725
$\|\|\widetilde{u}-u\|\|_2$	0.2379	0.1907
$\|\|\widetilde{v}-v\|\|_2$	0.1106	0.1023

TABLE 3. Case C: maximum pointwise error (L_∞) and average error (L_2) for two velocity components using the extended procedure and the local Kriging method.

mode. We see that the spatial modes are noisy although the first seven temporal modes are resolved according to the eigenspectrum shown in figure 14. This can be simply explained by equation (2.25) as follows. The induced error in the extended procedure $\delta\boldsymbol{u}$ is such that $\int_T \delta\boldsymbol{u}(\boldsymbol{x},t)\widetilde{\psi}_1(t)\,\mathrm{d}t/\widetilde{\lambda_1}$ is not neglegible with respect to $\widetilde{\boldsymbol{\Phi}}_1$. The perturbation $\delta\boldsymbol{\Phi}_1$ then affects the first mode as shown in figure 15, and similarly for mode 2. When we reconstruct the flow field we modulate the unresolved spatial modes by the resolved temporal modes and this leads to a noisy reconstruction in the case of very large gappiness.

3.3. *Comparison with other methods*

In order to appreciate the accuracy of the proposed reconstruction procedure we have compared it to the local Kriging method, the local linear interpolation method and also the standard Everson–Sirovich method for case A. In table 2 we summarize the maximum pointwise errors (denoted by $||\cdot||_\infty$) and also the average error (denoted by $||\cdot||_2$). For the local linear interpolation method we transform every triangular spectral element used in the mesh into a standard square element and subsequently employ linear interpolation on the Gauss–Lobatto quadrature points in an element-by-element fashion. For the local Kriging statistical predictor we use a Gaussian for the covariance kernel while the regression model corresponds to a second-order polynomial; see Lophaven, Nielsen & Sondergaard (2002) for details.

We found that the errors of the new method are lower than the 'optimal prediction' errors given by the local Kriging procedure, and one to two orders of magnitude better than the linear interpolation errors. The average error (L_2 norm) for the new method is an order of magnitude lower than the error resulted by applying the standard Everson–Sirovich method at $M_o = 8$; see also figure 8.

We have also performed a comparison for case C (listed in table 1) and the results are similar to case A, see table 3. Here we have only compared the extended procedure to the local Kriging method. The extended procedure results are better in maximum and average errors for both components of velocity.

The clear advantage of the new proposed method compared to local Kriging prediction is in reconstructing accurately the temporal modes. A comparison of the

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

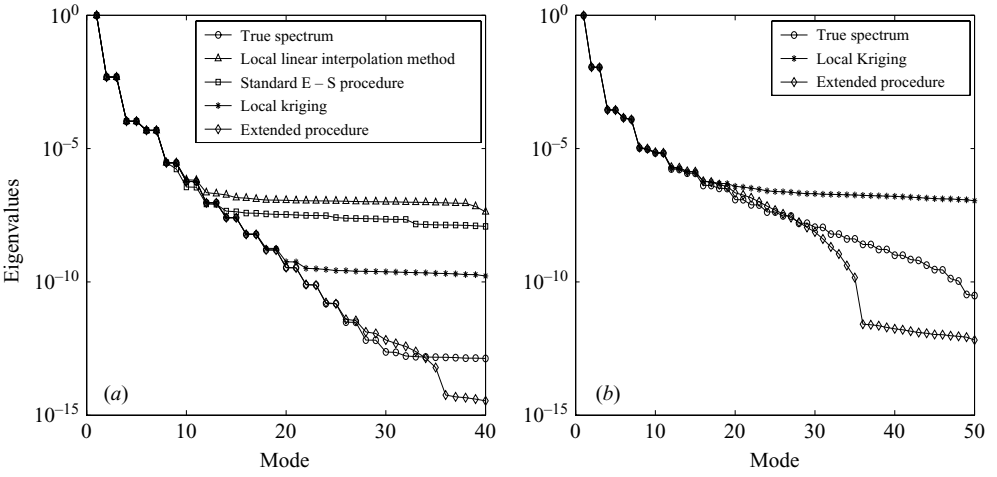


FIGURE 16. Eigenspectra for different reconstruction procedures. (*a*) case A; (*b*) case C.

corresponding eigenspectra for both methods is shown in figure 16. On plot (*a*) we show the comparison for the small degree of gappiness (case A) and on plot (*b*) the comparison for a large degree of gappiness (case C). We see that in both cases the extended procedure resolves many more modes than the local Kriging method. This, in turn, implies that the new extended procedure is more effective in dynamic simulations based on gappy experimental data, as for example in the simulations presented in Ma *et al.* (2003).

4. Summary

We have applied the proper orthogonal decomposition in reconstructing time-dependent gappy flow fields by extending a procedure first suggested by Everson & Sirovich (1995) for image processing applications. While the gappy data in this study were generated artificially by randomly omitting data points from DNS snapshots (MAR-type data) of flow past a cylinder, the same procedure can be applied to experimental data obtained by PIV or MRI. The original Everson–Sirovich procedure is based on the time-average values in the (spatio-temporal) gappy regions used as initial guess. It was applied to image denoising and facial reconstruction in Everson & Sirovich (1995); in Tan *et al.* (2003) it was applied to inviscid steady flow past an airfoil. However, the maximum number of accurately resolved modes seems to depend strongly on the initial guess for the missing data.

The extension we have developed employs the Everson–Sirovich procedure in an iterative manner starting from two modes for reconstruction and producing successively better estimates for the gappy regions. Upon convergence, a large portion of the exact eigenspectrum is resolved which is independent of the initial conditions, while the maximum possible spectral resolution is achieved for a given degree of spatio-temporal gappiness in a flow field. The specific new contributions of the current work are:

a new iterative POD/least-squares procedure that increases the maximum possible resolution by more than five times for gappiness up to 50%;

a new robust criterion for the selection of the optimum number of modes for the best reconstruction of the velocity field;

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

an analysis and verification of the difference in resolving a temporal mode versus resolving a spatial mode; this difference manifests itself at very large gappiness.

The robustness and effectiveness of the new iterative procedure comes at a price, however, since it is clearly computationally more expensive than the standard Everson–Sirovich procedure. However, it is hierarchical and can be terminated after the most energetic modes have been identified, thus reducing the cost to a desired level. Another potential problem is the loss of incompressibility condition in the reconstructed fields; this is true for both the original and the modified reconstruction. This, however, can be readily resolved by applying a Hodge decomposition of the reconstructed flow field, separating out the divergence-free components from the curl-free components. Computationally, this is equivalent to solving two Poisson equations for a two-dimensional flow field. Finally, if the gappiness is not random in space and time but it occurs, say, in a spatial region at all times then the current procedure cannot fill in that region. If appropriate, the symmetries of the POD modes can be used as shown in Ma *et al.* (2003) but in general statistical data imputation methods should be pursued in this case, see Little & Rubin (2002) and also Aubry *et al.* (2003).

This work was supported partially by NSF and ONR. We would like to thank Sirod Sirisup for helpful discussions. The first author expresses his sincere gratitude to G.K. for his hospitality during his visit at the Center for Fluid Mechanics, Brown University (RI,USA). He also would like to thank Prof. Sandro Salvigni and Dr. Beatrice Pulvirenti of University of Bologna for their support.

REFERENCES

Aubry, N. 1991 On the hidden beauty of the Proper Orthogonal Decomposition. *Theor. Comput. Fluid Dyn.* **2**, 339–352.

Aubry, N., Guyonnet, R. & Lima, R. 1995 Spatio-temporal symmetries and bifurcations via bio-orthogonal decomposition. *J. Nonlinear Sci.* **2**, 183–215.

Aubry, N., Guyonnet, R. & Stone, E. 1991 Spatio-temporal analysis of complex signals: Theory and applications. *J. Statist. Phys.* **64**, 683–739.

Aubry, N., Lian, W. & Titi, E. 1993 Preserving symmetries in the proper orthogonal decomposition. *SIAM J. Sci. Comput.* **14**, 483–505.

Aubry, N. & Lima, R. 1995*a* The dynamics of spatio-temporal modulations. *Chaos* **5**, 578–588.

Aubry, N. & Lima, R. 1995*b* Spatio-temporal and statistical symmetries. *J. Statist. Phys.* **81**, 793–828.

Aubry, N., Lima, R. & Rahibe, M. 2003 Breaking of space-time symmetries in modulated traveling waves. *Chaos* **13**, 541–551.

Cao, N.-Z. & Aubry, N. 1993 Numerical simulation of wake flow via a reduced system. *Proc. ASME Fluids Engineering Conference, Washington DC*.

Deane, A., Kevrekidis, I., Karniadakis, G. & Orszag, S. A. 1991 Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders. *Phys. Fluids* A **3**, 2337–2354.

Everson, R. & Sirovich, L. 1995 The Karhunen-Loeve transform of incomplete data. *J. Opt. Soc. Am.* A **12**, 1657.

Karniadakis, G. & Sherwin, S. 1999 *Spectral/hp Element Methods for CFD*. Oxford University Press.

Little, R. & Rubin, D. 2002 *Statistical Analysis with Missing Data*. Wiley Interscience.

Lophaven, S., Nielsen, H. & Sondergaard, J. 2002 DACE – A Matlab Kriging toolbox. *Informatics and Mathematical Modeling*. Technical University of Denmark.

Ma, X., Karamanos, G. & Karniadakis, G. 2000 Dynamics and low-dimensionality of the turbulent near-wake. *J. Fluid Mech.* **410**, 29–65.

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press

Ma, X., Karniadakis, G., Park, H. & Gharib, M. 2003 DPIV-driven simulation: A new computational paradigm. *Proc. R. Soc. Lond.* A **459**, 547–565.

Sirovich, L. 1987 Turbulence and the dynamics of coherent structures, Parts 1–3. *Q. Appl. Maths* **45**, 561–590.

Stein, M. 1999 *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.

Tan, B.-T., Willcox, K. & Damodaran, M. 2003 Applications of proper orthogonal decomposition for inviscid transonic aerodynamics. *AIAA Paper* 2003–4213.

Yates, F. 1933 The analysis of replicated experiments when the field results are incomplete. *Emp. J. Expl Agri.* **1**, 129–142.

https://doi.org/10.1017/S0022112004001338 Published online by Cambridge University Press