Dynamic reconstruction and data reconstruction for subsampled or irregularly sampled data

Jakub Krol[1] **and Andrew Wynn**[1,†]

[1]Department of Aeronautics, Imperial College London, London SW7 2AZ, UK

The Nyquist–Shannon criterion indicates the sample rate necessary to identify information with particular frequency content from a dynamical system. However, in experimental applications such as the interrogation of a flow field using particle image velocimetry (PIV), it may be impracticable or expensive to obtain data at the desired temporal resolution. To address this problem, we propose a new approach to identify temporal information from undersampled data, using ideas from modal decomposition algorithms such as dynamic mode decomposition (DMD) and optimal mode decomposition (OMD). The novel method takes a vector-valued signal, such as an ensemble of PIV snapshots, sampled at random time instances (but at sub-Nyquist rate) and projects onto a low-order subspace. Subsequently, dynamical characteristics, such as frequencies and growth rates, are approximated by iteratively approximating the flow evolution by a low-order model and solving a certain convex optimisation problem. The methodology is demonstrated on three dynamical systems, a synthetic sinusoid, the cylinder wake at Reynolds number $Re = 60$ and turbulent flow past the axisymmetric bullet-shaped body. In all cases the algorithm correctly identifies the characteristic frequencies and oscillatory structures present in the flow.

Key words: low-dimensional models, nonlinear dynamical systems

1. Introduction

Fluid flows of practical interest typically exhibit complex spatio-temporal behaviour, and the analysis of such flows frequently relies upon interrogating large collections of high-dimensional velocity field snapshots. A widely employed strategy is to attempt to extract coherent structures from the snapshots, that is, spatial features of dynamic importance to the flow that can be used to facilitate its analysis, modelling or control. For canonical flows, coherent structures such as vortex-shedding modes of bluff bodies flows (Choi and Kim 2008) or spatial streaks in wall-bounded flows (Orlandi 1994) are well known, and the suppression of such structures may motivate a flow control strategy. However, for less-studied flows, coherent structures may not be known *a priori* and must therefore be sought directly from the snapshot data, necessitating the development of automatic, data-driven, structure extraction algorithms.

Perhaps the most widely used structure extraction methodology is proper orthogonal decomposition (POD) (Holmes, Lumley & Berkooz 1996), which decomposes an

† Email address for correspondence: a.wynn@imperial.ac.uk

ensemble of snapshots into orthogonal flow fields ranked by their energetic content. However, POD does not make use of the fact that underlying data are sampled from a dynamical system, meaning that low-energy, low-ranked structures may still be of importance to the system's behaviour (Noack *et al.* 2003). If the data are sampled at a constant time step, however, the dynamic mode decomposition (DMD) (Schmid 2010) can be applied to incorporate temporal information when extracting modes. DMD can be viewed either as fitting a low-order linear model to the evolution of temporal POD coefficients (Wynn *et al.* 2013) or, alternatively, as an approximation to the underlying Koopman operator of the system (Rowley *et al.* 2009). DMD has received much recent interest both in terms of its application to analysis of different classes of flows (such as bluff body flows (Muld, Efraimsson & Henningson 2012; Bagheri 2013; Tissot *et al.* 2014), boundary layers (Grilli *et al.* 2012; Sayadi *et al.* 2014), cavity flows (Schmid 2010; Seena & Sung 2011; Vinha *et al.* 2016) or jet flows (Rowley *et al.* 2009; Schmid 2010; Schmid, Violato & Scarano 2012; Semeraro, Bellani & Lundell 2012)) but also in terms of its theoretical development. In particular, a number of generalisations of the DMD algorithm have recently been proposed, with each seeking to address various deficiencies of the algorithm. Examples of the recent developments include: establishing an appropriate way of selecting the most important DMD modes (Chen, Tu & Rowley 2012; Jovanovic, Schmid & Nichols 2014), finding an optimal low-order projection basis (Wynn *et al.* 2013), improving the robustness of the method to observation noise (Wynn *et al.* 2013; Hemati, Rowley & Nichols 2017), establishing a stronger link to the Koopman operator (Williams, Kevrekidis & Rowley 2015) or accounting for input/output systems (Proctor, Brunton & Kutz 2014).

A fundamental requirement of dynamic mode extraction algorithms is that they must be applied to a data set of flow snapshots sampled at a constant time step δt between subsequent snapshots. However, once δt has been chosen, the Nyquist sampling criterion (Nyquist 1928) indicates that an inherent limitation is placed upon the maximum frequency that can be extracted from a data set. This issue was discussed in Tu *et al.* (2014) in terms of current capabilities of PIV systems. The acquisition rate of a classical PIV system does not typically exceed 15 Hz, implying an upper limit upon the sampling rate which is not sufficient to investigate temporal characteristics of turbulent flows. Time-resolved PIV (TRPIV) systems exist that are capable of producing sampling rates in the range of 1–10 kHz (Adrian and Westerweel 2011), nonetheless, they still may not be sufficient to investigate frequencies of interest. Additionally, tomographic PIV (tomo-PIV) was introduced recently (Elsinga *et al.* 2006), allowing the measurement of three velocity components. In such a case, the number of measurements drastically increases, especially in a combination with TRPIV systems, leading to high computational cost of the data processing and consequently establishing a need to obtain dynamical information from potentially undersampled data.

An attempt to extend the DMD algorithm to regularly undersampled data sets is presented in Leroux & Cordier (2016), where a low-dimensional state space is constructed using POD modes, missing snapshots are reconstructed using an Expectation–Maximisation algorithm and DMD modes are estimated as a solution to a multiple regression problem. However, perhaps the most widely employed methodology to extract coherence from temporally downsampled data is compressive sampling (Candes & Tao 2006*a*,*b*; Donoho 2006). Now, as will be discussed in §2.2, compressive sampling is most effective at extracting sub-Nyquist frequency information in the situations where data snapshots are sampled randomly in time.

This is at odds with the requirement of DMD-based modal extraction methodologies of a constant separation time step. In Tu *et al.* (2014) this problem was addressed by first performing compressive sampling directly on a temporally downsampled velocity field data ensemble, then subsequently using DMD to extract coherent structures. However, applying the compressive sampling algorithm requires one to pre-specify a basis such that representation of the data in the chosen basis is approximately sparse (see §2.2). Such knowledge is often not available *a priori*. In this paper we propose an alternative to this strategy, the irregularly sampled optimal mode decomposition (is-OMD) algorithm. The motivation for our approach is to extend traditional modal extraction methods so that they can be applied to data sampled at non-constant time steps. Thus, the arbitrary choice of projection basis implied by a direct use of compressive sampling can be avoided, meaning that is-OMD provides a more flexible method for the analysis of high-dimensional temporally subsampled data ensembles. Further on, a proposed method makes use of already existing DMD-like algorithms, making is-OMD readily and easily implementable.

We discuss the performance of the is-OMD algorithm in terms of its ability to extract coherent spatial features with associated and accurate dynamical information, and whether such information can be used to reconstruct the instantaneous state of the flow from subsampled data. Section 2.1 provides an outline of the is-OMD algorithm followed by a rigorous stability analysis in §3. A robust extension to the algorithm in terms of estimation of mode shapes and instantaneous unknown snapshots is presented in §4. Section 5 demonstrates an example of a synthetic system for which is-OMD provides an improvement over compressive sampling in terms of frequency estimation, whilst §6 provides a practical discussion of implementation of the algorithm to the canonical example of the cylinder wake. The demonstration of the applicability of the is-OMD algorithm to dynamically more complicated turbulent axisymmetric wake is presented in §7.

2. Dynamic modes of downsampled and irregularly sampled data

Let $f(\boldsymbol{x}, t)$ denote the state of a dynamical system which depends upon a spatial variable $\boldsymbol{x}$ and temporal variable t. Suppose first that the system is sampled at p spatial locations $(x_i)_{i=1}^p$ and at equally spaced time steps $t_n = n\delta t$ to provide a data ensemble

$$\boldsymbol{w}_n = (f(x_i, t_n))_{i=1}^p \in \mathbb{R}^p, \quad n = 0, \ldots, N. \tag{2.1}$$

The DMD algorithm extracts dynamical information by approximating the evolution between subsequent snapshots. In particular, letting $\boldsymbol{U}_p \in \mathbb{R}^{p \times N}$ be the matrix whose columns are the POD modes of the first N snapshots $\{\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}\}$, DMD produces (Wynn *et al.* 2013) a matrix $\boldsymbol{S} \in \mathbb{R}^{N \times N}$ which minimises the residual error

$$\sum_{i=1}^N \|\boldsymbol{w}_{i+1} - \boldsymbol{U}_p \boldsymbol{S} \boldsymbol{U}_p^{\mathrm{T}} w_i\|^2. \tag{2.2}$$

In other words, DMD searches for the optimal linear model to describe the evolution of a set of temporal POD coefficients over the time step δt. Upon taking an eigenvalue decomposition $\boldsymbol{S} = \boldsymbol{P} \boldsymbol{\Delta} \boldsymbol{P}^{-1}$, the DMD modes of the system are defined to be the columns of $\boldsymbol{U}_p \boldsymbol{P}$, with each mode corresponding to an eigenvalue of $\boldsymbol{S}$ describing the mode's growth rate and frequency.

Now, increasing the time step δt between sample points may decrease the accuracy of the extracted dynamical information. For a simple illustrative example, figure 1
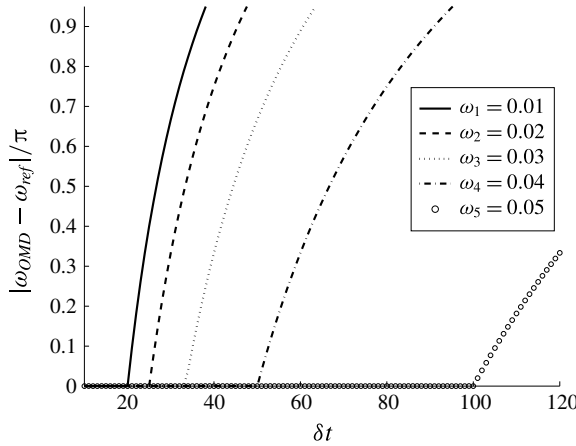
FIGURE 1. The error in frequency estimation for a range of uniform time steps between consecutive snapshots.

shows the error between the true frequencies $(\omega_j)_{j=1}^5 = (j\pi 10^{-2})_{j=1}^5$ and those identified by DMD from snapshots sampled from the one-dimensional sinusoidal system

$$f(x, t) = \sum_{i=1}^{5} \sin(x + \omega_i t), \quad 0 \leqslant x \leqslant 2\pi, t \geqslant 0, \tag{2.3}$$

for varying snapshot spacings $10 \leqslant \delta t \leqslant 120$. It can be seen that the error in each of the five system frequencies is zero until, respectively, δt exceeds 100, 50, 33, 25 and 20. This corresponds exactly to the Nyquist condition that the sampling rate must be at least twice the frequency to be identified, that is, $\delta t \leqslant \pi/\omega_j = 100/j$. This presents the natural question of whether it is possible to reduce this error and correctly identify a system's dynamical information in the case of undersampling. In signal processing, the established method for achieving such an aim is compressive sampling, which is typically applied to irregularly sampled data ensembles.

2.1. *Irregularly sampled data*

Consider the snapshots $\mathcal{W} = (\boldsymbol{w}_n)_{n=0}^N \subset \mathbb{R}^p$ described in (2.1), sampled with constant time step δt. An alternative to varying δt, as discussed above, is to instead view δt as a minimal time step representing the highest possible frequency of data collection, but assume that only a subset $\mathcal{U} \subset \mathcal{W}$ of snapshots may be used for analysis. We refer to $\mathcal{U}$ as the analysis data ensemble and define it by choosing a subset $\mathcal{J} \subset \{0, 1, \ldots, N\}$, re-labelling $\boldsymbol{u}_i := w_i$ for each index $i \in \mathcal{J}$, and setting $\mathcal{U} = \{\boldsymbol{u}_i : i \in \mathcal{J}\}$. Finally, let Δt_i be the sample times between successive snapshots in $\mathcal{U}$ which are, possibly non-constant, integer multiples of δt. The schematic representation of such a sampling strategy is presented in figure 2.

It will be assumed that $\mathcal{W}$ is sampled at a sufficiently high rate that it contains all dynamical information of interest concerning the state trajectory $f(\boldsymbol{x}, t)$. The question which we aim to address is to what extent $\mathcal{W}$ can be reconstructed from knowledge of the analysis data set $\mathcal{U}$ only. Compressive sampling presents a standard method of answering such a problem.
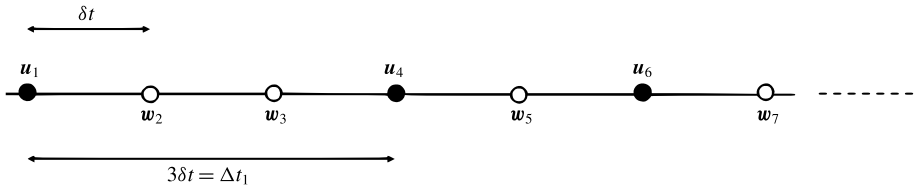
FIGURE 2. Downsampled and regularly sampled data.

2.2. *Compressive sampling*

Extracting coherent dynamical information from a downsampled data ensemble $\mathcal{U}$ may be possible if the underlying system is sparse, or compressible, when expressed in particular temporal basis. To describe this, first collect the high-frequency sampled snapshots $\mathcal{W}$ into a matrix

$$\boldsymbol{W} := \begin{bmatrix} \uparrow & & \uparrow \\ \boldsymbol{w}_0 & \cdots & \boldsymbol{w}_N \\ \downarrow & & \downarrow \end{bmatrix} \in \mathbb{R}^{p \times (N+1)}. \tag{2.4}$$

The state trajectory $\boldsymbol{W}$ is said to be compressible if there exists a matrix $\boldsymbol{\Xi} = (\xi_{ij}) \in \mathbb{R}^{(N+1)\times(N+1)}$ such that it can be approximately represented in the form

$$\boldsymbol{W} \approx \hat{\boldsymbol{W}}\boldsymbol{\Xi}, \tag{2.5}$$

where $\hat{\boldsymbol{W}} \in \mathbb{R}^{p \times N}$ is sparse in the sense that it has only k non-zero columns, with $k \ll N$. Letting $\hat{\boldsymbol{w}}_i$ denote the columns of $\hat{\boldsymbol{w}}$, and rewriting the decomposition (2.5) as

$$\boldsymbol{w}_i \approx \sum_{j=0}^{N} \hat{\boldsymbol{w}}_j \xi_{ji}, \quad i = 0, \ldots, N, \tag{2.6}$$

it is clear that the each column $\hat{\boldsymbol{w}}_j \in \mathbb{R}^p$ represents a spatial structure, while the respective row $(\xi_{ji})_{i=0}^{N}$ of $\boldsymbol{\Xi}$ describes its temporal evolution. For this discussion we make the assumption that a temporal basis $\boldsymbol{\Xi}$ is known for which a columnwise sparse solution $\hat{\boldsymbol{W}}$ to (2.5) exists.

Next, write the analysis data ensemble $\mathcal{U}$ in matrix form

$$\boldsymbol{U} = \begin{bmatrix} \uparrow & & \uparrow \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ \downarrow & & \downarrow \end{bmatrix} = \boldsymbol{W}\boldsymbol{\Theta}, \tag{2.7}$$

by defining an appropriate binary matrix $\boldsymbol{\Theta} \in \mathbb{R}^{N+1 \times m}$, $m = |\mathcal{J}|$, which encodes the particular time steps at which data are available. The aim of compressive sampling is to determine $\boldsymbol{W}$ from knowledge of $\boldsymbol{U}$ only. In view of (2.5) and (2.7), one way to achieve this is to seek a columnwise sparse solution $\hat{\boldsymbol{W}}$ to

$$\boldsymbol{U} = \hat{\boldsymbol{W}}\boldsymbol{\Xi}\boldsymbol{\Theta}, \tag{2.8}$$

which would then provide an approximation to the full signal via $\boldsymbol{W} \approx \hat{\boldsymbol{W}}\boldsymbol{\Xi}$.

Note that, since $m \leqslant N$, equation (2.8) is overdetermined, meaning that the imposition of sparsity is necessary to obtain a well-posed problem. The most widely used approach to sparsity promotion is to penalise the ℓ^1-norm of $\hat{\boldsymbol{W}}$ subject to the constraint (2.8). However, apart from in the scalar-valued case $p = 1$, this will impose entrywise sparsity rather than the desired columnwise sparsity of $\hat{\boldsymbol{W}}$. Methods to avoid this problem typically use greedy algorithms such as orthogonal matching pursuit (OMP) (Tropp & Gilbert 2007) or compressive sampling matching pursuit (CoSaMP) (Needell & Tropp 2009). Rather than minimising the ℓ^1-norm of $\hat{\boldsymbol{W}}$, such methodologies analyse the correlation of the matrix $\boldsymbol{\Xi}\boldsymbol{\Theta}$ with the downsampled data ensemble $\boldsymbol{U}$. They are iterative algorithms in which, upon each iteration, one sparse vector (a column of $\boldsymbol{\Xi}$) is added to the solution. Hence, the outputs of the algorithms are sparse, with the sparsity of the solution equal to the number of iterations.

An important fact is that the sampling strategy, encoded by $\boldsymbol{\Theta}$, plays a significant role in determining the performance of compressive sampling algorithms. In particular, a known requirement is that the chosen temporal basis $\boldsymbol{\Xi}$ has a low coherence with the sampling strategy $\boldsymbol{\Theta}$ (where coherence is defined $\max_{j,k} |\langle (\theta_{ji})_{i=0}^{N}, (\xi_{ik})_{i=0}^{N} \rangle|$) (Candes & Wakin 2008). A common way to achieve this is to use Bernoulli and Gaussian random measurements, which are known to be incoherent with many canonical bases, including the typical Fourier basis (Candes & Wakin 2008). That is, if periodic structures are sought in time, then an aperiodic sampling strategy is beneficial for extracting them from downsampled data. We reiterate that DMD-type modal extraction methods do not currently support sampling at non-constant time intervals, meaning that any potential benefits of such a strategy may not be realised using current techniques. We now expand upon this observation and propose an extension to DMD to enable its application to a general irregularly sampled data ensemble of the form $\mathcal{U}$ described in figure 2.

2.3. *Dynamic decomposition of irregularly sampled data*

The application of DMD to the regularly sampled data set $\mathcal{W}$, as discussed above, can be obtained by minimising the cost (2.2) with respect to the dynamics matrix $\boldsymbol{S}$. This cost function corresponds to projecting each snapshot $\boldsymbol{w}_i$ onto its POD coefficients via $\boldsymbol{U}_p^{\mathrm{T}}\boldsymbol{w}_i$, transforming these coefficients under the mapping $\boldsymbol{S}$, lifting the result back to create a modified flow field $\boldsymbol{U}_p\boldsymbol{S}\boldsymbol{U}_p^{\mathrm{T}}\boldsymbol{w}_i$, and finally comparing this flow field to the snapshot $\boldsymbol{w}_{i+1}$. Consequently, DMD can be viewed in terms of constructing an optimal model of the form $\boldsymbol{w}_i \mapsto \boldsymbol{U}_p\boldsymbol{S}\boldsymbol{U}_p^{\mathrm{T}}\boldsymbol{w}_i$ which best approximates the one-step snapshot evolution across the entire data ensemble. Solving the DMD optimisation problem is computationally undemanding: the calculation of $\boldsymbol{U}_p$ is simply $\mathrm{POD}(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1})$; while, since (2.2) is quadratic, an analytical solution for $\boldsymbol{S}$ can easily be obtained by differentiating the cost and solving the resulting linear equation (Wynn *et al.* 2013).

Now, if one attempts to apply a similar approach to the irregularly sampled data ensemble $\mathcal{U}$, a natural generalisation of (2.2) is to consider the modified cost function

$$\sum_{i=0}^{N} \|\boldsymbol{u}_{i+1} - \boldsymbol{U}_p\boldsymbol{S}^{n_i}\boldsymbol{U}_p^{\mathrm{T}}\boldsymbol{u}_i\|^2 = \sum_{i=0}^{N} \|\boldsymbol{w}_{i+n_i} - \boldsymbol{U}_p\boldsymbol{S}^{n_i}\boldsymbol{U}_p^{\mathrm{T}}\boldsymbol{w}_i\|^2, \qquad (2.9)$$

where $n_i := \Delta t_i/\delta t \in \mathbb{N}$ is the number of minimal time steps δt between each sample point in $\mathcal{U}$. In other words, if there are n_i minimal time steps between available data vectors $\boldsymbol{u}_{i+1}$ and $\boldsymbol{u}_i$ we seek to model the system evolution over an interval of length $\Delta t_i = n_i\delta t$ by the iterated linear mapping $(\boldsymbol{U}_p\boldsymbol{S}\boldsymbol{U}_p^{\mathrm{T}})^{n_i} = \boldsymbol{U}_p\boldsymbol{S}^{n_i}\boldsymbol{U}_p^{\mathrm{T}}$.
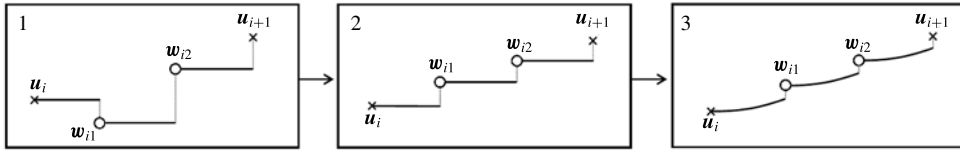
FIGURE 3. Schematic representation of a single iteration of the is-OMD algorithm. Given an initial guess for the dynamics $\boldsymbol{LML}^\mathrm{T}$ – represented by solid lines – over each time step and the slack variables (step 1), the slack variables are first updated (step 2) before being fixed to allow an update of $\boldsymbol{L}$ and $\boldsymbol{M}$ (step 3).

Such a formulation encounters two immediate problems. First, the DMD choice of $\boldsymbol{U}_p = \mathrm{POD}(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1})$ cannot be made since $\mathcal{U}$ may not contain all such data snapshots. Second, and more importantly, since in general $n_i > 1$, the modified cost (2.9) is at least of quartic order with respect to $\boldsymbol{S}$, meaning that an analytical solution is no longer available. To attempt to avoid these problems, consider the related optimisation problem

$$\min_{\hat{\boldsymbol{\omega}}_i, \boldsymbol{L}, \boldsymbol{M}} \sum_{i=0}^{N} \|\hat{\boldsymbol{\omega}}_{i+1} - \boldsymbol{LML}^\mathrm{T}\hat{\boldsymbol{\omega}}_i\|^2, \tag{2.10a}$$

$$\text{s.t. } \hat{\boldsymbol{\omega}}_i \in \mathbb{R}^p, \quad i = 0, \ldots, N, \tag{2.10b}$$

$$\hat{\boldsymbol{\omega}}_i = \boldsymbol{u}_i, \quad i \in \mathcal{J}, \tag{2.10c}$$

$$\boldsymbol{L}^\mathrm{T}\boldsymbol{L} = \boldsymbol{I}, \quad \boldsymbol{L} \in \mathbb{R}^{p \times r}, \quad \boldsymbol{M} \in \mathbb{R}^{r \times r}. \tag{2.10d}$$

Now, the variable space of (2.10) is considerably extended with respect to DMD. The low-order model $\hat{\boldsymbol{\omega}}_i \mapsto \boldsymbol{LML}^\mathrm{T}\hat{\boldsymbol{\omega}}_i$ approximating snapshot evolution over one minimal time step δt now has both its projective component $\boldsymbol{L}$ and dynamics matrix $\boldsymbol{M}$ as optimisation variables. This is the approach taken in the OMD algorithm (Wynn *et al.* 2013) and avoids the need to calculate POD modes. More importantly, higher powers of the dynamics matrix $\boldsymbol{M}$ are avoided by introducing 'slack' state variables $\hat{\boldsymbol{\omega}}_j \in \mathbb{R}^p$ corresponding to each underlying minimal time step $j\delta t$. The analysis data ensemble $\mathcal{U}$ is then employed by enforcing the constraint $\hat{\boldsymbol{\omega}}_i = \boldsymbol{u}_i$, $i \in \mathcal{J}$. Finally, again following the OMD formulation, the user-defined parameter $r \leqslant N - 1$ specifies the rank of the sought approximating dynamics $\boldsymbol{LML}^\mathrm{T}$. For completeness, we note that (2.10) reduces to the DMD algorithm in the case that $\mathcal{W} = \mathcal{U}$, $r = N - 1$ and $\boldsymbol{L} = \mathrm{POD}(\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1})$. The OMD algorithm is recovered in the case that $\mathcal{W} = \mathcal{U}$.

Despite the removal of higher orders of $\boldsymbol{M}$ from the cost function, (2.10) cannot be solved directly. Indeed, the inclusion of the slack state variables $\hat{\boldsymbol{\omega}}_i \in \mathbb{R}^p$ makes the problem overdetermined, while the products of decision variables – for example, in the $\boldsymbol{LML}^\mathrm{T}\hat{\boldsymbol{\omega}}_i$ term – indicate that the problem is non-convex. To begin to address these difficulties, we propose a two-step iterative solution method indicated schematically in figure 3: first, the model variables $\boldsymbol{L}$ and $\boldsymbol{M}$ are fixed and the slack variables $\hat{\boldsymbol{\omega}}_i$ are updated by solving a least-squares problem; second, the updated slack state variables are fixed, allowing $\boldsymbol{M}$ and $\boldsymbol{L}$ to be updated using the standard OMD algorithm applied to the regularly temporally sampled fields $\hat{\boldsymbol{\omega}}_i$.

The first of these steps – minimising (2.10*a*) with respect to $\hat{\omega}_i$ only, subject to constraints (2.10*b*), (2.10*c*) – is simply a least-squares problem. However, since typically $p = O(10^{5-6})$, the problem is memory-intensive and likely to be ill-conditioned. Instead, we only seek to minimise the model error over the subspace Im($\boldsymbol{L}$) $\subset \mathbb{R}^p$, where $\boldsymbol{L}$ is fixed within the slack variable update subiteration. This is achieved by considering the projected quadratic cost

$$\| \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}} \left(\hat{\omega}_{i+1} - \boldsymbol{L}\boldsymbol{M}\boldsymbol{L}^{\mathrm{T}}\hat{\omega}_i \right) \|^2 = \| \boldsymbol{L} \left(\boldsymbol{L}^{\mathrm{T}}\hat{\omega}_i - \boldsymbol{M}\boldsymbol{L}^{\mathrm{T}}\hat{\omega}_i \right) \|^2 = \| \boldsymbol{c}_{i+1} - \boldsymbol{M}\boldsymbol{c}_i \|^2, \quad (2.11)$$

where $\boldsymbol{c}_i := \boldsymbol{L}^{\mathrm{T}}\hat{\omega}_i \in \mathbb{R}^r$. Since $r \ll p$, minimising (2.11) with respect to $\boldsymbol{c}_i$ is advantageous to minimising (2.10*a*) with respect to $\hat{\omega}_i$. Consequently, the first step of the pairwise optimisation is to solve the constrained least-squares problem

$$\min_{\boldsymbol{c}_i} \sum_{i=0}^{N} \| \boldsymbol{c}_{i+1} - \boldsymbol{M}\boldsymbol{c}_i \|^2, \quad (2.12a)$$

$$\text{s.t. } \boldsymbol{c}_i = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{u}_i, \quad \text{if } i \in \mathcal{J}, \quad (2.12b)$$

$$\boldsymbol{c}_i \in \mathbb{R}^r, \quad (2.12c)$$

upon which the slack state variables are updated by lifting up via $\hat{\omega}_i \leftarrow \boldsymbol{L}\boldsymbol{c}_i$. After this update, the $\hat{\omega}_i$ are fixed and $\boldsymbol{L}, \boldsymbol{M}$ are updated via the OMD algorithm $[\boldsymbol{L}, \boldsymbol{M}] \leftarrow$ OMD($(\hat{\omega}_i)_{i=0}^{N}$). It is proved in appendix A that this two-step procedure is guaranteed to decrease the cost (2.10*a*) which, being bounded below, implies that it will converge asymptotically.

A subtle characteristic of the OMD algorithm (step 6 in algorithm 1) is that the cost function of the minimisation problem (2.13) is invariant under rotations of $\boldsymbol{L}$ (Wynn *et al.* 2013). In particular, for any orthogonal matrix $\tilde{\boldsymbol{R}}$ such that $\tilde{\boldsymbol{R}}^{\mathrm{T}}\tilde{\boldsymbol{R}} = \boldsymbol{I}$, the pairs $[\boldsymbol{L}, \boldsymbol{M}]$ and $[\boldsymbol{L}\tilde{\boldsymbol{R}}, \tilde{\boldsymbol{R}}^{\mathrm{T}}\boldsymbol{M}\tilde{\boldsymbol{R}}]$ have equivalent costs. The explanation of this behaviour is that OMD identifies a subspace Im($\boldsymbol{L}$) $\subset \mathbb{R}^p$ upon which a low-order representation of the system dynamics can be provided, and that each $\boldsymbol{L}\tilde{\boldsymbol{R}}$ is just a member of the same equivalence class describing this subspace. For visualisation purposes, we therefore propose in steps 10–12 of algorithm 1 to choose a rotation of $\boldsymbol{L}$ which ranks its columns by their energetic contribution to the known data.

The is-OMD procedure is summarised in algorithm 1 which outputs: (i) estimated intermediate snapshots $\boldsymbol{w}_i$, $i \notin \mathcal{J}$; (ii) a projection matrix $\boldsymbol{L}$ and (iii) an approximate dynamics matrix $\boldsymbol{M}$. The appropriate choice of initial projection and dynamics matrices, $\boldsymbol{L}_0, \boldsymbol{M}_0$, will be discussed in detail in §6.1 where it is demonstrated that a sensible choice is to set the initial dynamics matrix to be a small perturbation of the identity matrix with normal distribution of standard deviation σ, $(\boldsymbol{M}_0)_{ij} = N(0, \sigma^2) + \delta_{ij}$, and the projection space to be POD modes of the analysis data ensemble $\mathcal{U}$, $\boldsymbol{L}_0 = \text{POD}(\mathcal{U})$. Additionally, the beneficial impact of choosing a large value of r will be established in §6.2.

A natural variant of algorithm 1 is to use DMD in place of OMD in step 6, in which case $\boldsymbol{L}_k := \text{POD}((\boldsymbol{w}_i^k)_{i=0}^{N})$ would be fixed in step 6 within each iteration. Note that in this case, a final rotation of $\boldsymbol{L}$ in steps 10–12 is unnecessary. Finally, it should be noted that the notion of dynamic modes has not yet been defined in the context of algorithm 1. This will be discussed subsequently in §4.1.

Algorithm 1 is-OMD

1: **inputs** $\mathcal{U} = \{\boldsymbol{u}_i : i \in \mathcal{J}\}$, $\boldsymbol{L}_0 \in \mathbb{R}^{p \times r}$ satisfying $\boldsymbol{L}_0^T \boldsymbol{L}_0 = \boldsymbol{I}$, $\boldsymbol{M}_0 \in \mathbb{R}^{r \times r}$, $k = 1$, and $\epsilon > 0$.
2: **repeat**
3:

$$(\boldsymbol{c}_i^k) \leftarrow \underset{(\boldsymbol{c}_i)}{\arg\min} \sum_{i=1}^{N} \|\boldsymbol{c}_i - \boldsymbol{M}_{k-1}\boldsymbol{c}_{i-1}\|^2, \tag{2.13}$$
$$\text{s.t. } \boldsymbol{c}_i = \boldsymbol{L}_{k-1}^T \boldsymbol{u}_i \text{ if } i \in \mathcal{J}.$$

4: $\quad \boldsymbol{w}_i^k \leftarrow \boldsymbol{L}_{k-1} \boldsymbol{c}_i^k$, $i \notin \mathcal{J}$.
5: $\quad \boldsymbol{w}_i^k \leftarrow \boldsymbol{u}_i$, $i \in \mathcal{J}$.
6:

$$[\boldsymbol{L}_k, \boldsymbol{M}_k] \leftarrow \underset{\boldsymbol{L},\boldsymbol{M}}{\arg\min} \quad \sum_{i=1}^{N} \|\boldsymbol{w}_i^k - \boldsymbol{L}\boldsymbol{M}\boldsymbol{L}^T \boldsymbol{w}_{i-1}^k\|^2, \tag{2.14}$$
$$\text{s.t.} \quad \boldsymbol{L}^T \boldsymbol{L} = I,$$
$$\boldsymbol{L} \in \mathbb{R}^{p \times r}, \ \boldsymbol{M} \in \mathbb{R}^{r \times r}.$$

7:

$$S_k \leftarrow \sum_{i=1}^{N} \|\boldsymbol{w}_i^k - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^T \boldsymbol{w}_{i-1}^k\|^2. \tag{2.15}$$

8: $\quad k \leftarrow k + 1$
9: **until** $(S_k - S_{k-1}) < \epsilon$
10: $\tilde{\boldsymbol{R}} \boldsymbol{\Sigma} \overline{\boldsymbol{V}}^T \leftarrow \text{svd} \left(\boldsymbol{L}_k^T \boldsymbol{U} \right)$
11: $\boldsymbol{M} \leftarrow \tilde{\boldsymbol{R}}^T \boldsymbol{M}_k \tilde{\boldsymbol{R}}$
12: $\boldsymbol{L} \leftarrow \boldsymbol{L}_k \tilde{\boldsymbol{R}}$
13: **return** $(\boldsymbol{w}_i)_{i=0}^{N}$, $\boldsymbol{M} \in \mathbb{R}^{r \times r}$ and $\boldsymbol{L} \in \mathbb{R}^{p \times r}$.

---

3. Algorithmic considerations

We now discuss feasibility of the slack variable subproblem (2.13) and computational complexity of the algorithm 1. Since the model update subproblem (2.14) is an application of OMD or DMD, we refer to (Wynn *et al.* 2013) for a discussion of this step's performance. Furthermore a detailed proof demonstrating that the cost (S_k) is monotone decreasing (and hence, convergent) is presented in appendix A.

3.1. *Convexity of the slack variable update*

The equality constrains $\boldsymbol{c}_i = \boldsymbol{L}^T \boldsymbol{u}_i$, $i \in \mathcal{J}$ of (2.13) represent time steps at which data are available. Without loss of generality, we consider the problem in which $\mathcal{J} = \{0, N\}$, since the general optimisation problem can be broken into a number of independent subproblems of this form. Now, it is not difficult to prove that (2.13) can be written as the unconstrained least-squares problem

$$\min_{\boldsymbol{\xi}} \|\boldsymbol{A}\boldsymbol{\xi} - \boldsymbol{b}\|_2^2, \tag{3.1}$$

where

$$
\boldsymbol{A} = \begin{bmatrix} \boldsymbol{M} & 0 & \dots & 0 \\ -\boldsymbol{I} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \boldsymbol{M} \\ 0 & \dots & 0 & -\boldsymbol{I} \end{bmatrix} \in \mathbb{R}^{Nr \times (N-1)r}, \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{L}^{\mathrm{T}}\boldsymbol{u}_N \\ 0 \\ \vdots \\ 0 \\ -\boldsymbol{M}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{u}_1 \end{bmatrix} \in \mathbb{R}^{Nr}, \qquad (3.2a,b)
$$

and $\boldsymbol{\xi} = \left(\boldsymbol{c}_{N-1}^{\mathrm{T}} \cdots \boldsymbol{c}_1^{\mathrm{T}} \right)^{\mathrm{T}} \in \mathbb{R}^{(N-1)r}$. Consequently, (2.13) is convex and admits the solution $\boldsymbol{\xi} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ provided the matrix $\boldsymbol{A}$ is full rank. This is true, and follows from the easily shown fact that $\boldsymbol{A}\boldsymbol{\xi} = 0 \Leftrightarrow \boldsymbol{\xi} = 0$. Finally, we note that solving a series of independent subproblems has significant advantages in terms of computational memory savings and improved numerical stability due to a reduced condition number of $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ compared to the case where $\boldsymbol{\xi}$ and $\boldsymbol{b}$ contain $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{u}_i$ and $\boldsymbol{c}_i$ for all $i \in \mathcal{J}$.

3.2. *Computational complexity*

The computational cost of the algorithm 1 consists of that associated with the two optimisation subproblems; the least-squares problem (2.13) and the modal optimisation (2.14) where dynamic model is updated via OMD (or DMD). Solving the least-squares problem using, for example, the Moore–Penrose pseudoinverse of $\boldsymbol{A} \in \mathbb{R}^{Nr \times (N-1)r}$ has complexity $O(N^3 r^3)$. Moreover, as discussed in § 3.1 the full least-squares problem can be broken down into a number of smaller, independent, subproblems corresponding to N_i intermediate time steps, each of which can be solved at a low computation burden. For example, $(N_i = 10, r = 40)$ and $(N_i = 20, r = 60)$ can be solved in 0.01 s and 0.3 s on a standard desktop computer, respectively, and each subproblem could be parallelised, if required.

For the modal optimisation subproblem (2.14), a discussion of the computational performance of OMD and DMD can be found in (Wynn *et al.* 2013). Both the total number of slack variables N and the approximation rank r affect computation time in addition to the underlying dynamic characteristics of the specific system to which the each algorithm is applied. With respect to the examples considered subsequently: for the cylinder flow in § 6, the is-OMD algorithm converges with a tolerance of 10^{-2} after 80 iterations of algorithm 1, with each call of OMD in subproblem (2.14) requiring 8.0 s $(N = 500, r = 50)$, giving a total computation time for is-OMD of 640 s; for the application to PIV snapshots of turbulent flow past an axisymmetric bluff body in § 7, is-OMD converges to the same tolerance in only 10 iterations but with a higher cost of 13.8 s $(N = 120, r = 30)$ for each OMD subproblem, resulting in a total computation time of 138 s.

Finally, we note that since both OMD and algorithm 1 are iterative, a possible method to improve algorithmic efficiency would be to apply a (small) fixed number of OMD iterates at each subproblem (2.14). Alternatively, one may apply DMD in place of OMD in subproblem (2.14) to reduce the computational cost. This will be explored in future work. All computations were performed using MATLAB on a standard desktop PC with a 3.4 GHz quad-core Intel i7 processor and 16 GB RAM.

4. Mode extraction and approximation of intermediate snapshots

In order to describe the performance of algorithm 1, consider first a one-dimensional synthetic sinusoidal system

$$f(\boldsymbol{x}, t) = \sum_j A_j \sin(k_i \boldsymbol{x} + \omega_j t) \exp(\gamma_j t), \quad 0 \leqslant \boldsymbol{x} \leqslant 1, t > 0, \quad (4.1)$$

with exponential growth rates γ_j, amplitudes $A_j > 0$, frequencies $\omega_j > 0$ and wavenumbers $k_j > 0$. A reference data set $\mathcal{W} = \{\boldsymbol{w}_i\}_{i=1}^{900}$ is created where each snapshot

$$\boldsymbol{w}_i = \left(f\left(\frac{j}{N_x}, i\delta t \right) \right)_{j=0}^{N_x}, \quad i = 1, \ldots, 900, \quad (4.2)$$

is sampled at $N_x = 10^4$ equally spaced points.

Downsampled, analysis data sets $\mathcal{U} = \{\boldsymbol{w}_i\}_{i \in \mathcal{J}} \subset \mathcal{W}$ are created by defining a subsampling strategy $\mathcal{J} \subset \{1, \ldots, N\}$, and associated with each downsampled set is an average sampling time step

$$\Delta t_{av} := \frac{1}{|\mathcal{J}|} \sum_{i \in \mathcal{J}} \Delta t_i, \quad (4.3)$$

where Δt_i is the time step between subsequent elements of $\mathcal{U}$. For each frequency ω_j, let $\delta t_j^* := \pi / \omega_j$ denote the largest sampling time step which satisfies the Nyquist criterion for that particular frequency. Of interest will be the situations in which $\Delta t_{av} / \delta t_j^* > 1$, indicating cases where, on average, downsampling may prevent accurate frequency identification.

We consider data sampled from (4.1) in the case of two frequencies $\omega_1 = \pi/20$, $\omega_2 = 2\pi/25$, respective amplitudes $A_1 = 1$, $A_2 = 0.5$, wavenumbers $k_1 = 10$, $k_2 = 100$, growth rates $\gamma_1 = 0.01$, $\gamma_2 = 0.05$ and minimal time step $\delta t = 1$. The analysis data ensemble $\mathcal{U}$ is selected with an average sampling time step satisfying

$$\frac{\Delta t_{av}}{\delta t_1^*} = \frac{7}{8}, \quad \frac{\Delta t_{av}}{\delta t_2^*} = \frac{6}{5}, \quad (4.4a,b)$$

meaning that we would expect, according to the Nyquist criterion, to accurately extract ω_1 but not ω_2. Such a selection of an average sampling time step provides a downsampled data ensemble $\mathcal{U}$ of cardinality $|\mathcal{U}| = 45$, i.e. only 5 % of the time-resolved ensemble $\mathcal{W}$ is retained. Algorithm 1 is initialised with $(\boldsymbol{M}_0)_{ij} = N(0, \sigma^2) + \delta_{ij}$, that is, a small perturbation of the identity matrix with $\sigma = \sqrt{0.3}$, and $\boldsymbol{L}_0 = \text{POD}(\boldsymbol{U})$, where $\boldsymbol{U}$ is the analysis data ensemble in a matrix form.

Figure 4 presents the eigenvalues of the matrix $\boldsymbol{M}$ obtained from the is-OMD algorithm, demonstrating that the is-OMD algorithm accurately identifies the oscillatory and exponential parts of the dynamics corresponding to both resolved and undersampled frequencies. For results presented in figure 4 the proportional eigenvalue error is $O(10^{-6})$ for both λ_1 and λ_2. Additionally, figure 4 presents the most energetic columns of $\boldsymbol{L}$. These structures have the shape of sinusoids with an amplitude and wavenumber specified by A_j and k_j. Columns of $\boldsymbol{L}$ retain the prescribed values of k_j, albeit with a visible interaction between k_1 and k_2.

As discussed in § 2.3, the columns of $\boldsymbol{L}$ should be viewed as spanning a subspace of $\mathbb{R}^p$ appropriate to the system dynamics and, consequently, there is no direct
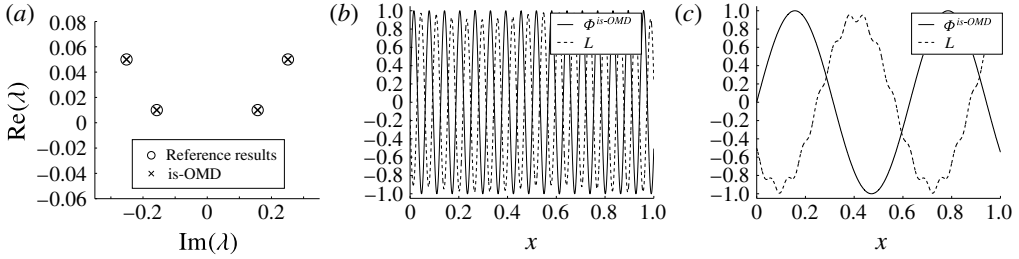
FIGURE 4. Dynamic matrix **M** obtained using the is-OMD algorithm for the synthetic sinusoidal system with exponential dynamics (*a*), structures present in the system corresponding to $k = 100$ (*b*) and to $k = 10$ (*c*) represented by most energetic columns of **L** and is-OMD modes $\boldsymbol{\Phi}^{is\text{-}OMD}$ obtained using algorithm 2.

correspondence between eigenvalues of **M** and columns of **L**. In the OMD algorithm, dynamic modes $\boldsymbol{\Phi}$ are defined as the embedding of the eigenvectors of **M** into a high-dimensional vector space using projection **L**, i.e. $\boldsymbol{\Phi} = \boldsymbol{LP}$, where $\boldsymbol{M} = \boldsymbol{P\Lambda P}^{-1}$, indicating that each mode corresponds directly to a particular eigenvalue. Now, despite the identification of accurate spectral information and projection basis vectors indicated in figure 4, it is the case that the columns of **LP** lack coherence. In particular, the lack of time-resolved data manifests itself in corrupting the eigenvectors of **M**. Although **L** and **M** may still be of sufficient quality to construct an approximate low-order model of the system, if one would like to define dynamic modes the following alternative mode extraction post-processing step is proposed.

4.1. *Dynamic mode extraction*

Given a data ensemble $(\boldsymbol{w}_i)_{i=0}^N$, a typical description (Rowley *et al.* 2009; Schmid 2010; Chen *et al.* 2012) of dynamic modes $\boldsymbol{\Phi}_i$ and eigenvalues λ_i is that they describe the evolution of the first $N - 1$ snapshots according to

$$\boldsymbol{w}_i = \sum_{j=1}^N \lambda_j^i \boldsymbol{\Phi}_j \Rightarrow \boldsymbol{W} = \boldsymbol{\Phi V}, \tag{4.5}$$

where $\boldsymbol{W} = [\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1}]$, $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_N]$ and **V** is the Vandermonde matrix

$$\boldsymbol{V} := \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{N-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_N & \lambda_N^2 & \cdots & \lambda_N^{N-1} \end{bmatrix}. \tag{4.6}$$

In particular, for **V** constructed using DMD eigenvalues, DMD modes can be interpreted as arising from the solution to the minimisation problem

$$\min_{\boldsymbol{\Phi}^{DMD}} \|\boldsymbol{W} - \boldsymbol{\Phi}^{DMD} \boldsymbol{V}\|. \tag{4.7}$$

In view of the observation that accurate spectral information may be identified by algorithm 1, we seek to adapt this approach to calculate dynamic modes from the

output of algorithm 1. In particular, consider the output $\boldsymbol{M}$ with eigenvalues λ_i obtained from a data ensemble with known data only at points $\mathcal{J} \subset \{0, \ldots, N\}$. From this, construct a reduced Vandermonde matrix $\boldsymbol{V}_{\mathcal{J}}$ containing r rows and only the columns whose indices are in $\mathcal{J}$. The dynamic modes associated with algorithm 1 are given in terms of the solution to the reduced minimisation problem

$$\min_{\boldsymbol{\Phi}} \| \boldsymbol{U} - \boldsymbol{\Phi} \boldsymbol{V}_{\mathcal{J}} \|^2. \tag{4.8}$$

This is a least-squares problem with solution, $\boldsymbol{\Phi}$, from which the dynamic is-OMD modes are defined as

$$\boldsymbol{\Phi}_j^{is\text{-}OMD} = \frac{\boldsymbol{\Phi}_j}{\| \boldsymbol{\Phi}_j \|}. \tag{4.9}$$

It is noted that computation of the dynamic modes through an inversion of $\boldsymbol{V}_{\mathcal{J}}$ can be restrictive in the sense that it is assumed that temporal evolution corresponding to coherent structures is highly periodic. In applications where this is not case an alternative approach may be required: an example is presented in § 7 where dynamic modes are obtained by introducing rate constraints to the low-dimensional dynamics.

A technical consideration is that an accurate solution to the problem (4.8) requires $\boldsymbol{V}$ to be full rank. However, it is known that Vandermonde matrices are typically ill-conditioned (Williams 2011). For this reason, it is also desirable to remove rows of the Vandermonde matrix corresponding to eigenvalues which are of comparative unimportance to the system dynamics. This can be achieved by taking a QR decomposition with column pivoting $\boldsymbol{V}_{\mathcal{J}}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{R}\tilde{\boldsymbol{P}}^{\mathrm{T}}$, where $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, $\boldsymbol{R}$ is an upper triangular matrix and $\tilde{\boldsymbol{P}}$ is a column permutation matrix such that absolute diagonal values of $\boldsymbol{R}$, $|r_{ii}|$, are decreasing. A tolerance, $0 < \tilde{t} < 1$, is chosen, and the number or rows retained, n, is set to $n = \max\{i : |r_{ii}| > \tilde{t}|r_{11}|\}$. The intention is that the condition number of the reduced matrix is decreased since rows corresponding to dynamically unimportant eigenvalues (which typically possess a large negative real part) are discarded. The number of removed rows will depend on the tolerance $\tilde{t}$, which has to be set such that majority of dynamically irrelevant eigenvalues are discarded but all of the dynamically relevant features are retained. An appropriate way to select $\tilde{t}$ is outlined in § 6.2, and the post-processing step is summarised in algorithm 2.

Algorithm 2 Dynamic Mode Post-Processing

1: **inputs** $\mathcal{U} = \{\boldsymbol{u}_i : i \in \mathcal{J}\}$, $[\lambda_1, \ldots, \lambda_r]^{\mathrm{T}} = \mathrm{eig}(\boldsymbol{M}_{\text{is-OMD}})$, $0 < \tilde{t} < 1$.
2: $\boldsymbol{V}_{\mathcal{J}} = (v_{\mathcal{J}})_{ji} \leftarrow \lambda_j^{i-1}$, for $i \in \mathcal{J}$ and $j = 1, \ldots, r$.
3: $\boldsymbol{Q}\boldsymbol{R}\tilde{\boldsymbol{P}}^{\mathrm{T}} \leftarrow \mathrm{qr}(\boldsymbol{V}_{\mathcal{J}}^{\mathrm{T}})$, such that $|r_{jj}| > |r_{ii}|$ for $j < i$.
4: $n \leftarrow \max\{i : |r_{ii}| > \tilde{t}|r_{11}|\}$.
5: $\tilde{\boldsymbol{P}} \leftarrow [\tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_n]$, where $\tilde{\boldsymbol{p}}_i$ are columns of $\tilde{\boldsymbol{P}}$.
6: $\boldsymbol{V}_{\mathcal{J}} \leftarrow \tilde{\boldsymbol{P}}^{\mathrm{T}} \boldsymbol{V}_{\mathcal{J}}$.
7: $\boldsymbol{\Phi} \leftarrow \arg\min_{\boldsymbol{\Phi}} \| \boldsymbol{U} - \boldsymbol{\Phi} \boldsymbol{V}_{\mathcal{J}} \|^2$.
8: $\boldsymbol{\Phi}_j^{is\text{-}OMD} \leftarrow \boldsymbol{\Phi}_j / \| \boldsymbol{\Phi}_j \|$
9: $\boldsymbol{\Lambda}^{is\text{-}OMD} \leftarrow [\lambda_1, \ldots, \lambda_r]^{\mathrm{T}} \tilde{\boldsymbol{P}}$
10: **return** $\boldsymbol{\Phi}^{is\text{-}OMD} \in \mathbb{R}^{p \times n}$, $\boldsymbol{\Lambda}^{is\text{-}OMD} \subset [\lambda_1, \ldots, \lambda_r]^{\mathrm{T}}$.

Using the eigenvalues obtained previously, the is-OMD modes of the system (4.1) are obtained from (4.8) for a Vandermonde matrix conditioned with $\tilde{t} = 0.01$.

Figure 4 shows modes corresponding to identified eigenvalues, which accurately match the prescribed spatial wavenumbers k_1 and k_2. Note the difference between the columns of L in figure 4, where there is a visible interaction between both wavenumbers, a phenomenon which is not evident in $\boldsymbol{\Phi}^{is\text{-}OMD}$. Although L is not used in the calculation of $\boldsymbol{\Phi}^{is\text{-}OMD}$, its inclusion in the variable space of algorithm 1 allows for greater freedom in the estimation, which may lead to more accurate λ_i and, hence, more accurate dynamical modes.

4.2. *Intermediate snapshot reconstruction*

Despite the fact that the intermediate snapshot variables w_i in algorithm 1 provide the freedom with which accurate spectral information can be obtained, these variables typically contain high-frequency content not present in the reference data. However, using the eigenvalues $\boldsymbol{\Lambda}^{is\text{-}OMD} = [\lambda_1, \ldots, \lambda_n]$ and dynamic modes $\boldsymbol{\Phi}^{is\text{-}OMD} = [\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_n]$ obtained from algorithm 2, it is possible to obtain an improved approximation of intermediate snapshots corresponding to time steps $i \notin \mathcal{J}$.

To do this, consider the interpretation of the modal decomposition output described through (4.5). At each of the known analysis data snapshots $(u_k)_{k \in \mathcal{J}}$ we consider the projection $(\boldsymbol{\Phi}^{is\text{-}OMD})^{\mathrm{T}} u_k =: (a_j)_{j=1}^r \in \mathbb{R}^r$ and subsequently extrapolate via

$$\tilde{w}_i = \sum_{j=1}^n \lambda_j^i a_j \boldsymbol{\Phi}_j, \quad k = 1, \ldots, n_k - 1, \tag{4.10}$$

where n_k is the number of minimal time steps δt between successive snapshots u_k and u_{k+1}.

The above technique is applied to the system (4.1) with the same parameters as in §4.1, with eigenvalues λ_j and dynamic modes $\boldsymbol{\Phi}_j$ obtained with algorithm 2, to produce intermediate snapshots $\tilde{w}_i$. Figure 5 presents the temporal evolution of the snapshot error $\epsilon_i = \|\tilde{w}_i - w_i^{ref}\|^2 / \|w_i^{ref}\|^2$, where it can be seen that due to the linear nature of the system, the error increases linearly when $\tilde{w}_i$ is estimated further away from known snapshots u_i. However, since the eigenvalues λ are estimated accurately, snapshot extrapolation gives a small error $\epsilon_i < 10^{-4}$.

5. A comparison with compressive sampling

The statistical performance of both is-OMD and compressive sampling in terms of the ability to extract frequency content from undersampled data will be analysed on the one-dimensional synthetic sinusoidal flow

$$f(\boldsymbol{x}, t) = \sum_j A_j \sin(k_j \boldsymbol{x} + \omega_j t), \quad 0 \leqslant \boldsymbol{x} \leqslant 1, t > 0, \tag{5.1}$$

with amplitudes $A_j > 0$, frequencies $\omega_j > 0$ and wavenumbers $k_j > 0$. Similarly to the system in §4, reference data set $\mathcal{W} = \{w_i\}_{i=1}^{900}$ is generated where each snapshot

$$w_i = \left(f\left(\frac{j}{N}, \mathrm{i}\delta t \right) \right)_{j=0}^{N_x}, \quad 0 = 1, \ldots, 900, \tag{5.2}$$

is sampled at $N_x = 10^4$ equally spaced points in $0 \leqslant \boldsymbol{x} \leqslant 1$. Algorithm 1 will be compared with the compressive sampling approach developed in Tu *et al.* (2014),
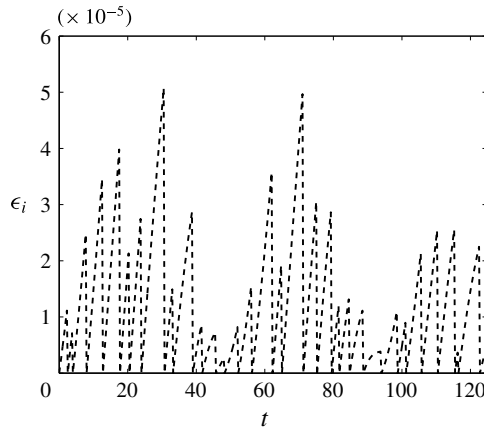
FIGURE 5. Temporal evolution of the snapshots reconstruction error $\epsilon_i = \|\tilde{\boldsymbol{w}}_i - \boldsymbol{w}_i^{ref}\|^2 / \|\boldsymbol{w}_i^{ref}\|^2$, where $\tilde{\boldsymbol{w}}_i$ is defined by (4.10).

which will be referred to here as CS-DMD. CS-DMD is implemented by first obtaining an approximation of the fully resolved data set, $\mathcal{W}$, from the analysis data set $\mathcal{U}$ using a CoSaMP algorithm (Needell & Tropp 2009). Subsequently, the DMD eigenvalues corresponding to $\mathcal{W}$ are obtained. Accuracy is described by the proportional error in extracted frequency

$$\epsilon_\lambda := \frac{|\omega_j - \mathrm{Im}(\lambda)|}{\omega_j}, \qquad (5.3)$$

where λ are the eigenvalues obtained by each of the two methods.

5.1. *Robustness of CS-DMD and is-OMD*

We consider data sampled from (5.1) in the case of two frequencies $\omega_1 = \pi/20$, $\omega_2 = 2\pi/25$, with respective amplitudes $A_1 = 1, A_2 = 0.5$, and minimal time step $\delta t = 1$. The analysis data ensemble $\mathcal{U}$ is selected with an average sampling time step satisfying

$$\frac{\Delta t_{av}}{\delta t_1^*} = \frac{7}{8}, \qquad \frac{\Delta t_{av}}{\delta t_2^*} = \frac{6}{5}, \qquad (5.4a,b)$$

as in § 4.

The first step of CS-DMD is to apply compressive sampling to $\mathcal{U}$ to form an approximation $\hat{\mathcal{W}}$ to the fully resolved data ensemble $\mathcal{W}$. Recall from § 2.2 that this is achieved by (i) choosing and fixing a temporal basis $\boldsymbol{\Xi}$, then (ii) successively adding a number, k say, of elements of this basis which best correlate with the available data. Motivated by the fact that the underlying data sampled from (5.1) are sparse in the Fourier basis, we set $\boldsymbol{\Xi}$ to be the inverse Fourier Transform and obtain approximations $\hat{\mathcal{W}}_k \approx \mathcal{W}$ corresponding to extracting $k = 4, 8$ and 20 dominant basis vectors. DMD is then applied to each $\hat{\mathcal{W}}_k$ to determine the identified eigenvalue information.

Since we do not want to assume *a priori* the number of dynamically important structures present in a given data set, it is desirable that the accuracy of the frequency
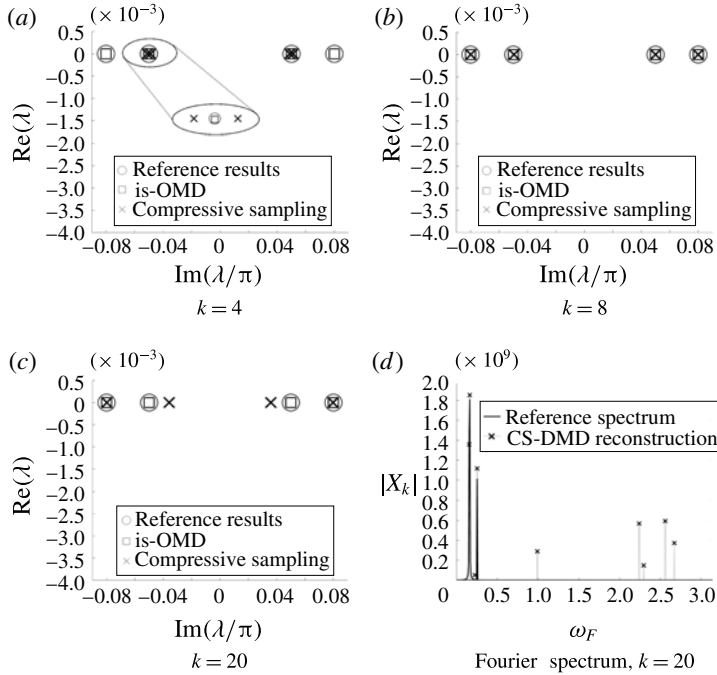
FIGURE 6. The DMD eigenvalue lattice obtained using snapshots reconstructed with CS-DMD for different numbers of sparse structures (*a–c*), with $\lambda_{is\text{-}OMD}$, is given for $r = 45$. The Fourier spectrum obtained by compressive sampling for $k = 20$ is shown in (*d*).

information provided by CS-DMD behaves predictably with the requested number of basis elements k. Figure 6 indicates that this is not the case. Indeed, the frequencies ω_i are accurately identified only in the case $k = 8$, with $k = 4$ failing to identify the higher frequency ω_2, while the choice $k = 20$ results in an underapproximation of ω_1.

For $k = 4$, inaccuracy arises from the discrete Fourier frequencies implied by the chosen basis Ξ. Since the fully resolved data ensemble contains 900 samples, the chosen basis contains only discrete frequencies $\omega_F = (2\pi n/900)_{n=0}^{900}$, k of which are used to express the data. For this particular example, no single ω_F coincides with $\omega_1 = \pi/20$. Instead there are multiple values of ω_F in a small neighbourhood of ω_1 and, since $A_1 > A_2$, the CoSamp algorithm applied with $k = 4$ (incorrectly) identifies a pair of frequencies ω_F close to ω_1. The result is improved for $k = 8$, which is sufficiently large to resolve both ω_1 and the higher frequency ω_2. However, for $k = 20$ it is shown in figure 6(*d*) that compressive sampling identifies Fourier frequencies ω_F not present in the underlying system, in addition to those close to the true values $\omega/\pi = 0.05, 0.08$. Upon application of DMD to $\mathcal{W}_k$, these extra frequencies result in an incorrect identification of ω_1.

Such a lack of robustness with respect to k is not evident in the application of algorithm 1 to this example, which, as can be seen in figure 6(*a–c*), correctly identifies the two underlying frequencies in the data using a model of order $r = 45$. The is-OMD results in figure 6(*a–c*) present only four eigenvalues of interest. The eigenvalues of relevance were chosen based on the energetic criterion similar to approaches in Rowley *et al.* (2009), Chen *et al.* (2012), Jovanovic *et al.* (2014), where the magnitudes of the DMD modes are examined. Here, the magnitude of a
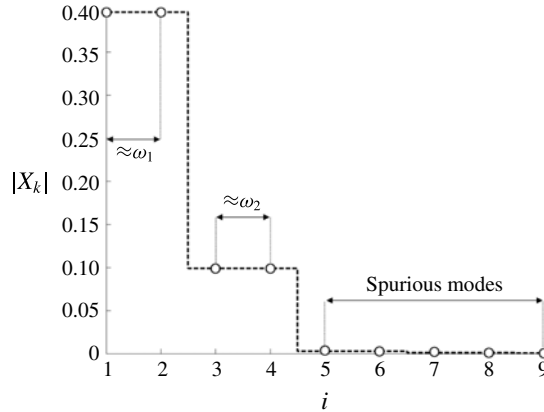
FIGURE 7. The relative contribution of the is-OMD modes.

solution to (4.8), $\|\boldsymbol{\Phi}_k\|_2$, is assessed and eigenvalues with the largest magnitude of the associated modes are selected. Note that, if $\boldsymbol{V}_{\mathcal{J}}$ contains dynamically unimportant eigenvalues with large decay rate, the associated modes will have an excessively large magnitude in order to compensate for the low value of λ_j^{i-1} at later time steps. Therefore, in order to obtain a meaningful ranking of modes based on the energy criterion, it is important to remove dynamically unimportant rows of $\boldsymbol{V}_{\mathcal{J}}$ using algorithm 2. Here, (4.8) is solved using $\boldsymbol{V}_{\mathcal{J}}$ conditioned with $\tilde{t} = 10^{-3}$ retaining nine rows of Vandermonde matrix, and consequently nine structures. Figure 7 shows a relative energetic contribution of each mode to the system dynamics defined as $|X_k| = \|\boldsymbol{\Phi}_k\|_2/\sum_j \|\boldsymbol{\Phi}_k\|_2$, demonstrating a clearly visible drop in $\|\tilde{\boldsymbol{\Phi}}_k\|_2$ between dynamically important and unimportant eigenvalues and allowing the identification of frequencies presented in figure 6(a–c).

Note that another variable prescribed *a priori* in the CS-DMD methodology is the rank r of the DMD approximation, which is upper bounded by the number of extracted basis vectors k. The CS-DMD results presented in figure 6 were obtained with DMD approximation of rank $r = 4$ as opposed to $r = 45$ of the is-OMD results. The reason for selecting this rank for the CS-DMD approximation can be explained if the maximal case $r = k$ is instead considered. With this choice, DMD eigenvalue information will replicate the Fourier frequencies which most accurately correlate with the analysis data ensemble, $\mathcal{U}$. Figure 8(a) presents the relative contribution of each mode for $r = k = 20$. Although the frequencies corresponding to the most dominant modes are close to the true values, it is nonetheless not clear from figure 8(a) how to choose a threshold which will allow frequencies of dynamical importance to be retained. A similar situation occurs in figure 8(c) for $r = k = 8$.

For $r = k = 8$ eigenvalues coincide with ω_F which lie in a small neighbourhood of ω_1 and ω_2 with a small error ϵ_λ corresponding to those frequencies. However, it is shown in figure 6(b) that if a combination $k = 8$, $r = 4$ is used, ϵ_λ decreases further. Based on these observations, in §5.2 where a statistical sample of sinusoidal flows is investigated, for CS-DMD methodology a combination of $k = 8$ and $r = 4$ is utilised.
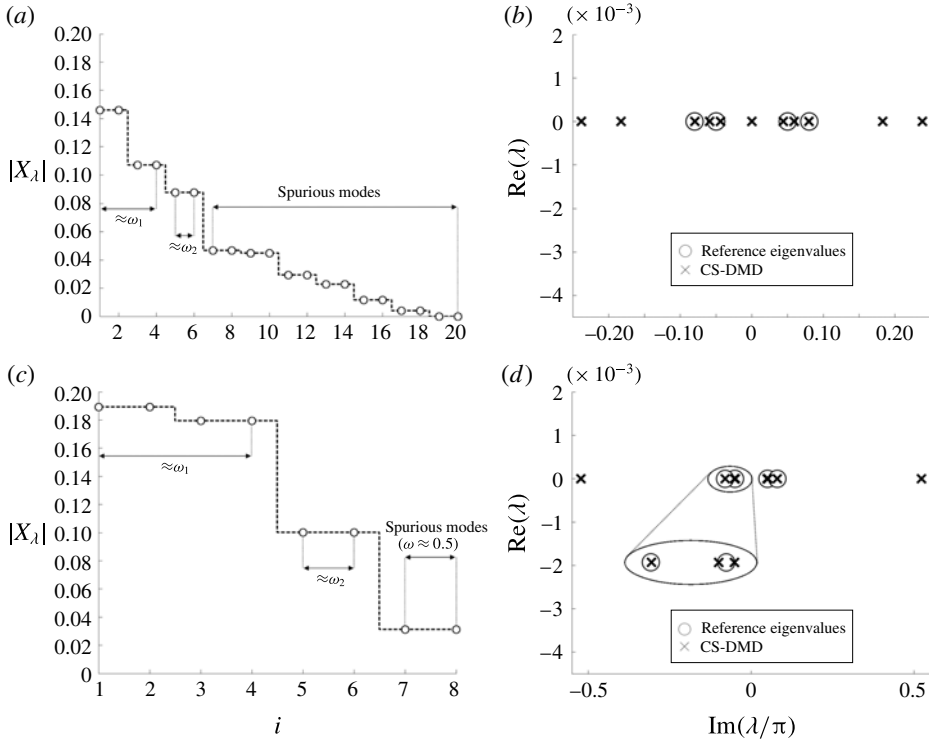
FIGURE 8. The output of the CS-DMD methodology for different values of k and r where $k = r$. The importance of modes and corresponding eigenvalue spectrum for $k = r = 20$ is shown in (a,b) while those corresponding to $k = r = 8$ are presented in (c,d).

5.2. *Frequency estimation analysis: statistical comparison between compressive sampling and is-OMD*

To study the differences in performance of is-OMD and compressive sampling, both methodologies were applied to the system (5.1) for a range of frequencies and average sampling rates. In particular, the lower frequency is fixed, as above, to $\omega_1 = \pi/20$, while the higher system frequency takes one of values

$$\omega_2(n) = \omega_1 + \frac{n\pi}{100}, \quad n = 1, \ldots, 20. \tag{5.5}$$

Subsequently, for each of value of $\omega_2(n)$ a number of downsampled analysis data ensembles $\mathcal{U}_{nm}$ are constructed whose average sampling rates satisfy

$$\frac{\Delta t_{av}(\mathcal{U}_{nm})}{\delta t_2^*} = 1 + \frac{m}{20}, \quad m = 1, \ldots, 55, \tag{5.6}$$

that is, each exceeds the Nyquist criterion to differing degrees. This is achieved by constructing fully resolved data sets

$$\mathcal{W}_{nm} = \left\{ f\left(x, i\left(\frac{1 + m/20}{1 + n/5} \right) \right) : 0 \leqslant x \leqslant 1, \ i = 1, \ldots, 900 \right\} \tag{5.7}$$
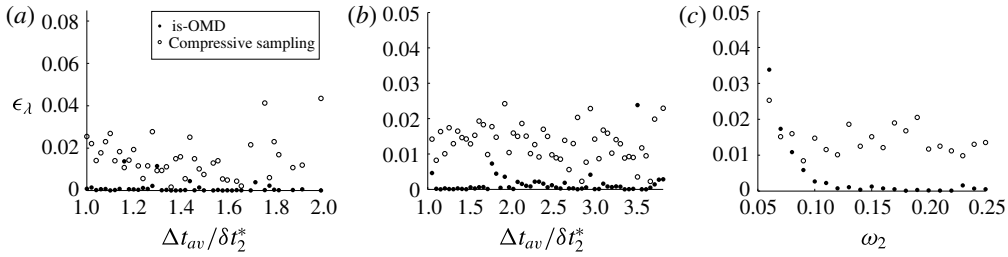
FIGURE 9. Average proportional frequency error ϵ_λ corresponding to ω_1 (*a*) and ω_2 (*b*) for each $\Delta t_{av}/\delta_2^*$; and corresponding to each ω_2 when ϵ_λ is averaged over a range of different downsampling ratios $\Delta t_{av}/\delta_2^*$ (*c*).

and sampling non-uniform analysis data ensembles $\mathcal{U}_{nm} \subset \mathcal{W}_{nm}$, each of cardinality $|\mathcal{U}_{nm}| = 45$. Note that the fully resolved data set is varied for each pair (n, m) in order that analysis data sets with different average downsampling rates (as a proportion of δt_2^*) can be constructed to possess the same proportion, 5 %, of elements to the fully resolved ensemble.

The CS-DMD and is-OMD algorithms are applied to each data set $\mathcal{U}_{nm}$. Motivated by § 5, we apply the CoSamp algorithm with sparse basis matrix $\boldsymbol{\Xi}$ as the inverse Fourier transform and set the sparsity level to $k = 8$. Algorithm 1 is initialised with $(\boldsymbol{M}_0)_{ij} = N(0, \sigma^2) + \delta_{ij}$, that is, a small perturbation of the identity matrix with $\sigma = \sqrt{0.3}$, and $\boldsymbol{L}_0 = \text{POD}(\mathcal{U}_{nm})$, with the modelling dimension set as $r = |\mathcal{U}_{nm}| = 45$.

Figures 9(*a*) and 9(*b*) show the proportional frequency error ϵ_λ corresponding to ω_1, ω_2, respectively, averaged over the different values of $\Delta t_{av}(\mathcal{U}_{nm})/\delta t_2^*$. Note that, since the data are sampled in such a way to achieve a particular value of $\Delta t_{av}/\delta t_2^*$, the corresponding downsampling ratio for δt_1^* is $\Delta t_{av}/\delta t_1^* = (\Delta t_{av}/\delta t_2^*)(\omega_1/\omega_2)$, meaning that $\Delta t_{av}/\delta t_1^*$ are not similarly grouped. Consequently, a discrete set of values is defined, $\mathcal{S} = \{1.02, 1.04, \dots, 1.92\}$, and for each element $\mathcal{S}_i$ results satisfying $|\Delta t_{av}/\delta t_1^* - \mathcal{S}_i| < 0.01$ are grouped together and averaged accordingly. It is clear from figure 9(*a*,*b*) that the proportional frequency error is reduced by application of algorithm 1 in comparison to CS-DMD.

Figure 9(*c*) shows the proportional frequency error ϵ_λ corresponding to ω_2, as ω_2 is varied, each averaged over the 55 different values of the downsampling ratio $\Delta t_{av}/\delta t_2^*$. The CS-DMD error is approximately constant, while to is-OMD error decays as $|\omega_1 - \omega_2|$ increases. It should be noted that such a pattern can be observed if the OMD algorithm is applied to fully time-resolved snapshots $\mathcal{W}$ which contain observation noise. This is, pairs of eigenvalues satisfying $|\text{Im}(\lambda_i) - \text{Im}(\lambda_j)| \ll 1$ are sensitive to observation noise.

6. Numerical example: cylinder wake

Algorithms 1 and 2 are applied to the numerical results of the circular cylinder flow at $Re = 60$ and their performance is analysed on the limit cycle regime. The numerical data were obtained using the in-house solver called Pantarhei, which was validated for the cylinder flow in Lu & Papadakis (2011, 2014). The code uses the finite volume method in an unstructured grid. The spatial terms of Navier–Stokes equations were discretised using a second-order central scheme and the system was advanced in time using a second-order, three-point backward scheme. The domain used for the simulations was 48D long and 32D wide, where D is the cylinder diameter.

The distance between cylinder location and upstream and downstream boundaries is respectively $16D$ and $32D$. The total number of cells was $62\,791$. The grid was non-uniform with the minimum grid spacing $h \approx D/60$ near the cylinder boundary. Uniform inflow ($U_\infty = 1$), convective outflow and symmetric boundary conditions on the lateral sides of the domain were imposed. Additionally, a no-slip condition was imposed on the cylinder boundary. Density and viscosity of the flow were modified accordingly to achieve the desired Reynolds number. This dynamical system exhibits a pair of vortices oscillating at a shedding frequency together with the series of higher harmonics; each oscillating at an integer multiple of the shedding frequency. These values can be extracted by applying the OMD algorithm to time-resolved data.

The performance of is-OMD will be assessed by comparing the eigenvalue information between eigenvalues retained in the post-processing algorithm 2, $\boldsymbol{\Lambda}^{is\text{-}OMD}$, and λ_{OMD} from the OMD algorithm applied to time-resolved data. Three aspects of algorithm initialisation and their impact on the extraction of dynamic information are discussed here: the initialisation of $\boldsymbol{L}_0$ and $\boldsymbol{M}_0$; the projection rank r; and sampling strategy $\mathcal{J}$. A downsampled data set $\mathcal{U}$ is extracted randomly from a reference data set $\mathcal{W}$ uniformly sampled at $\delta t = 0.25 s$ containing 30 snapshots per shedding period. Here, it is convenient to introduce the following quantities related to the sampling strategy: $\Delta t_{av} = \sum_{i \in \mathcal{J}} \Delta t_i / |\mathcal{J}|$, $\Delta t_{min} = \min\{\Delta t_i \mid i \in \mathcal{J}\}$ and $\Delta t_{max} = \max\{\Delta t_i \mid i \in \mathcal{J}\}$. Unless stated otherwise, $\mathcal{U}$ is sampled such that $\Delta t_{av} = 10\delta t$ without constraints on Δt_{min} and Δt_{max}. Applying a Nyquist criterion to data downsampled at a constant $\Delta t = 10\delta t$, such downsampling is sufficient to resolve the vortex-shedding frequency but not satisfactory if the objective is to identify higher-frequency modes.

6.1. *Initial condition analysis*

To assess the most appropriate properties of $\boldsymbol{L}_0$ and $\boldsymbol{M}_0$ to initialise is-OMD, three classes of initial condition are considered:

   (i) $\boldsymbol{L}_0$ is set to POD modes of the known data, $\boldsymbol{U}$, and an identity matrix chosen to approximate $\boldsymbol{M}_0$, i.e. $\boldsymbol{L}_0 = \text{POD}(\boldsymbol{U})$, $\boldsymbol{M}_0 = \boldsymbol{I}_r$;

  (ii) $\boldsymbol{L}_0 = \text{POD}(\boldsymbol{U})$ and an identity matrix with entries perturbed by noise used for $\boldsymbol{M}_0$: $(\boldsymbol{M}_0)_{ij} = N(0, \sigma) + \delta_{ij}$;

 (iii) OMD is applied to data undersampled with a constant time step Δt, $\overline{\boldsymbol{U}}$, to obtain $\boldsymbol{L}_0$ and $\boldsymbol{M}_0$, i.e. $[\boldsymbol{M}_0, \boldsymbol{L}_0] = \text{OMD}(\overline{\boldsymbol{U}})$.

The motivation for $[\boldsymbol{M}_0, \boldsymbol{L}_0] = \text{OMD}(\overline{\boldsymbol{U}})$ is that in systems exhibiting limit cycle behaviour, their dynamics does not qualitatively change with time. In such a case, two sets of data can be acquired, one with non-uniform Δt which is used as an input to is-OMD, $\boldsymbol{U}$, and a second one which is uniformly subsampled, $\overline{\boldsymbol{U}}$. In such a case $\overline{\boldsymbol{U}}$ is not sufficient to obtain all the dynamics of interest; however, the output of OMD can still be used for $\boldsymbol{L}_0$ and $\boldsymbol{M}_0$.

In order to consider the role of the initial conditions, the same downsampling strategy and approximation rank r are used for all three cases. In particular, $\mathcal{U}$ is sampled non-uniformly, with $\Delta t_{av} = 10\delta t$ and without constraints on Δt_{min} and Δt_{max}, while the projection rank is set to $r = 75$. The dimensions of both data sets $\mathcal{W}$, $\mathcal{U}$ are $\boldsymbol{W} \in \mathbb{R}^{p \times 750}$, $\boldsymbol{U} \in \mathbb{R}^{p \times 75}$, i.e. $|\mathcal{I}| = 750$ and $|\mathcal{J}| = 75$. The $\overline{\boldsymbol{U}} \in \mathbb{R}^{p \times 75}$ was sampled such that $\Delta t_{av} = 10\delta t$, which is sufficient to resolve a time scale of $2/3T$, where T is a shedding frequency. Note that cylinder flow contains harmonics with time scales of T/n for $n \in \mathbb{Z}^+$, meaning that $2/3T$ is not sufficient to resolve the subharmonic with a time scale of $T/2$.
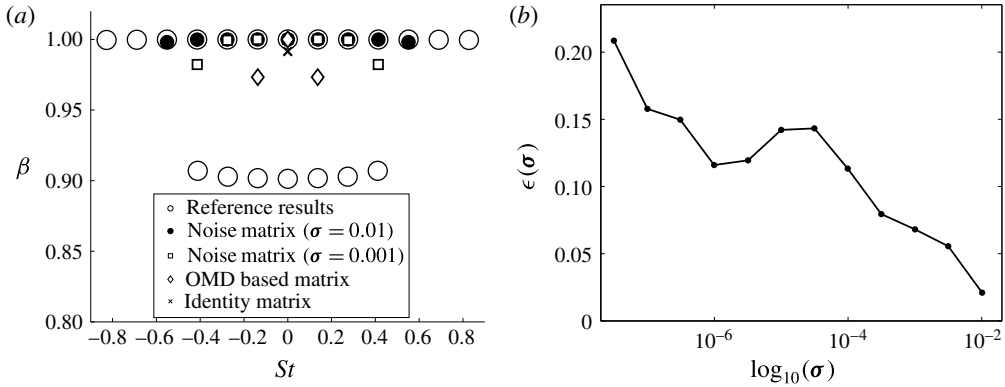
FIGURE 10. Eigenvalues of the dynamic matrix $\boldsymbol{M}$ obtained using the is-OMD algorithm for different initial conditions (*a*). The error in eigenvalue magnitude for identity matrices with $\boldsymbol{\sigma} \in [10^{-9}, 10^{-2}]$ (*b*), obtained for the projection rank $r = 75$.

The extracted dynamical information for $\boldsymbol{M}_0 = \boldsymbol{I}_r$, $[\boldsymbol{M}_0, \boldsymbol{L}_0] = \mathrm{OMD}(\overline{\boldsymbol{U}})$ and $(\boldsymbol{M}_0)_{ij} = N(0, \sigma) + \delta_{ij}$ with $\sigma = 10^{-2}, 10^{-3}$ are presented in figure 10. In all of the analysed cases, only the eigenvalues retained in algorithm 2 for $\tilde{t} = 0.3$ are presented. The Strouhal number is defined as $St = fD/U_\infty$, where D is the cylinder diameter and U_∞ is the inflow velocity. Assuming that flow field was non-dimensionalised using a cylinder diameter and an inflow velocity, the Strouhal number can be extracted from the phase information of each eigenvalue, $St = \mathrm{Im}(\lambda)/(2\pi)$. The magnitude of the eigenvalue provides the information about the growth rate, which is expressed in such a way so that growth rate, β, follows β^t, i.e. $\beta = \exp(\mathrm{Re}(\lambda))$. The reference eigenvalues in figure 10 were obtained by applying the OMD algorithm to $\boldsymbol{W}$ for $r = 30$, and consist of the asymptotic eigenvalues representing the limit cycle behaviour and eigenvalues at $\beta = 0.9$ characterising relaxation of the flow onto its limit cycle (Bagheri 2013). The primary objective is to assess the accuracy at which the is-OMD algorithm approximates the asymptotic eigenvalues, since the values at $\beta = 0.9$ correspond to dynamics at extremely low energy for the chosen data set $\boldsymbol{W}$.

In the case where $\boldsymbol{M}_0 = \boldsymbol{I}_r$, a majority of the eigenvalues of $\boldsymbol{M}$ are characterised by $\mathrm{Re}(\lambda) \ll 0$ and are discarded in algorithm 2 due to the choice of $\tilde{t}$. None of the frequencies present in the system are recovered. For $[\boldsymbol{M}_0, \boldsymbol{L}_0] = \mathrm{OMD}(\overline{\boldsymbol{U}})$, only frequencies within time-scale range of $\overline{\boldsymbol{U}}$ are correctly estimated (i.e. $St \in [-0.2, 0.2]$). Figure 10 demonstrates that noise must be added to $\boldsymbol{M}_0$ in order to extract high-frequency harmonics. It is interesting to note that as the standard deviation increases, more frequency information is recovered. In particular, for an order of magnitude increase in σ, an additional conjugate pair of eigenvalues is correctly identified and the growth rate of all eigenvalues, β, is more accurate.

In order to further demonstrate a beneficial influence of noise addition, a repeated analysis was performed for $(\boldsymbol{M}_0)_{ij} = N(0, \sigma) + \delta_{ij}$ with $\sigma \in [10^{-9}, 10^{-2}]$. Accuracy in the estimation of seven most dominant eigenvalues is analysed and the proportional error is defined as $\epsilon := \sum_{i=1}^{7} |\lambda_i^{true} - \lambda_i^{is\text{-}OMD}| / \sum_{i=1}^{7} |\lambda_i^{true}|$. Figure 10 demonstrates that ϵ decreases as σ increases. It is clear that, in order for the is-OMD algorithm to successfully recover dynamical information, a certain degree of flexibility has to be included in $\boldsymbol{M}_0$. For $\boldsymbol{M}_0 = \boldsymbol{I}_r$ and $[\boldsymbol{M}_0, \boldsymbol{L}_0] = \mathrm{OMD}(\overline{\boldsymbol{U}})$, the dynamics corresponding to dominant structures present in the system, i.e. constant mean mode and shedding

vortices, are strongly embedded in the initial matrix structure. In such cases the objective function can be significantly reduced in few iterations, indicating that such a $\boldsymbol{M}_0$ predetermines a local minimum of (2.10). This problem appears to be avoided simply by randomly perturbing the initial matrices. Note that it is possible to adjust a frequency range of the eigenvalue spectrum of $(\boldsymbol{M}_0)_{ij} = N(0, \sigma) + \delta_{ij}$ by adjusting a value of σ, where larger value of σ will cover larger range of frequencies. If the system exhibits primarily low-frequency dynamics it is desirable to set σ to a lower value such that an eigenvalue spectrum of $\boldsymbol{M}_0$ is in a neighbourhood of the expected range of the frequencies of interest.

6.2. *Rank size analysis*

It was demonstrated in § 5.1 that setting a high value of k in the CS-DMD approach induces Fourier modes corresponding to high-frequency oscillations not present in time-resolved data. It is therefore important to consider the impact of the equivalent parameter in the is-OMD algorithm, that is, the dimension r of Im($\boldsymbol{L}$). It will be demonstrated here that larger r improves accuracy in the recovered eigenvalues, which consequently leads to an improved accuracy of $\boldsymbol{\Phi}^{is\text{-}OMD}$ in algorithm 2. An analysis in terms of the eigenvalue and mode estimation is performed for $r = [25, 50, 75]$ using the same sampling strategy as in § 6.1 and setting $(\boldsymbol{M}_0)_{ij} = N(0, 0.3) + \delta_{ij}$, $\boldsymbol{L}_0 = \text{POD}(\boldsymbol{U})$.

Figure 11 presents eigenvalue lattices for different values of r and demonstrates a positive impact of increasing r on the accuracy of the dynamic estimation. When r increases, a progressively greater number of asymptotic eigenvalues is recovered and their growth rates β converge towards the correct value of 1. The number of modes which can be calculated is proportional to the number of recovered eigenvalues. For r of 25, 50 and 75 there are respectively one, three and four conjugate pairs of eigenvalues that are estimated with small error, and with growth rate $\beta > 0.95$. The dynamic modes are estimated as a solution to (4.8) which in turn depends upon $\boldsymbol{V}_\mathcal{J}$ being constructed using accurate eigenvalues. Figure 13 shows dynamic modes corresponding to eigenvalues with $\beta > 0.95$ obtained for $r = 75$, and it demonstrates that the mode shapes closely resemble the coherent structures obtained from time-resolved data (figure 12). The dynamic modes of similar precision can be obtained for other eigenvalues presented in 11, where $\beta > 0.95$, therefore mode shapes for $r = 25$ and $r = 50$ are omitted here.

Note that for $r = 25$ and $r = 75$ a pair of eigenvalues is recovered which contains an error in β. Figure 14 shows the corresponding dynamic modes. Although these structures are not as coherent as for eigenvalues where $\beta > 0.95$, there is nonetheless a resemblance to the structures present in the flow (figure 12). It is expected that when r is increased even further the eigenvalue accuracy improves, leading to better precision in the estimation of the dynamic modes. In practice, when $\boldsymbol{L}_0 = POD(\boldsymbol{U})$ and $(\boldsymbol{M}_0)_{ij} = N(0, \sigma) + \delta_{ij}$ are used, r is only limited by $|\mathcal{J}|$. This quantity is usually much higher than for the results presented here, and the choice of r would be limited only by the processing cost.

When unknown data sets are reconstructed using (4.10), the value of r impacts the accuracy of estimated snapshots, $\tilde{\boldsymbol{W}} = [\tilde{\boldsymbol{w}}_1, \ldots, \tilde{\boldsymbol{w}}_N]$ for $i \notin \mathcal{J}$, through the precision and number of reconstructed eigenvalues λ_j and dynamic modes $\boldsymbol{\Phi}_j$, and is expected to improve when r is increased. This fact is confirmed in figure 11, which presents the evolution of the error norm, $\epsilon_r = (1/|\mathcal{I}|) \sum_{i \in \mathcal{I}} \| \tilde{\boldsymbol{W}} - \boldsymbol{W}^{ref} \|^2 / \| \boldsymbol{W}^{ref} \|^2$ for $r \in [10, 70]$. There is a visible decay in the value of ϵ_r, where ϵ_r is less than 5 % for $r \geqslant 30$.
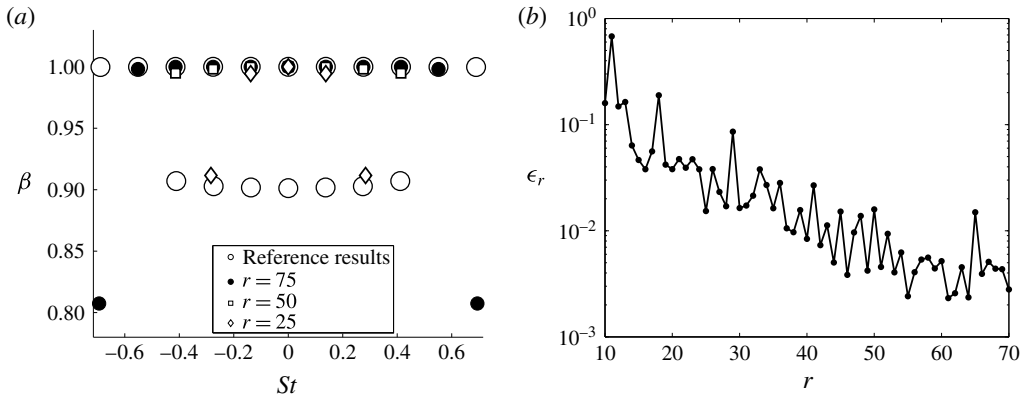
FIGURE 11. Eigenvalues of the dynamic matrix M for $r = [25, 50, 75]$ (a). The error $\epsilon_r = (1/|\mathcal{I}|) \sum_{i \in \mathcal{I}} \| \tilde{W} - W^{ref} \|^2 / \| W^{ref} \|^2$ for $(M_0)_{ij} = N(0, 0.3) + \delta_{ij}$ with different r (b).
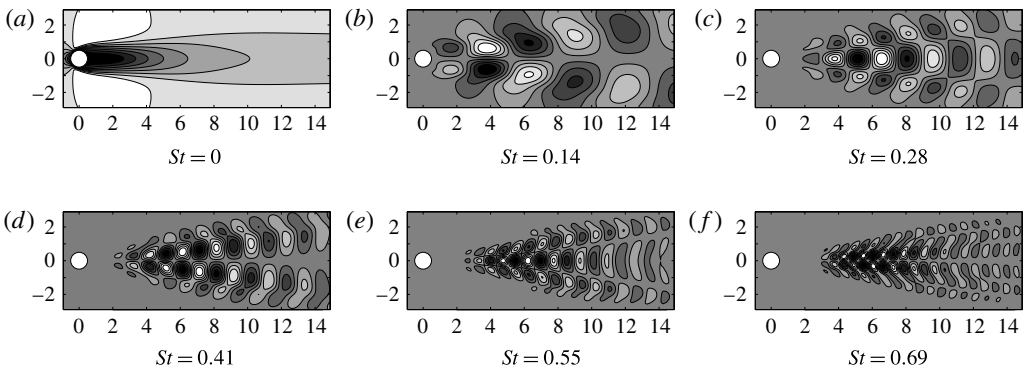

FIGURE 12. Modes $\boldsymbol{\Phi} = \boldsymbol{LP}$ obtained by applying the OMD algorithm to a time-resolved cylinder flow data.
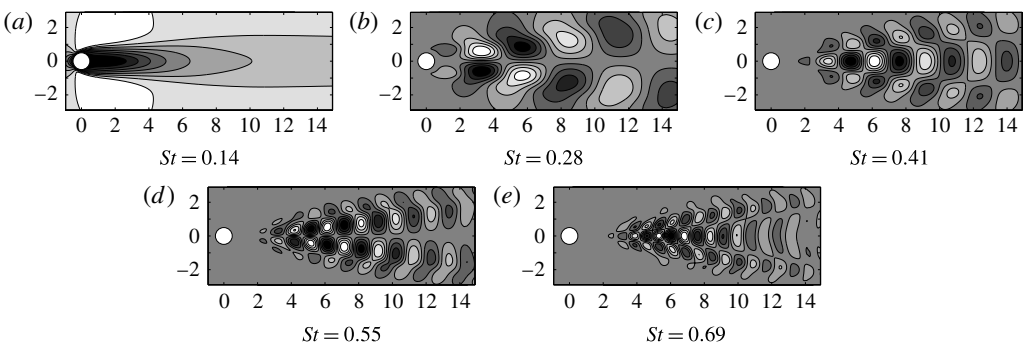

FIGURE 13. Modes $\boldsymbol{\Phi}^{is\text{-}OMD}$ corresponding to eigenvalues with $\beta > 0.95$ obtained using the is-OMD algorithm for $r = 75$.

Note that, despite being governed by nonlinear Navier–Stokes equations, cylinder flow in the limit cycle is known to be well approximated by a linear model consisting
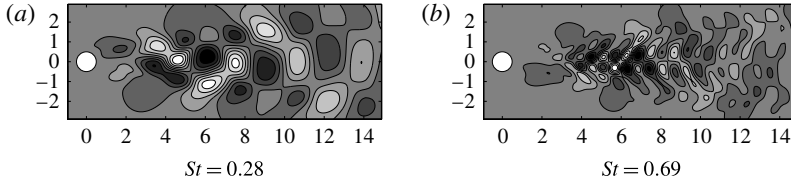
FIGURE 14. Modes $\boldsymbol{\Phi}^{is\text{-}OMD}$ corresponding to eigenvalues presented in figure 11 with $\beta < 0.95$; $St = 0.28$, $\beta = 0.91$ for $r = 25$ (a) and $St = 0.69$, $\beta = 0.81$ for $r = 75$ (b).
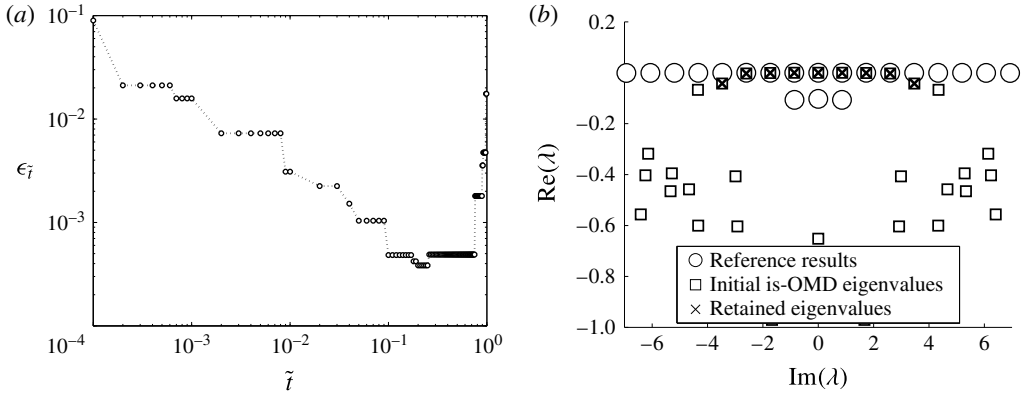

FIGURE 15. The error $\epsilon_{\tilde{t}} = (1/|\mathcal{J}|) \sum_{i \in \mathcal{J}} \|\tilde{\boldsymbol{w}}_{i+n_i} - \boldsymbol{u}_{i+n_i}\| / \|\boldsymbol{u}_{i+n_i}\|$ for different $\tilde{t}$ (a) and retained is-OMD eigenvalue spectrum for $\tilde{t} = 0.21$ (b).

of oscillating structures. For more dynamically complex systems, the value of ϵ_r is expected to have a larger magnitude than values presented in figure 11.

Finally, we note that the accuracy of $\tilde{\boldsymbol{W}}$ depends upon the value of λ_j used in (4.10). The number and choice of λ_j is determined based on the number of retained rows in $\boldsymbol{V}_{\mathcal{J}}$, which is conditioned with an appropriate value of $\tilde{t}$. When $\tilde{t}$ is small, a majority of λ_j in $\boldsymbol{V}_{\mathcal{J}}$ are retained, including dynamically unimportant values, introducing an error in $\tilde{\boldsymbol{W}}$. On the other hand, if $\tilde{t}$ is large, dynamically relevant λ_j are removed. In order to find an appropriate value of $\tilde{t}$, and consequently number of retained λ_j, let us compare snapshots estimated using (4.10) with reference measurements. Starting an extrapolation at each $i \in \mathcal{J}$ a snapshot is estimated at the time step when the next measurements is acquired, i.e. $\tilde{\boldsymbol{w}}_{i+n_i}$, where $n_i := \Delta t_i / \delta t$ is the number of minimal time steps δt between the next element of $\mathcal{U}$. Performing a comparison with all of the available measurements, a proportional error metric $\epsilon_{\tilde{t}} = (1/|\mathcal{J}|) \sum_{i \in \mathcal{J}} \|\tilde{\boldsymbol{w}}_{i+n_i} - \boldsymbol{u}_{i+n_i}\| / \|\boldsymbol{u}_{i+n_i}\|$ is defined, and $\tilde{t}$ is estimated such that the resulting $\tilde{\boldsymbol{w}}_{i+n_i}$ minimises $\epsilon_{\tilde{t}}$.

An example for the cylinder flow is-OMD results with $r = 70$ is presented in figure 15, which shows $\epsilon_{\tilde{t}}$ for $\tilde{\boldsymbol{w}}_{i+n_i}$ estimated with $\tilde{t} \in [10^{-4}, 1]$. For this particular case, $\epsilon_{\tilde{t}}$ is minimised at $\tilde{t} = 0.21$, where $\tilde{\boldsymbol{W}}$ is calculated using the nine most dominant eigenvalues (figure 15). Although $\tilde{t} = 0.21$ gives the smallest values of $\epsilon_{\tilde{t}}$, there is a clear trough visible for $\tilde{t} \in [0.01, 0.75]$, and setting $\tilde{t}$ to any value in that range will give a good estimation of snapshots at $i \notin \mathcal{J}$. In figure 11, each value of ϵ_r was obtained for $\tilde{\boldsymbol{W}}$ estimated with the smallest value of $\tilde{t}$.

6.3. *The importance of random sampling*

One of the key properties of the compressive sampling methodology, described in §2.2, is its reliance on non-uniform sampling (Candes & Tao 2006*b*). Here, it is demonstrated that is-OMD can also be improved by random sampling. Three different sampling strategies are analysed, with approximately the same $|\mathcal{J}|$ and Δt_{av} for different Δt_{min} and Δt_{max}. Depending on Δt_{min}, it is possible to use the Nyquist sampling criterion to estimate the range of different time scales that can be recovered as a multiple of shedding time scale T. Setting $(\boldsymbol{M}_0)_{ij} = N(0, 0.3) + \delta_{ij}$, $\boldsymbol{L}_0 = POD(\boldsymbol{U})$ and $r = 70$, is-OMD is applied to the following sampling strategies.

(i) Uniform sampling: $\mathcal{U}$ is sampled uniformly with $\Delta t = 10\delta t$. Such a sampling strategy corresponds to three snapshots per shedding period and is sufficient to resolve time scales greater than $2/3T$. In particular $|\mathcal{J}| = 79$ and $|\mathcal{I}| = 781$.

(ii) Random sampling (large variance): $\mathcal{U}$ is sampled non-uniformly such that $\Delta t_{min} = 3\delta t$ (0.2T) and $\Delta t_{max} = 17\delta t$ (1.15T), which results in $|\mathcal{J}| = 80$ and $|\mathcal{I}| = 783$. Together Δt_{min} and Δt_{max} encompass 3 % of $|\mathcal{J}|$.

(iii) Random sampling (small variance): $\mathcal{U}$ is sampled non-uniformly such that $\Delta t_{min} = 6\delta t$ (0.4T) and $\Delta t_{max} = 14\delta t$ (0.95T), which results in $|\mathcal{J}| = 79$ and $|\mathcal{I}| = 783$. Together Δt_{min} and Δt_{max} encompass 5 % of $|\mathcal{J}|$.

Figure 16 shows the eigenvalues retained using the algorithm 2 with $\tilde{t} = 0.1$ for all three cases. Recall that the limit cycle dynamics of the cylinder flow are characterised by structures oscillating with frequency n/T, where $n \in \mathbb{Z}^+$, and having corresponding time scales of T/n. For the cylinder data used at in the analysis, a uniform time step $\Delta t = 10\delta t$ is high enough to recover T but not sufficient to extract higher-frequency dynamics. Figure 16 shows that when the is-OMD algorithm is applied to uniformly sampled data, high-frequency dynamics are not recovered. When non-uniform sampling strategies are used, although $\Delta t_{av} = 10\delta t$, there is enough phase variation due to changes in Δt to extract further frequencies. Both random strategies recover the dynamics within ranges determined by the Nyquist sampling criterion corresponding to Δt_{min}, which are indicated in figure 16 with dotted lines. In both cases an additional pair of eigenvalues corresponding to frequencies beyond recoverable bounds is identified. Additionally, it can be observed that higher deviation from Δt_{av} positively impacts the accuracy of the recovered dynamics, not only in terms of number of recovered frequencies but also in the estimation of β.

7. Application to the turbulent wake of an axisymmetric bluff body

The ability of the is-OMD algorithm to recover the frequency information from subsampled data was demonstrated in §§5 and 6 for systems with dominant periodicity. For a more challenging test case, we consider the application of is-OMD to PIV snapshots of the turbulent wake of an axisymmetric bluff body. The experimental set-up is detailed in Oxlade *et al.* (2015) and consists of a bullet-shaped model with base diameter $D = 0.1965$ m and a length-to-diameter ratio of $L/D = 6.48$. The inflow velocity is $U_\infty = 15$ m s^{-1}, which gives the diameter Reynolds number of $Re_D = 1.88 \times 10^5$. The wake velocity is measured inside a PIV interrogation window extending 1.72D in the streamwise direction and being 1.5D wide. A schematic representation of the experimental set-up is shown in figure 17. Time-resolved data are collected in a series of independent experiments lasting 3.79 s, each with a frequency of 720 Hz.

It is known Sevilla & Martínez-Bazán (2004), Grandemange *et al.* (2012), Rigas *et al.* (2014) that a large-scale anti-symmetric structure corresponding
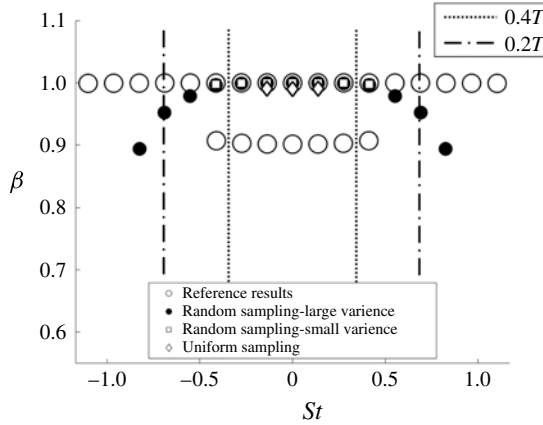
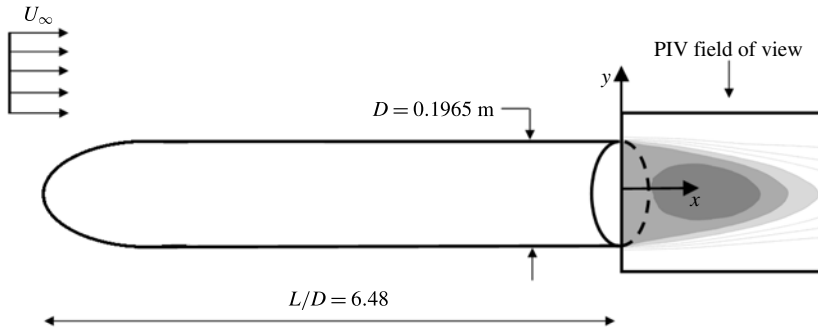FIGURE 16. Eigenvalue lattice for different sampling strategies.

FIGURE 17. Schematic representation of the experimental model and PIV field of view.

to vortex shedding is present in the wake and oscillates at the approximate vortex-shedding frequency $St_{sh} \approx 0.22$. This is confirmed by examining the power spectrum in figure 18(*h*) of the streamwise velocity, averaged over the region $D \leqslant x \leqslant 1.72D$, $-0.375D \leqslant y \leqslant 0.375D$ in which the shedding is expected to be observed. It is possible to identify this frequency and its associated structure by applying the OMD algorithm to the time-resolved data. Indeed, applying OMD to a data ensemble of $N = 1000$ snapshots, sampled at 720 Hz and setting $r = 30$ produces an OMD mode with frequency $(D/2\pi U_\infty)\mathrm{Im}(\lambda) = 0.21$ whose real and imaginary parts are shown in figure 18(*a*,*d*).

It should be noted that, even when applied to the highly resolved data ensemble, the frequency content is somewhat sensitive to the particular implementation of the OMD algorithm. For example, varying the length of the data ensemble $100 \leqslant N \leqslant 1000$ and the model rank $20 \leqslant r \leqslant 200$ provides OMD eigenvalues λ satisfying $0.18 \leqslant (D/2\pi U_\infty)\mathrm{Im}(\lambda) \leqslant 0.26$. Such behaviour is not unexpected due to the relatively broad peak in the spectrum observed in figure 18(*h*), and we note that similar sensitivity was observed when applying the more widely employed DMD algorithm.

We now undersample the data to determine if is-OMD is able to extract the dominant vortex-shedding frequency at $St_{sh} \approx 0.22$. In particular, it is assumed that only every 12th snapshot is available, resulting in a data ensemble sampled at 60 Hz. Following § 6.3 we apply the is-OMD algorithm to an analysis ensemble $\boldsymbol{U}$ formed
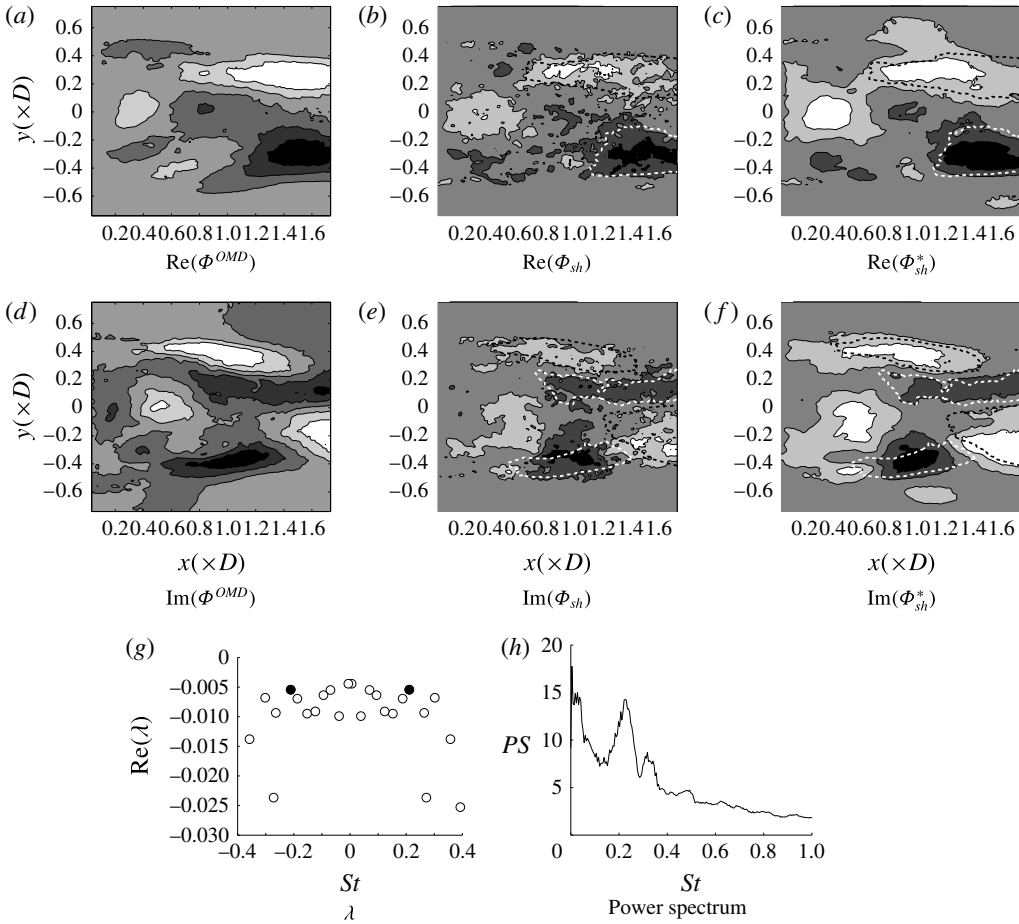
FIGURE 18. The OMD mode $\boldsymbol{\Phi}^{OMD}$ corresponding to $St_D \approx 0.22$ obtained from time-resolved data is shown in (*a*,*d*); the corresponding is-OMD mode $\boldsymbol{\Phi}_{sh}$ in (*b*,*e*); and its projection $\boldsymbol{\Phi}_{sh}^*$ onto first 30 POD modes of $\boldsymbol{U}$ in (*c*,*f*). Eigenvalues of the corresponding dynamic matrix $\boldsymbol{M}^{is\text{-}OMD}$ are shown in (*g*) with the pair of shedding eigenvalues highlighted; (*h*) shows the mean power spectrum of streamwise velocities in a selected wake region.

by taking a random sample of the 60 Hz data. Specifically, we let the time-step snapshots in $\boldsymbol{U}$ be one of the values $\Delta t = j/60$, where $j \in \{1, 2, 3\}$, with a uniform probability of choosing any of the three possible values of Δt. This approach provides an undersampled data ensemble satisfying $\Delta t_{av}/\delta t_{Nyq} \approx 1.12$ – that is, the average time step Δt_{av} is greater than the one specified by the Nyquist criterion corresponding to the St_{sh}.

Frequency identification: the spectrum $\lambda(\boldsymbol{M})$ obtained by is-OMD is shown in figure 18(*g*) and contains an eigenvalue pair corresponding to the shedding frequency at $St = 0.23$. This eigenvalue pair was selected based upon the proportion of data spanned by each of the modes (see appendix B) and, in this case, the shedding eigenvalue pair is ranked as the third most important. Two other dynamically significant eigenvalue pairs correspond to low-frequency structures with $St = 0.006$ and $St = 0.07$, an order of magnitude smaller than St_{Sh}. It is interesting to note that wake structures are known to possess frequency content at approximately these values

(Rigas *et al.* 2014), specifically a random reorientation of the shedding reflectional plane ($St \approx 0.002$) and an axisymmetric pulsation of the vortex cores ($St \approx 0.06$).

Mode extraction: for this example, the contribution of the shedding mode to the total energy of the wake is approximately 75 % lower than that of the cylinder wake studied in §6. Furthermore, although a (relatively wide) spectral peak at $St_{sh} \approx 0.22$ is visible in figure 18(*h*), a projection of the shedding structure onto the underlying data will not produce a purely periodic signal. For this reason, obtaining dynamic modes through an inversion of the Vandermonde matrix, which essentially assumes periodic evolution of the temporal coefficients between known snapshots, is overly restrictive. Instead, it is more appropriate to impose a looser rate constraint upon the unknown mode coefficients c_i appearing in the least-squares problem of the is-OMD algorithm. Specifically, by replacing step 3 of Algorithm 1 by

$$c_i^k = \underset{c_i \in \mathbb{R}^r}{\arg \min} \sum_{i=1}^{N} \|c_i - M c_{i-1}\|^2, \tag{7.1a}$$

$$\text{s.t.} \quad c_i = L^T u_i, \quad \text{if } i \in \mathcal{J}, \tag{7.1b}$$

$$|(c_{i+1})_n - (c_i)_n| \leqslant \gamma \max_{j \in \mathcal{J}} |(c_j)_n|, \quad n = 1, \dots, r, \tag{7.1c}$$

where $(c_i)_n$ denotes the *n*th element of $c_i \in \mathbb{R}^r$ and $\gamma > 1$ is a parameter. In other words, a restriction is placed upon the coefficients c_i to ensure that between each minimal time step, each of its elements does not vary by more than a proportion γ of its largest observed value on the known data. Subsequently, we let $\Phi = LP$, where $M = P^{-1} \Lambda P$.

The real and imaginary parts of the mode Φ_{Sh} associated with $St = 0.23$ are presented in figure 18(*b,e*) and were obtained by setting $\gamma = 3.75$. Although less coherent than the OMD modes obtained from time-resolved data, both $\text{Re}(\Phi_{sh})$ and $\text{Im}(\Phi_{sh})$ possess anti-symmetric and spatially out-of-phase structures present in the modes obtained from time-resolved data. For comparison, the spatial locations of the anti-symmetric structures visible in figure 18(*a,d*) are overlaid and indicate reasonable agreement between the OMD and is-OMD modes. Furthermore, coherency of the mode can be improved through a projection $\Phi_{Sh}^* = U_p U_p^T \Phi_{Sh}$, where $U_p \subset \text{POD}(U)$, suppressing the spatial noise present in the structure. An example of Φ_{Sh}^* obtained through a projection using first 30 POD modes of U is presented in figure 18(*c,f*). We emphasise that the constraint (7.1*c*) is significantly less restrictive than the one imposed in the Vandermonde-based post-processing Algorithm 2. An interesting direction of future research will be to determine the optimal form of constraint required for effective spectral and mode information, dependent upon the intermittency of the underlying data ensemble.

Statistical analysis: To quantify the improvement of is-OMD over OMD applied to downsampled data, a statistical analysis akin to that presented in §5.2 is performed. The experimental data are undersampled using a range of uniform time steps δt to obtain data sets sampled at frequencies in the range 18–240 Hz, which in terms of δt_{Nyq} is equivalent to $0.141 \leqslant \delta t / \delta t_{Nyq} \leqslant 1.875$. Subsequently, randomly sampled data ensembles U are created in an analogous manner to the one used to obtain results in figure 18(*b,g*), i.e. the time step between known snapshots is $\Delta t = j \delta t$, where $j \in \{1, 2, 3\}$, which gives the average time step $\Delta t_{av} \approx 2 \delta t$. The is-OMD algorithm is applied to U and the eigenvalue corresponding to the shedding structure is identified using the methodology described in appendix B. Note that, due to the chosen sampling strategy, U contains 50 % of data used to obtain the OMD approximation. Finally, the error $\epsilon_\lambda = |\lambda_{sh} - \lambda| / |\lambda_{sh}|$ is computed, where λ corresponds to the eigenvalues
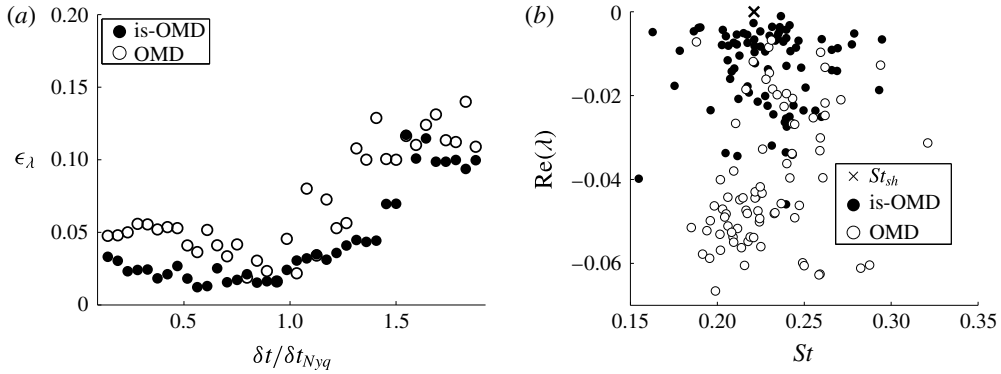
FIGURE 19. (*a*) Proportional eigenvalue error ϵ_λ for a range of $\delta t/\delta t_{Nyq}$; (*b*) the eigenvalues obtained for $\delta t/\delta t_{Nyq} < 1$, where the x-axis represents $St = (D/2\pi U_\infty)\text{Im}(\lambda)$.

obtained using either OMD algorithm applied to the uniformly downsampled data or the is-OMD algorithm applied to $\boldsymbol{U}$. The reference shedding eigenvalue λ_{sh} is defined as $\lambda_{sh} = i(2\pi U_\infty/D)St_{sh}$.

The eigenvalues obtained for $\delta t/\delta t_{Nyq} < 1$ and the values of ϵ_λ for the considered range of $\delta t/\delta t_{Nyq}$ are presented in figure 19. Due to the identified sensitivity of the OMD algorithm to the approximation rank r and of the is-OMD algorithm to the analysis data ensemble, figure 19(*a*) presents an average value of ϵ_λ corresponding to results obtained for $r \in [20, 30, \ldots, 60]$ in the case of the OMD algorithm and, in the case of the is-OMD algorithm, for five different values of $\boldsymbol{U}$ each satisfying $\Delta t_{av} \approx 2\delta t$. Figure 19(*a*) demonstrates that the is-OMD algorithm provides a visible improvement for all values of δt except when $\delta t/\delta t_{Nyq} \approx 1$, where both methods are similarly accurate. Further insight into the differences between both results can be obtained from figure 19(*b*), where the OMD output gives eigenvalues which are concentrated around a lower frequency than $\text{Im}(\lambda_{sh})$ and have more negative growth rate than is-OMD results. It is reiterated that the is-OMD results were obtained using only half of the data used to obtain the OMD approximation with $\Delta t_{av} > \delta t$, implying a beneficial impact of the randomness in the data acquisition on the accuracy of the eigenvalue estimation.

8. Conclusions

Here, a methodology was presented which provides a way to extract dynamical information from data sampled at irregular time steps, below the Nyquist–Shannon criterion. Information concerning temporal dynamics is obtained using the proposed is-OMD algorithm, which can be interpreted as solving an OMD optimisation problem where the majority of snapshots are treated as unknowns. It is assumed that no *a priori* information about the dynamics is available and at each iteration two problems are solved in a pairwise manner; modal decomposition (using DMD or OMD) to update an estimate of the dynamics and reconstruction of intermediate snapshot with a least-squares technique. Further, mode recovery is improved by assuming that the system can be approximated using a Vandermonde matrix and by solving an appropriate minimisation problem. The is-OMD methodology's performance is analysed on synthetic sinusoidal flow and on flow past a circular cylinder at a low Reynolds number, showing improvement over compressive-sampling-based mode
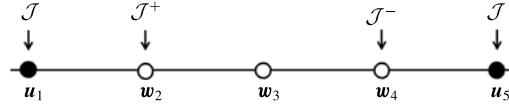
FIGURE 20. A representation of an allocation of indices to different sets in the is-OMD problem.

extraction methodologies. Furthermore, the analysis on the turbulent flow demonstrated a potential of the is-OMD algorithm to be applicable to more complicated dynamical systems.

Acknowledgements

Appendix A. Monotonicity of (S_k)

To facilitate the analysis let $\mathcal{I} = \{0, 1, \ldots, N\}$ denote all time steps of interest, recall that $\mathcal{J} \subset \mathcal{I}$ denotes points of known data and introduce two new subsets: $\mathcal{J}^-$ contains time step indices immediately preceding those in $\mathcal{J}$; and $\mathcal{J}^+$ which contains indices representing time steps immediately after those in $\mathcal{J}$. This assignment is indicated in figure 20. It is assumed for simplicity that $\mathcal{J}^- \cap \mathcal{J}^+ = \emptyset$ and we note that if $|\mathcal{J}| = m$ then $|\mathcal{J}^- \cup \mathcal{J}^+| = 2(m-1)$. Recalling that $\boldsymbol{w}_i^k = \boldsymbol{u}_i, i \in \mathcal{J}$ and letting $\mathcal{J}^0 := \mathcal{I} \setminus (\mathcal{J} \cup \mathcal{J}^+ \cup \mathcal{J}^-)$, the residual S_k can be written

$$
\begin{aligned}
S_k &= \sum_{i \in \mathcal{J}^0} \left\| \boldsymbol{w}_i^k - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{w}_{i-1}^k \right\|^2 + \sum_{i \in \mathcal{J}^+} \left\| \boldsymbol{w}_i^k - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{u}_{i-1} \right\|^2 \\
&\quad + \sum_{i \in \mathcal{J}^-} \left\| \boldsymbol{u}_{i+1} - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{w}_i^k \right\|^2 .
\end{aligned}
\tag{A 1}
$$

THEOREM 1. *Let S_k be the residual (2.15) from algorithm 1. Then $S_k \geqslant S_{k+1} \geqslant 0$ and there exists $S^* \geqslant 0$ such that $S_k \to S^*, k \to \infty$.*

Proof. For each $k \geqslant 0$, $\boldsymbol{L}_k \boldsymbol{L}_k^{\mathrm{T}} : \mathbb{R}^p \to \mathbb{R}^p$ is a projection. Estimating the first two terms of (A 1) from below by their projections onto $\mathrm{Im}(\boldsymbol{L}_k)$ and decomposing the final term into contributions from the image and kernel of $\boldsymbol{L}_k$ give

$$
\begin{aligned}
S_k &\geqslant \sum_{i \in \mathcal{J}^0} \| \boldsymbol{L}_k \boldsymbol{L}_k^{\mathrm{T}} \left(\boldsymbol{w}_i^k - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{w}_{i-1}^k \right) \|^2 + \sum_{i \in \mathcal{J}^+} \| \boldsymbol{L}_k \boldsymbol{L}_k^{\mathrm{T}} \left(\boldsymbol{w}_i^k - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{u}_{i-1} \right) \|^2 \\
&\quad + \sum_{i \in \mathcal{J}^-} \left[\| \boldsymbol{L}_k \boldsymbol{L}_k^{\mathrm{T}} \left(\boldsymbol{u}_{i+1} - \boldsymbol{L}_k \boldsymbol{M}_k \boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{w}_i^k \right) \|^2 + \| \left(\boldsymbol{I} - \boldsymbol{L}_k \boldsymbol{L}_k^{\mathrm{T}} \right) \boldsymbol{u}_{i+1} \|^2 \right] .
\end{aligned}
\tag{A 2}
$$

Now, since $\boldsymbol{L}_k^{\mathrm{T}} \boldsymbol{L}_k = \boldsymbol{I}$, it follows that $\| \boldsymbol{L}_k \boldsymbol{v} \| = \| \boldsymbol{v} \|$, for any $\boldsymbol{v} \in \mathbb{R}^r$. Hence,

$$S_k \geqslant \sum_{i\in\mathcal{J}^0} \|\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^k - \mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_{i-1}^k\|^2 + \sum_{i\in\mathcal{J}^+} \|\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^k - \mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{u}_{i-1}\|^2$$
$$+ \sum_{i\in\mathcal{J}^-} \left[\|\mathbf{L}_k^{\mathrm{T}}\mathbf{u}_{i+1} - \mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^k\|^2 + \| \left(\mathbf{I} - \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\right) \mathbf{u}_{i+1}\|^2 \right]. \tag{A 3}$$

Recalling that $\mathbf{c}_i^k = \mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^k$ and writing $\mathbf{d}_i = \mathbf{L}_k^{\mathrm{T}}\mathbf{u}_i$,

$$S_k \geqslant \sum_{i\in\mathcal{J}^0} \|\mathbf{c}_i^k - \mathbf{M}_k\mathbf{c}_{i-1}^k\|^2 + \sum_{i\in\mathcal{J}^+} \|\mathbf{c}_i^k - \mathbf{M}_k\mathbf{d}_{i-1}\|^2$$
$$+ \sum_{i\in\mathcal{J}^-} \left[\|\mathbf{d}_{i+1} - \mathbf{M}_k\mathbf{c}_i^k\|^2 + \| \left(\mathbf{I} - \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\right) \mathbf{u}_{i+1}\|^2 \right]. \tag{A 4}$$

Now, the first three terms of (A 4) are exactly of the form of the optimisation problem (2.13) of algorithm 1. Hence, by optimality,

$$S_k \geqslant \sum_{i\in\mathcal{J}^0} \|\mathbf{c}_i^{k+1} - \mathbf{M}_k\mathbf{c}_{i-1}^{k+1}\|^2 + \sum_{i\in\mathcal{J}^+} \|\mathbf{c}_i^{k+1} - \mathbf{M}_k\mathbf{d}_{i-1}\|^2$$
$$+ \sum_{i\in\mathcal{J}^-} \left[\|\mathbf{d}_{i+1} - \mathbf{M}_k\mathbf{c}_i^{k+1}\|^2 + \| \left(\mathbf{I} - \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\right) \mathbf{u}_{i+1}\|^2 \right]. \tag{A 5}$$

Note that values of $\mathbf{d}_i$ do not change between subsequent iterations due to constraints imposed in the optimisation problem (2.13).

The subsequent step 4 in the algorithm consists of approximating time-resolved snapshots $\mathbf{w}_i$ with the 'backwards' embedding using the columns of $\mathbf{L}_k$. Again, using the identity $\|\mathbf{L}_k\mathbf{v}\|^2 = \|\mathbf{v}\|^2$, for $\mathbf{v} \in \mathbb{R}^r$,

$$S_k \geqslant \sum_{i\in\mathcal{J}^0} \|\mathbf{L}_k\mathbf{c}_i^{k+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{c}_{i-1}^{k+1}\|^2 + \sum_{i\in\mathcal{J}^+} \|\mathbf{L}_k\mathbf{c}_i^{k+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{d}_{i-1}\|^2$$
$$+ \sum_{i\in\mathcal{J}^-} \left[\|\mathbf{L}_k\mathbf{d}_{i+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{c}_i^{k+1}\|^2 + \| \left(\mathbf{I} - \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\right) \mathbf{u}_{i+1}\|^2 \right]. \tag{A 6}$$

In step 4 of Algorithm 1, $\mathbf{w}_i^{k+1} = \mathbf{L}_k\mathbf{c}_i^{k+1}$, and, hence, it is true that $\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^{k+1} = \mathbf{c}_i^{k+1}$, which implies that the above expression is equivalent to

$$S_k \geqslant \sum_{i\in\mathcal{J}^0} \|\mathbf{w}_i^{k+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_{i-1}^{k+1}\|^2 + \sum_{i\in\mathcal{J}^+} \|\mathbf{w}_i^{k+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{u}_{i-1}\|^2$$
$$+ \sum_{i\in\mathcal{J}^-} \left[\|\mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{u}_{i+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^{k+1}\|^2 + \| \left(\mathbf{I} - \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\right) \mathbf{u}_{i+1}\|^2 \right]. \tag{A 7}$$

Again, using the fact that $\|\mathbf{v}\|^2 = \|\mathbf{L}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{v}\|^2 + \|(\mathbf{I} - \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}})\mathbf{v}\|^2$ for any $\mathbf{v} \in \mathbb{R}^p$, the final two terms in the above inequality can be combined to show that

$$S_k \geqslant \sum_{i\in\mathcal{J}^0} \|\mathbf{w}_i^{k+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_{i-1}^{k+1}\|^2 + \sum_{i\in\mathcal{J}^+} \|\mathbf{w}_i^{k+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{u}_{i-1}\|^2$$
$$+ \sum_{i\in\mathcal{J}^-} \|\mathbf{u}_{i+1} - \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}}\mathbf{w}_i^{k+1}\|^2. \tag{A 8}$$

Finally, invoking optimality of the [L, M]-update optimisation subproblem (2.14) gives

$$
\begin{aligned}
S_k &\geqslant \sum_{i \in \mathcal{J}^0} \| w_i^{k+1} - L_{k+1} M_{k+1} L_{k+1}^{\mathrm{T}} w_{i-1}^{k+1} \|^2 + \sum_{i \in \mathcal{J}^+} \| w_i^{k+1} - L_{k+1} M_{k+1} L_{k+1}^{\mathrm{T}} u_{i-1} \|^2 \\
&\quad + \sum_{i \in \mathcal{J}^-} \| u_{i+1} - L_{k+1} M_{k+1} L_{k+1}^{\mathrm{T}} w_i^{k+1} \|^2 \\
&= S_{k+1}.
\end{aligned}
\tag{A 9}
$$

Hence, $(S_k)_{k \geqslant 0}$ is a monotone decreasing sequence which is bounded below by 0. Therefore, there exists $S^* \geqslant 0$ such that $S_k \to S^*$, $k \to \infty$. □

Appendix B. Scaling of the modes and eigenvalues

To provide a ranking of the OMD modes, recall that OMD seeks matrices $L \in \mathbb{R}^{p \times r}$, $M \in \mathbb{R}^{r \times r}$ which minimise the sum of squares of the residuals $r_i = w_{i+1} - LML^{\mathrm{T}} w_i$, given a data ensemble [$w_1 \cdots w_N$]. This provides an approximation

$$
\begin{aligned}
\begin{bmatrix} w_2 & \cdots & w_N \end{bmatrix} &\approx LML^{\mathrm{T}} \begin{bmatrix} w_1 & \cdots & w_{N-1} \end{bmatrix} \\
&= LP \cdot \Lambda P^{-1} L^{\mathrm{T}} \begin{bmatrix} w_1 & \cdots & w_{N-1} \end{bmatrix} \\
&=: \Phi \mathcal{A},
\end{aligned}
\tag{B 1}
$$

where the columns of $\Phi = LP$ are the OMD modes, $M = P \Lambda P^{-1}$ is an eigendecomposition of M and $\mathcal{A} = (\alpha_{ij}) := \Lambda P^{-1} L^{\mathrm{T}} [w_1 \cdots w_{N-1}] \in \mathbb{R}^{r \times N}$. Noting that the coefficients multiplying ith OMD mode correspond to the ith row of $\mathcal{A}$, we define the modal ranking in terms of the constants

$$
\sum_{j=1}^{N} |\alpha_{ij}|^2.
\tag{B 2}
$$

Such a scaling incorporates the energetic contribution of each mode to the underlying data ensemble in addition to the magnitude of each mode's corresponding eigenvalue, implying, for example, that rapidly decaying modes will be penalised.

REFERENCES

ADRIAN, R. J. & WESTERWEEL, J. 2011 *Particle Image Velocimetry*. Cambridge University Press.

BAGHERI, S. 2013 Koopman-mode decomposition of the cylinder wake. *J. Fluid Mech.* **726**, 596–623.

CANDES, E. J. & TAO, T. 2006*a* Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory* **52**, 5406–5425.

CANDES, E. J. & WAKIN, M. B. 2008 An introduction to compressive sampling. *IEEE Signal Process. Magazine* **25** (2), 21–30.

CANDES, R. J., J., E. & TAO, T. 2006*b* Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52** (2), 489–509.

CHEN, K. K., TU, J. H. & ROWLEY, C. W. 2012 Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses. *J. Nonlinear Sci.* **22** (6), 887–915.

CHOI, W. P., JEON, H. & KIM, J. 2008 Control of flow over a bluff body. *Annu. Rev. Fluid Mech.* **40** (1), 113–139.

DONOHO, D. L. 2006 Compressed sensing. *IEEE Trans. Inf. Theory* **52** (4), 1289–1306.

ELSINGA, G. E., SCARANO, F., WIENEKE, B. & VAN OUDHEUSDEN, B. W. 2006 Tomographic particle image velocimetry. *Exp. Fluids* **41**, 933–947.

GRANDEMANGE, M., GOHLKE, M., PAREZANOVIĆ, V. & CADOT, O. 2012 On experimental sensitivity analysis of the turbulent wake from an axisymmetric blunt trailing edge. *Phys. Fluids* **24**, 035106.

GRILLI, M., SCHMID, P. J., HICKEL, S. & ADAMS, N. A. 2012 Analysis of unsteady behaviour in shockwave turbulent boundary layer interaction. *J. Fluid Mech.* **700**, 16–28.

HEMATI, M. S., ROWLEY, C. W. & NICHOLS, J. W. 2017 De-biasing the dynamic mode decomposition for applied Koopman spectral analysis. *Theor. Comput. Fluid Dyn.* doi:10.1007/s00162-017-0432-2.

HOLMES, P., LUMLEY, J. L. & BERKOOZ, G. 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press.

JOVANOVIC, M. R., SCHMID, P. J. & NICHOLS, J. W. 2014 Sparsity-promoting dynamic mode decomposition. *Phys. Fluids* **26**, 024103.

LEROUX, R. & CORDIER, L. 2016 Dynamic mode decomposition for non-uniformly sampled data. *Exp. Fluids* **57** (5), 94.

LU, L. & PAPADAKIS, G. 2011 Investigation of the effect of external periodic flow pulsation on a cylinder wake using linear stability analysis. *Phys. Fluids* **23**, 094105.

LU, L. & PAPADAKIS, G. 2014 An iterative method for the computation of the response of linearised Navier–Stokes equations to harmonic forcing and application to forced cylinder wakes. *Intl J. Numer. Meth. Fluids* **74**, 794–817.

MULD, T. W., EFRAIMSSON, G. & HENNINGSON, D. S. 2012 Flow structures around a high-speed train extracted using proper orthogonal decomposition and dynamic mode decomposition. *Comput. Fluids* **57**, 87–97.

NEEDELL, D. & TROPP, J. A. 2009 Cosamp: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **26**, 301–321.

NOACK, B. R., AFANASIEV, K., MORZYNSKI, M., TADMOR, G. & THIELE, F. 2003 A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.* **497**, 335–363.

NYQUIST, H. 1928 Certain topics in telegraph transmission theory. *Proc. IEEE* **90** (2), 280–305.

ORLANDI, P. 1994 On the generation of turbulent wall friction. *Phys. Fluids* **6**, 634–641.

OXLADE, A. R., MORRISON, J. F., QUBAIN, A. & RIGAS, G. 2015 High-frequency forcing of a turbulent axisymmetric wake. *J. Fluid Mech.* **770**, 305–318.

PROCTOR, J. L., BRUNTON, S. L. & KUTZ, J. N. 2014 Dynamic mode decomposition with control.

RIGAS, G., OXLADE, A. R., MORGANS, A. S. & MORRISON, J. F. 2014 Low-dimensional dynamics of a turbulent axisymmetric wake. *J. Fluid Mech.* **755**, R5.

ROWLEY, C. W., MEZIC, I., BAGHERI, S., SCHLATTER, P. & HENNINGSON, D. S. 2009 Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127.

SAYADI, T., SCHMID, P. J., NICHOLS, J. W. & MOIN, P. 2014 Reduced-order representation of near-wall structures in the late transitional boundary layer. *J. Fluid Mech.* **748**, 278–301.

SCHMID, P. J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.

SCHMID, P. J., VIOLATO, D. & SCARANO, F. 2012 Decomposition of time-resolved tomographic piv. *Exp. Fluids* **52**, 1567–1579.

SEENA, A. & SUNG, H. J. 2011 Dynamic mode decomposition of turbulent cavity flows for self-sustained oscillations. *Intl J. Heat Fluid Flow* **32**, 1098–1110.

SEMERARO, O., BELLANI, G. & LUNDELL, F. 2012 Analysis of time-resolved piv measurements of a confined turbulent jet using POD and Koopman modes. *Exp. Fluids* **53**, 1203–1220.

SEVILLA, A. & MARTÍNEZ-BAZÁN, C. 2004 Vortex shedding in high reynolds number axisymmetric bluff-body wakes: local linear instability and global bleed control. *Phys. Fluids* **16**, 3460–3469.

TISSOT, G., CORDIER, L., BENARD, N. & NOACK, B. R. 2014 Model reduction using dynamic mode decomposition. *C. R. Méc* **342** (67), 410–416.

TROPP, J. A. & GILBERT, A. C. 2007 Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **53**, 4655–4666.

TU, J. H., ROWLEY, C. W., KUTZ, J. N. & SHANG, J. K. 2014 Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. *Exp. Fluids* **55**, 1805.

VINHA, N., MESEGUER-GARRIDO, F., DE VICENTE, J. & VALERO, E. 2016 A dynamic mode decomposition of the saturation process in the open cavity flow. *Aerosp. Sci. Technol.* **52**, 198–206.

WILLIAMS, G. 2011 *Linear Algebra with Applications*. Jones and Bartlett Publishers.

WILLIAMS, M. O., KEVREKIDIS, I. G. & ROWLEY, C. W. 2015 A datadriven approximation of the Koopman operator: extending dynamic mode decomposition. *J. Nonlinear Sci.* **25**, 1307–1346.

WYNN, A., PEARSON, D. S., GANAPATHISUBRAMANI, B. & GOULART, P. J. 2013 Optimal mode decomposition for unsteady flows. *J. Fluid Mech.* **733**, 473–503.