A dynamic mode decomposition technique for the analysis of non–uniformly sampled flow data

Binghua Li [a,c], Jesús Garicano-Mena [a,b,*], Eusebio Valero [a,b]

[a] *ETSI Aeronáutica y del Espacio, Universidad Politécnica de Madrid, Madrid, Spain*
[b] *Center for Computational Simulation (CCS), Boadilla del Monte, Spain*
[c] *School of Aeronautics and Astronautics, Zhejiang University, China*

ARTICLE INFO

ABSTRACT

A novel Dynamic Mode Decomposition (**DMD**) technique capable of handling non–uniformly sampled data is proposed. As it is usual in **DMD** analysis, a linear relationship between consecutive snapshots is made. The performance of the new method, which we term θ-**DMD**, is assessed on three different, increasingly complex datasets: a *synthetic* flow field, a $Re_D = 60$ flow around a cylinder cross section, and a $Re_\tau = 200$ turbulent channel flow. For the three datasets considered, whenever the dataset is uniformly sampled, the θ-**DMD** method provides comparable results to the original **DMD** method. Additionally, the θ-**DMD** is still capable of recovering relevant flow features from non–uniformly sampled databases, whereas **DMD** cannot. The proposed tool opens the way to conduct **DMD** analyses for non–uniformly sampled data, and can be useful *e.g.*, when confronted with experimental datasets with *missing data*, or when facing numerical datasets generated using adaptive time-integration schemes.

1. Introduction

Over the last twenty years, and in whatever the area of technological interest, there has been a massive increase in the quality and volume of data available. The field of fluid dynamics is not an exception to this trend: both the accurate numerical computations performed on high–performance computing facilities (*e.g.* Direct Numerical Simulation) and experimental campaigns leveraging multiple sensors with fast sampling data-acquisition systems make the collection of large datasets a trivial task.

The surge in data availability is certainly welcome, though it also comes hand in hand with its own associated challenges, *e.g.* the data processing becomes more computing intensive, and discerning the traces of the physical phenomena of interest from irrelevant features can be difficult.

The fluid dynamics community has not been idle in the face of these challenges. On the contrary, a plethora of methods capable of obtaining *information* from flow data (*e.g. data-driven*) has been developed, thoroughly tested and assessed over the last years.

The most prevalent data–driven flow analysis strategy is perhaps the Proper Orthogonal Decomposition (**POD**, [1–4]) and derivations. This family of methods –sometimes also referred to as Principal Component Analysis or Karhunen-Loève decomposition– operate on sequences of *snapshots*, that is, either experimental measurements or numerical solutions ac-

quired at successive time instants. An optimal representation of the sequence is provided by the **POD**, as features identified by **POD** are orthogonal to each other [4]. Recently, techniques to perform Spectral **POD** have been described in References [5–7]. Randomized [8] and multi–resolution [9] implementations of **POD** are also available.

An alternative, more recent family of strategies builds on the Dynamic Mode Decomposition (**DMD**) method [10,11]. **DMD** techniques, when applied on flow snapshot sequences provide structures that are not orthogonal to each other, but which oscillate harmonically at specific frequencies. **DMD** can be regarded from a number of complementary perspectives [12]: *e.g.*, from that of the Koopman analysis [10,13], or from that of global stability analysis [11,14]; **DMD** can also be connected to Discrete Fourier Transform analysis [15] and with Spectral **POD**, as discussed in [6]; and with Resolvent Analysis [16, 17]. Randomized variants of **DMD** are presented in [8,18,19]. **DMD** techniques have been applied to a wide array of flow problems, both in numerical [10,20,21] and experimental [22–27] settings. Though early **DMD** applications focused mostly on transitional flows; recent investigations addressing the **DMD** analysis of turbulent flows are reported in [28–31].

From the previous discussion, both **POD** and **DMD** methods can be considered firmly established data analysis tools, as the number of available monographs [1,4,12,32,33] and reviews [14,34–37] confirm.

One common trait shared by nearly of the methods so far considered is the prerequisite of a uniformly sampled temporal sequence. However, one might be confronted with data sequences that are not uniformly sampled, (*i.e.*, datasets where subsequent snapshots are not uniformly equiseparated in time). Such situations arise, *e.g.*, whenever a numerical dataset has been generated using adaptive time-integration schemes, or when faced with *gappy* experimental measurements.

Such non-uniformly sampled datasets have been addressed already by Guéniat and collaborators in [38], through the Non–Uniform **DMD** (or **NU**-**DMD**) method. The **NU**-**DMD** method accomplishes the **DMD** factorization of not temporally equiseparated data sequences by posing an optimization problem that resorts to the Nelder–Mead minimization algorithm. More recently, and on a related note, Le Clainche and her collaborators assess the sensitivity to sampling time of the **HODMD** method [31].

In this contribution we propose a derived **DMD** method to address non–uniformly sampled datasets. This method, a derivation from [39], has two differences with most of the established **DMD** methods: first, the linear relationship between consecutive snapshots is made on the numerically discretized counterpart of a linearized state-equation; and second, a dimensionally–reduced view of the state data is considered. The performance of the new method, which we call θ-**DMD** for reasons to be discussed in section 2, is assessed on three different, increasingly complex datasets: a *synthetic* flow field, a $Re_D = 60$ flow around a cylinder cross section, and a $Re_\tau = 200$ turbulent channel flow. For the three datasets considered, we demonstrate that whenever the dataset is uniformly sampled, the θ-**DMD** method provides comparable results to the original **DMD** method. Additionally, the θ-**DMD** is still capable of recovering relevant flow features from non–uniformly sampled databases, whereas **DMD** cannot.

The rest of this paper is organized as follows: Section 2 introduces the θ-**DMD** method. Next, Section 3 introduces the testcases considered, describes the results obtained and assesses thoroughly the performance of the θ-**DMD** method in comparison with the **DMD** method when confronted with non–uniformly sampled datasets. Finally, Section 4 presents the conclusions of this contribution.

2. Methodology

2.1. The **DMD** technique

In this work we consider Dynamic Mode Decomposition (**DMD**) methods as data-driven feature extraction techniques. These techniques analyse the state evolution of a dynamical system in time, under the assumption of a linear relationship between consecutive system states (the *snapshots*). Consider the general non-linear dynamical system in discrete time, given by:

$$\mathbf{v}_{j+1} = \mathcal{G}\left(\mathbf{v}_j\right) \text{ with } j = 0, 1, 2, \ldots \tag{1}$$

and its linearized counterpart around the state $\mathbf{v}_0$:

$$\mathbf{u}_{j+1} = \mathcal{A}^d \mathbf{u}_j \text{ with } j = 0, 1, 2, \ldots, \tag{2}$$

where $\mathcal{A}^d$ can be interpreted as a state transition matrix [16].

Classical **DMD** algorithms (among others, [11,35,40,41]) attempt to approximate the previous relationship without constructing explicitly the linear operator $\mathcal{A}^d$. Consider a sequence of n_p-dimensional instantaneous states (*e.g.*, flow fields) indexed from 1 to n_s, which can be arranged as a *data matrix*:

$$\mathcal{U}_1^{n_s} = \left[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_s-1}, \mathbf{u}_{n_s}\right] \in \mathbb{R}^{n_p \times n_s}. \tag{3}$$

The subindex and superindex identify, respectively, the first and last time instants of the sequence. The data is ordered in time, and separated by a *constant* sampling time interval Δt^s such that: $t_{j+1} = t_j + \Delta t^s$ for all $j = 1, \ldots, n_s - 1$.

The **DMD** techniques considered in this work follow [33], and thus depart from the factorization of the input data matrix $\mathcal{U}_1^{n_s}$ using the (economy–sized) Singular Value Decomposition (**SVD**) technique:

$$\mathcal{U}_1^{\mathbf{n_s}} \overset{SVD}{=} \mathbf{L}_0 \, \mathbf{S}_0 \, \mathbf{R}_0^T = \sum_{j=1}^{\min(n_p, n_s)} \sigma_j \, \mathbf{l}_{0,j} \cdot \mathbf{r}_{0,j}^T. \tag{4}$$

Matrix $\mathbf{S_0}$ contains as diagonal entries the non–negative and decreasing singular values $\sigma_{0,j}$, whereas the real matrices $\mathbf{L}_0$ and $\mathbf{R}_0$, which are orthogonal, have as their columns $\mathbf{l}_{0,j}$ and $\mathbf{r}_{0,j}$ the left and right singular vectors.

Since $\mathbf{l}_{0,j} \in \mathbb{R}^{n_p}$, matrix $\mathbf{L}_0$ is related to the *spatial dimension* of the dataset and thus $\mathbf{L}_0$ is sometimes referred as the *topos* matrix. In the same manner, $\mathbf{r}_{0,j} \in \mathbb{R}^{n_s}$, and thus related to the *temporal dimension*; matrix $\mathbf{R}_0$ is called accordingly the *chronos* matrix. Typical flow applications, be they numerical or experimental, lead to data matrices with $n_s \ll n_p$, termed Tall & Skinny (**TS**) matrices.

Thanks to the optimality properties of **SVD**, it is possible to obtain rank r_0 reduced approximations to the data set simply by choosing $r_0 \leq \min(n_p, n_s)$.

In this work we consider the **DMD** algorithm in [11], but applied to the reduced scaled chronos matrix $\mathcal{C}_1^{n_s} \equiv \widehat{\mathbf{S}}_0 \widehat{\mathbf{R}}_0^T = \sum_{j=1}^{r_0} \sigma_{0,j} \mathbf{e}_j \cdot \mathbf{r}_{0,j}^T$, where $\mathbf{e}_j$ is the j-th unit vector in $\mathbb{R}^{r_0}$. The partial subsequences $\mathbf{X} \equiv \mathcal{C}_1^{n_s-1}$ and $\mathbf{Y} \equiv \mathcal{C}_2^{n_s}$ are built, and the following linear relationship is presumed:

$$\mathbf{Y} = \mathcal{A}^d \, \mathbf{X}. \tag{5}$$

Next, the economy-sized Singular Value Decomposition (**SVD**) of the first subsequence is performed:

$$\mathbf{X} \overset{SVD}{=} \mathbf{L_1} \, \mathbf{S_1} \, \mathbf{R_1}^T. \tag{6}$$

The same remarks on non-negativity, orthogonality and optimality of reduced representations ($\mathbf{X} = \sum_{j=1}^{r_1} \sigma_{1,j} \mathbf{e}_j \cdot \mathbf{r}_{1,j}^T$) apply as before.

Using the **SVD** of the **X** matrix into Eq. (5) allows to build a reduced matrix $\widetilde{\mathbf{A}}$, defined as:

$$\widetilde{\mathbf{A}} \equiv \mathbf{L}_1^T \, \mathbf{A}^d \, \mathbf{L}_1 = \mathbf{L}_1^T \, \mathbf{Y} \, \mathbf{R}_1 \, \mathbf{\Sigma}_1^{-1}. \tag{7}$$

The reduced matrix $\widetilde{\mathbf{A}}$ is the projection of the matrix $\mathbf{A}^d$ onto the space linearly generated by the columns of $\mathbf{L}_1$ [11]. The **DMD** method operates under the assumption that the projected matrix $\widetilde{\mathbf{A}}$ conveys most of the information codified into operator $\mathbf{A}^d$.

Once the reduced matrix $\widetilde{\mathbf{A}}$ has been calculated, its (right) eigenvalue decomposition:

$$\widetilde{\mathbf{A}} \, \mathbf{\Psi} = \mathbf{\Psi} \, \mathbf{\Lambda}_\mu, \tag{8}$$

offers the reduced **DMD** modes $\boldsymbol{\psi}_i$ as the columns of $\mathbf{\Psi}$; the corresponding eigenvalues μ_i (the diagonal entries of $\mathbf{\Lambda}_\mu$) indicate the temporal growth rates ($\mathrm{Re}(\mu_i)$) and angular pulsation ($\mathrm{Im}(\mu_i)$).

The projected eigenmodes (namely, the dynamic modes) of matrix $\mathbf{A}^d$ are recovered as $\mathbf{\Phi} = \mathbf{L}_0 \mathbf{L}_1 \mathbf{\Psi}$. Note also that the growth rates and frequencies in the complex half-plane can be recovered from the eigenvalues as:

$$\lambda_i = \log(\mu_i)/\Delta t^s. \tag{9}$$

Finally, note that the **DMD** decomposition allows to reconstruct the original data sequence as:

$$\mathbf{v}(t_j) = \sum_{i=1}^{n_s-1} \alpha_i \, \boldsymbol{\phi}_i \mu_i^j. \tag{10}$$

The expression above can be recast in matrix form as:

$$\mathcal{U}_1^{n_s-1} = \mathbf{\Phi} \, \mathbf{D}_\alpha \, \mathbf{V}_\mu, \tag{11}$$

where $\mathbf{V}_\mu$ is a Vandermonde matrix whose columns are generated by the successive powers of the column vector $\left[\mu_1^j, \ldots, \mu_{n_s-1}^j\right]^T$, with $j = 0, \ldots, n_s - 1$; and $\mathbf{D}_\alpha$ is a diagonal matrix whose non-zero entries are to be determined. In this contribution, the amplitudes α_i are computed, following [40], from the minimization problem in the Frobenius norm:

$$\min_{\alpha_i} \|\mathcal{U}_1^{n_s-1} - \mathbf{\Phi} \, \mathbf{D}_\alpha \, \mathbf{V}_\mu\|_F^2. \tag{12}$$

2.2. The θ-DMD technique

Alternatively, one might consider the generic non-linear dynamical system in continuous time:

$$\dot{\mathbf{v}} = \mathcal{F}(\mathbf{v}). \tag{13}$$

Linearizing the expression above around the state $\mathbf{v}_0$ leads to the following system of ordinary differential equations:

$$\dot{\mathbf{u}} = \mathcal{A}^c \mathbf{u}, \tag{14}$$

where $\mathcal{A}^c$ is the Jacobian $\frac{\partial \mathcal{F}}{\partial \mathbf{v}}(\mathbf{v_0})$.

Note that the exact solution to Eq. (14) is [42,16]:

$$\mathbf{u}(t) = e^{\mathcal{A}^c t} \mathbf{u}(0), \tag{15}$$

which for two successive snapshots $\mathbf{u}_j$, $\mathbf{u}_{j+1}$ reads:

$$\mathbf{u}_{j+1} = \underbrace{e^{\mathcal{A}^c \Delta t^s}}_{\mathcal{A}^d} \mathbf{u}_j. \tag{16}$$

As for the relationship between the eigenvectors of matrix $\mathcal{A}^c$ and those of its exponential, a linear algebra theorem [43] guarantees that both coincide if $\mathcal{A}^c$ is diagonalizable.

We consider now the finite-difference discretization of Equation (14). The derivative term is discretized using forward Euler, whereas a θ-method is applied to the right-hand side. These two choices lead to the following relationship:

$$\frac{\mathbf{u}_{j+1} - \mathbf{u}_j}{\Delta t_j^s} = \mathcal{A}^c \left[(1 - \theta) \mathbf{u}_j + \theta \mathbf{u}_{j+1} \right] \text{ with } j = 1, \ldots, n_s - 1. \tag{17}$$

Note how we are not assuming a constant sampling time in this case, and hence $\Delta t_j^s \equiv t_{j+1} - t_j$ for all $j = 1, \ldots, n_s - 1$.

The relation given by Eq. (17) applies also to the *reduced* snapshots, *i.e.* the columns of $\mathcal{C}_1^{n_s}$:

$$\frac{\mathbf{c}_{j+1} - \mathbf{c}_j}{\Delta t_j^s} = \mathcal{A}^c \left[(1 - \theta) \mathbf{c}_j + \theta \mathbf{c}_{j+1} \right] \text{ with } j = 1, \ldots, n_s - 1. \tag{18}$$

Equation (18) can be arranged as:

$$\underbrace{\left(\mathbf{I} - \theta \Delta t_j^s \mathcal{A}^c \right)}_{\mathbf{P}} \mathbf{c}_{j+1} = \underbrace{\left(\mathbf{I} + (1 - \theta) \Delta t_j^s \mathcal{A}^c \right)}_{\mathbf{Q}} \mathbf{c}_j, \tag{19}$$

or, in matrix form $\mathbf{P}\mathcal{C}_2^{n_s} = \mathbf{Q}\mathcal{C}_1^{n_s - 1}$.

The matrix inversion required by Eq. (19) above can be avoided if we consider the alternative formulation:

$$\mathcal{C}_1^{n_s} \mathbf{M}_\delta = \mathcal{A}^c \mathcal{C}_1^{n_s} \mathbf{M}_\theta, \tag{20}$$

where matrices $\mathbf{M}_\delta, \mathbf{M}_\theta \in \mathbb{R}^{n_s \times (n_s - 1)}$ are defined as:

$$\mathbf{M}_\delta = \begin{bmatrix} \frac{-1}{\Delta t_1} & 0 & \cdots & \cdots & 0 & 0 \\ \frac{1}{\Delta t_1} & \frac{-1}{\Delta t_2} & \cdots & \cdots & 0 & 0 \\ 0 & \frac{1}{\Delta t_2} & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & \frac{-1}{\Delta t_{n_s-1}} & 0 \\ 0 & 0 & \cdots & \cdots & \frac{1}{\Delta t_{n_s-1}} & \frac{-1}{\Delta t_{n_s}} \\ 0 & 0 & \cdots & \cdots & 0 & \frac{1}{\Delta t_{n_s}} \end{bmatrix} \text{ and } \mathbf{M}_\theta = \begin{bmatrix} \theta & 0 & \cdots & \cdots & 0 & 0 \\ 1 - \theta & \theta & \cdots & \cdots & 0 & 0 \\ 0 & 1 - \theta & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & \theta & 0 \\ 0 & 0 & \cdots & \cdots & 1 - \theta & \theta \\ 0 & 0 & \cdots & \cdots & 0 & 1 - \theta \end{bmatrix}.$$

Defining $\mathbf{X} \equiv \mathcal{C}_1^{n_s} \mathbf{M}_\theta$ and $\mathbf{Y} \equiv \mathcal{C}_1^{n_s} \mathbf{M}_\delta$ allows to rewrite Eq. (20) simply as:

$$\mathbf{Y} = \mathcal{A}^c \mathbf{X}. \tag{21}$$

From here on, application of the **DMD** technique in section 2.1 allows to retrieve dynamically relevant information about the system. We term this novel approach, capable of handling non–uniformly sampled state sequences, the θ-**DMD** technique.
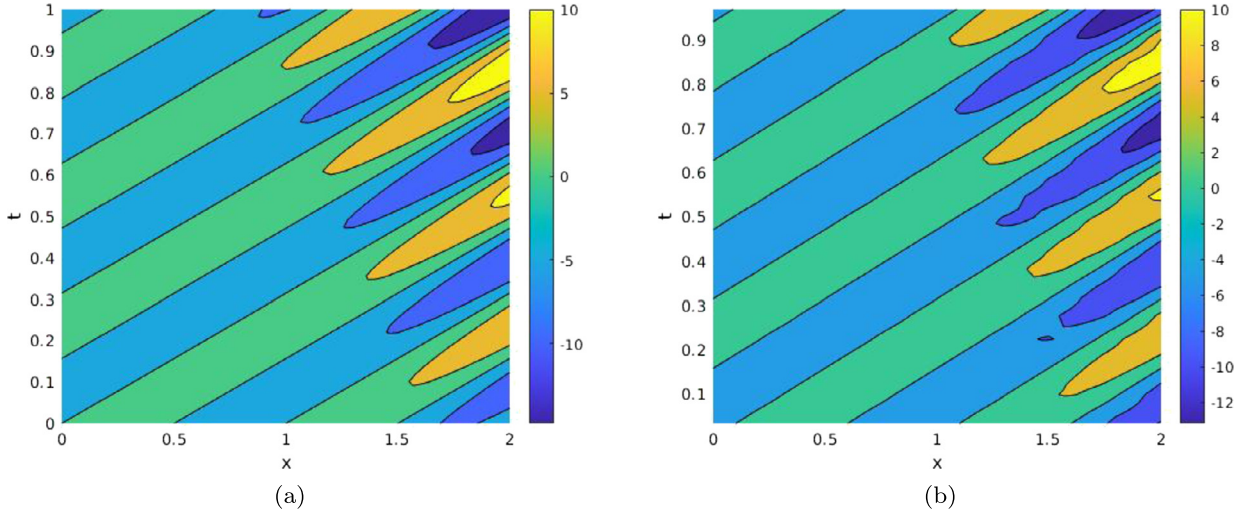
Fig. 1. Synthetic dataset: in 1a, temporal samples are equiseparated; in 1b, temporal dimension sampled randomly according to a uniform $U(0, 1)$ distribution. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

In the next section, we conduct a thorough assessment of the θ-**DMD** and its capabilities, in comparison with the baseline **DMD** algorithm.

Note that the θ-**DMD** approach in Eq. (21) is formally similar to the **DMD** technique in Equation (5). However, whereas Eq. (5) codes a direct relationship between consecutive system states, Eq. (21) establishes an input-output relationship between a weighted average of consecutive states and its difference [41,16].

Finally, the θ-**DMD** approach in Eq. (21) leads to a minor adjustment regarding the reconstruction step in Eq. (11): a factorization relationship $\mathbf{X} = \mathbf{\Phi} \, \mathbf{D}_\alpha \, \mathbf{V}_\mu$ still holds, but now $\lambda_i = \mu_i$; and matrix $\mathbf{V}_\mu$ is not a Vandermonde matrix, but a circulant matrix given as:

$$
\mathbf{V}_\mu =
\begin{bmatrix}
\exp(\mu_1 t_1) & \exp(\mu_1 t_2) & \vdots & \vdots & \exp(\mu_1 t_{n_s-1}) \\
\exp(\mu_2 t_1) & \exp(\mu_2 t_2) & \vdots & \vdots & \exp(\mu_2 t_{n_s-1}) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\exp(\mu_{n_s} t_1) & \exp(\mu_{n_s} t_2) & \vdots & \vdots & \exp(\mu_{n_s} t_{n_s-1})
\end{bmatrix}
\tag{22}
$$

3. Results

As we discussed in Section 2, the goal of this work is to assess the capabilities of the θ-**DMD** method regarding the identification of flow features from non–uniformly sampled temporal sequences. In order to obtain conclusions that are as informative as possible, we have considered three testcases, encompassing different flow regimes [44].

The first testcase pertains to a one-dimensional field with spatio-temporal dependence. Despite its simplicity, this flow is very useful nevertheless, as it allows to calibrate the parameter θ.

The second testcase is the $Re_D = 60$ flow field around the mid-section of a very long cylinder. This flow is laminar, but the fact that $Re_D > Re_{D,c} \approx 45$ makes this problem rich enough to study with **DMD**/θ-**DMD**.

Finally, the $Re_\tau = 200$ turbulent channel flow field along two indefinitely long plates is considered. The multiscale nature of the problem makes it a challenging test for both the **DMD** and the θ-**DMD** techniques.

All the computations described have been performed on a computer equipped with an 4-core Intel(R) Core(TM) i5-3570K CPU at 3.40 GHz, a cache memory of 6144 kB and 8.0 GB of RAM.

3.1. Results – synthetic dataset

As the first testcase, we consider a synthetic field given by Equation (23).

$$
u(x, t) = u_s(1 + \xi) \sin(2\pi \kappa_s x - \omega_s t) \, exp(\sigma_s t + \gamma_s x).
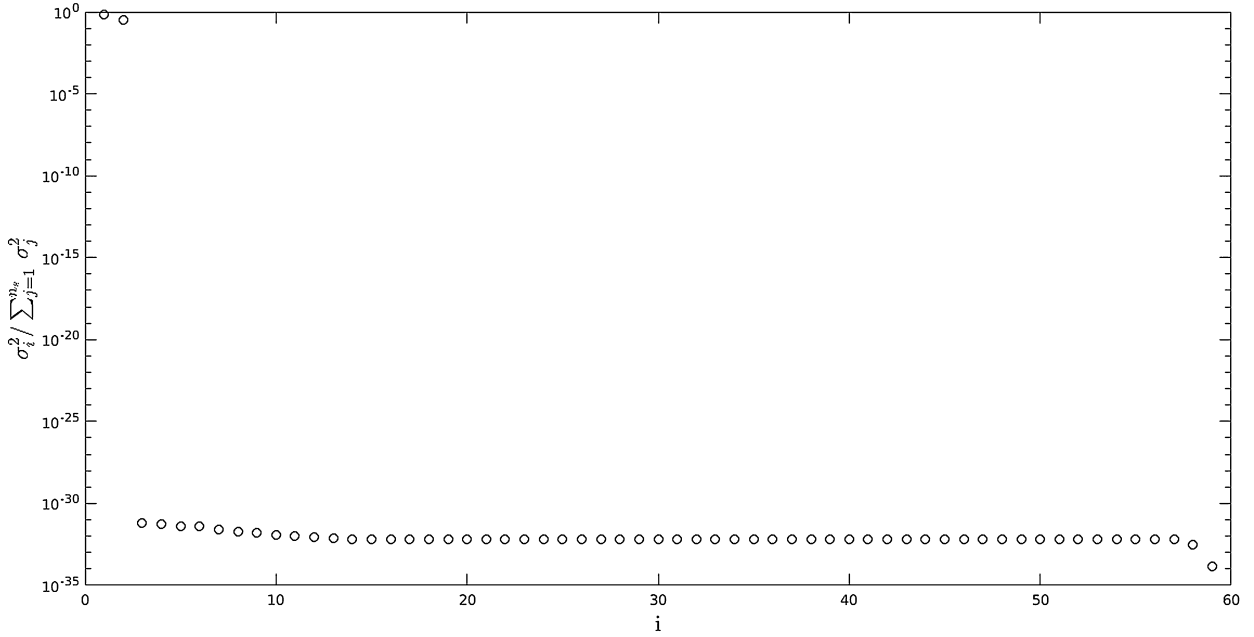\tag{23}
$$

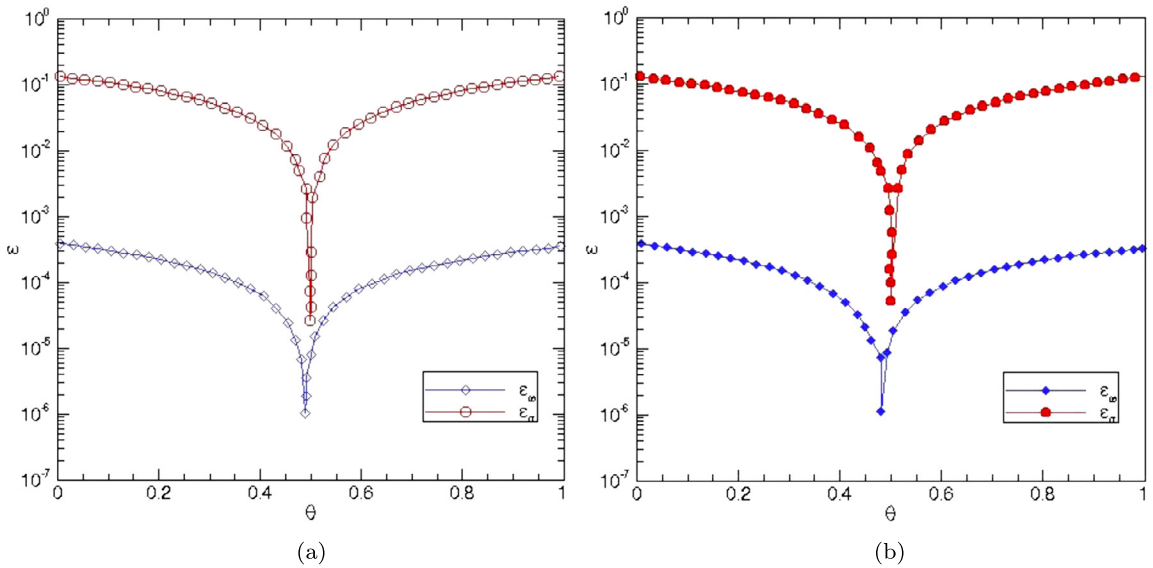**Fig. 2.** Synthetic dataset: singular values.

Fig. 3. Synthetic dataset: effect of parameter $\theta \in (0.1)$ in the accuracy of the results. In 3a, for uniformly sampled database; in 3b, for non-equiseparated (randomly sampled) database.

This toy model is one-dimensional, presents both spatial and temporal variations and has been already investigated in references [22,38,44]. Following Reference [38], we consider angular pulsation $\omega_s = 20$ and temporal growth rate $\sigma_s = 0.75$. The initial amplitude u_s, the wavenumber κ_s and the spatial growth rate γ_s are all set to 1; white multiplicative noise $\xi \sim U([-1, 1])$ is considered, with a Noise to Signal Ratio ($NSR = \max |\xi/u_s|$) up to 5%. The spatio-temporal domain is discretized with $n_p = 2000$ equispaced points in $[0, 2]$, whereas two different temporal discretizations for the interval $[0, 1]$ are considered: a) on the one hand, $n_s \in [61, 2001]$ equiseparated temporal samples (see Fig. 1a); b) on the other hand, $n_s' \in [61, 2001)$ temporal samples taken at instants that follow a uniform $U(0, 1)$ distribution are considered (see Fig. 1b).

We begin by investigating to which extent can the θ-**DMD** method identify the correct ω_s and σ_s, be it from the uniformly sampled or from the non–equiseparated in time datasets. According to Fig. 2, a rank 2 **SVD** reconstruction is
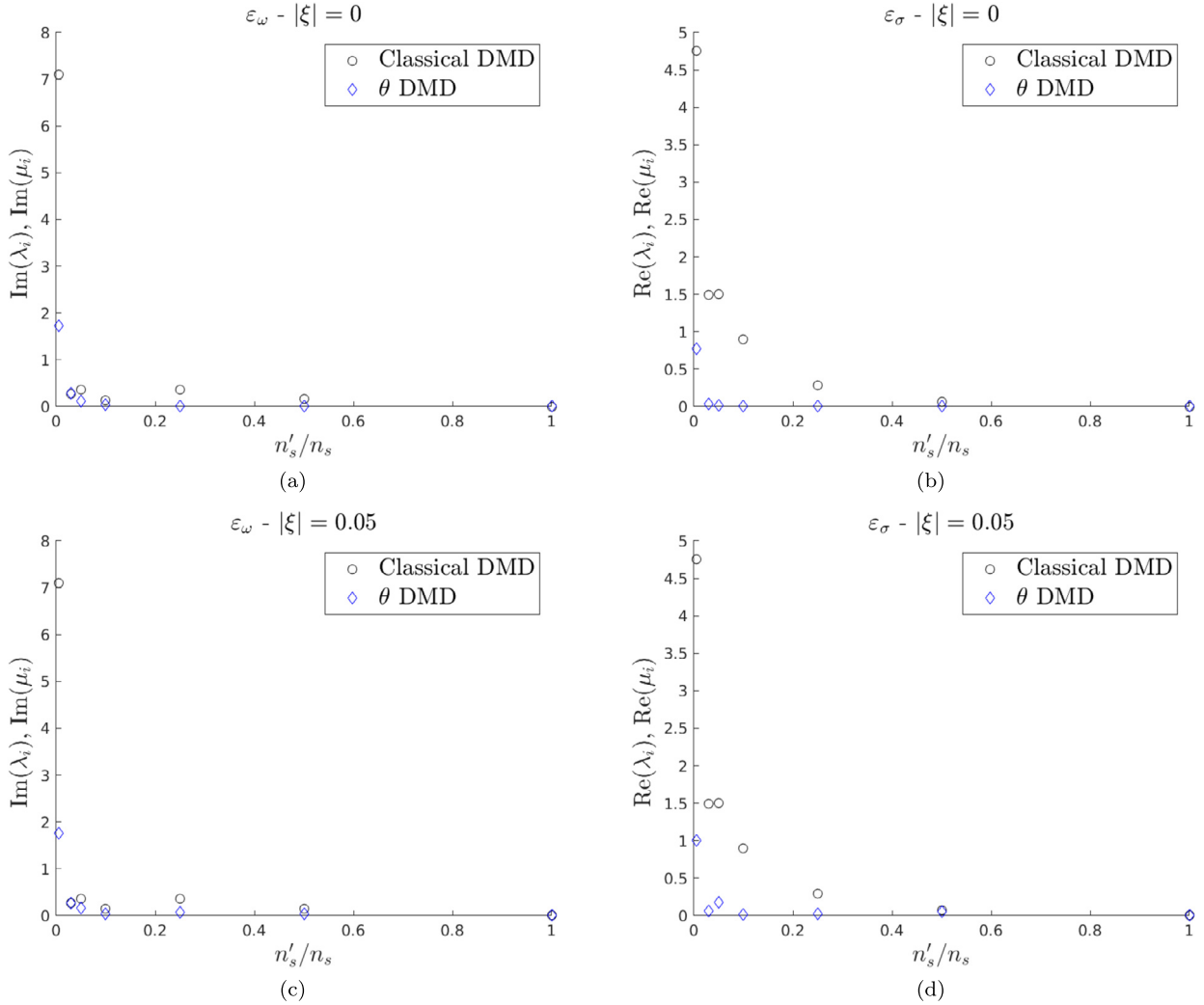
Fig. 4. Synthetic dataset: the effect of random temporal subsampling and noise level for **DMD** and θ-**DMD** analyses, using $n_p = 2000$. Figs. 5a and 5b, clean dataset $|\xi| = 0$; Figs. 5c and 5d, perturbed dataset with $|\xi| = 5e - 2$.

adequate for this dataset. Fig. 3 shows the accuracy of the θ-**DMD** method for different values of the parameter θ, using as metrics the errors $\varepsilon_\omega = \frac{|\omega_\theta - \omega_s|}{|\omega_s|}$ and $\varepsilon_\sigma = \frac{|\sigma_\theta - \sigma_s|}{|\sigma_s|}$. First, errors ε_σ are systematically larger than ε_ω, a behaviour already observed in [38,44]. It is also evident the noticeable error decrease as $\theta \to 0.5$, consistent with the increased order of accuracy associated to a *Crank-Nicolson* temporal discretization. In view of these results, we will consider $\theta = 0.5$ for the rest of our study.

Figs. 4 and 5 investigate the influence of n'_s (the randomly decimated dataset length) and the noise level on the results obtained by classical **DMD** and θ-**DMD** methods. Figs. 4a and 4b present the errors in growth rate and frequency, for a clean dataset ($|\xi| = 0$, see Eq. (23)); Figs. 4c and 4d show the same errors but for the dataset perturbed by noise with $|\xi| = 5e - 2$. Irrespective of the level of noise, if the number of temporal samples is large enough, both classical **DMD** and θ-**DMD** attain low error in the identification of the growth and frequency. As n'_s decreases, the errors arise progressively; typically, the error of the θ-**DMD** is lower than that of the **DMD**. In the noise–free case, over the range $n'_s/n_s \in (0.05, 0.25)$, θ-**DMD** outperforms largely **DMD** in the identification of the growth rate; in the noise–contaminated case, θ-**DMD** shows acceptable results up to 0.083 (*i.e.*, $n'_s = 61$) Finally, whenever n'_s/n_s is very low, neither of the methods offer acceptable error levels. The behaviour reported so far is independent of the n_p considered, see Fig. 5, which repeats the experiment using $n_p = 60$.

To conclude, let us compare the reconstruction offered by the **DMD** and θ-**DMD** methods for both uniformly and non–uniformly sampled versions of the database for the challenging $n'_s = 61$ case. The contraposition of Figs. 6a and 6b shows that, as long as the data samples are temporally equiseparated, no appreciable differences in the reconstruction are discernible.
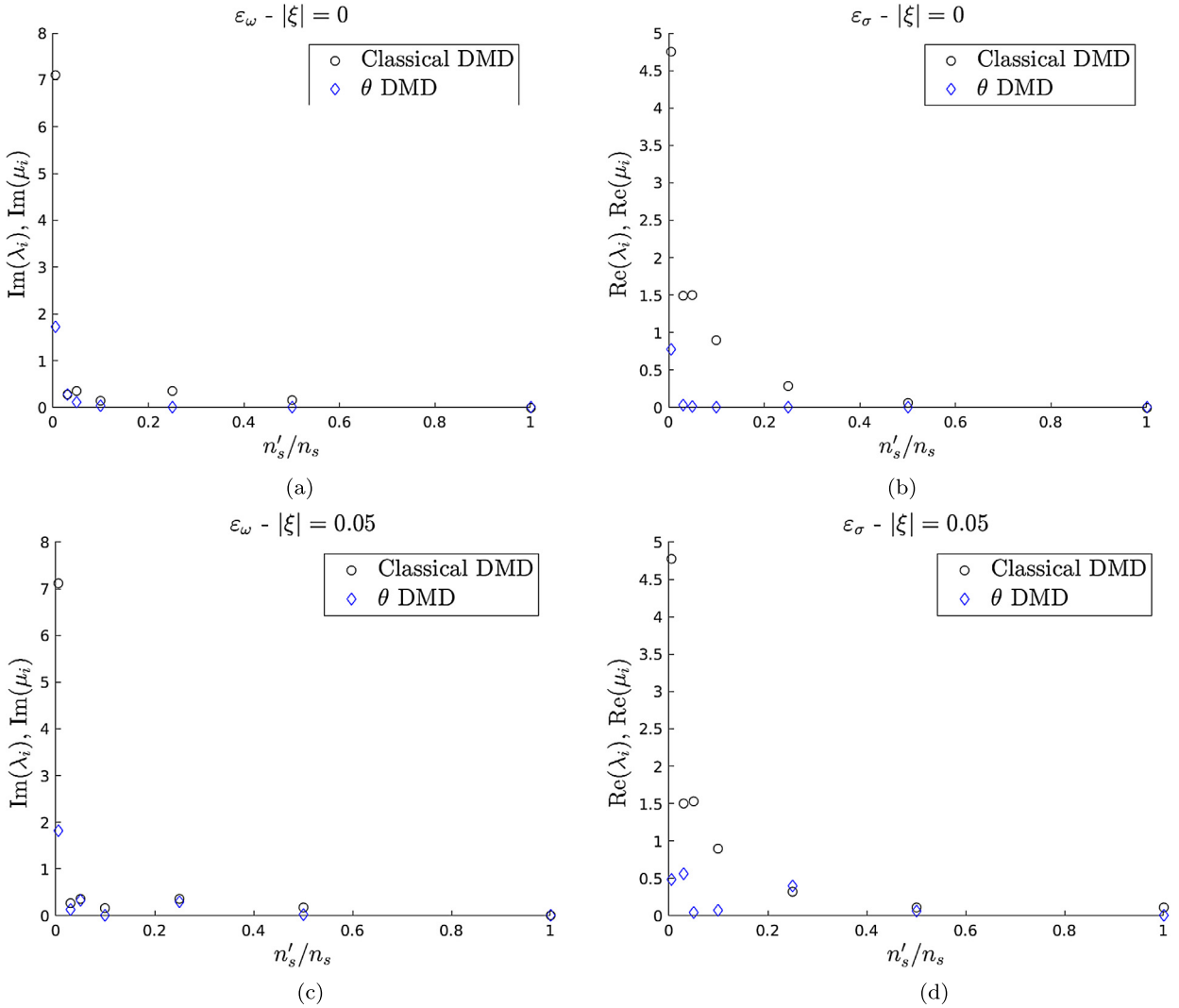
Fig. 5. Synthetic dataset: the effect of random temporal subsampling and noise level for **DMD** and θ-**DMD** analyses, using $n_p = 60$. Figs. 5a and 5b, clean dataset $|\xi| = 0$; Figs. 5c and 5d, perturbed dataset with $|\xi| = 5e - 2$.

The situation changes dramatically when non-equiseparated samples are considered, see Figs. 6c-6d. As seen in Fig. 6c, the reconstruction step using **DMD** is compromised when not equiseparated temporal samples are used. The results for θ-**DMD** appear in Fig. 6d: the reconstruction accuracy is remarkable (even for this low n'_s). The observations made so far (namely, that **DMD** and θ-**DMD** show a comparable behaviour on temporally equiseparated datasets; and that θ-**DMD** outperforms **DMD** method for non-equiseparated snapshots) extend to the increasingly complex datasets analysed in the rest of the manuscript.

3.2. Results – $Re_D = 60$ cylinder flow dataset

The **2D** flow around the cross-section of an infinitely long cylinder is a classical configuration, extensively used as validation testcase, see *e.g.* [15,45,44]. This flow presents a Hopf bifurcation occurring at $Re_D \approx 46$; conditions slightly above this critical value are interesting: the flow behaviour, despite being laminar, includes unsteady vortex shedding.

In this work, following [45,44], we focus on a $Re_D = 60$ flow. The flow is periodic, with dominant frequency $\omega \approx 56.55$. This corresponds to a Strouhal number of $St = 0.137$, which is consistent with the correlations in [46].

We consider the dataset generated in [44], consisting of $n_s = 800$ temporally equiseparated velocity snapshots. The sampling time is $\Delta t = 0.0025$, and the grid employed has $n_p = 36,474$ points. Since only the horizontal velocity component u_x is included in the study, the dataset can be arranged as a $n_p \times n_s$ matrix.

The comparison of both the **DMD** and the θ-**DMD** methods has been accomplished in two phases. First, the complete dataset (*i.e.*, $n_s = 800$ snapshots) has been analysed with **DMD** and θ-**DMD**. The results of these analyses are summarized in
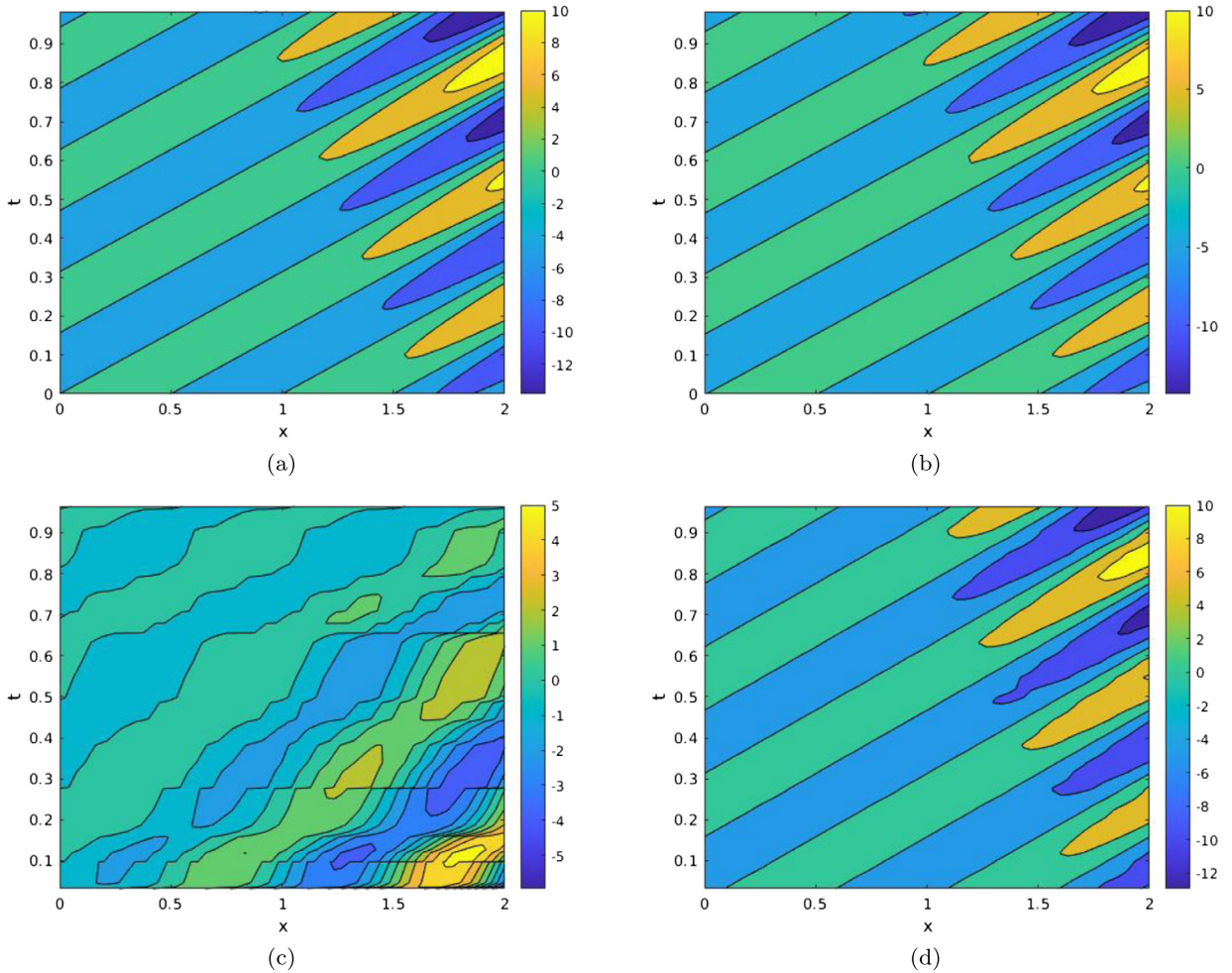
Fig. 6. Synthetic dataset, reconstructed using Eq. (11): in Figs. 6a-6c, results for **DMD**; in Figs. 6b-6d, results for θ-**DMD**. Figs. 6a-6b, shows results for uniformly sampled time; Figs. 6c-6d, shows results for non–uniformly sampled time.

Fig. 7 and Fig. 8. Fig. 7a shows the non-dimensionalized singular values of the dataset. Notice the fast decay of the singular values, which is more apparent in Fig. 7c. This fast decay is consistent with the highly organized nature of the flow, which is laminar. Also, Fig. 7c suggests taking rank $r_0 = 20$ in Eq. (4). Fig. 7e shows the non-dimensionalized amplitudes that both the **DMD** and the θ-**DMD** methods identify using rank $r_1 = 7$. We observe that the larger amplitude coincides with the mode representing the average; also, both methods identify modes at the same frequencies.

Fig. 8 shows the spatial structure of the oscillatory modes retrieved by **DMD** and θ-**DMD**. The modes have very close frequency and growth rate, and are visually indistinguishable from each other. This supports the statement that, as long as the dataset is uniformly sampled in time, **DMD** and θ-**DMD** will provide comparable results.

The second phase of our study consists in analysing a reduced version of the cylinder dataset. This time, $n'_s = 200$ snapshots are randomly chosen from the original dataset, which implies that the snapshots are not temporally equiseparated anymore. Recall that **DMD** requires a fixed Δt in order to convert from the discrete to the continuous eigenvalues in Eq. (9). In that case, we use the averaged temporal separation $\overline{\Delta t}$. Before discussing the results, let us note in advance that, once we break the temporal equiseparation assumption, we do not expect the baseline **DMD** algorithm to provide reasonable results anymore. If we show the **DMD** results is because they are informative *in comparison with* θ-**DMD** results.

The results from this second phase of the study are summarized by Figs. 7 and 9. Fig. 7b shows the non-dimensionalized singular values of the dataset. The singular values present again a fast decay, which can be seen better in Fig. 7d. Also, Fig. 7d suggests taking rank $r_0 = 7$ in Eq. (4). Finally, Fig. 7f shows the non-dimensionalized amplitudes that both the **DMD** and the θ-**DMD** methods (rank $r_1 = 7$) identify. The frequencies identified by the θ-**DMD** method match relatively well the dominant frequencies retrieved from the original dataset (with errors $\varepsilon < 5\%$), whereas the frequencies identified by **DMD** are partly wrong: the higher frequency mode is not retrieved anymore, whereas a mode appears at an intermediate frequency ≈ 94. Fig. 9 shows the spatial structures of the modes. The comparison of the pairs Fig. 8b-9b, Fig. 8d-9d and Fig. 8f-9f shows the noticeable similitude from the θ-**DMD** modes obtained from both data sequences. Finally, the **DMD**
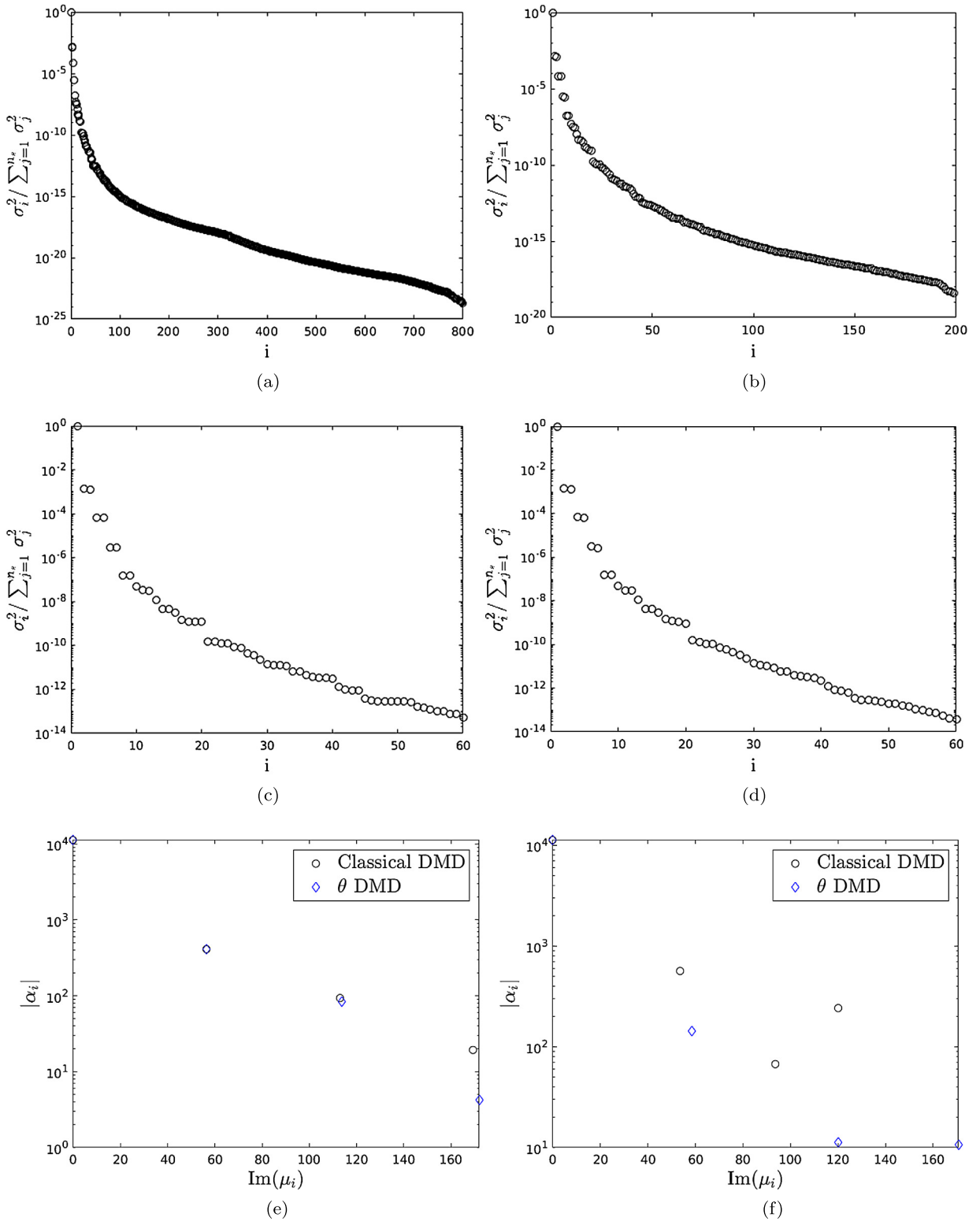
Fig. 7. $Re_D = 60$ dataset: **DMD** and θ-**DMD** analyses on two temporal sequences. Results for the complete ($n_s = 800$ snapshots) sequence in Figs. 7a, 7c and 7e; results for a subsequence built by randomly selecting $n'_s = 200$ snapshots from the original sequence in Figs. 7b, 7d and 7f.
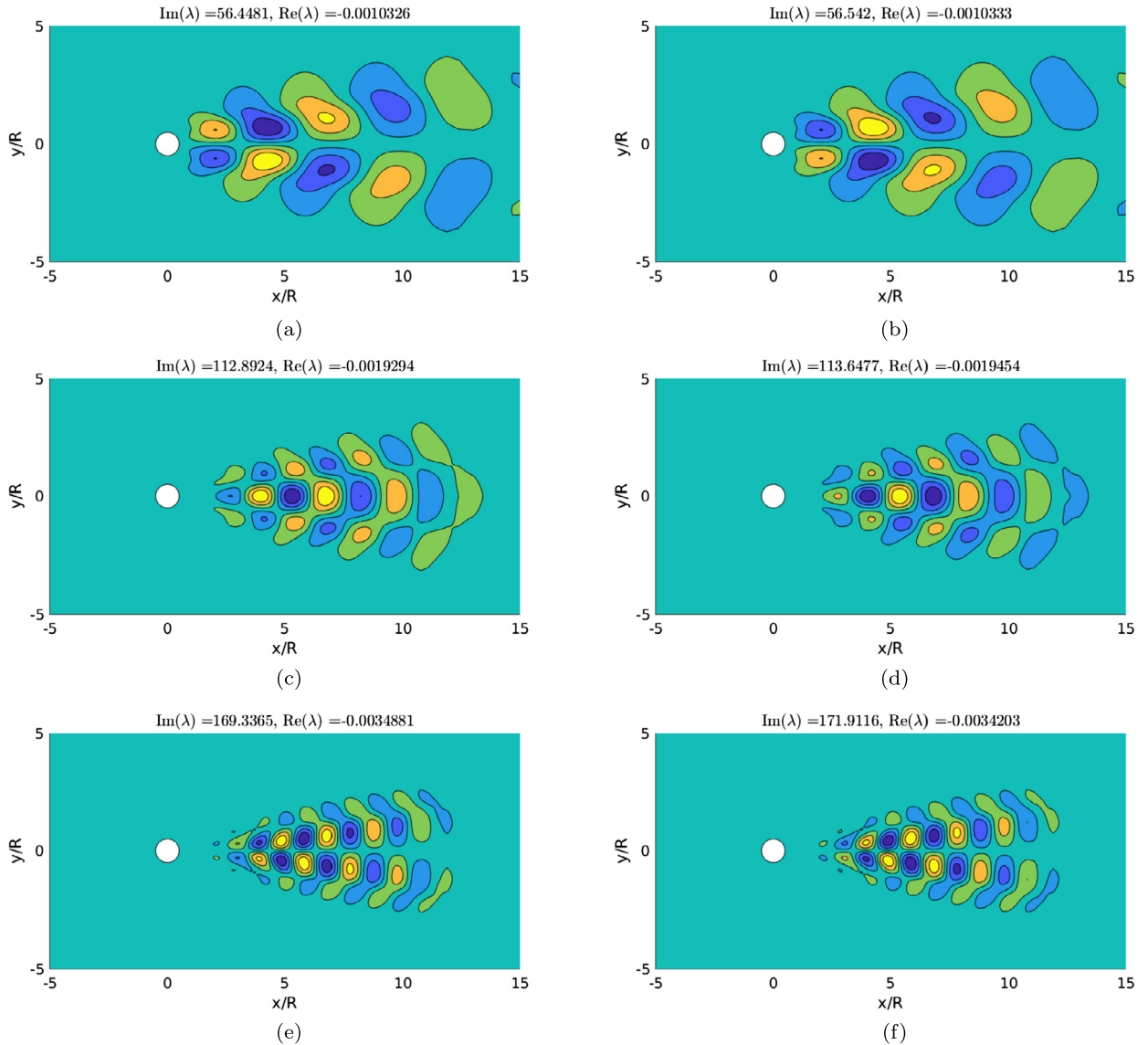
Fig. 8. $Re_D = 60$ dataset: modes obtained for the complete $n_s = 800$ snapshots sequence. Baseline **DMD** results in 8a, 8c, 8e; θ-**DMD** modes in 8b, 8d, 8f.

mode at frequency ≈ 120 shows an asymmetric spatial structure, whereas the mode at the intermediate frequency ≈ 94 presents the structure resembling the mode in Fig. 8c.

3.3. Results – $Re_\tau = 200$ turbulent channel flow dataset

We finally compare the performance of the **DMD** and the θ-**DMD** methods when applied to a turbulent channel flow database with $Re_\tau \approx 200$ [47].

The channel flow here considered has been generated using the incompressible **DNS** solver described in [48]. The code solves for the wall-normal components of velocity v and vorticity η. These quantities are Fourier-transformed (dealiased using the *2/3* rule) along the homogeneous directions, and discretized using explicit compact finite-differences along the wall normal direction. The streamwise u and spanwise w velocity components are both retrieved using the continuity equation with the relation $\eta = \frac{\partial w}{\partial x} - \frac{\partial u}{\partial z}$. Time integration is achieved by an explicit third order, low-storage Runge–Kutta method combined with an implicit second–order Crank–Nicolson scheme for the non–linear terms. The simulation has been conducted under the assumption of constant flow rate. The database characteristics are summarized in Table 1. In total, 1200 flow snapshots were stored, separated in time by $\Delta t^s = 0.156$.

The turbulent database allows to identify an statistically converged averaged solution $\vec{U}(y)$. For our study we consider the temporal sequence formed by the perturbation velocity vectors $\vec{u}'(x, y, z, t_j) = \vec{u}(x, y, z, t_j) - \vec{U}(y)$. As discussed in
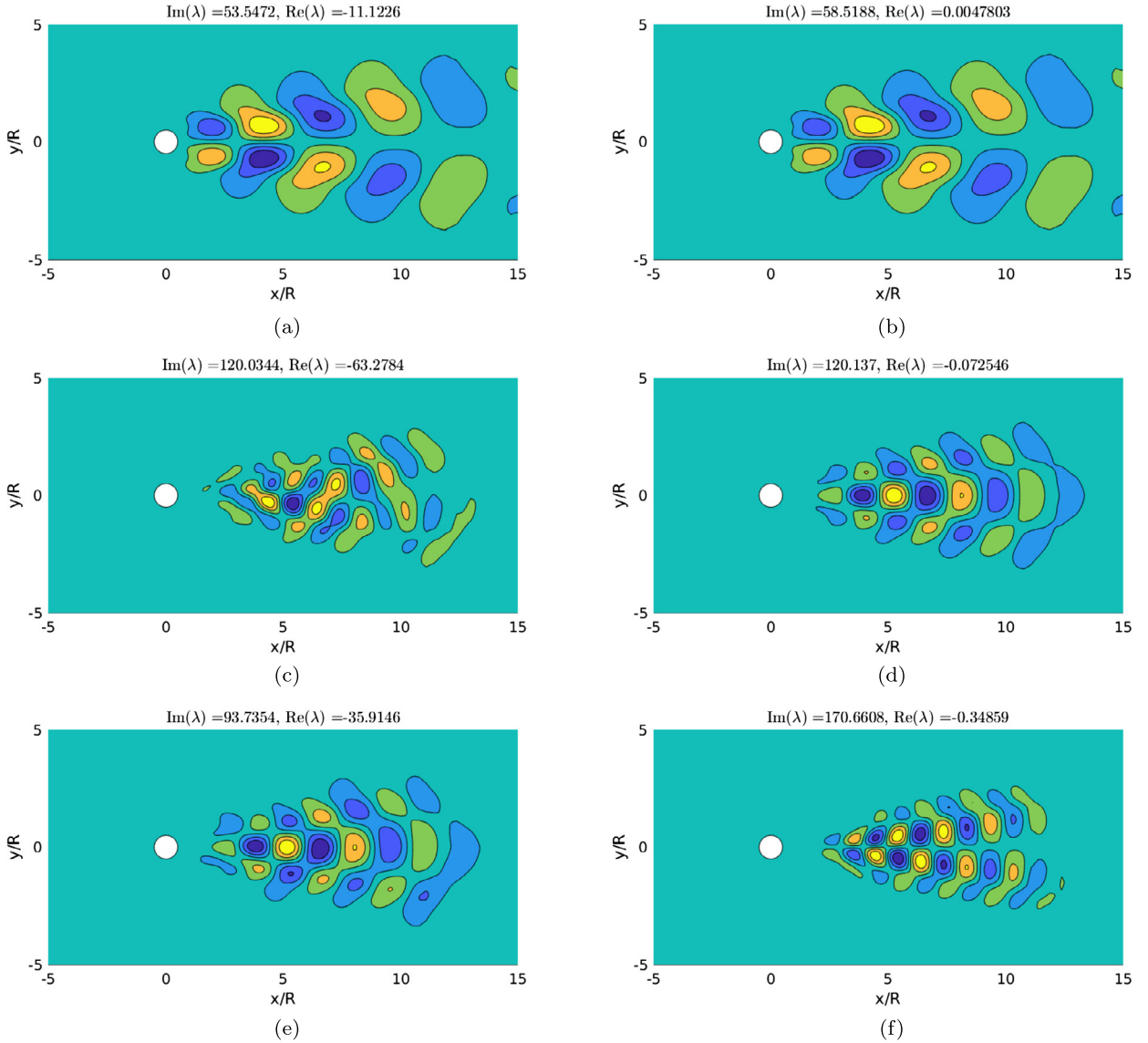
Fig. 9. $Re_D = 60$ dataset: modes obtained from a subsequence built by randomly selecting $n'_s = 200$ snapshots from the original sequence. Baseline **DMD** results in 9a, 9c, 9e; θ-**DMD** modes in 9b, 9d, 9f.

Table 1
$Re_\tau = 200$ turbulent channel flow: database characteristics.

L_x/δ	L_y/δ	L_z/δ	Re_c	u_τ	n_x	n_y	n_z	Δt^s
π	2	$\pi/2$	3678.7	0.042	96	101	96	0.156

[30], a simplified database that still represents the turbulent physics is obtained by removing every other point along the homogeneous x and z directions and retaining only either the region corresponding to $y^+ < 50$ from the *lower* half of the domain. The resulting database can be recast as a $n_p \times n_s$ matrix, with $n_p = 117, 504$ and $n_s = 1200$; This dataset is large enough to pose a tough challenge for most workstations. Techniques to accomplish the **SVD**/**DMD** decompositions of such databases are discussed in [20,30,44].

The comparison of the **DMD** and the θ-**DMD** methods is organized again in two steps: first, **DMD** and θ-**DMD** have been applied both to the complete dataset (*i.e.*, $n_s = 1200$ snapshots). The results of this comparison are summarized in Fig. 10 and Fig. 11.

Fig. 10a shows the non-dimensionalized singular values of the complete dataset. In this case, the decay of the singular values is much slower than in the other cases considered. This behaviour reflects the complex nature of the turbulent
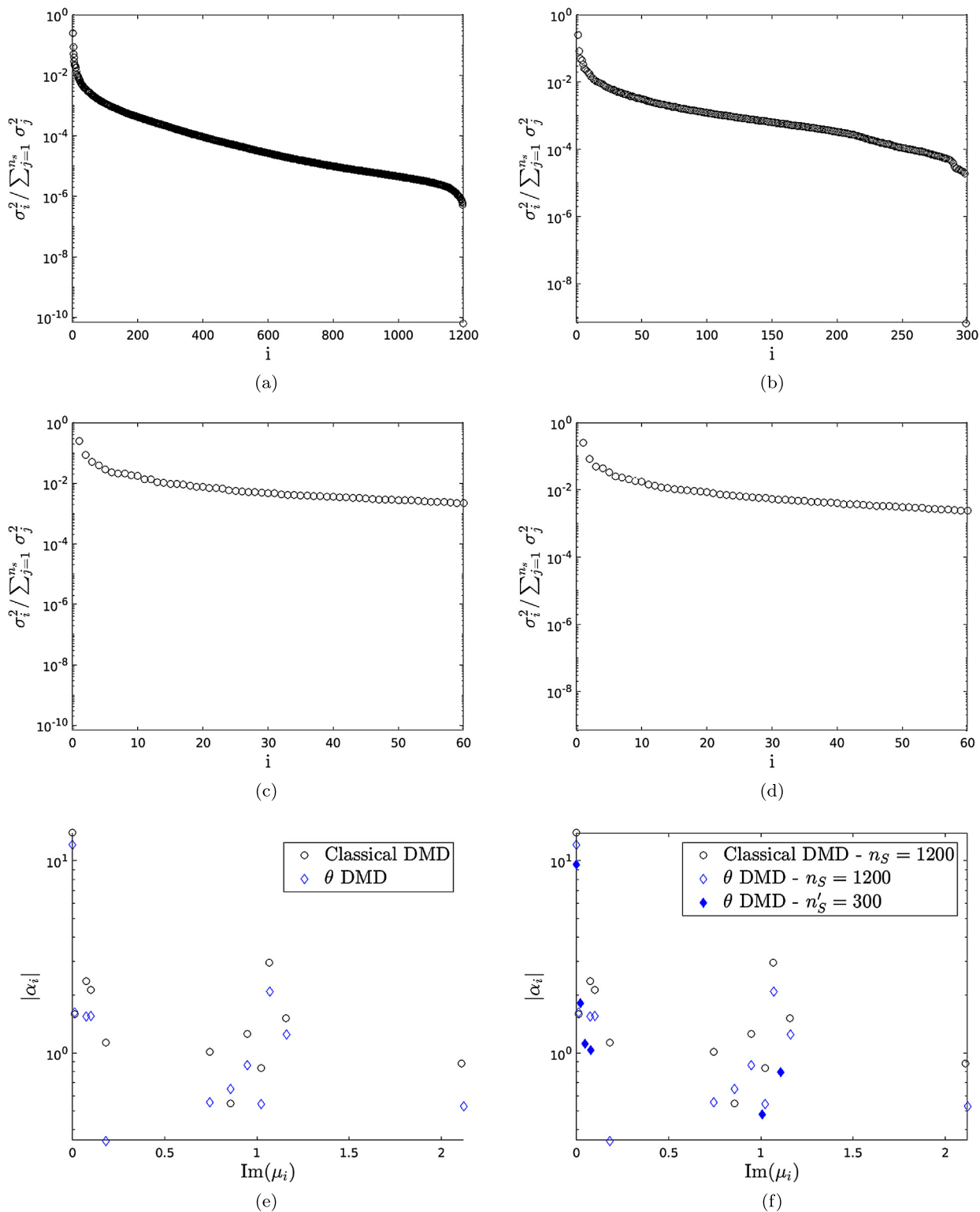
Fig. 10. $Re_\tau = 200$ dataset: **DMD** and θ-**DMD** analyses on two temporal sequences. Results for the complete ($n_s = 1200$ snapshots) sequence in Figs. 10a, 10c and 10e; results for a subsequence built by randomly selecting $n'_s = 300$ snapshots from the original sequence in Figs. 10b, 10d and 10f.
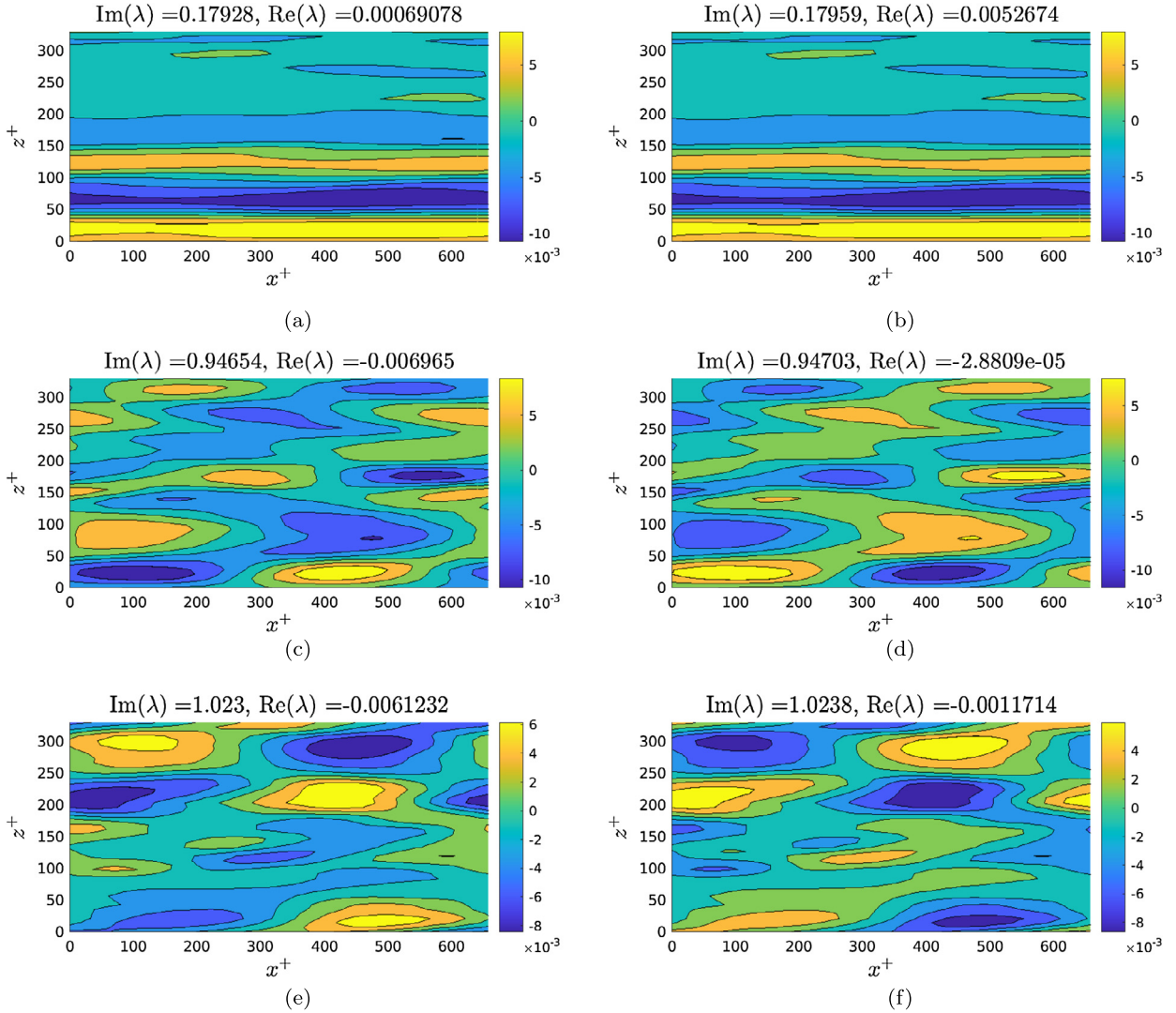
Fig. 11. $Re_\tau = 200$ dataset: modes obtained for the complete $n_s = 1200$ snapshots sequence. Baseline **DMD** results in 11a, 11c, 11e; θ-**DMD** modes in 11b, 11d, 11f. Note that modes in 11c and 11d, and in 11e and 11f, are in phase opposition.

channel flow, a statistically stationary phenomenon involving multiple temporal and spatial interacting scales. In view of Fig. 10c, we take $r_0 = 44$ in Eq. (4).

Fig. 10e compares the non-dimensionalized amplitudes identified by **DMD** and θ-**DMD** using rank $r_1 = 23$. Again, both methods identify modes at the same frequencies.

Fig. 11 shows the near wall ($y^+ \approx 15$) spatial structure of the oscillatory modes retrieved. Once again, the modes have very close frequency and growth rate, and are practically indistinguishable from each other. This confirms that **DMD** and θ-**DMD** identify the same modes as long as the dataset is uniformly sampled in time.

The second analysis step considers a reduced version of the dataset obtained by randomly choosing $n'_s = 300$ snapshots. Only the θ-**DMD** method is applied this time, since –*cf.* section 3.2– **DMD** modes on non–uniformly sampled datasets cannot be trusted.

The results from this second phase of the study are summarized by Figs. 10 and 12. Fig. 10b shows the non-dimensionalized singular values of the dataset, which again decay progressively, see also Fig. 10d.

Finally, Fig. 10f shows the non-dimensionalized amplitudes identified by θ-**DMD** (ranks $r_0 = 40$ and $r_1 = 11$) from the non-uniformly sampled subsequence in comparison with the amplitudes obtained by the **DMD** and θ-**DMD** methods on the complete $n_s = 1200$ data sequence. We observe how θ-**DMD** identifies the same modes captured by either method applied to the original dataset.

Fig. 12 shows the spatial structures of the modes obtained from the non–uniformly sampled $n'_s = 300$ dataset. The comparison of the pairs Fig. 11b-12a, Fig. 11d-12b and Fig. 11f-12c shows the noticeable similitude between θ-**DMD** modes obtained from both data sequences.
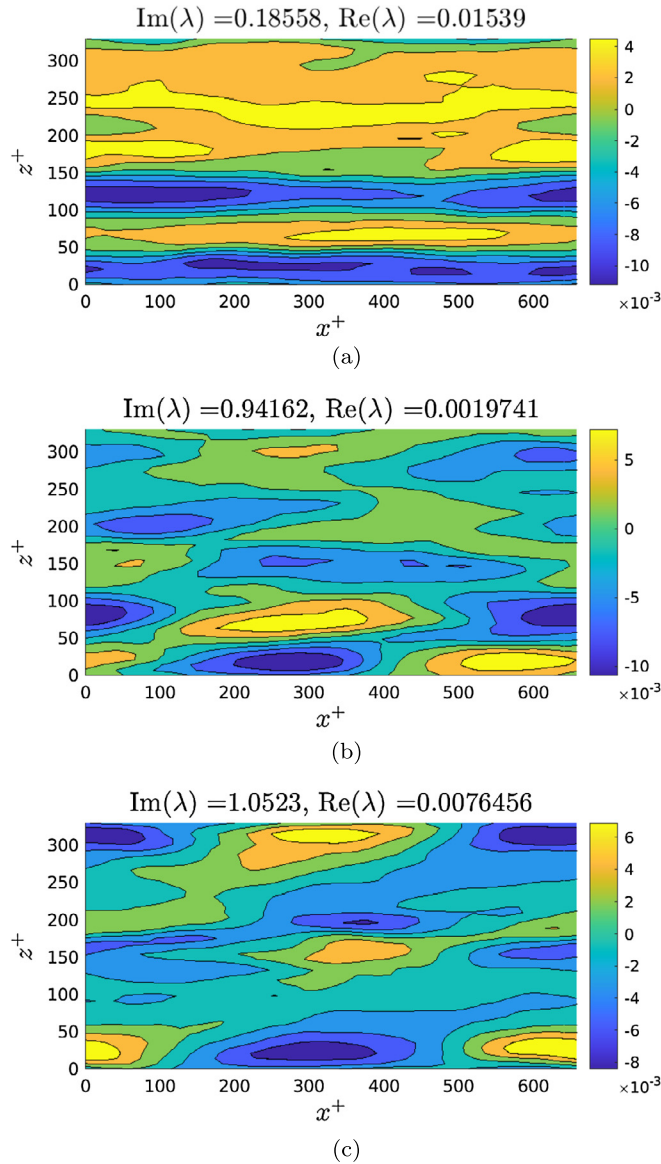
Fig. 12. $Re_\tau = 200$ dataset: θ-**DMD** modes obtained from a subsequence built by randomly selecting $n'_s = 300$ snapshots from the original sequence.

4. Conclusions

In this work we have proposed a novel Dynamic Mode Decomposition (**DMD**) technique capable of handling non–uniformly sampled data. As it is usual in **DMD** analysis, a linear relationship between consecutive snapshots is made. However, the technique here proposed differs from established **DMD** methods on two major aspects: first, the linear relationship assumption is established on the numerically discretized counterpart of a linearized state-equation; and second, a dimensionally–reduced view of the state data is considered.

The data matrix dimensionality reduction is accomplished by simply exploiting the optimality properties of the **SVD** technique. In this work we have resorted to standard (*lapack*) **SVD**, but nothing prevents to resort to other implementations, *e.g.* memory distributed **SVD** based on the **TSQR** algorithm [20]. Alternative methods could be considered to conduct the spatial dimension reduction, *e.g.* the variational autoencoders/convolutional neural networks approach suggested in [49].

The performance of the new method, which we term θ-**DMD**, is assessed on three different, increasingly complex datasets: a *synthetic* flow field, a toy model consisting on a one-dimensional both temporally and spatially unstable; a $Re_D = 60$ flow around a cylinder cross section, laminar but presenting unsteady periodic vortex shedding, and a $Re_\tau = 200$ turbulent channel flow, a complex flow presenting multiple spatio-temporal interacting scales.

For the three datasets considered, whenever the dataset is uniformly sampled, the θ-**DMD** method systematically provides comparable results to the original **DMD** method.

Additionally, the θ-**DMD** is still capable of recovering relevant flow features from non–uniformly (sub–)sampled databases, whereas (understandably) **DMD** cannot.

The tool here proposed opens the way to conduct **DMD** analyses for non–uniformly sampled data, and therefore might become useful *e.g.*, when confronted with experimental datasets with *missing data*, or when facing numerical datasets generated using adaptive time-integration schemes.

Finally, the θ-**DMD** technique here presented is formulated by establishing a linear relationship between an input (**X**) and output (**Y**) matrices. These matrices are obtained by post–multiplying the reduced chronos matrix $\mathcal{C}_1^{n_s}$ with matrices $\mathbf{M}_\delta$ and $\mathbf{M}_\theta$, respectively. Future work will explore whether alternative definitions for the $\mathbf{M}_\delta$, $\mathbf{M}_\theta$ operators yield to alternative algorithms with improved properties.

CRediT authorship contribution statement

Binghua Li: Conceptualization, Software, Validation, Writing – review & editing. **Jesús Garicano-Mena:** Conceptualization, Software, Validation, Writing – original draft, Writing – review & editing. **Eusebio Valero:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] J.L. Lumley, Stochastic Tools in Turbulence, Academic Press, 1970.
[2] L. Sirovich, Turbulence and the dynamics of coherent structures, Q. Appl. Math. 45 (3) (1987) 561–590.
[3] G. Berkooz, P. Holmes, J. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annu. Rev. Fluid Mech. 25 (1) (1993) 539–575.
[4] S. Volkwein, Proper Orthogonal Decomposition: Theory and Reduced-Order Modelling, Lecture Notes, August 2013.
[5] M. Sieber, C.O. Paschereit, K. Oberleithner, Spectral proper orthogonal decomposition, J. Fluid Mech. 792 (2016) 798–828, https://doi.org/10.1017/jfm.2016.103.
[6] A. Towne, O. Schmidt, T. Colonius, Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis, J. Fluid Mech. 847 (2018) 821–867.
[7] S. Derebail Muralidhar, B. Podvin, L. Mathelin, Y. Fraigneau, Spatio-temporal proper orthogonal decomposition of turbulent channel flow, J. Fluid Mech. 864 (2019) 614–639.
[8] N. Halko, P.G. Martinsson, J.A. Tropp, Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev. 53 (2) (2011) 217–288, https://doi.org/10.1137/090771806.
[9] M.A. Méndez, M. Balabane, J.-M. Buchlin, Multi-scale proper orthogonal decomposition of complex fluid flows, J. Fluid Mech. 870 (2019) 988–1036, https://doi.org/10.1017/jfm.2019.212.
[10] C. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D. Henningson, Spectral analysis of nonlinear flows, J. Fluid Mech. 641 (2009) 115–127.
[11] P. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.
[12] J.N. Kutz, S.L. Brunton, B.W. Brunton, J.L. Proctor, Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems, SIAM, 2016.
[13] I. Mezić, Analysis of fluid flows via spectral properties of the Koopman operator, Annu. Rev. Fluid Mech. 45 (2013) 357–378.
[14] K. Taira, S.L. Brunton, S.T.M. Dawson, C.W. Rowley, T. Colonius, B.J. McKeon, O.T. Schmidt, S. Gordeyev, V. Theofilis, L.S. Ukeiley, Modal analysis of fluid flows: an overview, AIAA J. 55 (12) (2017) 4013–4041, https://doi.org/10.2514/1.J056060.
[15] K. Chen, J. Tu, C. Rowley, Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses, J. Nonlinear Sci. 22 (6) (2012) 887–915.
[16] M.R. Jovanović, From bypass transition to flow control and data-driven turbulence modeling: an input–output viewpoint, Annu. Rev. Fluid Mech. 53 (1) (2021) 311–345, https://doi.org/10.1146/annurev-fluid-010719-060244.
[17] B. Herrmann, P.J. Baddoo, R. Semaan, S.L. Brunton, B.J. McKeon, Data-driven resolvent analysis, J. Fluid Mech. 918 (2021) A10, https://doi.org/10.1017/jfm.2021.337.
[18] D. Bistrian, I. Navon, Randomized dynamic mode decomposition for nonintrusive reduced order modelling, Int. J. Numer. Methods Fluids 112 (1) (2017) 3–25.
[19] N.B. Erichson, L. Mathelin, J.N. Kutz, S.L. Brunton, Randomized dynamic mode decomposition, SIAM J. Appl. Dyn. Syst. 18 (4) (2019) 1867–1891, https://doi.org/10.1137/18M1215013.
[20] T. Sayadi, P. Schmid, Parallel data-driven decomposition algorithm for large-scale datasets: with application to transitional boundary layers, Theor. Comput. Fluid Dyn. 30 (5) (2016) 415–428.

[21] T. Sayadi, P. Schmid, J. Nichols, P. Moin, Reduced-order representation of near-wall structures in the late transitional boundary layer, J. Fluid Mech. 748 (2014) 278–301.

[22] D. Duke, J. Soria, D. Honnery, An error analysis of the dynamic mode decomposition, Exp. Fluids 52 (2) (2012) 529–542.

[23] P. Schmid, D. Violato, F. Scarano, Decomposition of time-resolved tomographic PIV, Exp. Fluids 52 (6) (2012) 1567–1579.

[24] A. Cammilleri, F. Gueniat, J. Carlier, L. Pastur, E. Memin, F. Lusseyran, G. Artana, POD-spectral decomposition for fluid flow analysis and model reduction, Theor. Comput. Fluid Dyn. 27 (2013) 787–815.

[25] S. Le Clainche, J. Vega, J. Soria, Higher order dynamic mode decomposition of noisy experimental data: the flow structure of a zero-net-mass-flux jet, Exp. Therm. Fluid Sci. 88 (Supplement C) (2017) 336–353.

[26] N. Groun, M. Villalba-Orero, E. Lara-Pezzi, E. Valero, J. Garicano-Mena, S. Le Clainche, Higher order dynamic mode decomposition: from fluid dynamics to heart disease analysis, Comput. Biol. Med. 144 (2022) 105384, https://doi.org/10.1016/j.compbiomed.2022.105384, https://www.sciencedirect.com/science/article/pii/S0010482522001767.

[27] N. Groun, M. Villalba-Orero, E. Lara-Pezzi, E. Valero, J. Garicano-Mena, S.L. Clainche, A novel data-driven method for the analysis and reconstruction of cardiac cine MRI, arXiv:2205.12097, 2022.

[28] A. Cassinelli, M. de Giovanetti, Y. Hwang, Streak instability in near-wall turbulence revisited, J. Turbul. 18 (5) (2017) 443–464.

[29] T. Grenga, J. MacArt, M. Mueller, Dynamic mode decomposition of a direct numerical simulation of a turbulent premixed planar jet flame: convergence of the modes, Combust. Theory Model. 22 (4) (2018) 1–17.

[30] J. Garicano-Mena, B. Li, E. Ferrer, E. Valero, A composite dynamic mode decomposition analysis of turbulent channel flows, Phys. Fluids 31 (11) (2019) 115102.

[31] S. Le Clainche, D. Izbassarov, M. Rosti, L. Brandt, O. Tammisola, Coherent structures in the turbulent channel flow of an elastoviscoplastic fluid, J. Fluid Mech. 888 (2020) A5, https://doi.org/10.1017/jfm.2020.31.

[32] S.L. Brunton, J. Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control, SIAM, 2020.

[33] S. Le Clainche, J. Vega, Higher order dynamic mode decomposition, SIAM J. Appl. Dyn. Syst. 16 (2) (2017) 882–925.

[34] K. Taira, M.S. Hemati, S.L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S.T.M. Dawson, C.-A. Yeh, Modal analysis of fluid flows: applications and outlook, AIAA J. 58 (3) (2020) 998–1022, https://doi.org/10.2514/1.J058462.

[35] C. Rowley, S. Dawson, Model reduction for flow analysis and control, Annu. Rev. Fluid Mech. 49 (1) (2017) 387–417.

[36] P.J. Schmid, Dynamic mode decomposition and its variants, Annu. Rev. Fluid Mech. 54 (1) (2022) 225–254, https://doi.org/10.1146/annurev-fluid-030121-015835.

[37] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. 52 (1) (2020) 477–508, https://doi.org/10.1146/annurev-fluid-010719-060214.

[38] F. Guéniat, L. Mathelin, L.R. Pastur, A dynamic mode decomposition approach for large and arbitrarily sampled systems, Phys. Fluids 27 (2) (2015) 025113, https://doi.org/10.1063/1.4908073.

[39] B. Li, J. Garicano-Mena, E. Ferrer, E. Valero, Feature extraction algorithms applied to turbulent channel flow databases: the effect of agglomeration strategies, in: 8th European Conference for Aeronautics and Aerospace Sciences (EUCASS), 2019.

[40] M.R. Jovanović, P.J. Schmid, J.W. Nichols, Sparsity-promoting dynamic mode decomposition, Phys. Fluids 26 (2) (2014) 024103, https://doi.org/10.1063/1.4863670.

[41] J. Tu, C. Rowley, D. Luchtenburg, S. Brunton, J. Kutz, On dynamic mode decomposition: theory and applications, J. Comput. Dyn. 1 (2) (2014) 391–421.

[42] W. Boyce, R. Diprima, D. Meade, Elementary Differential Equations and Boundary Value Problems, Wiley, 2017.

[43] J. Liesen, V. Mehrman, Linear Algebra, Springer, 2015.

[44] B. Li, J. Garicano-Mena, Y. Zheng, E. Valero, Dynamic mode decomposition analysis of spatially agglomerated flow databases, Energies 13 (9) (2020) 2134.

[45] J. Kou, W. Zhang, An improved criterion to select dominant modes from dynamic mode decomposition, Eur. J. Mech. B, Fluids 62 (2017) 109–129.

[46] A. Roshko, On the development of turbulent wakes from vortex streets, TN 1191, NACA, 1954.

[47] M. Quadrio, B. Frohnapfel, Y. Hasegawa, Does the choice of the forcing term affect flow statistics in DNS of turbulent channel flow?, in: Vortical Structures and Wall Turbulence, Eur. J. Mech. B, Fluids 55 (2016) 286–293.

[48] P. Luchini, M. Quadrio, A low-cost parallel implementation of direct numerical simulation of wall turbulence, J. Comput. Phys. 211 (2) (2006) 551–571.

[49] H. Eivazi, S. Le Clainche, S. Hoyas, R. Vinuesa, Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows, arXiv:2109.01514, 2021.