CrossMark

J. Fluid Mech. (2014), *vol.* 759, *pp.* 404–431. © Cambridge University Press 2014
doi:10.1017/jfm.2014.566

A data-assimilation method for Reynolds-averaged Navier–Stokes-driven mean flow reconstruction

Dimitry P. G. Foures[1], Nicolas Dovetta[2], Denis Sipp[3,†] and Peter J. Schmid[4]

[1]DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, UK

[2]LadHyX, Ecole Polytechnique, 91128 Palaiseau, France

[3]ONERA-DAFE, 8 rue des Vertugadins, 92190 Meudon, France

[4]Department of Mathematics, Imperial College London, London SW7 2AZ, UK

(Received 12 November 2013; revised 12 August 2014; accepted 24 September 2014; first published online 4 November 2014)

We present a data-assimilation technique based on a variational formulation and a Lagrange multipliers approach to enforce the Navier–Stokes equations. A general operator (referred to as the measure operator) is defined in order to mathematically describe an experimental measure. The presented method is applied to the case of mean flow measurements. Such a flow can be described by the Reynolds-averaged Navier–Stokes (RANS) equations, which can be formulated as the classical Navier–Stokes equations driven by a forcing term involving the Reynolds stresses. The stress term is an unknown of the equations and is thus chosen as the control parameter in our study. The data-assimilation algorithm is derived to minimize the error between a mean flow measurement and the measure performed on a numerical solution of the steady, forced Navier–Stokes equations; the optimal forcing is found when this error is minimal. We demonstrate the developed data-assimilation framework on a test case: the two-dimensional flow around an infinite cylinder at a Reynolds number of $Re = 150$. The mean flow is computed by time-averaging instantaneous flow fields from a direct numerical simulation (DNS). We then perform several 'measures' on this mean flow and apply the data-assimilation method to reconstruct the full mean flow field. Spatial interpolation, extrapolation, state vector reconstruction and noise filtering are considered independently. The efficacy of the developed identification algorithm is quantified for each of these cases and compared with more traditional methods when possible. We also analyse the identified forcing in terms of unsteadiness characterization, present a way to recover the second-order statistical moments of the fluctuating velocities and finally explore the possibility of pressure reconstruction from velocity measurements.

Key words: general fluid mechanics, turbulence modelling, variational methods

† Email address for correspondence: denis.sipp@onera.fr

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

1. Introduction

In a large variety of experimental scientific domains where measurements are performed, a major challenge has to be faced: no matter the type of data acquisition, the measured quantities are only a very sparse representation of the real, and therefore inaccessible, field. This sparsity can manifest itself in several forms depending on the specific circumstances. For example, in fluid mechanics, measurements are often under-resolved in time or space, or only contain partial information about the total state. The measured field can be thought of as a low-order representation of the real field. Moreover, the measured fields are commonly contaminated by noise and thus deviate from the true values. Techniques that aim at a full reconstruction of the state vector from available (limited) data are referred to as inverse methods. The data-assimilation approach took its roots in the domain of weather forecasting where predicting the future evolution of both atmospheric and oceanic flows is based on the space–time extrapolation of unequally distributed data of different types measured at weather stations all around the world.

Several types of estimation problems can be distinguished. For example, the full state (for instance, pressure, velocity, temperature, etc.) can be estimated at a fixed time everywhere in space, or the future state-vector trajectory can also be predicted. Mathematically, both of these cases fall into the same category of inverse problems, where full information is retrieved from low-order, limited measurements. Many different methods have been developed to achieve accurate data reconstruction, ranging from simple interpolation techniques to more sophisticated approaches which take advantage of the underlying governing equations of the system. The formulation of such a problem using a variational formulation has been thoroughly studied in the meteorological community (Le Dimet & Talagrand 1986; Courtier 1997; Mohammadi & Pironneau 2004). A review of various data-assimilation techniques used in meteorology is given in Ghil & Malanotte-Rizzoli (1991). Recently, the interest in data-assimilation has reached fluid experimentalists, who wish to extract a maximum amount of information from their measurements. This method can therefore be applied to improve the quality of a particle image velocimetry (PIV) or magnetic resonance imaging (MRI) acquisition. This includes spatial refinement, extension of the measured fields beyond their domain of acquisition, and the reconstruction of unmeasured flow field quantities from measured ones. In this context, it is worth mentioning the 'gappy POD' method proposed by Everson & Sirovich (1995), based on incomplete proper orthogonal decomposition, which has been used successfully to reconstruct missing PIV snapshots (see Gunes, Sirisup & Karniadakis 2006).

This paper focuses on data-assimilation using variational methods and the enforcement of the governing equations with Lagrange multipliers (or adjoint variables). This approach is widely used in the flow optimization community for finding optimal perturbations, forcings or control strategies, to cite but a few applications (Hill 1995; Gunzburger 2000; Luchini & Bottaro 2001; Schmid 2007; Sipp *et al.* 2010). The adjoint variables can be interpreted as sensitivities and therefore yield valuable information on the impact of any kind of changes in the constraints that apply to the system. A variable then has to be chosen as a design parameter (or control parameter) which will be optimized in order to identify an extremum of a cost functional measuring the deviation from the simulated state vector to the dataset of measurements. The control parameters can be of various kinds, e.g. an initial condition (Lundvall, Kozlov & Weinerfelt 2006), a physical parameter, or a material characteristic (Avril *et al.* 2008). The outcome of such data-assimilation techniques is twofold: both the state and the design parameter are identified optimally, as the error

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

between the model-based solution and the true measurements reaches a minimum. Beyond reconstructing the state vector, the application of such a methodology leads to an improvement of the model through the identification of the optimal control variable.

Variational techniques for fluid flow estimation from image sequences have been introduced recently to consistently combine image measurements with constraints expressing that the fluid behaves as a continuum material (Heitz, Mémin & Schnorr 2010). The objective is to add physical constraints to the measurements (here snapshots of optical intensity) to improve the quality of the flow reconstruction in terms of velocity components. Classical PIV algorithms rely on correlation techniques to extract the velocity fields. More physical constraints have already been considered: optical flow model (Ruhnau *et al.* 2005), Stokes equations (Ruhnau & Schnorr 2007) and even time-dependent vorticity transport equations (Ruhnau, Stahl & Schnorr 2007).

In the present paper, contrary to these authors, we directly consider snapshots with velocity components (for example, obtained with a PIV technique). Also, we do not aim at reconstructing a series of flow snapshots obtained at successive times but rather the time-average of these snapshots, the mean flow, and their second-order statistics, the Reynolds stresses. For this, we will use as a regularization the full Reynolds-averaged Navier–Stokes (RANS) equations. It is important to note that this choice of regularization operator (or kernel) is not unique. However, we choose the RANS equations as we believe they describe the physics of the problem accurately. The computational cost should therefore remain reasonable, even in three-dimensional configurations, since only steady-state solutions of the RANS equations and adjoint solutions, which do not involve time, need to be evaluated numerically. The approach employed in this paper can be applied to any unsteady (not necessarily turbulent) flow. Such a flow can, in a first instance, be described by its first statistical moment, the mean flow. Even if the original flow can only be fully understood using both the mean and the unsteady components, we can gain some information about the flow by replacing the full unsteady terms by the second-order momentum, i.e. the Reynolds stress tensor. The goal of the present study is to investigate the possibilities of state-vector reconstruction from sparse mean flow measurements. We presume that the mean (or time-averaged) flow satisfies the RANS equations. In this set of equations, the Reynolds stress tensor appears as an additional unknown, and its definition in terms of the mean quantities is known as the closure problem. However, in our case, this unknown is chosen as a design variable (sometimes referred to as the control parameter) in an optimization process and will be considered as an unknown forcing term in the standard, steady Navier–Stokes equations. We thus identify the full mean flow from sparse data measurements (taken from a direct numerical simulation (DNS)), together with the corresponding optimal forcing, that ensures the averaged flow to be a solution of the RANS equations.

This article presents the mathematical framework of variational data assimilation and applies it to the specific case of time-averaged quantities of unsteady flows. We present, in § 2, the governing equations of the problem as well as the theory for the data-assimilation procedure. Once the optimization algorithm has been derived, we define the test case in § 3. The geometry is presented as well as the base and mean flows around a cylinder at a Reynolds number of $Re = 150$. We then present, in § 4, the results in terms of interpolation and extrapolation of data measurements. In the same section, a more realistic case is considered where the magnitude of the velocity field is measured and the full state vector is retrieved. The ability of the

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

developed optimization algorithm to filter out measurement noise is also investigated. The identified forcing is used in order to characterize the unsteadiness of the flow in § 5. Finally we stress the versatility as well as the many possible improvements of the presented method in § 6 and draw our conclusions.

2. Data assimilation of flow measurements

The presented technique aims at finding a solution of a parameterized model equation, that optimally matches the data measurements. The section starts by introducing the considered model and the type of measure performed. The optimization procedure is then developed in a further section. Finally, uniqueness and other properties of the optimal solution will be briefly discussed.

2.1. *Mean flow considerations*

Any unsteady (laminar or turbulent) flow can be described following the Reynolds decomposition, where the total flow $(u, v, w, p)^\top$ is taken as a sum of a steady term $(\overline{u}, \overline{v}, \overline{w}, \overline{p})^\top$ (the time average) and an unsteady term $(u', v', w', p')^\top$ (the fluctuations around the mean); the $\overline{\cdot}$ operation denotes the average in time. Any fluctuation term q' naturally satisfies the property $\overline{q'} = 0$, and we furthermore have $\partial_t \overline{q} = 0$. By time averaging the Navier–Stokes equations for the total flow, we obtain the so-called steady RANS equations, which read

$$\overline{\boldsymbol{u}} \cdot \boldsymbol{\nabla} \overline{\boldsymbol{u}} + \boldsymbol{\nabla} \overline{p} - Re^{-1} \boldsymbol{\nabla}^2 \overline{\boldsymbol{u}} = \boldsymbol{f}^*, \tag{2.1a}$$

$$\boldsymbol{\nabla} \cdot \overline{\boldsymbol{u}} = 0. \tag{2.1b}$$

In the case of the RANS equations, the forcing term can be expressed as

$$\boldsymbol{f}^* = -\boldsymbol{\nabla} \cdot \boldsymbol{R}, \quad \text{with } R_{ij} = \overline{\boldsymbol{u}_i' \boldsymbol{u}_j'} \tag{2.2}$$

with $\boldsymbol{R}$ being the Reynolds stress tensor, which represents the flow unsteadiness.

By using the incompressibility condition $\partial_i \boldsymbol{u}_i' = 0$, we can write $f_i^* = \overline{\boldsymbol{u}_j' \partial_j \boldsymbol{u}_i'}$. This expression shows that the forcing has to vanish $\boldsymbol{f}^* = 0$ on solid walls, where no-slip boundary conditions apply. Also, we note that $\boldsymbol{\nabla} \cdot \boldsymbol{f}^* \neq 0$ in the bulk of the flow.

A challenge in turbulence research (and, more generally, in any investigation of unsteady flow behaviour) is to model this second-order moment of the velocity field by linking it to the mean flow. This issue is often referred to as the closure problem. This tensor, however, does not explicitly depend on the mean flow, and $-\boldsymbol{\nabla} \cdot \boldsymbol{R}$ can therefore be considered as an independent volumetric forcing term $\boldsymbol{f}^*$ applied to the standard, steady Navier–Stokes equations, as written in (2.1). This forcing term will be considered as a design variable in the following optimization procedure.

2.2. *Data-assimilation and error measure*

The starting point of our reconstruction algorithm is the measured quantity which will be later on referred to as the target field and denoted $\overline{\boldsymbol{m}}$. Motivated by PIV/MRI flow reconstruction, we choose not to consider measurements of the pressure field and only focus on field reconstruction based on velocity-only measurements. The measured field is obtained by the application of a low-rank projection operator $\mathscr{M} : V \to M$ which maps the velocity vectorial field onto a finite-dimensional vector. We denote by V the

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

space of vectorial fields and by M the measure space. This mapping, or measure, can be expressed as

$$\overline{\boldsymbol{m}} = \mathscr{M}\left(\overline{\boldsymbol{u}}\right), \tag{2.3}$$

where $\overline{\boldsymbol{u}} = (\overline{u}, \overline{v}, \overline{w})^{\top}$ is the mean-velocity vector. This operator $\mathscr{M}$ defines the type of data acquisition performed. This measure corresponds to a discrete, low-order representation of the real solution $\overline{\boldsymbol{u}}$. The operator $\mathscr{M}$ also defines the type of quantity observed, and accounts for the spatial quality and location of the measure.

We consider a flow field $(\tilde{u}, \tilde{v}, \tilde{w}, \tilde{p})^{\top}$ to be a solution of the steady, forced Navier–Stokes equations, without any assumption on the forcing term $\boldsymbol{f}$,

$$\tilde{\boldsymbol{u}} \cdot \boldsymbol{\nabla}\tilde{\boldsymbol{u}} + \boldsymbol{\nabla}\tilde{p} - Re^{-1}\nabla^2\tilde{\boldsymbol{u}} = \boldsymbol{f}, \tag{2.4a}$$

$$\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{u}} = 0. \tag{2.4b}$$

This system of equations is completed by homogeneous Dirichlet boundary conditions on solid walls, non-homogeneous Dirichlet boundary conditions for the inflow and appropriate outflow boundary conditions. These boundary conditions will be stated more explicitly when we define the geometry of the example chosen to demonstrate the method. By tuning the forcing term $\boldsymbol{f}$ appropriately, we seek the flow $(\tilde{\boldsymbol{u}}, \tilde{p})^{\top}$ that will best match the measurements $\overline{\boldsymbol{m}}$. The solution of (2.4) with $\boldsymbol{f} = 0$ will be classically referred to as the base-flow solution.

The true mean flow is assumed to satisfy (2.1), with $\boldsymbol{f}^*$ as the true forcing (which is directly derived from the true Reynolds stress tensor). The goal is to find the optimal forcing $\boldsymbol{f}_{opt}$ such that the corresponding velocity field $\tilde{\boldsymbol{u}}_{opt}$ is compatible with the measured quantity $\overline{\boldsymbol{m}}$. To find this optimal forcing, we have to define the distance (error) between the observed quantity $\overline{\boldsymbol{m}}$ and the corresponding measure $\tilde{\boldsymbol{m}}$ of the reconstructed field $\mathscr{M}(\tilde{\boldsymbol{u}})$. The error is thus defined as

$$\mathscr{E}\left(\tilde{\boldsymbol{u}}\right) = \tfrac{1}{2}\left\|\overline{\boldsymbol{m}} - \mathscr{M}\left(\tilde{\boldsymbol{u}}\right)\right\|_M^2 = \tfrac{1}{2}\left\|\Delta\boldsymbol{m}\right\|_M^2 = \tfrac{1}{2}\left\langle\Delta\boldsymbol{m}, \Delta\boldsymbol{m}\right\rangle_M, \tag{2.5}$$

where $\|\cdot\|_M$ is the norm associated with the scalar product $\langle\cdot,\cdot\rangle_M$ which acts on the measure space. Moreover, we implicitly defined in this expression $\Delta\boldsymbol{m}$ as the error field between the real and simulated measure evaluated at each measurement location. The goal is to reduce the error functional $\mathscr{E}$ as much as possible by adjusting the forcing $\boldsymbol{f}$, until a minimum value is reached, at which the optimal forcing $\boldsymbol{f}_{opt}$ emerges.

In this article, the reference velocity field $\overline{\boldsymbol{u}}$ and the measurements are obtained by DNS, thus allowing us to consider the full velocity error field:

$$\Delta\boldsymbol{u} = \overline{\boldsymbol{u}} - \tilde{\boldsymbol{u}}. \tag{2.6}$$

The field $\Delta\boldsymbol{u}$ is a vector field. In a real experimental situation, such an error field cannot be evaluated since the flow $\overline{\boldsymbol{u}}$ is by definition not accessible. By minimizing the error on the measure $\mathscr{E}$, we expect to also decrease the norm of the true error $\Delta\boldsymbol{u}$ and thus reconstruct the field $\overline{\boldsymbol{u}}$.

Note finally that the problem of identifying the optimal forcing $\boldsymbol{f}_{opt}$ is not yet a well-posed problem; in fact, inspection of (2.4) shows that infinitely many solutions exist at this stage. More specifically, starting with a forcing term $\boldsymbol{f}$, a modified forcing $\boldsymbol{f}' = \boldsymbol{f} + \boldsymbol{\nabla}\phi$ will lead to the same solution $\tilde{\boldsymbol{u}}$ with the appropriate modification in the pressure term $(\tilde{p}' = \tilde{p} + \phi)$, ensuring the incompressibility of the velocity field. In compact form, we can write

$$\tilde{\boldsymbol{u}}\left(\boldsymbol{f}, \tilde{p}\right) = \tilde{\boldsymbol{u}}\left(\boldsymbol{f} + \boldsymbol{\nabla}\phi, \tilde{p} + \phi\right). \tag{2.7}$$

In the next section, we will add constraints on $\boldsymbol{f}$ so as to define properly the minimization problem.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

2.3. *Comments on the choice of the reconstruction operator*

In the previous section, we chose the forced Navier–Stokes equations (as presented in (2.4)) as the underlying governing equations for the data-assimilation procedure. However, it is legitimate to ask whether another, simpler kernel (modelling equation) could be used in order to reconstruct the field. An obvious simplification would be to consider the following Stokes operator:

$$\boldsymbol{\nabla}\tilde{p} - Re^{-1}\nabla^2\tilde{\boldsymbol{u}} = \boldsymbol{g}, \tag{2.8a}$$

$$\boldsymbol{\nabla}\cdot\tilde{\boldsymbol{u}} = 0. \tag{2.8b}$$

If such a model was to be chosen, the identified forcing $\boldsymbol{g}_{opt}$ would contain all nonlinear properties of the reconstructed flow. Indeed, if the same minimum of the error functional $\mathscr{E}$ is reached using both methods (Navier–Stokes and Stokes), we can write

$$\boldsymbol{g}_{opt} = \boldsymbol{f}_{opt} - \boldsymbol{u}_{opt}\cdot\boldsymbol{\nabla}\boldsymbol{u}_{opt}. \tag{2.9}$$

It thus seems possible to obtain the same solution with the simpler Stokes kernel. Yet, we verified that the optimization procedure is not well-posed in such a case and that the descent algorithm has difficulties decreasing the objective functional despite numerous iterations (we found that $\mathscr{E}$ decreased by one order of magnitude using 6000 iterations with the Stokes kernel while a decrease by 7 orders of magnitude using 2000 iterations was achieved with the Navier–Stokes kernel, see figure 3 below). We conclude that the optimization space corresponding to $\boldsymbol{f}$ appears to be more convex than the optimization space of $\boldsymbol{g}$. Hence, based on our experiment, for the descent algorithm to be efficient, it is advantageous to treat the convective phenomena by the Navier–Stokes kernel while having the optimization procedure search for the Reynolds stress term $\boldsymbol{f}$.

We conclude that choosing a kernel for the optimization procedure that accurately captures the underlying physics leads to a well-posed optimization problem.

2.4. *Variational formulation*

An optimal forcing $\boldsymbol{f}$ is sought such that the error $\mathscr{E}$ reaches a minimum. The chosen functional $\mathscr{E}$, however, does not explicitly depend on $\boldsymbol{f}$; rather, the dependence on $\boldsymbol{f}$ arises implicitly as $\tilde{\boldsymbol{u}}$ is a solution of (2.4). Therefore, in order to account for this constraint in the optimization, we have to define an augmented functional which not only measures the error $\mathscr{E}$ but also ensures that the flow equations are satisfied. This new functional is called the Lagrangian functional $\mathscr{L}$ and can be formulated as

$$\mathscr{L}(\boldsymbol{f}, \tilde{\boldsymbol{u}}, \tilde{p}, \tilde{\boldsymbol{u}}^\dagger, \tilde{p}^\dagger) = \mathscr{E}(\tilde{\boldsymbol{u}}) - \left\langle \tilde{\boldsymbol{u}}^\dagger, \tilde{\boldsymbol{u}}\cdot\boldsymbol{\nabla}\tilde{\boldsymbol{u}} + \boldsymbol{\nabla}\tilde{p} - Re^{-1}\nabla^2\tilde{\boldsymbol{u}} - \boldsymbol{f}\right\rangle - \left\langle \tilde{p}^\dagger, \boldsymbol{\nabla}\cdot\tilde{\boldsymbol{u}}\right\rangle, \tag{2.10}$$

where $\langle\cdot,\cdot\rangle$ represents the spatial scalar product

$$\langle \boldsymbol{a}, \boldsymbol{b}\rangle = \int_\Omega \boldsymbol{a}\cdot\boldsymbol{b}\,\mathrm{d}\Omega, \tag{2.11}$$

with $\boldsymbol{a}$ and $\boldsymbol{b}$ denoting arbitrary (possibly vectorial) functions of space. To this scalar product is associated the L^2-norm defined as

$$\|\boldsymbol{c}\|_2 = \sqrt{\langle \boldsymbol{c}, \boldsymbol{c}\rangle}, \tag{2.12}$$

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

with $\boldsymbol{c}$ denoting again an arbitrary function of space. Now that we have embedded the constraint in the Lagrangian functional, the forcing appears explicitly in the functional to optimize. However, unconstraining the problem comes at the expense of introducing new (*a priori* unknown) variables, the adjoint state variables $(\tilde{u}^\dagger, \tilde{v}^\dagger, \tilde{w}^\dagger, \tilde{p}^\dagger)^\top$ which are the Lagrange multipliers enforcing the incompressible Navier–Stokes equations.

We are looking for a minimum of the cost functional $\mathscr{L}$, which means that all of the partial functional derivatives of $\mathscr{L}$ have to vanish. We note that enforcing a vanishing first variation with respect to adjoint variables is equivalent to the constraint (2.4). Enforcing the variation with respect to direct variables to be zero yields the adjoint Navier–Stokes equations

$$-\tilde{\boldsymbol{u}} \cdot \boldsymbol{\nabla}\tilde{\boldsymbol{u}}^\dagger + \tilde{\boldsymbol{u}}^\dagger \cdot \boldsymbol{\nabla}\tilde{\boldsymbol{u}}^\top - \boldsymbol{\nabla}\tilde{p}^\dagger - Re^{-1}\nabla^2\tilde{\boldsymbol{u}}^\dagger = \frac{\delta\mathscr{E}}{\delta\tilde{\boldsymbol{u}}}, \tag{2.13}$$

$$\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{u}}^\dagger = 0 \tag{2.14}$$

together with an appropriate set of boundary conditions, stemming from the vanishing of the boundary terms in the functional derivative. Again, the boundary conditions will be detailed later, when the definition of the test-case geometry is presented. The superscript $^\top$ denotes the transpose operation, in the above expression applied to the gradient of the adjoint velocity field $\tilde{\boldsymbol{u}}$.

We observe that the adjoint equations are forced by the functional derivative of the error functional $\mathscr{E}$ with respect to $\tilde{\boldsymbol{u}}$. The forcing will therefore depend on the type of the selected physical measure $\mathscr{M}$, as well as the associated scalar product used to define the error. We can derive the formal expression for the forcing by using the definition of the adjoint of an operator with respect to a scalar product. For example, for any $\boldsymbol{v} \in V$ (V being the space of the velocity vectors), $\boldsymbol{n} \in M$ (M being the space of the measure) and any operator $\mathscr{N} : V \to M$ we can write

$$\langle \mathscr{N}(\boldsymbol{v}), \boldsymbol{n} \rangle_M = \left\langle \boldsymbol{v}, \mathscr{N}^\dagger(\boldsymbol{n}) \right\rangle, \tag{2.15}$$

where the scalar product on the right-hand side is a scalar product on the space of velocity fields V, and $\mathscr{N}^\dagger : M \to V$ denotes the adjoint operator of $\mathscr{N}$ with respect to these scalar products. With this property, we can derive the following expression for the forcing term in the adjoint equation

$$\frac{\delta\mathscr{E}}{\delta\tilde{\boldsymbol{u}}} = -\frac{\delta\mathscr{M}^\dagger}{\delta\tilde{\boldsymbol{u}}}\,\Delta\boldsymbol{m}. \tag{2.16}$$

At this stage, the operator $((\delta\mathscr{M})/(\delta\tilde{\boldsymbol{u}}))^\dagger$ has yet to be defined; the dependence of the forcing term on the error $\Delta\boldsymbol{m}$ is nonetheless evident: the adjoint momentum equation is linearly forced by the error measure $\Delta\boldsymbol{m}$.

The derivation of the gradient with respect to the forcing $\boldsymbol{f}$ is straightforward. By noting that the partial derivative of $\mathscr{L}$ with respect to $\boldsymbol{f}$ is indeed the total derivative of $\mathscr{E}$, the gradient of the error functional with respect to the forcing can be expressed as

$$\boldsymbol{\nabla}_{\boldsymbol{f}}\mathscr{E} = \tilde{\boldsymbol{u}}^\dagger, \tag{2.17}$$

where $\boldsymbol{u}^\dagger$ is solution of (2.13) with appropriate boundary conditions and $\boldsymbol{u}$ is solution of (2.4) with the specified boundary conditions and the considered $\boldsymbol{f}$.

With the gradient of the error with respect to the forcing known, we need to employ a descent algorithm to minimize the error and identify both the optimal

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

forcing $\boldsymbol{f}_{opt}$ and the associated recovered field $\tilde{\boldsymbol{u}}_{opt}$. We will use a conjugate-gradient descent method along with a line-search algorithm. We also need to choose an initial guess $\boldsymbol{f}_g$ in order to start the optimization procedure. We decide to start from a forcing $\boldsymbol{f}_g$ that is divergence-free ($\boldsymbol{\nabla}\cdot\boldsymbol{f}_g = 0$) and zero at the no-slip walls ($\boldsymbol{f}_g = 0$). For example, $\boldsymbol{f}_g = 0$ verifies these conditions.

In the following, we will use $\boldsymbol{f}_{opt}$ to denote the solution that is obtained at the end of the minimization process. It is worth noting that, using a gradient-based approach, this forcing can be expressed as a linear combination of gradients (evaluated at different locations in the optimization space). Therefore, the identified forcing will naturally satisfy $\boldsymbol{\nabla}\cdot\boldsymbol{f}_{opt} = 0$ and $\boldsymbol{f}_{opt} = 0$ on solid boundaries (because $\tilde{\boldsymbol{u}}^{\dagger}$ and the initial guess $\boldsymbol{f}_g$ satisfy these conditions). Therefore, the use of the present iterative gradient-based method combined with the above-mentioned choice of the initial condition implicitly imposes additional constraints on the choice of the forcing $\boldsymbol{f}$ that minimizes the cost-functional $\mathscr{E}$. Uniqueness of the resulting solution $\boldsymbol{f}_{opt}$ and its relation to the true forcing $\boldsymbol{f}^*$ will be discussed in § 2.6. Finally, the flow field $(\tilde{\boldsymbol{u}}_{opt}, \tilde{p}_{opt})$ will refer in the following to the solution of (2.4) with the forcing $\boldsymbol{f}_{opt}$.

2.5. *Type of measure*

As detailed in the previous section, the error functional $\mathscr{E}$ is entirely defined by the operator $\mathscr{M}$ which is at the heart of the data-assimilation technique. Here, we will consider operators that can be decomposed into two operators $\mathscr{P}$ and $\mathscr{Q}$, respectively, describing the projection from the true solution (having an infinite number of degrees of freedom, space V) to a low-rank representation of the field, i.e. the sampled data points (with, in practice, a finite number of degrees of freedom, space M), and the projection from the vector of velocities to whatever quantity is indeed observed. According to this definition, we can write

$$\mathscr{M}(\boldsymbol{u}) = \mathscr{P}(\mathscr{Q}(\boldsymbol{u}))\,. \tag{2.18}$$

The simplest operator $\mathscr{Q}$ one can imagine is the identity, indicating that the measure is performed on all of the components of the velocity field, such as for PIV measurements. However, many other measures are possible, for example observation of a single component of the velocity field. In this case, we would have $\mathscr{Q}(\boldsymbol{u}) = u$ which would be a scalar field containing the streamwise velocity only. In any case, for vectorial or scalar measured quantities alike, we will use

$$\mathscr{Q}(\boldsymbol{u}) = \boldsymbol{q}. \tag{2.19}$$

At this stage, the vector $\boldsymbol{q}$ still belongs to an infinite-dimensional space. The operator $\mathscr{P}$, acting on $\boldsymbol{q}$, describes the 'geometrical' features of the measure. This operator is responsible for the discretization of the continuous field $\boldsymbol{q}$ onto a finite-dimensional vector containing all of the data points; the projection operator $\mathscr{P}$ defines the spatial sampling quality and the spatial extent of the measure. The measure can, for instance, be the evaluation of $\boldsymbol{q}$ at a finite number N of locations in the domain

$$\left[\mathscr{P}(\boldsymbol{q})\right]_i = \boldsymbol{q}(\boldsymbol{x}_i) = \int_{\Omega} \boldsymbol{q}(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{x}_i)\,\mathrm{d}\Omega\,, \tag{2.20}$$

where $\delta(\boldsymbol{x}_i)$ is the delta function centred on the coordinate points $\boldsymbol{x}_i = (x_i, y_i, z_i)^{\top}$.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

However, a more general type of measure would be a weighted, local average of the real field over small elements of the domain Ω_i such as

$$\left[\mathscr{P}(\boldsymbol{q})\right]_i = \int_\Omega \boldsymbol{q}(\boldsymbol{x}) b_i(\boldsymbol{x})\,\mathrm{d}\Omega, \tag{2.21}$$

where b_i is the weight function associated with element Ω_i. The above-mentioned case with point-wise measure would correspond to $b_i = \delta(\boldsymbol{x}_i)$. An average over each measurement cell Ω_i can be described by the weight function $b_i(\boldsymbol{x}) = H(\Omega_i)$, where $H(\Omega_i)$ is equal to $1/V_{\Omega_i}$ (V_{Ω_i} being the volume of the element Ω_i, $V_{\Omega_i} = \int_{\Omega_i} \mathrm{d}\Omega_i$) for $\boldsymbol{x} \in \Omega_i$ and zero everywhere else. At this point, we decide to stay as general as possible by not specifying the basis functions b_i. The measurement mesh, and therefore the projection operator $\mathscr{P}$ is entirely defined by the basis functions $(b_i)_{i\in[\![1,N]\!]}$.

In the case of a discrete measure operator $\mathscr{P}$, the space M is a finite-dimensional space. In this case, the scalar product on the space M will be defined as the classical vectorial dot product (the sum of the component-wise products). According to the previous definition, and using the usual vectorial scalar product for $\langle\cdot,\cdot\rangle_M$, we can express the error defined in (2.5) as

$$\mathscr{E}(\tilde{\boldsymbol{u}}(\boldsymbol{f})) = \frac{1}{2}\sum_{i=0}^{N} \Delta m_i^2, \quad \text{with } \Delta m_i = \overline{m}_i - \int_\Omega \boldsymbol{q} b_i\,\mathrm{d}\Omega, \tag{2.22}$$

and then compute its derivative as

$$\left\langle \frac{\delta\mathscr{E}}{\delta\tilde{\boldsymbol{u}}}, \delta\tilde{\boldsymbol{u}} \right\rangle = -\sum_{i=0}^{N} \Delta m_i \int_\Omega \frac{\delta\mathscr{Q}}{\delta\tilde{\boldsymbol{u}}} \delta\tilde{\boldsymbol{u}} b_i\,\mathrm{d}\Omega. \tag{2.23}$$

After rearranging the integral and the sum, we find

$$\frac{\delta\mathscr{E}}{\delta\tilde{\boldsymbol{u}}} = -\sum_{i=0}^{N} \frac{\delta\mathscr{Q}}{\delta\tilde{\boldsymbol{u}}} b_i \Delta m_i. \tag{2.24}$$

This expression is in the space of velocities V and corresponds to the forcing term in the adjoint equations; it is proportional to the error made at each measure location (over the domain Ω_i in the case of averaged measures). If the point-wise measurement had been used instead, we would have a sum of Dirac distributions centred on each measurement point. We see that the averaged evaluation (2.21) is more regular than the discrete one because the field $\delta\mathscr{E}/\delta\tilde{\boldsymbol{u}}$ is piecewise continuous. Moreover, it is mathematically closer to a real physical measure.

Moreover, from (2.16) and (2.24), we identify

$$\frac{\delta\mathscr{M}}{\delta\tilde{\boldsymbol{u}}}^{\dagger} \Delta\boldsymbol{m} = \sum_{i=0}^{N} \frac{\delta\mathscr{Q}}{\delta\tilde{\boldsymbol{u}}} b_i \Delta m_i. \tag{2.25}$$

2.6. *Relation between $\boldsymbol{f}_{opt}$ and $\boldsymbol{f}^*$*

The link between the optimal solution $\boldsymbol{f}_{opt}$ with the true forcing $\boldsymbol{f}^*$ is an important issue to be discussed. We start by recalling that on the one hand we have $\nabla\cdot\boldsymbol{f}_{opt} = 0$ with $\boldsymbol{f}_{opt} = 0$ on solid-wall boundaries (see § 2.4), while on the other hand we have

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

$\nabla \cdot \boldsymbol{f}^* \neq 0$ with $\boldsymbol{f}^* = 0$ on the walls (see § 2.1). It is therefore tempting to compare $\boldsymbol{f}_{opt}$ with the divergence-free part of $\boldsymbol{f}^*$ and discuss their relation.

The real forcing $\boldsymbol{f}^*$ (as expressed in (2.2)) can be decomposed into

$$\boldsymbol{f}^* = \boldsymbol{f}^*_s + \nabla\phi, \tag{2.26}$$

where $\boldsymbol{f}^*_s$ is a divergence-free part of $\boldsymbol{f}^*$ and $\nabla\phi$ a potential component. By taking the divergence of this equation and setting $\nabla \cdot \boldsymbol{f}^*_s = 0$, we find

$$\nabla \cdot \boldsymbol{f}^* = \nabla^2\phi. \tag{2.27}$$

In order to uniquely define the above decomposition, boundary conditions have to be chosen for ϕ on the solid walls. In order for $\boldsymbol{f}^*_s$ to be closest to the identified forcing $\boldsymbol{f}_{opt}$, we would like to set $\nabla\phi = 0$ at the no-slip walls. Yet, for a Poisson equation, we cannot prescribe simultaneously the tangential and normal components of $\nabla\phi$ to be zero on the no-slip walls and it is only possible to impose the normal component to zero:

$$\partial_n\phi = \nabla\phi \cdot \boldsymbol{n} = 0, \tag{2.28}$$

with $\boldsymbol{n}$ as the outward normal to the domain. Solving (2.27) with this homogeneous Neumann condition then allows us to find ϕ and therefore to fully identify the projected forcing $\boldsymbol{f}^*_s$ from (2.26). The normal (with respect to the no-slip boundary) component of $\boldsymbol{f}^*_s$ therefore vanishes on the no-slip walls, but its tangential component does not. Hence, the divergence-free part of $\boldsymbol{f}^*$, i.e. $\boldsymbol{f}^*_s$, is not expected to be exactly equal to $\boldsymbol{f}_{opt}$, since on the no-slip walls their tangential component is not equal. Yet, we will see below (see § 4.1 in the case of full-state information identification) that these fields are nearly identical in the whole space except in the vicinity of the solid walls. Also, we will check that the iterative process converges toward a very small value of $\mathscr{E}(\tilde{\boldsymbol{u}}_{opt})$.

The reconstructed pressure will also be different from the real pressure $\bar{p}$. In fact, since $\boldsymbol{f}_{opt} \simeq \boldsymbol{f}^* - \nabla\phi$, it is seen that

$$\tilde{p}_{opt} \simeq \bar{p} - \phi, \tag{2.29}$$

where $\tilde{p}_{opt}$ is the pressure recovered by our algorithm and $\bar{p}$ is the true pressure as defined in (2.1). The reconstructed pressure $\tilde{p}_{opt}$ is therefore an augmented pressure which includes the potential part of the forcing ϕ. We will discuss the possibilities of pressure reconstruction in § 5.3.

It may appear striking to the reader to look for an optimal solution $\boldsymbol{f}_{opt}$ in a space which, by definition, does not contain the true forcing $\boldsymbol{f}^*$. However, in the case of a kernel based on incompressible equations, the sole knowledge of velocity measurements automatically induces that the gradient of the error $\mathscr{E}$ with respect to the forcing $\boldsymbol{f}$ is divergence-free. Therefore, an optimization procedure based on a gradient with an incompressibility constraint can only find an optimal solution in the space of the divergence-free functions.

In order to obtain some gradient information in a wider space (containing non-divergence-free functions), we can, along with velocity data, consider pressure measurements as input variables to the data-assimilation algorithm. In that case, the adjoint field is no longer incompressible, since the cost functional now depends also on $\tilde{p}$; we then have

$$\nabla \cdot \tilde{\boldsymbol{u}}^\dagger = -\frac{\delta\mathscr{E}}{\delta\tilde{p}}. \tag{2.30}$$

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

The gradient $\nabla_f \mathscr{E}$ (see (2.17)) and consequently the identified forcing are no longer divergence-free, and therefore matching both the solenoidal and irrotational components of the forcing becomes possible.

In this study, we restrict ourselves to a target field exclusively composed of velocity variables, as it is of both experimental and theoretical interest.

2.7. *Comments on uniqueness and regularization issues*

Reconstructing a steady flow field from only limited velocity measurements poses mathematical and algorithmic difficulties which originate from the chosen measurement norm $\| \cdot \|_M = \sqrt{\langle \cdot, \cdot \rangle_M}$. This (semi-)norm describes the error and thus the cost functional of our variational principle and is based on evaluating the velocity field at only a few available measurement points. Therefore, it is conceivable that flow fields could be recovered that accurately reproduce the measurement values (with a vanishing error $\mathscr{E}$) but yield unphysical behaviour in the rest of the domain. In fact, an infinite number of recovered flow fields with $\mathscr{E} = 0$ are feasible. One has to keep in mind, however, that the augmented Lagrangian $\mathscr{L}$ enforces the governing model equations over the entire computational domain, as reflected in the use of the scalar product $\langle \cdot, \cdot \rangle$ to incorporate the physics-based constraints given by the RANS model. Even though the descent algorithm aims to minimize the error $\mathscr{E}$ (given in the $_M$-norm), during each iterative step the candidate solutions for the recovered flow field $\tilde{\boldsymbol{u}}$ are subjected to adjustments, stemming from the adjoint solution $\tilde{\boldsymbol{u}}^\dagger$, to conform to the driven RANS equations (and boundary conditions) within the entire computational domain. The complete descent algorithm applied to this mixed-norm optimization problem is hence imposing a regularization component to the space of all possible flow-field solutions, which penalizes flow fields that may match the measurement values, but are not solutions of the driven RANS model. Despite this positive influence, uniqueness of the converged solution can neither be guaranteed nor assumed. This is to be expected from an iterative descent algorithm searching for a minimum of a nonlinear cost functional. A rigorous analysis concerning convergence to a global minimum, uniqueness of the converged solution or the regularizing influence of the governing and adjoint equations is beyond the scope of this study. Despite these open questions, we will show in the next sections that physically relevant flow fields can be recovered successfully from only few (and limited) measurements; in this respect, the presented technique should prove helpful and valuable in extracting flow information that goes beyond available measurements.

3. Flow around a cylinder

We apply the data-assimilation method described above to a simple test case: the two-dimensional flow around an infinite circular cylinder. Even though the theory was presented for a general three-dimensional flow, it applies straightforwardly to a two-dimensional case. The geometry, as well as the corresponding boundary conditions are presented in figure 1. We restate the boundary conditions of the direct system (2.4) for clarity:

$$\left.\begin{array}{ll} \tilde{u} = 1, \quad \tilde{v} = 0 & \text{at the inlet,} \\ \tilde{u} = 0, \quad \tilde{v} = 0 & \text{on the cylinder's surface,} \\ \partial_y \tilde{u} = 0, \quad \tilde{v} = 0 & \text{on symmetry boundaries,} \\ Re^{-1} \partial_x \tilde{u} - \tilde{p} = 0, \quad \partial_x \tilde{v} = 0 & \text{at the outlet.} \end{array}\right\} \tag{3.1}$$

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

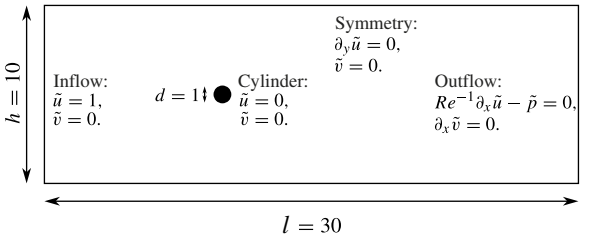


FIGURE 1. Sketch of the geometry of two-dimensional flow around an infinite cylinder. The full computational domain, the characteristic lengths and the applied boundary conditions are shown. The cylinder's diameter, the height and length of the domain d, h and l are given in non-dimensional units.

The boundary conditions of the adjoint system (2.13) are obtained via the integration-by-parts step as explained in § 2.4. These conditions read

$$\left.\begin{array}{l} \tilde{u}^\dagger = 0, \quad \tilde{v}^\dagger = 0 \quad \text{at the inlet,} \\ \tilde{u}^\dagger = 0, \quad \tilde{v}^\dagger = 0 \quad \text{on the cylinder's surface,} \\ \partial_y \tilde{u}^\dagger = 0, \quad \tilde{v}^\dagger = 0 \quad \text{on symmetry boundaries,} \\ Re^{-1}\partial_x \tilde{u}^\dagger + \tilde{p}^\dagger = -\tilde{u}\tilde{u}^\dagger, \quad Re^{-1}\partial_x \tilde{v}^\dagger = -\tilde{u}\tilde{v}^\dagger \quad \text{at the outlet.} \end{array}\right\} \tag{3.2}$$

This flow has been shown (see Jackson 1987) to undergo a first transition (more precisely, a supercritical Hopf bifurcation) at a critical Reynolds number of $Re_c \simeq 46$, above which the base flow (solution of the system (2.4) with $\boldsymbol{f} = 0$) is no longer stable. Beyond this threshold, the flow becomes time periodic, and vortices are shed from the back of the cylinder. After this transition, we can define any mean quantity by averaging its instantaneous value over a finite number of periods in. By doing so, we can compute the mean-velocity field $\overline{\boldsymbol{u}}$ as well as the different components of the Reynolds stress tensor $\overline{u_i' u_j'}$, defined in (2.2).

As we mentioned earlier, the optimization method employed in the remainder of this article requires the definition of an initial guess $\boldsymbol{f}_g$ for the forcing. We choose $\boldsymbol{f}_g = 0$, which means that the first computed flow $\tilde{\boldsymbol{u}}$ is the base flow $\boldsymbol{u}_b$.

For our simulations, we choose a Reynolds number of $Re = 150$ and compute the base flow $\boldsymbol{u}_b$ (using a classical Newton method), the mean flow $\overline{\boldsymbol{u}}$ (using time averaging) and the various components of the Reynolds stress tensor. These fields are computed with high accuracy using a finite-element method, using FreeFem++ software (see www.freefem.org), on a mesh of $N \simeq 1.7 \times 10^5$ degrees of freedom. The base flow is represented in figure 2(*a*,*b*); the mean flow is displayed in figure 2(*c*,*d*) and finally the initial error is plotted in figure 2(*e*,*f*).

4. Navier–Stokes-driven field reconstruction

To validate the presented data-assimilation method, several measure operators $\mathscr{M}$ are considered, each of them chosen to demonstrate the efficacy of the method in various reconstruction scenarios. In the two following sections the considered measure is the full velocity vector and the error on both components of the velocity is estimated. Spatial interpolation and extrapolation are investigated. In a further section, we consider a more challenging and realistic case where only the velocity magnitude is measured, on a relatively coarse mesh. We demonstrate the ability of

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

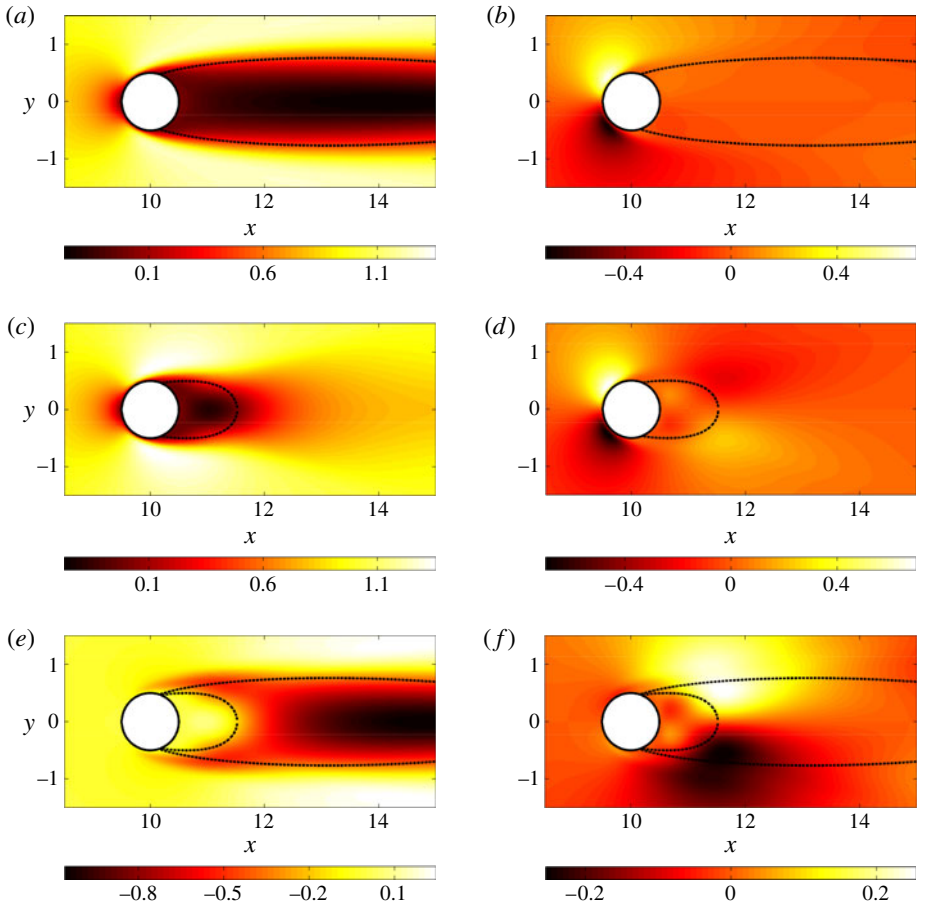


FIGURE 2. Flow around an infinite cylinder at $Re = 150$. (*a*) and (*b*) Base flow $\boldsymbol{u}_b$; (*c*) and (*d*) time-averaged flow $\overline{\boldsymbol{u}}$; (*e*) and (*f*) Difference (evaluated on both components) $\Delta\boldsymbol{u}$, measuring the error committed when approximating the mean flow by the base flow. The left column represents the x-component of the corresponding vector field, while the right column displays the y-component. The dashed-lines refer for each vector-field to the extent of the recirculation bubble. The dashed lines of (*a–d*) have been reproduced in (*e*) and (*f*) for comparison.

the method to not only reconstruct the measured quantity but also to identify the full velocity vector. The robustness of the identification algorithm when the measure is corrupted by noise is also assessed.

4.1. *Full-state information identification*

As a starting point, we will choose the operator $\mathscr{M}$ to be the identity operator on V, such that the assimilated field is the full, continuous velocity field (in practice, it is a discrete field, evaluated on the finite-element grid). In this case, we choose $\langle\cdot,\cdot\rangle_M \equiv \langle\cdot,\cdot\rangle$ and $\|\cdot\|_M \equiv \|\cdot\|_2$. We present the convergence curves for the case $\mathscr{M} = \mathscr{I}$ for both the cost functional $\mathscr{E}$ and the norm of the gradient $\nabla_f\mathscr{E}$ in figure 3. It is seen that both quantities decrease by several orders of magnitude, indicating that the error $\mathscr{E}$ becomes extremely weak in the case of full-state information identification.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

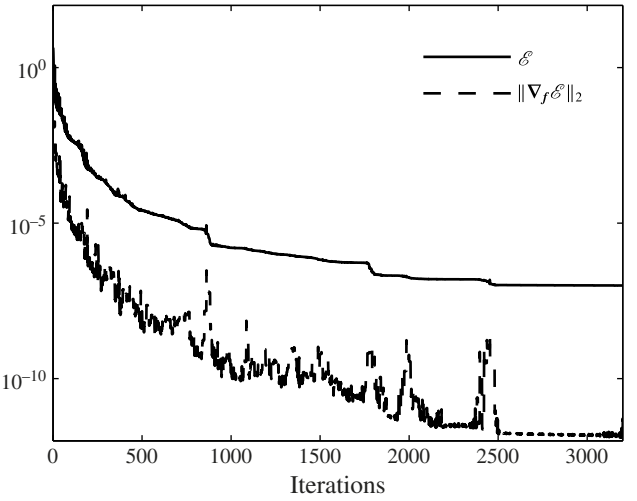


FIGURE 3. Convergence curves for full-identification case ($\mathscr{M} = \mathscr{I}$). The solid line represents $\mathscr{E}$,and the dashed line represents $\|\nabla_f \mathscr{E}\|_2$.

We plot in figure 4(*a*,*b*) both components of the true forcing $\boldsymbol{f}^*$ (computed via a DNS) and in figure 4(*e*,*f*) those related to the identified forcing $\boldsymbol{f}_{opt}$. There exists a strong correlation for the streamwise component, but no similarity can be found for the cross-stream component. This was expected because the identification algorithm yields a forcing $\boldsymbol{f}_{opt}$ which is divergence-free while $\boldsymbol{f}^*$ is not. The divergence-free component of the latter forcing, $\boldsymbol{f}_s^*$, is computed following the procedure explained in § 2.6 and we plot the results in figure 4(*c*,*d*). We can see that the identified forcing $\boldsymbol{f}_{opt}$ is indeed matching the projected forcing $\boldsymbol{f}_s^*$. However, some discrepancies appear on the cylinder's surface, which are due to the different tangential values of these two fields on the cylinder surface.

Differences between $\boldsymbol{f}_{opt}$ and the projection of $\boldsymbol{f}^*$ over the space of divergence-free fields are best analysed by comparing the curl of these two fields, since the curl of a potential field is zero. We can observe in figure 4(*g*,*h*) that the z-component of these fields is nearly the same everywhere except in the vicinity of the cylinder boundary. We therefore have $\boldsymbol{f}_{opt} = \boldsymbol{f}_s^*$ almost everywhere, which validates our optimization procedure.

4.2. *Partial-state information identification*

In the case of a partial-state information identification, we still minimize $\mathscr{E}$ as defined in (2.5), expecting that this minimization results in a decrease of the real error $\Delta\boldsymbol{u}$ (see (2.6)). We therefore need to decide how to measure the real error $\Delta\boldsymbol{u}$. We choose to use two norms: the L^2-norm (normalized by $\sqrt{V_\Omega}$ to remove any dependence on the chosen area of integration) and the L^∞-norm. The first will give information about the average error while the second will yield the error at the worst reconstructed location. We therefore define the following residuals (measuring the real errors):

$$\left.\begin{aligned} r_2 &= \frac{1}{\sqrt{V_\Omega}}\,\|\Delta\boldsymbol{u}\|_2 = \sqrt{\frac{1}{V_\Omega}\int_\Omega |\Delta\boldsymbol{u}|^2\ \mathrm{d}\Omega}, \\ r_\infty &= \|\Delta\boldsymbol{u}\|_\infty = \lim_{p\to+\infty}\left(\int_\Omega |\Delta\boldsymbol{u}|^p\ \mathrm{d}\Omega\right)^{1/p}. \end{aligned}\right\} \tag{4.1}$$

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

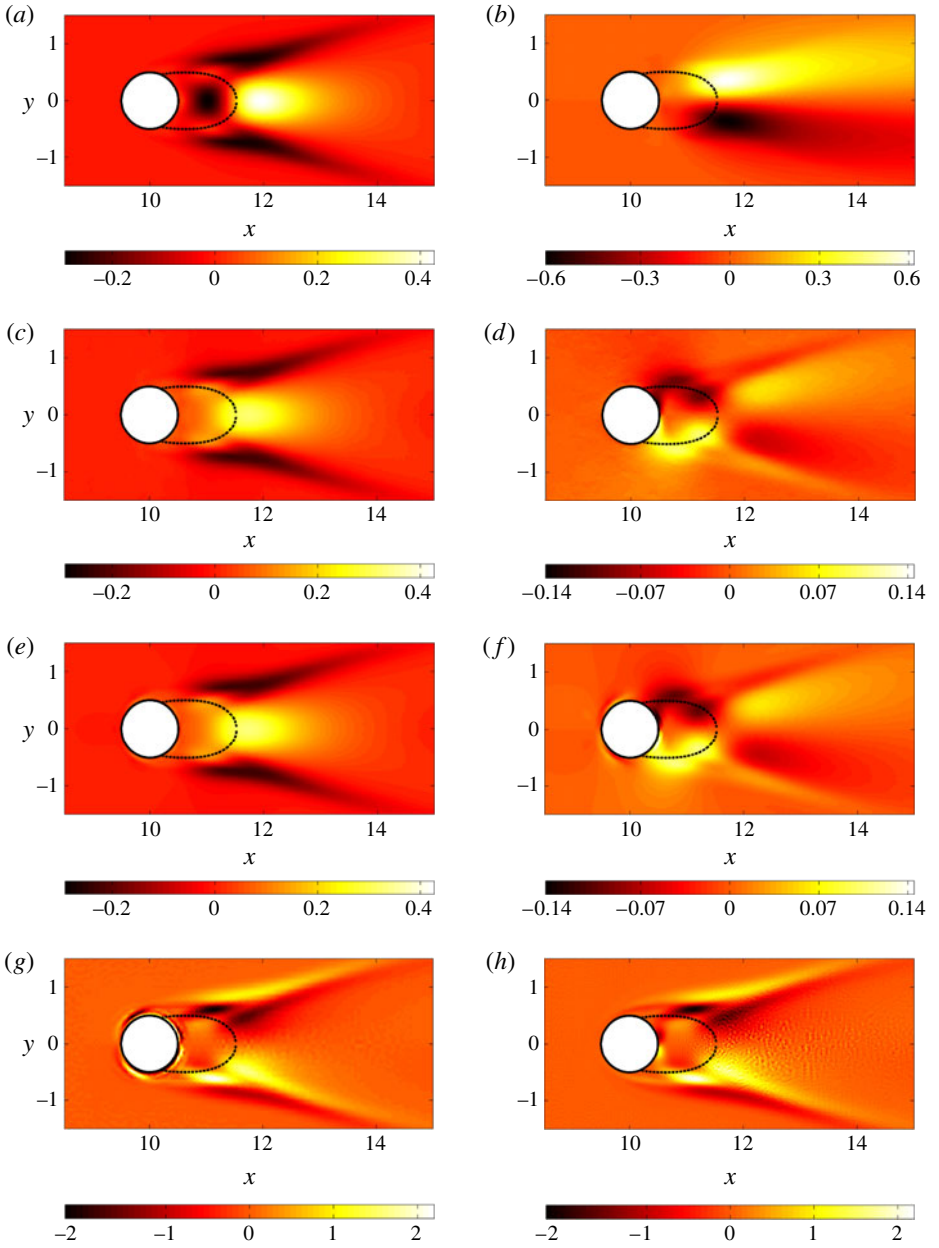


FIGURE 4. Averaged forcing term $\boldsymbol{f}^*$ displaying the x and y components, respectively, in (a) and (b). Projected forcing $\boldsymbol{f}_s^*$ (see § 2.6 for its definition) for both components, in (c) and (d). Optimally identified forcing $\boldsymbol{f}_{opt}$ for both components, in (e) and (f). Note that the colour scales are identical for all streamwise components, but vary for the cross-stream components. However, (d) and (f) have the same colour scale to allow comparison. Parts (g) and (h) represent the z-component of $\boldsymbol{\nabla} \times \boldsymbol{f}^*$ and $\boldsymbol{\nabla} \times \boldsymbol{f}_{opt}$, respectively. The difference is mainly located close to the cylinder's surface. The dashed lines refer for each vector field to the extent of the recirculation bubble.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

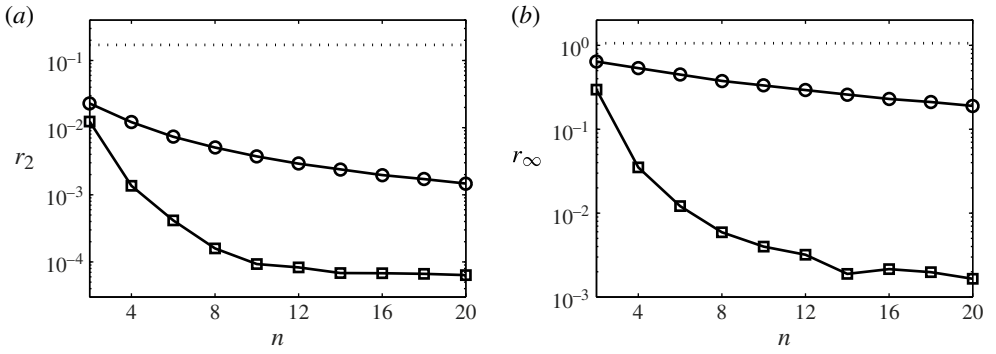


FIGURE 5. Results for the interpolation case: (*a*) r_2 error; (*b*) r_∞ error as defined in (4.1). The norm of the error is plotted against the line density of nodes n. Circle symbols represent the spline interpolation case (naïve approach) and square symbols represent the data-assimilation results. The error of approximating the mean flow by the base flow is represented with a dashed line. For $n \simeq 10$, the reconstruction of the field is efficient and a further increase of resolution yields only little improvement.

4.2.1. *Interpolation*

Interpolation consists of reconstructing the field in between the available, measured data points. In what follows, the measure (as defined in (2.21)) is an average over each element of a rectangular mesh (called measurement mesh) and is reconstructed on the computational mesh which has a very high resolution. Here, $\mathscr{Q}$ is the identity operator, i.e. the two components of the velocity vector are measured:

$$\mathscr{Q}(\tilde{\boldsymbol{u}}) = \tilde{\boldsymbol{u}}. \tag{4.2}$$

In order to characterize the efficacy of the field reconstruction method, the discretization projection operator $\mathscr{P}$ defined in (2.21) is considered, with the b_i being the basis functions of a mesh composed of squares paving the whole computational domain. The different measurement meshes have different line density (density per unit length) of points $n \in [2, 20]$. For instance, the case $n = 2$ corresponds to 60×20 measurements meshes. An integrated measure is performed on each of the elements composing the measurement mesh, according to the expression (2.21). The identification algorithm is applied to all of the above measurement meshes and, for each optimum found, the errors r_2 and r_∞ as defined in (4.1) are plotted against the line-density of points n in figure 5.

First, it is encouraging that even for the worst case considered ($n = 2$), the reconstructed mean flow corresponds to a relatively low error. The measure $\mathscr{M}(\overline{\boldsymbol{u}})$ performed in a particularly under-resolved case ($n = 2$) as well as the corresponding reconstructed field $\tilde{\boldsymbol{u}}_{opt}$ and error $\Delta\boldsymbol{u}$ are plotted in figure 6. From this figure, we can confirm that the data-assimilation algorithm successfully identifies the mean field. Indeed, for this extremely coarse case, the L^2 error is $r_2 = 0.012$, which is very small considering the amount of information initially given. However, the maximum error is $r_\infty = 0.297$. We have found an average error of 1.2 %, and of 29.7 % at the worst reconstructed location. Both of these errors are relative to the unit inflow velocity. We see in figure 6(*e*,*f*) that the error is mainly located in the near-wake of the cylinder. The worst-reconstructed area corresponds to the end of the recirculation bubble whose reattachment point location does not exactly match the real one (extracted from the

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

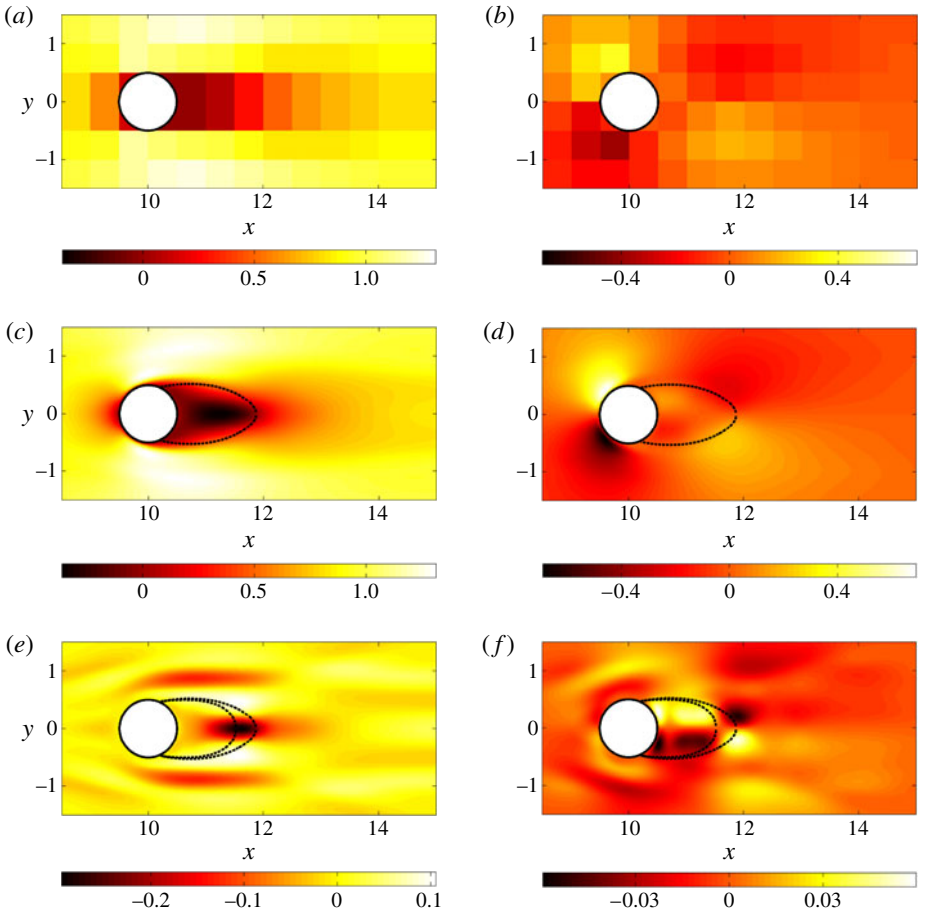


FIGURE 6. Measures of the two velocity components: (*a*) $\mathscr{M}(\overline{u})$; (*b*) $\mathscr{M}(\overline{v})$. The measurement corresponds to a rectangular mesh of line density of points $n = 2$, and the measurement window is the full flow field. Reconstructed fields: (*c*) $\tilde{u}$; (*d*) $\tilde{v}$. Error fields: (*e*) Δu; (*f*) Δv. The dashed lines in (*c*) and (*d*) refer to the extent of the recirculation bubble of the reconstructed field. These dashed lines have been reproduced in (*e*) and (*f*), along with those of the true mean flow.

mean flow $\overline{\boldsymbol{u}}$); its length is underestimated by the identification algorithm. For more accurate measures (increasing n), the error decreases: the more information is available initially, the more accurately the field is reconstructed. For the finest meshes, we can reduce the errors to $6.36 \times 10^{-3}\,\%$ in L^2-norm and $0.17\,\%$ in L^∞-norm.

Spline interpolation is a non-physical, but common way to obtain highly refined fields from measurements on a coarse grid. To assess the improved quality of data-assimilation interpolation, figure 5 also presents the interpolation error for a classical spline interpolation reconstruction. We note that this simple interpolation technique is producing better results than a base-flow approximation, but is far from being as accurate as the presented algorithm. The data-assimilation method always outperforms spline interpolation and can lead to more accurate results by two orders of magnitude.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

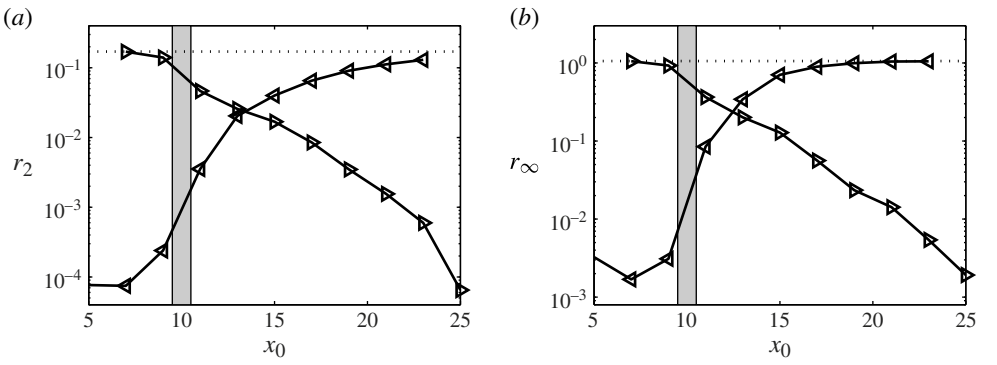


FIGURE 7. Results for the extrapolation case: (*a*) r_2 error; (*b*) r_∞ error as defined in (4.1). The error $\|\Delta \boldsymbol{u}\|$ is plotted against the streamwise coordinate x_0 defining the extent of the identification window. Right-oriented triangles are associated with the measurement window $W_1 = [5,\ x_0]$, while the left-oriented triangles correspond to the identification window $W_2 = [x_0,\ 25]$. The error of approximating the mean flow by the base flow is represented by a dashed line. The grey rectangle represents the location of the cylinder.

4.2.2. *Extrapolation*

In this section, the operator defining the measured quantity is still the identity ($\mathscr{Q} = \mathscr{I}$), i.e. the full velocity vector is assimilated on a given measurement mesh. However, the projection operator $\mathscr{P}$ is now designed to investigate the extrapolation capabilities of the method. The used mesh is the finer one from the previous section (the line density of points is $n = 20$), but locations, where measurements are performed, are limited in space. More precisely, the assimilation (or identification) window has the height of the computational domain but varies in length. For a given abscissa x_0, we consider two different identification windows: the first starting at $x = 5$ and ending at $x = x_0$, and the second starting at $x = x_0$ and ending at $x = 25$. The two identification windows are thus $W_1 = [5, x_0]$ and $W_2 = [x_0, 25]$.

The results are presented in figure 7. The first series of simulations for the identification window $W_1 = [5, x_0]$ shows that the error is decreasing with the length of the identification window. The algorithm starts to produce accurate results as soon as the zone around the cylinder is included in the assimilation window (from $x_0 = 11$). In particular, if the data assimilated does not contain any measures where the mean flow and base flow are different (unsteady zones), the identification will not be able to match the mean flow, since hardly any information is provided. However, as soon as some unsteady zones are assimilated, the algorithm is able to reconstruct not only the identified zone, but the full flow field. This demonstrates that extrapolation is a true capability of the presented data-assimilation method. The second series of simulations corresponds to the identification window $W_2 = [x_0, 25]$. In this case, even for small identification windows (starting from $x_0 = 23$), the mean and base flows are different in the considered zone. Therefore, the extrapolation is relatively effective, even with little information, as we can see by considering the L^2-norm plot of the error (figure 7*a*). However, the L^∞ residual r_∞ remains rather large (see figure 7*b*).

The results of the extrapolation for the identification window $W_2 = [13,\ 25]$ (corresponding to $x_0 = 13$) are presented in figure 8. We can see that only limited information is available downstream of the cylinder; in particular, the location of the reattachment point of the recirculation bubble is outside the identification window.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

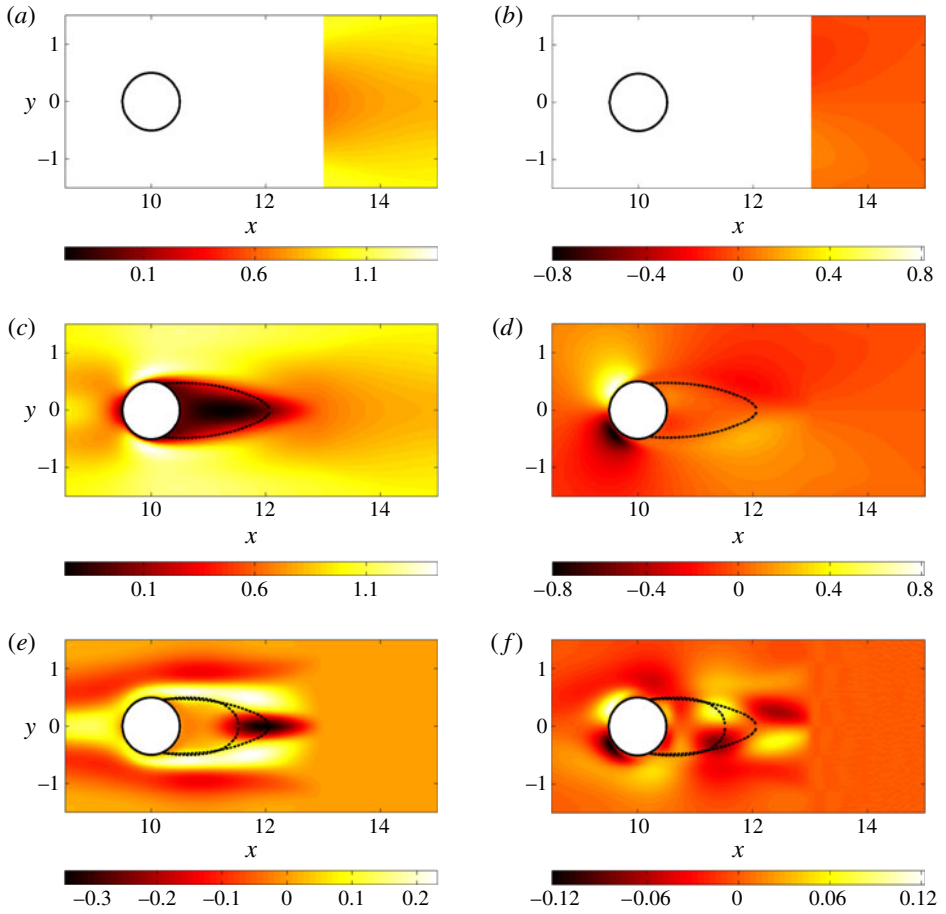


FIGURE 8. Measures of the two velocity components: (*a*) $\mathscr{M}(\overline{u})$; (*b*) $\mathscr{M}(\overline{v})$. The measurement corresponds to a rectangular mesh of line density of points $n = 20$ and the measurement window is $W_2 = [13, 25]$ ($x_0 = 13$). Reconstructed fields: (*c*) $\tilde{u}$; (*d*) $\tilde{v}$. Error fields: (*e*) Δu; (*f*) Δv. The dashed lines in (*c*) and (*d*) refer to the extent of the recirculation bubble of the reconstructed field. These dashed lines have been reproduced in (*e*) and (*f*), along with those of the true mean flow.

Despite this lack of information, the reconstructed field $\tilde{\boldsymbol{u}}$ is very similar to the mean flow $\overline{\boldsymbol{u}}$. However, some non-negligible errors in the reconstruction appear immediately upstream of the identification window (nearly no error can be found within the identification window). Interestingly, the incoming flow field has been modified from a uniform flow to a y-dependent flow field (the uniform inflow boundary condition is still satisfied further upstream), as we can see in figure 8(*c*), illustrating that several types of solutions can be found depending on the quality of the measure. In fact, it is likely that the cost functional is multimodal (several local minima may exist), associated with different flow fields, one of which is the physical solution of the problem. There is however no guarantee, as demonstrated here, of finding this physical solution. Nevertheless, even if the mean flow is not exactly matched, the reconstructed field yields far more information about the nature of the flow than is contained in the initial information we measured.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

The results of the extrapolation procedure indicate that the field is best reconstructed as long as the recirculation bubble is part of the measurement window. This finding is reminiscent of the results by Marquet, Sipp & Jacquin (2008) and Pralits, Brandt & Giannetti (2010) who showed that the recirculation bubble is the most sensitive region regarding the overall dynamics of the flow field.

4.2.3. *State vector reconstruction*

In order to approach real cases for which this method is likely to be employed, we proceed from the full two-component velocity matching to a scalar-field matching based on the velocity magnitude. The operator defining the measured quantity is therefore

$$\mathscr{Q}(\boldsymbol{u}) = \tfrac{1}{2}\,|\boldsymbol{u}|^2 = \tfrac{1}{2}\left(u^2 + v^2\right). \tag{4.3}$$

The projection operator $\mathscr{P}$ is chosen such that the measure domain is the entire domain and the line density of points of the measures is $n = 6$. The measured field $\tilde{\boldsymbol{m}}$ is plotted in figure 9(*a*). The reconstructed field from data assimilation of the measurements is presented in figure 9(*b*,*c*).

We observe that even though only velocity magnitude on a rather coarse mesh has been used as an input to our data-assimilation algorithm, it is possible to retrieve a good approximation of the full component-wise mean-velocity field. Both the streamwise and normal velocity components have been recovered accurately, with a small error concentrated around the recirculation zone in the near-wake of the cylinder.

4.3. *Noise reduction*

Experimental measurements are often contaminated by noise, and for an application of data assimilation to experimental data the ability of the technique to deal with measurement noise must be studied. We consider the same measure operator $\mathscr{M}$ as presented in the previous section: velocity magnitude measured on a mesh of line density $n = 6$. We thus have $\boldsymbol{m}^* = \mathscr{M}(\overline{\boldsymbol{u}})$ where the asterisk indicates a noiseless quantity. We add to this measure a noise component

$$\overline{\boldsymbol{m}} = \boldsymbol{m}^* + \eta\xi\,, \tag{4.4}$$

with ξ a random N-dimensional vector and η a real number controlling the amplitude of the added noise. The vector ξ is constructed with a uniform distribution chosen such that $|\xi_i| < \sqrt{\|\boldsymbol{m}^*\|_M{}^2/N}$. The level of noise is governed by the real number η; for example, $\eta = 0.1$ corresponds to a noise-to-signal ratio of 10 %, while for $\eta = 1$ we have a 100 % noise-to-signal ratio. These ratios depend on the norm used to define the noise, in our case, the L^2-norm. We present in figure 10 the results of data assimilation as a function of the noise level η. The error is, as expected, an increasing function of the noise level. The algorithm however captures the flow features very well, even for large noise amplitudes. The measure, the reconstructed field and the error are plotted in figure 11 for $\eta = 1$. The identified field $\tilde{\boldsymbol{u}}$ is also relatively noisy, but the flow has been remarkably reconstructed from a low-resolution, scalar and noisy measure. Moreover, the recirculation bubble is accurately reconstructed, together with the main features of the flow.

The evolution, during the iterative process, of both the cost functional $\mathscr{E}$ and the real L^2 residual r_2 for the two cases, $\eta = 0$ (no noise) and $\eta = 1$ (noise level of 100 %), are plotted in figure 12. We can see that for both cases, the cost functional $\mathscr{E}$

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

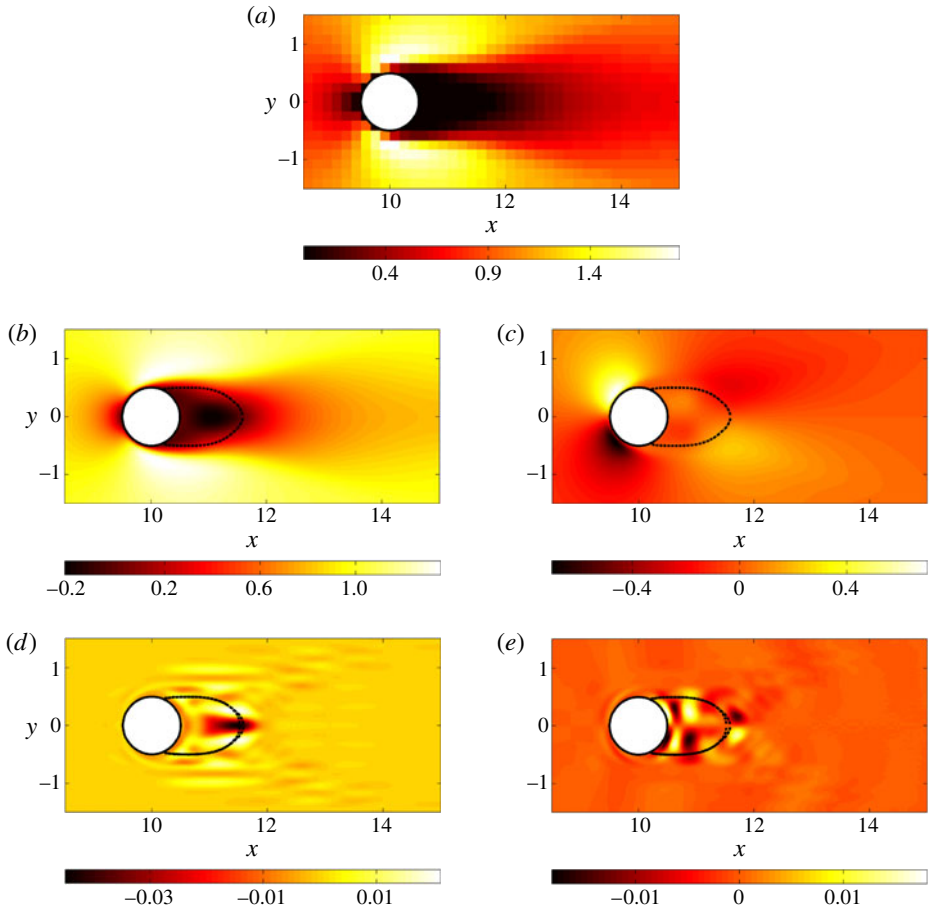


FIGURE 9. (*a*) Velocity magnitude measures on a mesh corresponding to $n=6$. (*b*) and (*c*) Reconstructed field $\boldsymbol{u}_{opt}$. (*d*) and (*e*) Two components of the error $\Delta\boldsymbol{u}$. The dashed lines in (*b*) and (*c*) represent the extent of the recirculation bubble for the reconstructed field. These lines have been reproduced in (*d*) and (*e*), along with those of the true recirculation bubble.

decreases monotonically (due to the gradient descent), but reaches an asymptote at a higher value for the noisy case, confirming that it is not possible to match the measurement noise with a solution of the Navier–Stokes equations. The Navier–Stokes operator thus acts as a filter and cannot provide a fit matching the noise-corrupted measurement. Analysing the evolution of r_2, we see that in the noiseless case the real error monotonically decreases, while it starts to decrease and then increases again, after approximately 200 iterations, for the noisy measurements case. This means that after the turning point, the algorithm tries to match the measurement noise, which drives the solution away from the physical solution $\overline{\boldsymbol{u}}$. In a real situation, this convergence curve is not accessible, and we therefore cannot terminate the iterative identification when the error is lowest. It is however useful to evaluate when to stop the reconstruction based on a comparison between the norm of the gradient of the cost functional and an estimation of the measurement noise level.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

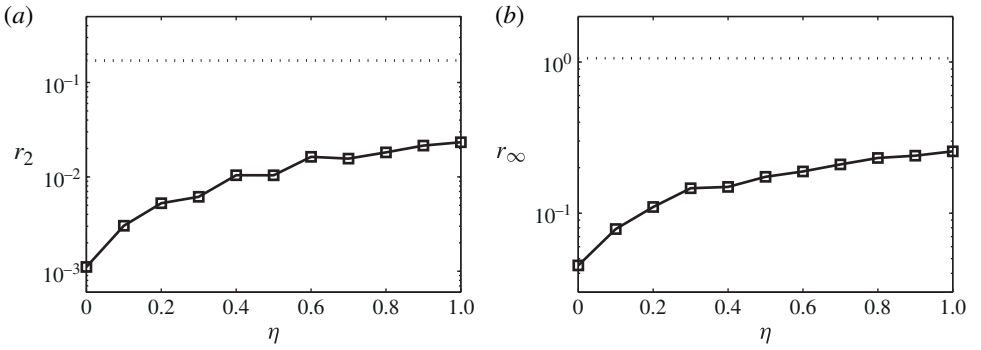


FIGURE 10. Results for the noise reduction case: (*a*) r_2 error; (*b*) r_∞ error as defined in (4.1). The error (measured relative to the base flow error) $\|\Delta \boldsymbol{u}\|$ is plotted (square symbols) against the noise strength η in the measured signal. The error made by approximating the mean flow by the base flow is represented with a dashed line.

5. Additional flow reconstruction

5.1. *Drag evaluation*

Another way to assess the quality of the reconstructed fields is to compare the cylinder drag of the mean field with the drag of the reconstructed field. The drag coefficient C_D induced by a field $(\boldsymbol{u}, p)^\top$ on the cylinder is defined as

$$C_D = \underbrace{2\int_{\mathscr{C}} Re^{-1}\left(\boldsymbol{\nabla u} + \boldsymbol{\nabla u}^\top\right)\boldsymbol{n}\cdot\boldsymbol{e}_x \,\mathrm{d}l}_{C_V} + \underbrace{2\int_{\mathscr{C}} -p\boldsymbol{n}\cdot\boldsymbol{e}_x \,\mathrm{d}l}_{C_P}, \tag{5.1}$$

where $\mathscr{C}$ represents the cylinder's surface and $\mathrm{d}l$ the integration element along this contour. The factors 2 stem from the normalization by the dynamic pressure $\frac{1}{2}\rho U^2$, which simplifies to $\frac{1}{2}$ in the non-dimensional variables. As seen in (5.1), we can divide the drag force into two components: the viscous drag C_V and the pressure drag (or form drag) C_P. We expect the C_V-component to be accurately predicted (as it only depends on the velocity vector which is properly reconstructed), and the C_P-component to exhibit a more appreciable mismatch (as the true pressure field $\bar{p}$ is only recovered up to a scalar field ϕ linked to the potential part of the Reynolds stress term, as explained in (2.29)).

We display in table 1 the different drag coefficients obtained for the true mean flow $(\bar{\boldsymbol{u}}, \bar{p})^\top$ and for several reconstruction scenarios. We note that, as expected, the viscous drag is accurately predicted whereas the relative error on the pressure drag is always greater than 10 %. Note that for larger Reynolds numbers, the viscous drag is expected to decrease and, therefore, the relative total error to increase.

5.2. *Unsteadiness characterization*

In the previous sections, we have exploited the reconstruction abilities of the data-assimilation technique by analysing the reconstructed state vector for several cases of interest. In this section, we explore the possibilities of unsteadiness characterization offered by the identified forcing $\boldsymbol{f}_{opt}$ in the case of full-state information identification, where the global optimal has been reached.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

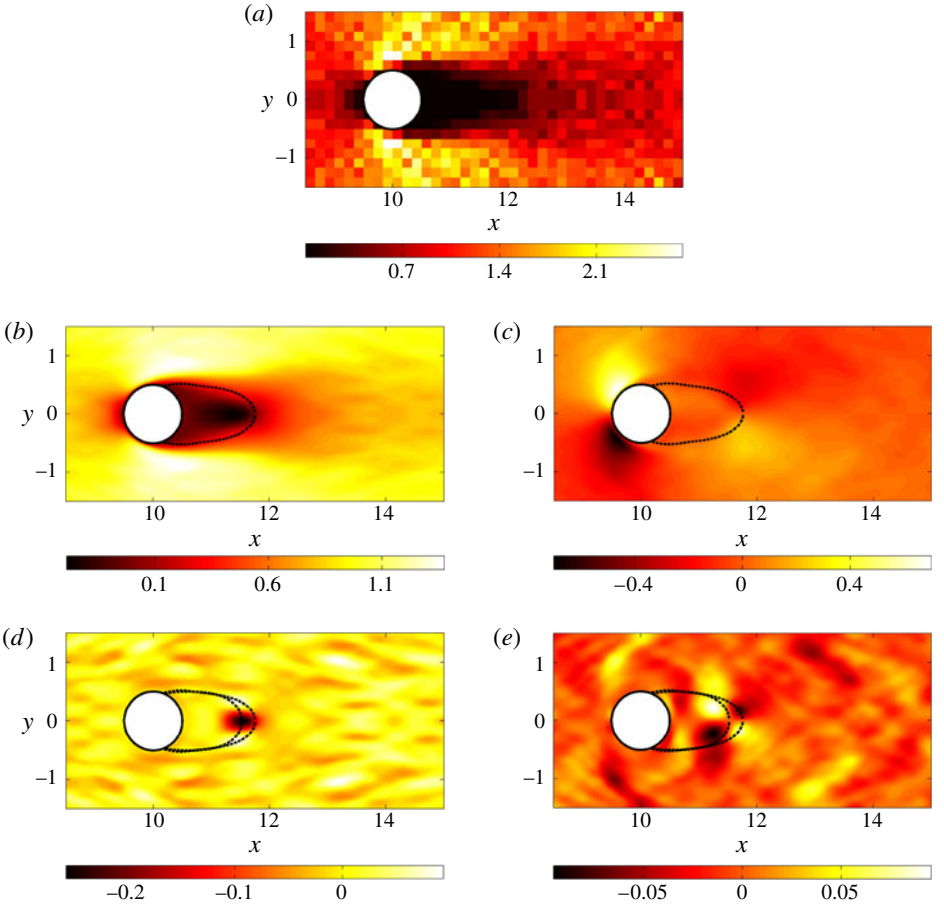


FIGURE 11. (*a*) Velocity magnitude measures on a mesh corresponding to $n = 6$. The noise level has been set to $\eta = 1$, i.e. we have a 100 % noise-to-signal ratio according to our definition. (*b*) and (*c*) Reconstructed field $\boldsymbol{u}_{opt}$. (*d*) and (*e*) Two components of the error $\Delta\boldsymbol{u}$. The dashed lines in (*b*) and (*c*) refer to the extent of the recirculation bubble of the reconstructed field. These dashed lines have been reproduced in (*d*) and (*e*), along with those of the true mean flow.

As covered in § 2.6, the algorithm only identifies the solenoidal part of the true forcing $\boldsymbol{f}^* = -\boldsymbol{\nabla}\cdot\boldsymbol{R}$. However, it is the full Reynolds stress tensor $\boldsymbol{R}$, rather than the components of its divergence, that would be most valuable. In a two-dimensional setting, this tensor can be expressed as

$$\boldsymbol{R} = \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix} + k\boldsymbol{I}, \tag{5.2}$$

with $\alpha = \frac{1}{2}(\overline{u'^2} - \overline{v'^2})$, $\beta = \overline{u'v'}$ and $k = \frac{1}{2}(\overline{u'^2} + \overline{v'^2})$ denoting the fluctuating kinetic energy. These quantities vanish on the cylinder's surface due to the no-slip boundary condition.

It should then be possible to optimally reconstruct a tensor of this particular structure from the identified forcing $\boldsymbol{f}_{opt}$. As we explained in § 2.6, the forcing $\boldsymbol{f}^*$ can

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

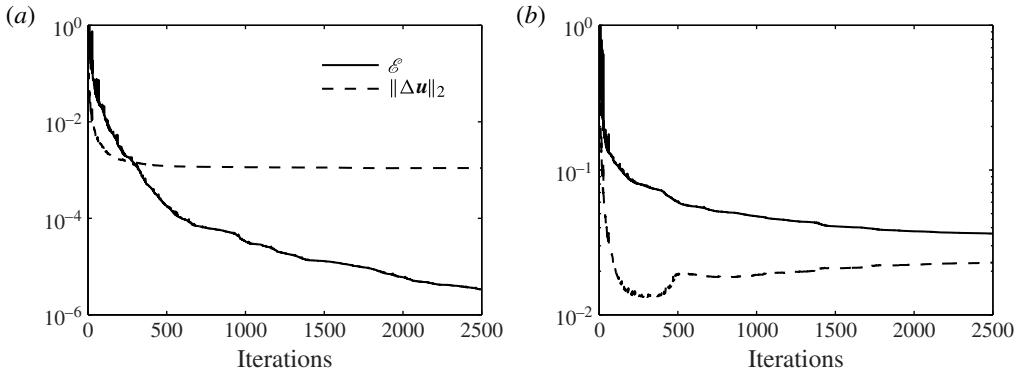


FIGURE 12. (*a*) Convergence curves for the noiseless case ($\eta = 0$) and (*b*) for the noisiest case ($\eta = 1$). We note that for the case without noise, both the cost functional $\mathscr{E}$ and the real residual r_2 are monotonically decreasing. For the noisy case, however, even though $\mathscr{E}$ decreases as a result of the optimization, around the 200th iteration, the real residual r_2 starts to increase again, signifying that from this point forward the optimization actually leads to a deterioration of the identified optimal mean flow $\tilde{\boldsymbol{u}}$.

	Viscous drag (C_V)	Pressure drag (C_P)	Total drag (C_D)
Mean flow	0.30 (—)	1.13 (—)	1.43 (—)
Base flow	0.25 (18.3 %)	0.86 (23.3 %)	1.11 (22.3 %)
Full-state identification	0.30 (1.3 %)	1.00 (11.0 %)	1.30 (9.0 %)
Interpolation ($n = 2$)	0.29 (3.7 %)	0.95 (15.7 %)	1.24 (13.1 %)
Extrapolation ($x_0 = 13$)	0.30 (2.2 %)	0.98 (12.6 %)	1.28 (10.4 %)
Velocity magnitude ($\eta = 0$)	0.30 (2.9 %)	1.00 (11.2 %)	1.30 (9.4 %)
Velocity magnitude ($\eta = 1$)	0.29 (5.4 %)	1.00 (10.7 %)	1.29 (9.6 %)

TABLE 1. Drag coefficients for the mean flow obtained through DNS simulation, for the base flow, for the full-state identification and for several reconstruction scenarios corresponding to the cases displayed in figures 6, 8, 9 and 11. The relative error with respect to the mean flow is displayed in parentheses for each case.

be decomposed into a solenoidal part and a potential component as $\boldsymbol{f}^* = \boldsymbol{f}^*_s + \boldsymbol{\nabla}\phi$. Furthermore, we also have $\boldsymbol{f}^*_s \simeq \boldsymbol{f}_{opt}$ (strict equality is not true because of the boundary condition issue discussed in § 2.6). In summary, we have

$$\boldsymbol{f}^* = -\begin{pmatrix} \partial_x\alpha + \partial_y\beta \\ \partial_x\beta - \partial_y\alpha \end{pmatrix} - \boldsymbol{\nabla} k \tag{5.3a}$$

$$\simeq \boldsymbol{f}_{opt} + \boldsymbol{\nabla}\phi. \tag{5.3b}$$

Then taking the curl of this expression, the gradient terms drop and we are left with one equation relating α, β and the identified forcing $\boldsymbol{f}_{opt}$. Such an identification, however, is under-constrained as we have only one equation for two unknowns. If we further assume that $-\boldsymbol{\nabla} k \simeq \boldsymbol{\nabla}\phi$ (see the next section for justification in the cylinder case), we obtain

$$-\begin{pmatrix} \partial_x\alpha + \partial_y\beta \\ \partial_x\beta - \partial_y\alpha \end{pmatrix} = \boldsymbol{f}_{opt}. \tag{5.4}$$

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

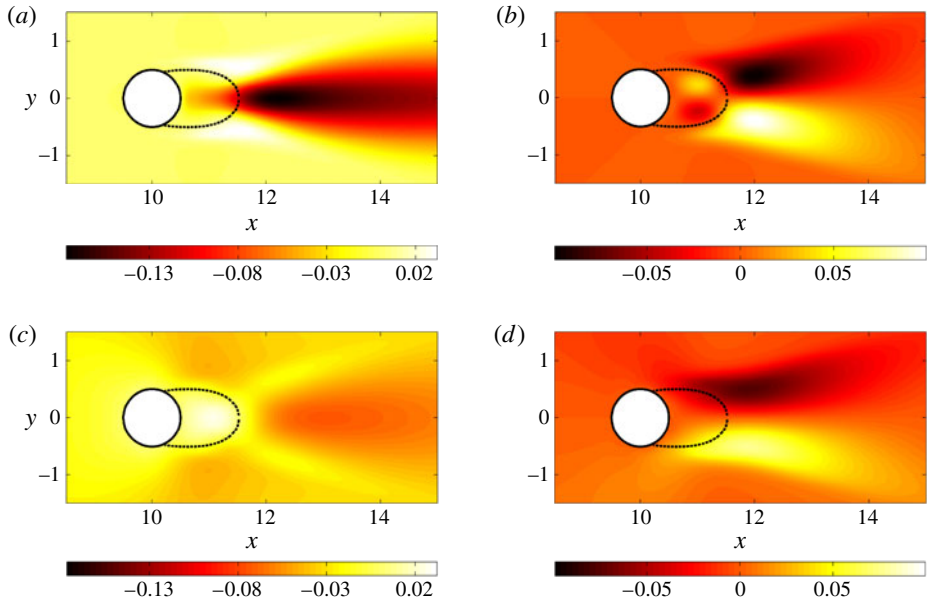


FIGURE 13. Results for (*a*) α and (*b*) β of the real reduced Reynolds stress tensor, and reconstructed tensor components (*c*) α_{opt} and (*d*) β_{opt}. The dashed lines indicate the extent of the recirculation bubble of the true mean flow.

This yields two equations for two unknowns, along with the boundary conditions $\alpha = 0$ and $\beta = 0$ on the boundaries. Yet, recalling that the optimal forcing is divergence-free ($\boldsymbol{\nabla} \cdot \boldsymbol{f}_{opt} = 0$), we obtain a constraint on α and β, which is non-physical so that the identified α and β obtained by solving (5.4) will, in general, not match their experimental values. Instead, we propose a weaker constraint to relate the left- and right-hand sides of (5.4). We decide in the following to minimize the cost functional

$$\mathscr{J}(\alpha, \beta) = \int_{\Omega} ((-\partial_x \alpha - \partial_y \beta - f_x)^2 + (-\partial_x \beta + \partial_y \alpha - f_y)^2)\, \mathrm{d}\Omega, \tag{5.5}$$

where $\boldsymbol{f}_{opt} = (f_x, f_y)$. This minimization problem is solved by setting the gradient of this cost functional to zero. This leads to the two independent equations

$$\nabla^2 \alpha = -\partial_x f_x + \partial_y f_y, \tag{5.6}$$

$$\nabla^2 \beta = -\partial_y f_x - \partial_x f_y, \tag{5.7}$$

which have to be solved subject to the boundary conditions stated above. The results are presented in figure 13.

We see in this figure that the reconstruction of the coefficients which characterize the unsteadiness of the flow fails for α but yields relatively good results for β.

5.3. *Pressure reconstruction*

The component of the reconstructed field that we have not yet analysed, is the pressure $\tilde{p}_{opt}$. We know that the optimally identified pressure $\tilde{p}_{opt}$ is the difference between the real mean pressure $\overline{p}$ and the potential ϕ (see (2.29)). It is interesting to compare the

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

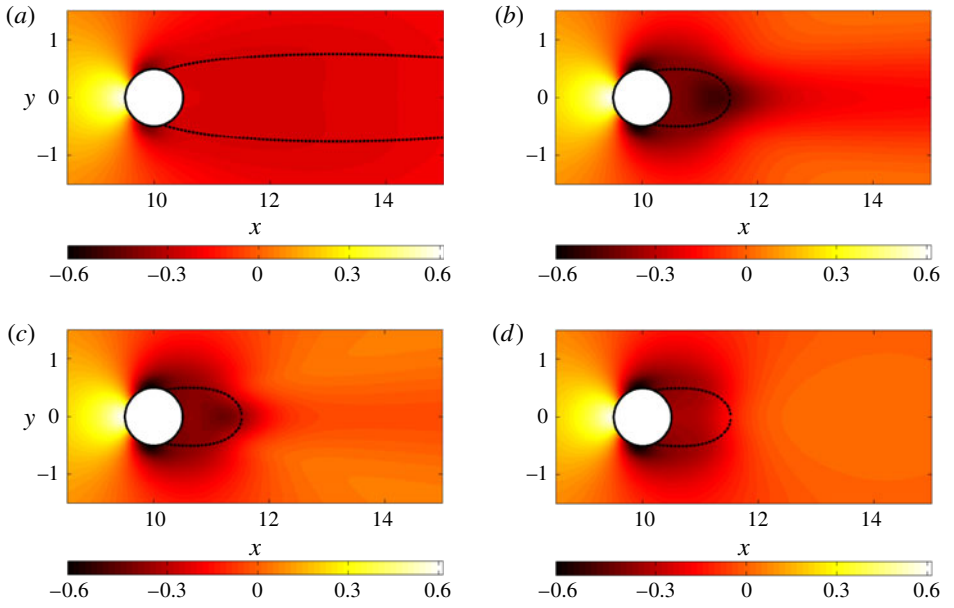


FIGURE 14. (*a*) Base-flow pressure p_b; (*b*) mean flow pressure $\overline{p}$; (*c*) total pressure $p_T = \overline{p} + k$; (*d*) reconstructed pressure $\tilde{p}_{opt}$. The dashed-lines indicate the extent of the recirculation bubbles, (*a*) in the case of the base flow, (*b–d*) in the case of the true mean flow.

reconstructed pressure field $\tilde{p}_{opt}$ to the base-flow pressure p_b (for sake of comparison), to the real mean pressure $\overline{p}$ and to the total pressure $p_T = \overline{p} + k$ (sum of the real mean pressure and the fluctuating kinetic energy k). This is achieved in figure 14.

Analysing the base-flow pressure p_b and the real mean pressure $\overline{p}$, we note that the pressure drop is centred on the respective recirculation bubbles. However, the real mean pressure also displays a constant pressure drop in the wake of the cylinder. If we consider the total pressure p_T, we can see that this wake effect disappears and the pressure drop appears to be mainly located within the recirculation bubble. When compared with the reconstructed pressure $\tilde{p}_{opt}$, we can see that the two pressure fields p_T and $\tilde{p}_{opt}$ are very similar, even though the reconstructed pressure is slightly overestimated near the end of the recirculation bubble, around the stagnation point. The difference between these two pressure fields is weak, so that $-k \simeq \phi$, therefore justifying the assumption $-\nabla k \simeq \nabla \phi$ of the previous section. It is remarkable that the pressure reconstructed from velocity measurements only $\tilde{p}_{opt}$, is rather close to the total pressure p_T, and also not that far from the real pressure $\overline{p}$. In order to improve the identification of the real mean-pressure field, it is mandatory to incorporate pressure measurements within the data-assimilation algorithm. This is briefly discussed below.

6. Concluding remarks and extensions

Despite recent progress in experimental measuring techniques, many data from experiments are still limited in terms of resolution, in terms of access to specific regions of interest and in terms of quantities that can be measured reliably. We have developed and introduced a data-assimilation technique for the recovery of mean flow fields that are solutions of the RANS equations and match the available data

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

points. The algorithm, based on an iterative, direct-adjoint optimization approach, has been validated on two-dimensional flow past a circular cylinder at a supercritical Reynolds number ($Re = 150$). The method has been used in an interpolatory mode (recovering mean-velocity fields from much coarser measurements), in an extrapolatory mode (extending the mean-velocity field into domains where no measurements have been taken), and in a state-vector reconstruction mode (determining mean-velocity components from measurements of only the magnitude). In all cases, we have observed a satisfactory recovery of the mean-velocity field that proves substantially more accurate than more naïve approaches based on higher-order interpolation that disregard physical constraints on the mean fields. Moreover, the influence of noise in the input data fields has been assessed, and the method has been found to be robust, even in the face of a noise-to-signal ratio of 100 %.

Together with the assimilated mean-velocity field, we also obtain the forcing vector field that renders the error between the available measurements and the recovered mean-velocity field minimal. From this forcing, information about the unsteadiness and the pressure field can be extracted. However, this information is limited to the solenoidal part of the Reynolds stress tensor and the non-kinetic part of the pressure. Despite this limitations, it has to be kept in mind that only steady (averaged) information has entered the data-assimilation algorithm. Yet, we were able to gain insight, albeit partial and incomplete, on the second-order moments of the unsteady flow. This is quite remarkable, given the limited input information.

Various extensions and variations of the presented algorithms are conceivable. For large-Reynolds-number applications, the current numerical technique can be adapted to aid in the solving of (2.4) (which can be challenging for large Reynolds numbers when using Newton's iteration method). Due to the averaging, we know that the mean flow will be similar to a base flow at a lower Reynolds number. We therefore propose to artificially decrease the Reynolds number in (2.4). Ideally one would choose the Reynolds number yielding a base flow as close as possible to the measured mean flow. We therefore have $Re = Re^* - \Delta Re$ with Re^* as the real flow Reynolds number and ΔRe ($0 < \Delta Re < Re^*$) as the adjustment performed. We can start with a zero initial guess ($\boldsymbol{f}_g = 0$) and identify the optimal forcing $\boldsymbol{f}_{opt}$. The identified total forcing (for the Navier–Stokes equation with the given Reynolds number Re^*) takes the form $\boldsymbol{f}^*_{opt} = \boldsymbol{f}_{opt} + (\Delta Re/ReRe^*)\nabla^2\tilde{\boldsymbol{u}}_{opt}$, where we have added the contribution due to the artificial decrease of Reynolds number. We note that this method does not violate the divergence-free forcing condition. This technique can be thought of as a preconditioning method allowing us to bypass the problem of solving the Navier–Stokes equations for high Reynolds numbers. It can also be interpreted as a preconditioning step for the optimization. The initial guess $\boldsymbol{f}_g$ could also be obtained by a RANS model (based, for example, on an eddy-viscosity closures or more elaborate Reynolds stress models). Indeed, by so proceeding we avoid the costly (and probably ill-posed) computation of a high-Reynolds-number base flow. Moreover, it is also possible to compare the Reynolds stresses issued from the turbulence model with those of the optimal solution compatible with the experimental observations. This approach would be interesting and useful in assessing the performance of turbulence closures in more complex flow configurations.

The cost functional, i.e. the error between measured and the model-predicted data can be augmented by a weight function which directs more emphasis towards regions that are of more relevance. For example, shear layers or regions close to walls can be singled out as more important and as having a larger contribution to the cost functional. Also, pressure measurements could be considered, leading to a more well-posed optimization problem as explained in § 2.6. Some of these extensions will be

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press

pursued in a future study. In its current form, the data-assimilation method is a very valuable tool for the experimentalist to recover more information about the flow than is directly measurable. In this sense, it takes the measured data beyond their intrinsic information content and extends the scope of current measurement capabilities.

REFERENCES

AVRIL, S., BONNET, M., BRETELLE, A.-S., GRÉDIAC, M., HILD, F., IENNY, P., LATOURTE, F., LEMOSSE, D., PAGANO, S., PAGNACCO, E. & PIERRON, F. 2008 Overview of identification methods of mechanical parameters based on full-field measurements. *Exp. Mech.* **48** (4), 381–402.

COURTIER, P. 1997 Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123** (544), 2449–2461.

EVERSON, R. & SIROVICH, L. 1995 Karhunen–Loeve procedure for gappy data. *J. Opt. Soc. Amer.* A **12** (8), 1657–1664.

GHIL, M. & MALANOTTE-RIZZOLI, P. 1991 Data assimilation in meteorology and oceanography. *Adv. Geophys.* **33**, 141–266.

GUNES, H., SIRISUP, S. & KARNIADAKIS, G. E. 2006 Gappy data: to krig or not to krig? *J. Comput. Phys.* **212** (1), 358–382.

GUNZBURGER, M. D. 2000 Adjoint equation-based methods for control problems in incompressible, viscous flows. *Flow Turbul. Combust.* **65** (3–4), 249–272.

HEITZ, D., MÉMIN, E. & SCHNORR, C. 2010 Variationnal fluid flow measurements from image sequences: synopsis and perspectives. *Exp. Fluids* **48** (3), 369–393.

HILL, D. C. 1995 Adjoint systems and their role in the receptivity problem for boundary layers. *J. Fluid Mech.* **292**, 183–204.

JACKSON, C. P. 1987 A finite-element study of the onset of vortex shedding in flow past variously shaped bodies. *J. Fluid Mech.* **182** (1), 23–45.

LE DIMET, F.-X. & TALAGRAND, O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A**, 97–110.

LUCHINI, P. & BOTTARO, A. 2001 Linear stability and receptivity analyses of the stokes layer produced by an impulsively started plate. *Phys. Fluids* **13**, 1668–1678.

LUNDVALL, J., KOZLOV, V. & WEINERFELT, P. 2006 Iterative methods for data assimilation for burgers's equation. *J. Inverse Ill-Posed Probl.* **14** (5), 505–535.

MARQUET, O., SIPP, D. & JACQUIN, L. 2008 Sensitivity analysis and passive control of the cylinder flow. *J. Fluid Mech.* **615**, 221–252.

MOHAMMADI, B. & PIRONNEAU, O. 2004 Shape optimization in fluid mechanics. *Annu. Rev. Fluid Mech.* **36**, 255–279.

PRALITS, J., BRANDT, L. & GIANNETTI, F. 2010 Instability and sensitivity of the flow around a rotating circular cylinder. *J. Fluid Mech.* **650**, 513–536.

RUHNAU, P., KOHLBERGER, T., NOBACH, H. & SCHNORR, C. 2005 Variationnal optical flow estimation for particle image velocimetry. *Exp. Fluids* **38**, 21–32.

RUHNAU, P. & SCHNORR, C. 2007 Optical stokes flow estimation: an imaging-based control approach. *Exp. Fluids* **42**, 61–78.

RUHNAU, P., STAHL, A. & SCHNORR, C. 2007 Variational estimation of experimental fluid flows with physics-based spatio-temporal regularization. *Meas. Sci. Technol.* **18**, 755–763.

SCHMID, P. J. 2007 Nonmodal stability theory. *Annu. Rev. Fluid Mech.* **39**, 129–162.

SIPP, D., MARQUET, O., MELIGA, PH. & BARBAGALLO, A. 2010 Dynamics and control of global instabilities in open flows: a linearized approach. *Appl. Mech. Rev.* **63**, 030801.

https://doi.org/10.1017/jfm.2014.566 Published online by Cambridge University Press