DOI: 10.13009/EUCASS2019-111

DOI: ADD DOINUMBER HERE

Feature Extraction Algorithms Applied to Turbulent Channel Flow Databases: The Effect of Agglomeration Strategies

B. Li[1,2,†]*, J. Garicano-Mena*[2]*, E. Ferrer*[2] *and E. Valero*[2]
[1]*Center for Engineering and Scientific Computation, and School of Aeronautics and Astronautics, Zhejiang University*
38 Zheda road, Hangzhou, Zhejiang 310027 PR China
[2]*ETSI Aeronutica y del Espacio, Universidad Politcnica de Madrid*
Pza. Cardenal Cisneros 3, 28040, Madrid, España
libh@zju.edu.cn · jesus.garicano.mena@upm.es
[†]Corresponding author

Abstract

In this contribution, we consider strategies that allow to perform a Dynamic Mode Decomposition analysis of data at a reduced computational and memory cost. We depart from the classical `DMD`[21] and Non-Uniform `DMD`.[10] Additionally, we present a novel DMD based strategy which can also process datasets non-uniformly sampled.
The three DMD based methods are compared with respect to their numerical performance and their capabilities of handling shortened possibly non-uniformly sampled databases and to accommodate spatial agglomeration strategies, on three cases: a synthetic system, a $Re_D = 60$ flow around 2D cylinder, and a $Re_\tau \approx 200$ turbulent channel flow.

1. Introduction

Dynamic mode decomposition (DMD) techniques, which emerged about ten years ago with the contributions of Schmid[21] and Rowley and collaborators[17] , have arisen as prominent feature identification methods in the field of fluid dynamics. Any of the multiple variants of the DMD method allows to retrieve meaningful flow structures from either experimental or numerical flow data in a purely data-driven (*i.e.*, equation free) manner. Examples of successful application of any of different implementations of the method to a great variety of flow data can be found in.[3,5,6,9,13,20,22]

As we shall review in section 2, DMD strategies rely on the algebraic manipulation of flow data (*snapshots*) acquired at either successive times or spatial stations. Two relevant steps in many DMD variants are, in the first place, the singular value decomposition (SVD) of the input data sequence. The outcome of the SVD step is then leveraged to assemble a matrix eigenvalue problem. Solving the eigenvalue problem provides the dynamic modes and its temporal/spatial behaviour.

As the number and size of flow snapshots increases –*i.e.*, as we attempt to analyze more complex flows– the singular value and the eigenvalue decomposition steps become computationally more expensive. Several alternatives to handle larger data sequences exist. One possible avenue is to exploit parallelism: Sayadi and Schmid introduced in[19] a memory distributed parallel DMD algorithm, which they used to analyze laminar-turbulent transition in a flat plate. Alternatively, Grenga *et al.* [9] described another parallel DMD algorithm, with which they analyzed a turbulent combustion problem.

Another family of strategies relies on reducing, in a principled manner, the size of the input data sequence. Randomized DMD (rDMD) algorithms (see *e.g.*[1,7]) exploit the principles of stochastic nature described in[11] to compute a near-optimal low-rank DMD decomposition of the input data sequence. Alternatively, Guéniat *et al.* [10] propose a variant of DMD –the Non-Uniform DMD or NU-DMD algorithm– that attempts to represent the input data sequence at a reduced cost, in certain optimal sense. The NU-DMD algorithm brings two potential opportunities to reduce the computing cost: the first, is the possibility to represent the temporal dynamics of the data sequence with a reduced number of snapshots. The second opportunity comes from the possibility of identifying degrees of freedom (DoFs) that represent the dynamics of many other DoFs: that is, several DoFs are *agglomerated* into that one that, in an optimal

DOI: 10.13009/EUCASS2019-111

sense, represents them. In[10] , these points are singled out by application of the K-means algorithm.

In this contribution, we focus on the NU-DMD algorithm of Guéniat *et al.* . Specifically, we will investigate the effect of considering alternative agglomeration strategies. Additionally, we present a novel DMD strategy, which we term θ-DMD. The performance of this strategy is evaluated in the case of snapshots non-uniformly separated in time. The effect of considering different agglomeration strategies with θ-DMD is also analysed.

The performance of the different DMD variants considered is evaluated on three different test cases: on a synthetic system proposed in,[10] on the two-dimensional $Re_D = 60$ flow around a cylinder[13] and on the turbulent channel flow at $Re_\tau \approx 200$.[8]

This article is organized as follows: next section describes the different DMD algorithms considered. Section 3 presents then the performance of the different DMD strategies applied. Finally, section 4 discusses the conclusions attained.

2. Dynamic Mode Decomposition Strategies

We proceed now to present the different DMD based feature detection strategies considered in this work. In order to ease the exposition, we introduce first some concepts that will assist during the upcoming subsections. Consider a sequence of instantaneous flow fields, or data *snapshots* $\mathbf{v}(t_j)$, indexed from 1 to N. Very often, these snapshots are solutions of a (possibly) non-linear system given by:

$$\dot{\mathbf{v}}(t) = \mathbf{f}\left(\mathbf{v}(t), t\right). \tag{1}$$

Whatever the origin of the snapshots, a *data matrix* can be assembled from them:

$$\mathbf{V}_1^N = \{\mathbf{v}(t_1), \mathbf{v}(t_2), .., \mathbf{v}(t_N)\} \in \mathbb{R}^{n_p \times N}. \tag{2}$$

Two successive snapshots are separated by the (non-necessarily uniform) time interval Δt_j such that $t_{j+1} = t_j + \Delta t_j$ for all $j = 1, .., N-1$. In the case of linear stability analysis and within the exponential growth region, it is possible to define a linear operator $\mathbf{A}$ (*i.e.*, a numerical approximation of the linearized Navier–Stokes operator) such that $\mathbf{v}(t_{j+1}) = \mathbf{A}\,\mathbf{v}(t_j)$. For non-linear systems, $\mathbf{A}$ approximates the Koopman operator. Eq. (2) can then be rewritten as a Krylov sequence (see[18]):

$$\mathbf{V}_1^N = \{\mathbf{v}(t_1), \mathbf{A}\mathbf{v}(t_1), .., \mathbf{A}^{N-1}\mathbf{v}(t_1)\}. \tag{3}$$

Note that Eq. (3) can be equated to Eq. (2):

$$\mathbf{A}\{\mathbf{v}(t_1), \mathbf{v}(t_2), .., \mathbf{v}(t_{N-1})\} = \{\mathbf{v}(t_2), \mathbf{v}(t_3), .., \mathbf{v}(t_N)\}, \tag{4}$$

which in matrix form writes as:

$$\mathbf{A}\mathbf{V}_1^{N-1} = \mathbf{V}_2^N. \tag{5}$$

2.1 Dynamic mode decomposition

Let us discuss first the well established DMD algorithm, for which several equivalente implementations exist. Here, we follow mainly[12,21] .The DMD based analysis of data sequence in Eq. 2 begins with the Singular Value Decomposition (SVD) of the subsequence $\mathbf{V}_1^{N-1} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^H$. Superscript H denotes conjugate transposition, matrix $\mathbf{\Sigma}$ is a diagonal matrix with entries σ_i the singular values, and the left singular vectors – the columns of $\mathbf{U}$ – can be related to the POD modes of the input data sequence[16] . Inserting the SVD of the snapshot matrix into Eq. (5) yields $\mathbf{A}\mathbf{U}\mathbf{\Sigma}\mathbf{W}^H = \mathbf{V}_2^N$. Some algebra leads to the reduced matrix $\widetilde{\mathbf{A}} = \mathbf{U}^H\mathbf{A}\mathbf{U}$, the projection of the matrix $\mathbf{A}$ onto the space contained in $\mathbf{U}$.

Both the reduced DMD modes $\mathbf{y}_i$ and the associated eigenvalues/Ritz values μ_i (*i.e.*, growth rates $Re(\mu_i)$ and frequencies $Im(\mu_i)$ mapped to the unit circle) are obtained by solving the eigenvalue problem $\widetilde{\mathbf{A}}\mathbf{y}_i = \mu_i\mathbf{y}_i$. Approximated eigenmodes of the matrix $\mathbf{A}$ are recovered through projection onto the original space, *i.e.* $\boldsymbol{\phi}_i = \mathbf{U}\mathbf{y}_i$. Provided $\Delta t_j = \Delta t$, the growth rates and frequencies in the complex half-plane are given by as $\lambda_i = \log(\mu_i)/\Delta t$.

The DMD decomposition allows to reconstruct the snapshots as:

$$\mathbf{v}(t_j) = \sum_{i=1}^{N-1} \alpha_i\, \boldsymbol{\phi}_i\, \mu_i^k = \mathbf{\Phi}\,\mathbf{D}_\alpha\mathbf{V}_{and}, \tag{6}$$

where $\mathbf{D}_\alpha$ is the diagonal matrix with the amplitudes α_i and $\mathbf{V}_{and}$ is a Vandermonde matrix built using the Ritz values. Note that the amplitudes are identified through a minimization problem in the Fröbenius norm, as detailed in.[12]

DOI: 10.13009/EUCASS2019-111

2.2 Non-Uniform Dynamic Mode Decomposition

We present now the Non-Uniform DMD (NU-DMD) strategy of Guéniat *et al.*[10] This approach allows to process data sequences with snapshots non-equally separated in time, *i.e.* $\Delta t_j = \Delta_k$ if $j \neq k$ in general. Departing from Eq. 6, one can account for non-uniform temporal spacing by grouping together the amplitudes α_i and the temporal dependence $\mu_i^j = e^{\lambda_i j \cdot \Delta t}$ into another quantity $\nu_i^{t_j} \equiv \alpha_i e^{\lambda_i (t_j - t_1)}$. Additionally, assume that $N_r \leq N-1$ modes are retained. In that case, one obtains:

$$v(t_j) \approx \boldsymbol{\phi}_1 \, \nu_1^{t_j} + \boldsymbol{\phi}_2 \, \nu_2^{t_j} + \ldots + \boldsymbol{\phi}_{N_r} \, \nu_{N_r}^{t_j}., \quad \forall t_j \in \mathbb{R}. \tag{7}$$

Equation above can be recast in matrix form as $\mathbf{V}_1^{N-1} \approx \mathbf{M}\mathbf{N} + \mathbf{R}_{es}$. Note then that one is interested in those ν_is that minimize the residual matrix $\mathbf{R}_{es}$, *i.e.*:

$$\mathbf{R}_{es} = \mathbf{V}_1^{N-1}(\mathbf{I} - \mathbf{N}^{+}\mathbf{N}). \tag{8}$$

According to[10] , an effective strategy to solve the previous problem is to select a priori N_r and then minimize:

$$\nu \in \underset{\widetilde{\nu} \in \mathbb{C}^{N_r}}{arg\,min} \left\| \mathbf{V}_1^{N-1}(\mathbf{I} - \mathbf{N}(\widetilde{\nu})^{+}\mathbf{N}(\widetilde{\nu}) \right\|_F . \tag{9}$$

Actually, since unitary matrices do not affect the minimization problem above, it is computationally more efficient to perform the QR decomposition of the data sequence, $\mathbf{V}_1^{N-1} \overset{QR}{=} \mathbf{Q}\,\mathbf{R}$, and then minimize:

$$\nu \in \underset{\widetilde{\nu} \in \mathbb{C}^{N_r}}{arg\,min} \left\| \mathbf{R}(\mathbf{I} - \mathbf{N}(\widetilde{\nu})^{+}\mathbf{N}(\widetilde{\nu}) \right\|_F \tag{10}$$

using derivative-free Nelder-Mead method. The dynamic modes are given through the Moore-Penrose pseudo-inverse, $\mathbf{M} \approx \mathbf{V}_1^{N-1}\mathbf{N}^{+}$. Finally, note how if $N_r = N-1$ and $\Delta j = \Delta t$ in Eq. 7 gives back the standard DMD method.

2.3 θ-Dynamic Mode Decomposition

In this section, we introduce an alternative DMD strategy also capable of handling snapshots that are not equiseparated in time. This novel strategy actually departs from the θ-like discretization (hence the name θ-DMD) of the linearized version of Eq. 1, namely:

$$\frac{\mathbf{v}_{j+1} - \mathbf{v}_j}{\Delta t_j} = \mathcal{A}\Big((1-\theta)\,\mathbf{v}_{j+1} + \theta\,\mathbf{v}_j \Big), \tag{11}$$

where $\mathcal{A}$ is the Jacobian matrix of the forcing term $\mathbf{f}$. Trivial algebra leads to:

$$\Big(1 - (1-\theta)\,\Delta t_j\,\mathcal{A}\Big)\,\mathbf{v}_{j+1} = \Big(1 + \theta\,\Delta t_j\,\mathcal{A}\Big)\,\mathbf{v}_j, \tag{12}$$

and to the matrix counterpart:

$$\underbrace{(\mathbf{I} - (1-\theta)\,\mathbf{D}_2\mathcal{A})}_{\mathbf{B}}\,\mathbf{V}_2^N = \underbrace{(\mathbf{I} + \theta\mathbf{D}_1\mathcal{A})}_{\mathbf{C}}\,\mathbf{V}_1^{N-1}, \tag{13}$$

where matrix $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrix whose non-zero entries are non-uniform temporal steps with multiples of θ and $1-\theta$, respectively. The θ-DMD method as stated in Eq. 13, requires the inversion of operator $\mathbf{B}$. In principle, this is possible provided that the dynamics described by Eq. 1 is stable. However, the actual implementation of the method employed in this work avoids the inversion operation by posing the following relation:

$$\mathbf{Y} = \mathbf{T}\,\mathbf{X}, \tag{14}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are those matrices whose columns are, respectively, $\mathbf{x_j} = (1-\theta)\mathbf{v}_{j+1} + \theta\mathbf{v}_j$ and $\mathbf{y_j} = (\mathbf{v}_{j+1} - \mathbf{v}_j) \cdot \Delta t_j{}^{-1}$, for $j = 1, \ldots, N$. From there on, the DMD algorithm is applied as usual. We will use $\theta = \frac{1}{2}$ throughout this work. Finally, note how setting $\theta = 1$ and $\Delta t_j = \Delta t$ in Eq. 14 gives back the standard DMD method.

DOI: 10.13009/EUCASS2019-111

2.4 Agglomeration strategies

Guéniat *et al.* introduced in[10] the K-means algorithm as an agglomeration strategy, with the objective of finding, in a principled manner, space-decimated subsets of the input data-sequence whereas maintaining dominant temporal information. In this contribution, we will study the effect of the K-means strategy in combination with the three DMD strategies considered.
Additionally, we will consider alternatives to the K-means algorithm, provided by the numpy/scipy packages. First, the mini-Batch K-means[15] which uses random sampling in every iteration to compute the distance function; this makes the algorithm more efficient than the K-means, at least when dealing with large clustering problems.
To have a wider view of the agglomeration strategies, we consider another two widespread clustering algorithms. The first one, HDBSCAN, uses unsupervised learning to find clusters in the dataset. It's a hierarchical clustering algorithm which is stable at varying density clusters.[2] Agglomerative Clustering[15] is also used, which is similar to HDBSCAN but more effective when dealing with multi-scale datasets.

3. Results

3.1 Synthetic system with set frequency and growth rate

We perform a first comparison of the three different DMD methods described in section 2 by considering the *synthetic system* introduced in Duke *et al.*[6] . This system consists of spatio-temporal data, given by:

$$u(x,t) = u_0(1+\epsilon)sin(2\pi\kappa x - \omega t)exp(\sigma t + \gamma x). \tag{15}$$

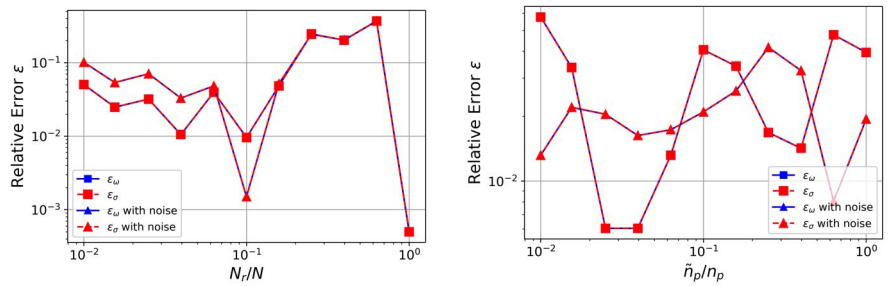
where the frequency $f \equiv \omega/2\pi$ and both spatial γ and temporal σ growth rates can be chosen. We consider, following,[10] angular frequency $\omega = 20$ and temporal growth rate $\sigma = 0.75$. The wavenumber κ, the initial amplitude u_0 and the spatial growth rate γ are all set to 1. Then, $n_p = 2001$ equispaced points in $[0, 2]$ and $n_S = 2000$ equiseparated samples in the time interval $[0, 1]$ are considered. The robustness of the decomposition methods is probed by introducing white multiplicative noise $\epsilon \sim U([-1, 1])$, in an attempt to mimic the deviation in the real data from an experiment. In this work we have considered error levels $max\,|\epsilon|$ to 0.0% and 5.0%. The corresponding results, for both DMD and θ-DMD, are shown in Fig. 1.

From Fig. 1, we observe how for a *compressed* number of snapshots –*i.e.*, a case where $N_r/N < 1$–, the θ-DMD shows a uniform behaviour of the relative error ε when changing the number of snapshots (N_r/N, both logarithm scaled). This behavior is relatively robust to the noise introduced; actually the capture of the growth rate σ is seen to improve in the highly compressed cases. Regarding the standard DMD, its behavior with N_r/N is erratic at first, and stabilizes at a relatively high error level. However, the introduction of noise has barely any influence to the effect of DMD. The θ-DMD results show a great robustness against noise, and only at high reduction rates an effect of the noise is observed. The K-means clustering algorithm (as implemented in SciPy) has been employed to group similarly behaved spatial points; these results are discussed when commenting Fig. 2.

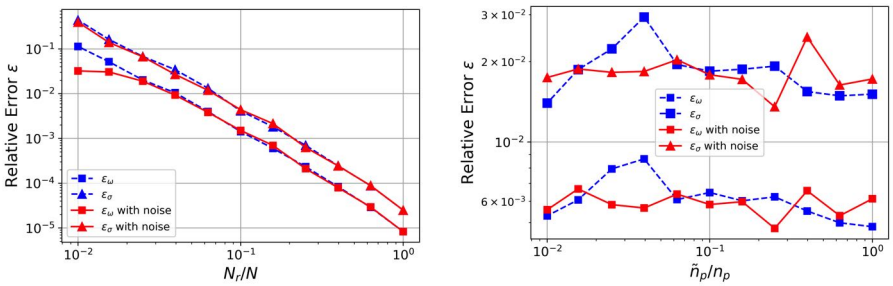
Regarding NU-DMD, we just consider the relative error distribution on the synthetic system without noise. However, we take the opportunity to evaluate the error behaviour with yet another clustering algorithm, the mini-Batch K-means. These results are summarized in Fig. 2.
Comparing Figures 1 and 2, one observes how the error behavior with temporal compression is worse than that of DMD or θ-DMD. This is specially so for the error in growth rate: when $N_r/N = 10^{-2}$ the error level is simply unacceptable. In the reduced cases, the capture of the growth rate σ are more unstable with even unacceptable relative errors on some cases. The introduction of mini-Batch K-means algorithm decreases slightly both the frequency and growth rate errors. However, no obvious effects on the monotonicity are retrieved. In general, θ-DMD provides higher accuracy and less erratic error trends for both temporal compression and spatial reduction.

We take the opportunity to perform, in combination with NU-DMD, a more extensive testing of the agglomeration strategies. As described before, K-means, mini-Batch K-means, HDBSCAN and Agglomerative Clustering have been considered. The test has been perform on a data subsequence spanning $5T$ time units, where T is the dominant period related to the main frequency $f = 2\pi/\omega = \pi/10$ with 50% randomly compression and progressive reduction. The number of the snapshots retained is $N_r = 200$ and the initial points are still kept the same.
From Fig. 3, we observe that HDBSCAN shows good capturing ability compared to the K-means, which with very less number of snapshots for this case compared to that of above. When changing of the size of the snapshots, the relative

DOI: 10.13009/EUCASS2019-111

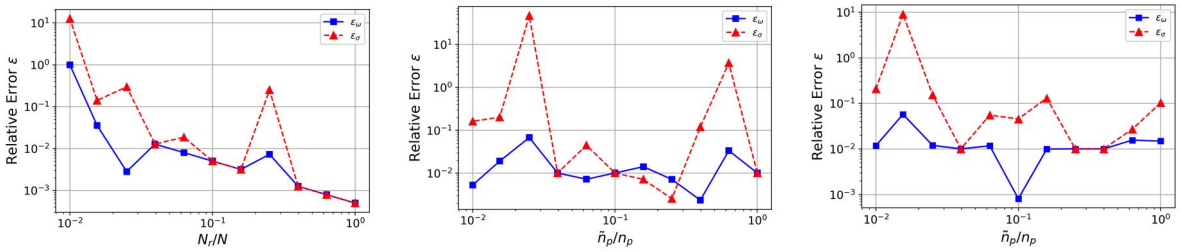


(a) Standard DMD, randomly sampled snapshots (b) Standard DMD, K-means



(c) θ-DMD, randomly sampled snapshots (d) θ-DMD, K-means

Figure 1: DMD and θ-DMD results. Relative errors $\varepsilon_f = \left|f - \hat{f}\right| / \hat{f}$ and $\varepsilon_\sigma = |\sigma - \hat{\sigma}| / \hat{\sigma}$ committed on capturing the natural frequency $\hat{f}$ and growth rate $\hat{\sigma}$, with temporal compression $N_r/N < 1$ and spatial reduction/agglomeration $n'_p/n_p < 1$.



(a) Randomly sampled snapshots (b) K-means (c) mini-Batch K-means

Figure 2: NU-DMD results, with temporal compression and spatial reduction/agglomeration.

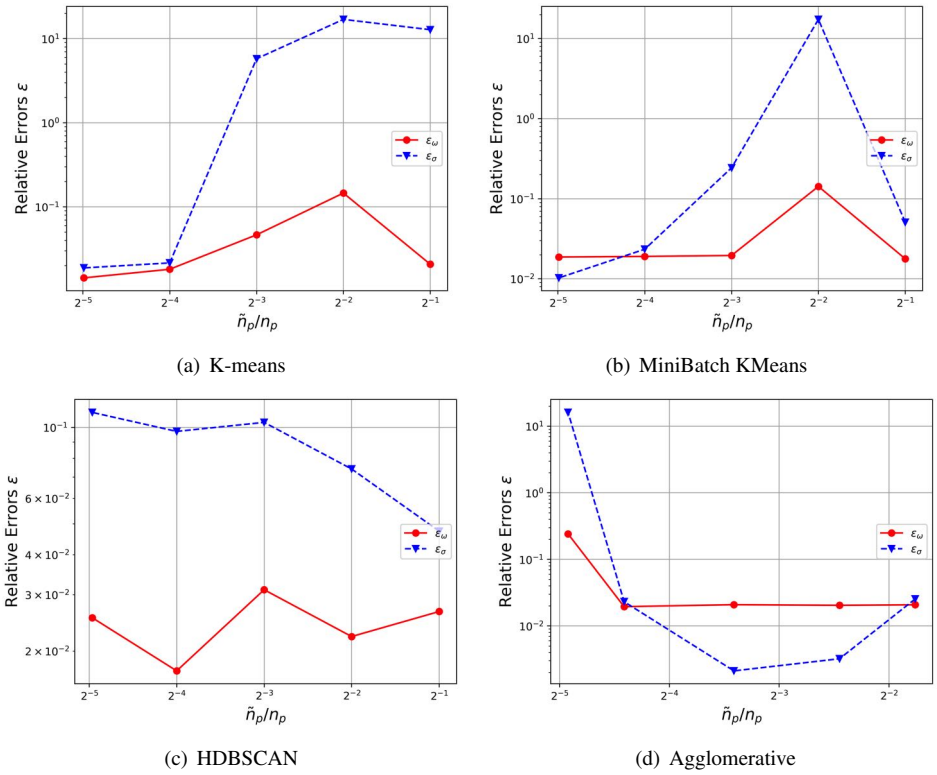
error ϵ showed peaks with 1/4 reduction in the mini-Batch K-means algorithm strategy, and with 1/32 reduction in the Agglomerative Clustering strategy. However, the performances of K-means strategies both showed a shape decline after the reduction of 1/4. This phenomenon seems to indicate that for different kinds of techniques, and also for the different scale of dynamic systems, the agglomerative strategies may need more detailed investigation inside for the performance and robustness in complicated problems.

3.2 Two-dimensional cylinder flow at $Re = 60$

The 2D flow around the cross-section of an infinitely long cylinder is a typical configuration which has been extensively studied by the fluid dynamics community. DMD practitioners are no exception, see *e.g.*[4,13] . Since this flow presents a Hopf bifurcation occurs at $Re_D \approx 46$, conditions slightly above this critical value are interesting: they evince a laminar yet rich behaviour, including unsteady vortex shedding, see Fig. 4. In this work we consider, following[13] a $Re_D = 60$ flow.

Temporally equispaced snapshots corresponding to a 2D $Re_D = 60$ flow around a cylinder have been generated by the DLR TAU solver, a 2^{nd} order finite volume flow solver. For this data sequence, $\Delta t = 0.0025$ and the total non-dimensional simulation time is $T = 6.0$. The size (*i.e.*, the number of DoFs) of the snapshots is $n_p = 36474$, and the

DOI: 10.13009/EUCASS2019-111



(a) K-means (b) MiniBatch KMeans

(c) HDBSCAN (d) Agglomerative

Figure 3: Relative error of NU-DMD with four agglomerative strategies

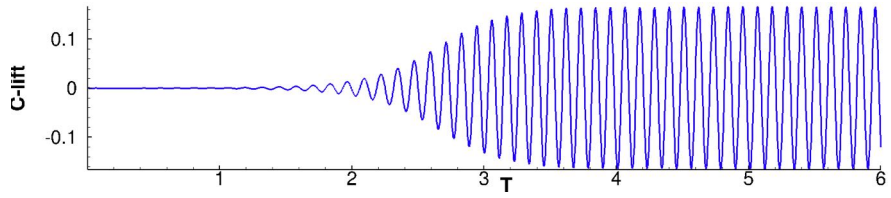


Figure 4: Evolution of lift coefficient C_l from equilibrium steady solution to the limit cycle.

total number of snapshots is 2400. The limit circle stage ($t \in [4.0, 6.0]$) is considered in the analyses to be presented next; this implies that $N = 800$ snapshots were finally selected.

3.2.1 Preliminary study of the cylinder flow using fast Fourier transform

First of all, fast Fourier transform (FFT) was applied to the flow parameters and the flow field data in an attempt to identify *significant* frequencies present in the data. We will focus on these frequencies when we perform the different DMD analyses. As visible in Fig. 5, three relevant frequencies are identified, $f_{main} = 9.02,\ 18.05,\ 27.07$. As it is evident, these three frequencies are, practically, integer multiples of the lower one.

3.2.2 Standard DMD analysis

A standard DMD analysis was applied to the data sequence. Those modes oscillating at frequencies being an (approximately) integer multiple of the fundamental one where selected and classified in decreasing order of the moduli $\|\alpha\|_i$s: 15 DMD modes were selected, see their corresponding frequencies, growth ratios and spatial in Fig. 6.

From the 15 dominant dynamic modes, that indexed as 0 and with frequency $f = 0$ represents time average of the data sequence. The main frequency mode pair (indexes 1-2) shows a growth rate close to zero, which implies it has a persistent nature in time (quasi-stability). Double and triple frequency modes (pairs 3-4 and 12-13) are also found.

DOI: 10.13009/EUCASS2019-111

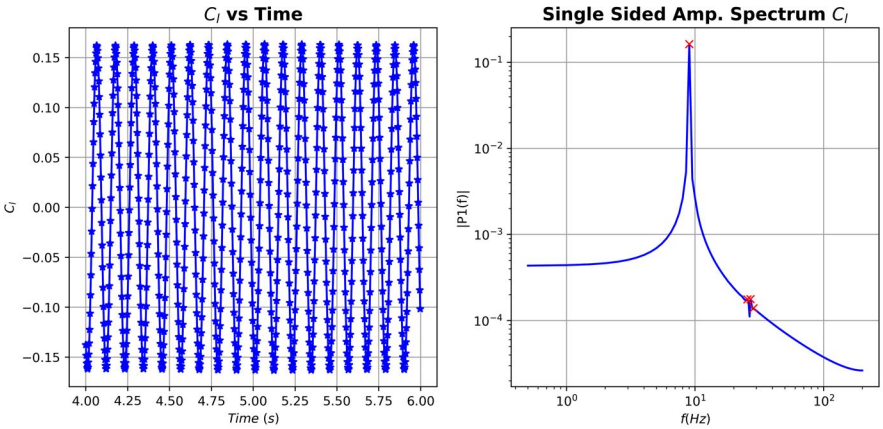
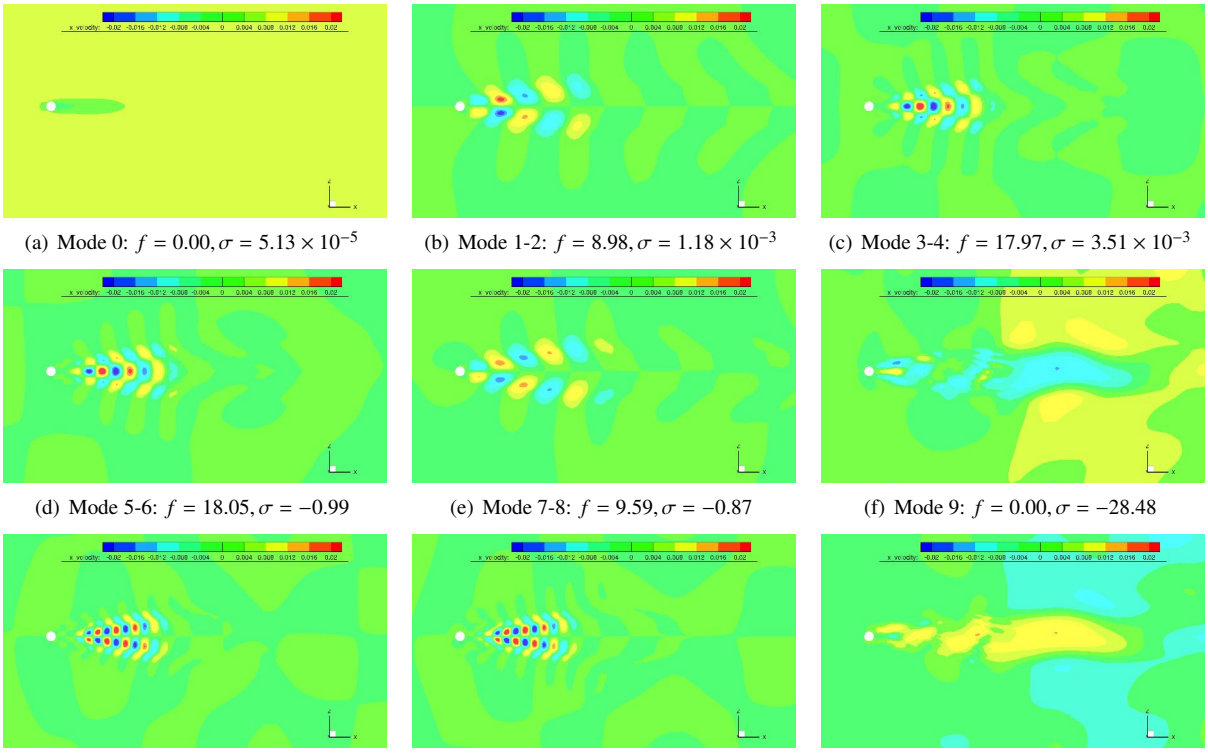


Figure 5: C_l *vs* time and associated FFT spectrum.



(a) Mode 0: $f = 0.00, \sigma = 5.13 \times 10^{-5}$ (b) Mode 1-2: $f = 8.98, \sigma = 1.18 \times 10^{-3}$ (c) Mode 3-4: $f = 17.97, \sigma = 3.51 \times 10^{-3}$

(d) Mode 5-6: $f = 18.05, \sigma = -0.99$ (e) Mode 7-8: $f = 9.59, \sigma = -0.87$ (f) Mode 9: $f = 0.00, \sigma = -28.48$

(g) Mode 10-11: $f = 27.14, \sigma = -1.75$ (h) Mode 12-13: $f = 26.95, \sigma = 1.61 \times 10^{-3}$ (i) Mode 14: $f = 0.0, \sigma = -15.85$

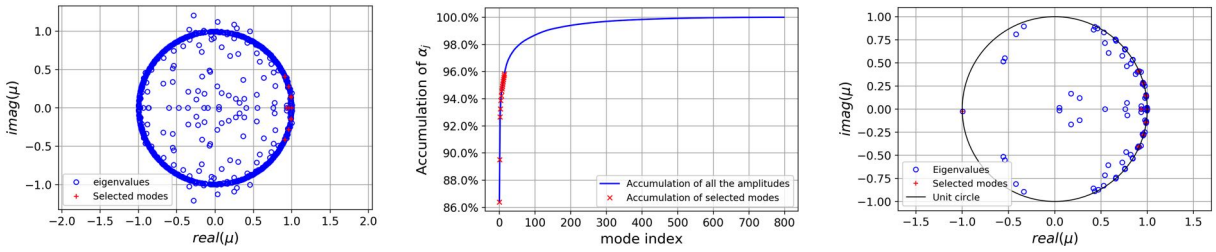
Figure 6: The 15 main DMD modes and their corresponding frequencies and growth rates.

Fig. 7 represents both the distribution of 15 modes studied in the complex plane μ and the percentage of the accumulation of amplitudes over the 799 amplitudes retrieved by DMD: observe how these 15 accumulate almost 96% of the total sum of amplitude moduli.

3.2.3 NU-DMD analysis

An analysis similar as described for the standard DMD is also conducted with the NU-DMD method. We instruct the optimizer to retrieve 100 modes, and from those we identify the most relevant $N_{md} = 15$ dynamic modes and eigenvalues. One word of caution: as the eigenvalues retrieved by NU-DMD come from an optimization process, slight asymmetries about the real axis exist. The dominant modes appear by complex-conjugated pairs, but their frequencies and growth rates may present slightly differences. These asymmetries come from the finite precision with which the Nelder-Mead algorithm operates. In order to provide clearer view, we will treat similar mode pairs as one pair and, for the sake of

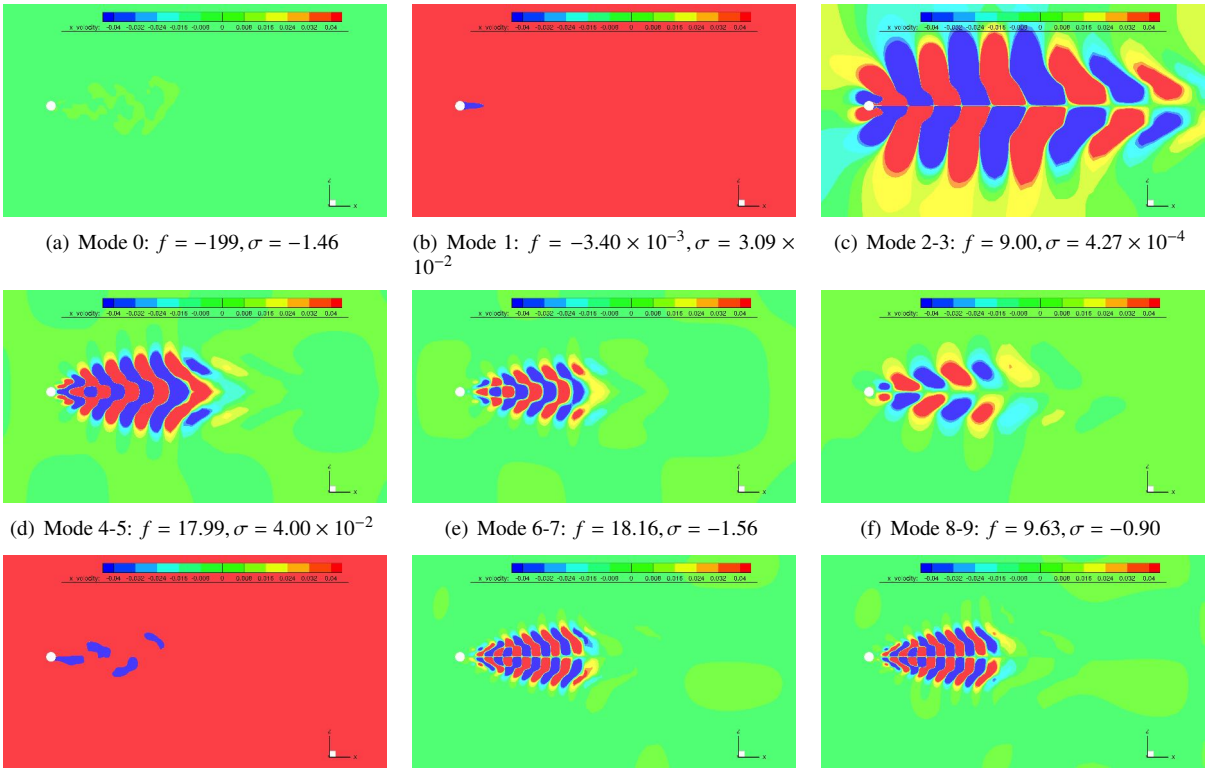
DOI: 10.13009/EUCASS2019-111



(a) The eigenvalues by DMD (b) The accumulation of amplitude α by DMD (c) The eigenvalues by NU-DMD

Figure 7: The spectrum of 15 main modes of 2D cylinder flow by DMD and NU-DMD

brevity, show one of them if they are apparently coupled.



(a) Mode 0: $f = -199, \sigma = -1.46$ (b) Mode 1: $f = -3.40 \times 10^{-3}, \sigma = 3.09 \times 10^{-2}$ (c) Mode 2-3: $f = 9.00, \sigma = 4.27 \times 10^{-4}$

(d) Mode 4-5: $f = 17.99, \sigma = 4.00 \times 10^{-2}$ (e) Mode 6-7: $f = 18.16, \sigma = -1.56$ (f) Mode 8-9: $f = 9.63, \sigma = -0.90$

(g) Mode 10: $f = -7.32 \times 10^{-5}, \sigma = -27.22$ (h) Mode 11-12: $f = 27.17, \sigma = -2.75$ (i) Mode 13-14: $f = 26.99, \sigma = -0.19$

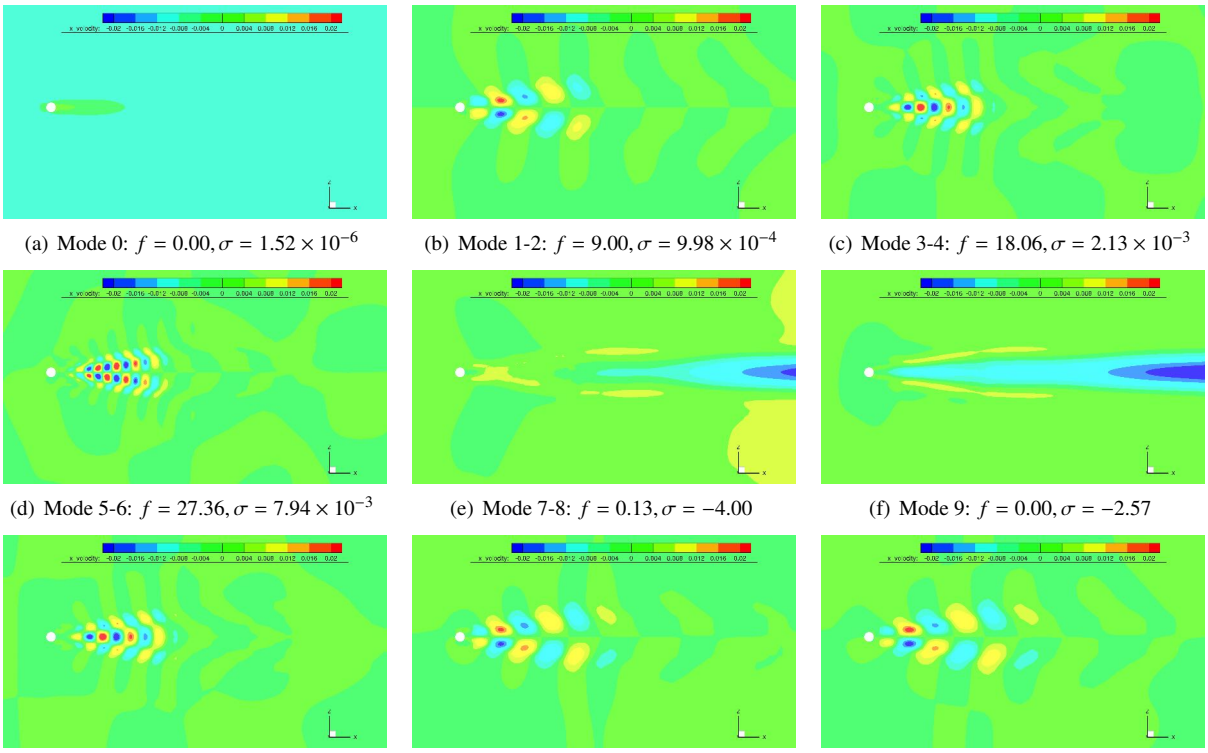
Figure 8: The 15 main modes and their corresponding frequencies and growth rates by NU-DMD.

The dominant modes produced by NU-DMD, see Fig. 8, also capture the average mode (index 1, with nearly zero frequency and growth rate), the main frequency mode (index 2), the double frequency mode (indexes 4-5) and the triple frequency mode (but with negative growth rate). Note that NU-DMD, the mode ordering is less well established than for standard DMD: for the specific frequencies we are focusing on, NU-DMD may not extract precisely the modes we want first (*e.g.*, the mode 0) or it may capture modes with similar physical characteristics, *i.e. repeated*.
From the eigenvalue spectrum, Fig. 7(c), we observe that NU-DMD captured a mode whose eigenvalue close to the left peak of the unit radius circle (corresponding to mode 0). The modes remaining are quite similar to those provided by DMD, with NU-DMD modes 1-14 being similar to modes 0-13 of DMD.

3.2.4 θ-DMD analysis

The θ-DMD algorithm is employed to conduct an analysis equivalent to that described in 3.2.2. Again, the 15 main modes with their physical characteristics are shown in Fig. 9. And the corresponding λ distribution compared with that of DMD and the amplitude accumulation curve in fig. 10.

DOI: 10.13009/EUCASS2019-111



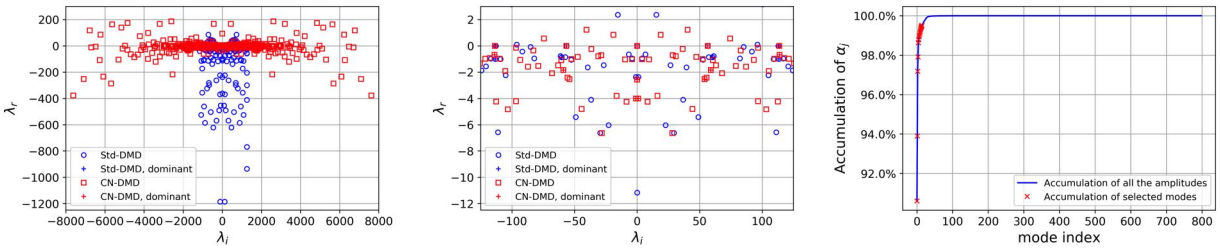
(a) Mode 0: $f = 0.00, \sigma = 1.52 \times 10^{-6}$ (b) Mode 1-2: $f = 9.00, \sigma = 9.98 \times 10^{-4}$ (c) Mode 3-4: $f = 18.06, \sigma = 2.13 \times 10^{-3}$

(d) Mode 5-6: $f = 27.36, \sigma = 7.94 \times 10^{-3}$ (e) Mode 7-8: $f = 0.13, \sigma = -4.00$ (f) Mode 9: $f = 0.00, \sigma = -2.57$

(g) Mode 10-11: $f = 18.13, \sigma = -0.69$ (h) Mode 12-13: $f = 9.39, \sigma = -1.83$ (i) Mode 14-15: $f = 8.59, \sigma = -2.53$

Figure 9: The 15 main modes and their corresponding frequencies and growth rates by θ-DMD.

The first mode directly corresponds to the averaged flow; the next three mode pairs oscillate at three dominant frequencies identified by the preliminary FFT analysis. The overall λ distribution shows the broadband characteristics compared to DMD and the dominant modes tally well with that of DMDin the partial view. The accumulation curve of the amplitude α, shows how those 4 mode pairs provide 98%, demonstrating the beneficial capturing ability of the θ-DMD method.



(a) The λ distribution compared to DMD (b) Partial view of (a) (c) The accumulation of amplitude α

Figure 10: The spectrum of 15 main modes of 2D cylinder flow by θ-DMD.

3.2.5 Investigation of temporal compression and spatial clustering

In this section we compare the sensitivity and robustness of the methods described in sections 2.1-2.3 to both temporal compression and spatial agglomeration. In order to provide a comparison that is simultaneously as complete and as far as possible, we proceed as follows: for both NU-DMD and θ-DMD, a variable fraction (from 100% to 1%) of the 800 snapshots in the data sequence are chosen. A sequence of snapshots of the same length is fed to the standard DMD method, but in that case the snapshots are chosen equidistributed in time. Regarding the spatial clustering, the K-means algorithm is applied at agglomeration ratios n'_p/n_p also ranging from 100% to 1%.
In order to evaluate the behaviour with temporal compression and spatial clustering, we propose an indicator accounting for the 5 most common frequencies identified by FFT (the continuous signal and integer multiples of $f = 9.02$). The sum

DOI: 10.13009/EUCASS2019-111

of the relative error of every target frequency was calculated as the indicator of the performance of each decomposition methods. For the zero frequency mode, the error is:

$$E_0 = min(\left|\hat{f}_0 - f_j\right|), \quad j = 0, 1, \cdots, N_r, \tag{16}$$

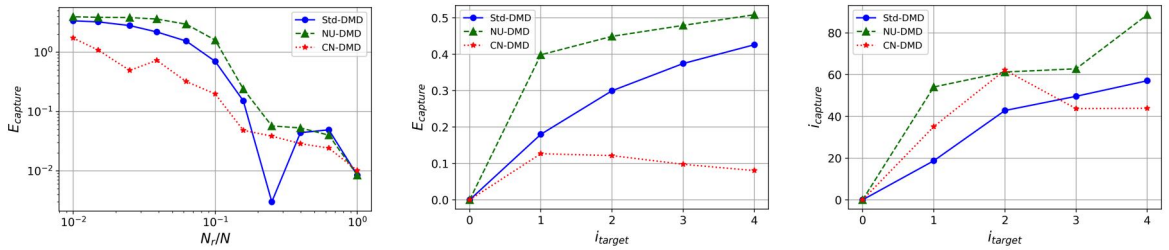
whereas for the other target frequencies, the error is:

$$E_i = min(\left|\hat{f}_i - f_j\right| / \hat{f}_i), \quad i = 1, 2, \cdots, 4, \;\; j = 0, 1, \cdots, N_r. \tag{17}$$

Therefore, the overall indicator is:

$$E_{cap} = \sum_{i=0}^{4} E_i = min(\left|\hat{f}_0 - f_j\right|) + \sum_{i=1}^{4} min(\left|\hat{f}_i - f_j\right| / \hat{f}_i). \tag{18}$$

In previous equation, $\hat{f}_i$, $(i = 0, 1, \cdots, 4)$ are the target frequencies we want to capture. Moreover, the average of every target frequency relative error and the corresponding mode index were also collected to find the detailed effectiveness on every capturing of the five targets in different case conditions.



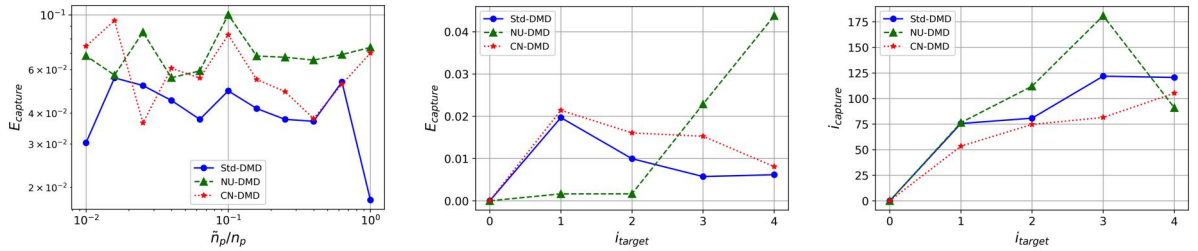
(a) Relative errors of capturing target frequencies over compressing snapshots

(b) Relative errors of on capturing every target frequencies

(c) The index of mode on capturing every target frequencies

Figure 11: Compressed cases of randomly sampled snapshots (with $n_p = 36474$)

From Fig. 11(a), one observes how the error indicator retrieved by the θ-DMD algorithm when increasing the temporal compression is, systematically, the lowest. Fig. 11(b) shows how θ-DMD has the average error indicator is the lowest for every target frequency over all the cases. Finally, in Fig. 11(c) one sees how the DMD method is the one identifying earlier the first three target modes; θ-DMD method identifies faster the 4^{th} and 5^{th} target frequencies, on average.



(a) Relative errors of capturing target frequencies with clustered snapshots

(b) Relative errors on capturing every target frequencies

(c) The index of mode on capturing every target frequencies

Figure 12: Compressed cases of clustered snapshot size (with $N_r = 200$)

Regarding the spatial agglomeration effects, if one considers $N_r = 200$ snapshots (1/4 to the original, see Fig. 12) one observes how DMD gives better performance almost systematically, and is also the fastest in identifying the five target frequencies on average. Surprisingly, NU-DMD works much better on the first three modes. Regarding the indexes of the target modes captured, the θ-DMD methods provides the earliest identification.

3.3 The turbulent channel flow

We finally investigate the performance of the different DMD methods considered when applied to a turbulent channel flow database. This turbulent database has been computed by an incompressible DNS solver;[14] its characteristics are summarized in the following Table 1. In total, 1200 flow snapshots were stored. These snapshots are separated in time by $\Delta t = 0.0122$, which corresponds to $\Delta t^+ \approx 0.1$.

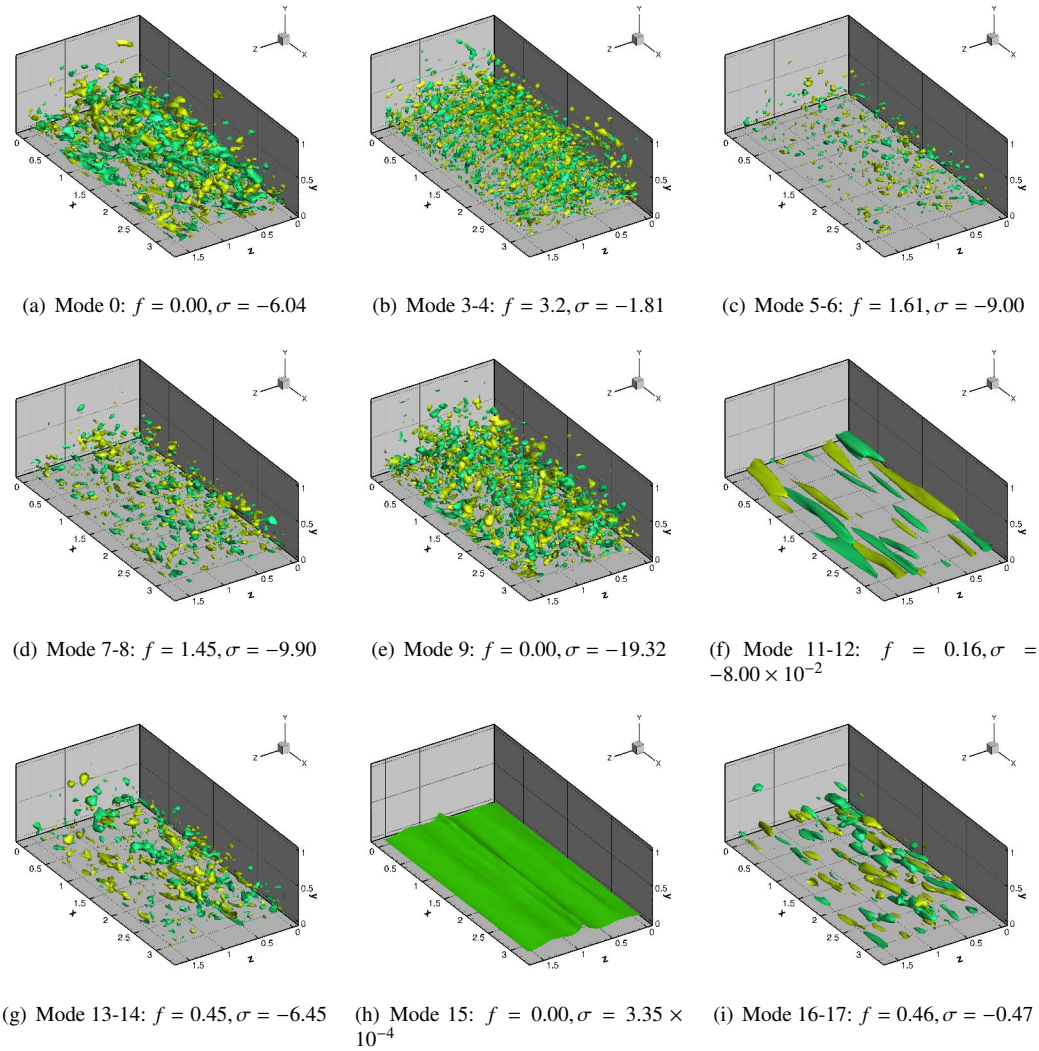
DOI: 10.13009/EUCASS2019-111

Table 1: Channel flow database descriptions.

L_x/δ	L_x/δ	L_x/δ	n_x	n_y	n_z	Δx^+	Δy^+_{min}	Δz^+	Re_c	Δt^s
2π	2	π	192	129	192	1.64×10^{-4}	0.95	8.20×10^{-5}	3678.7	0.122

3.3.1 Standard DMD analysis

In the first place, a standard DMD based analysis of the channel flow was conducted on the channel. For this analysis, we focus on several representative modes; we report their corresponding frequencies and growth rates in Fig. 13. The DMD structures are visualized using iso-surfaces of the streamwise velocity ($u = \pm 0.05$).



(a) Mode 0: $f = 0.00, \sigma = -6.04$ (b) Mode 3-4: $f = 3.2, \sigma = -1.81$ (c) Mode 5-6: $f = 1.61, \sigma = -9.00$

(d) Mode 7-8: $f = 1.45, \sigma = -9.90$ (e) Mode 9: $f = 0.00, \sigma = -19.32$ (f) Mode 11-12: $f = 0.16, \sigma = -8.00 \times 10^{-2}$

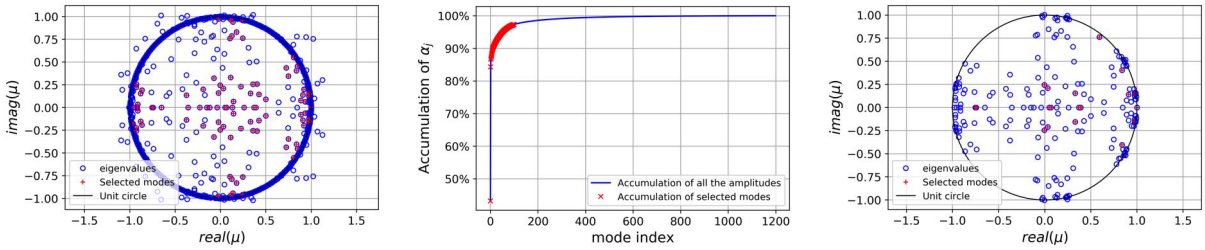
(g) Mode 13-14: $f = 0.45, \sigma = -6.45$ (h) Mode 15: $f = 0.00, \sigma = 3.35 \times 10^{-4}$ (i) Mode 16-17: $f = 0.46, \sigma = -0.47$

Figure 13: The 15 main DMD modes and their corresponding frequencies and growth rates

In Fig. 13, we present the modes with largest amplitudes. For frequencies $f = 0.45$, $f = 1.45$, $f = 1.61$ and $f = 3.2$, the spatial mode structure is relevant mainly on the middle region of the flow domain. Mode pair $11 - 12$, with frequency $f = 0.16$, shows an elongated structure. Regarding the eigenvalues spectrum and the cumulative amplitudes sum, many of the relevant modes are located inside the unit radius circle (*e.g.*, they decrease in time), see Figs. 13; the three of the modes at zero frequency provide about 86%.

The distribution of the eigenvalues of the modes considered and the percentage of the sum of amplitudes over all are both showed in Figs. 13 and 14.

DOI: 10.13009/EUCASS2019-111

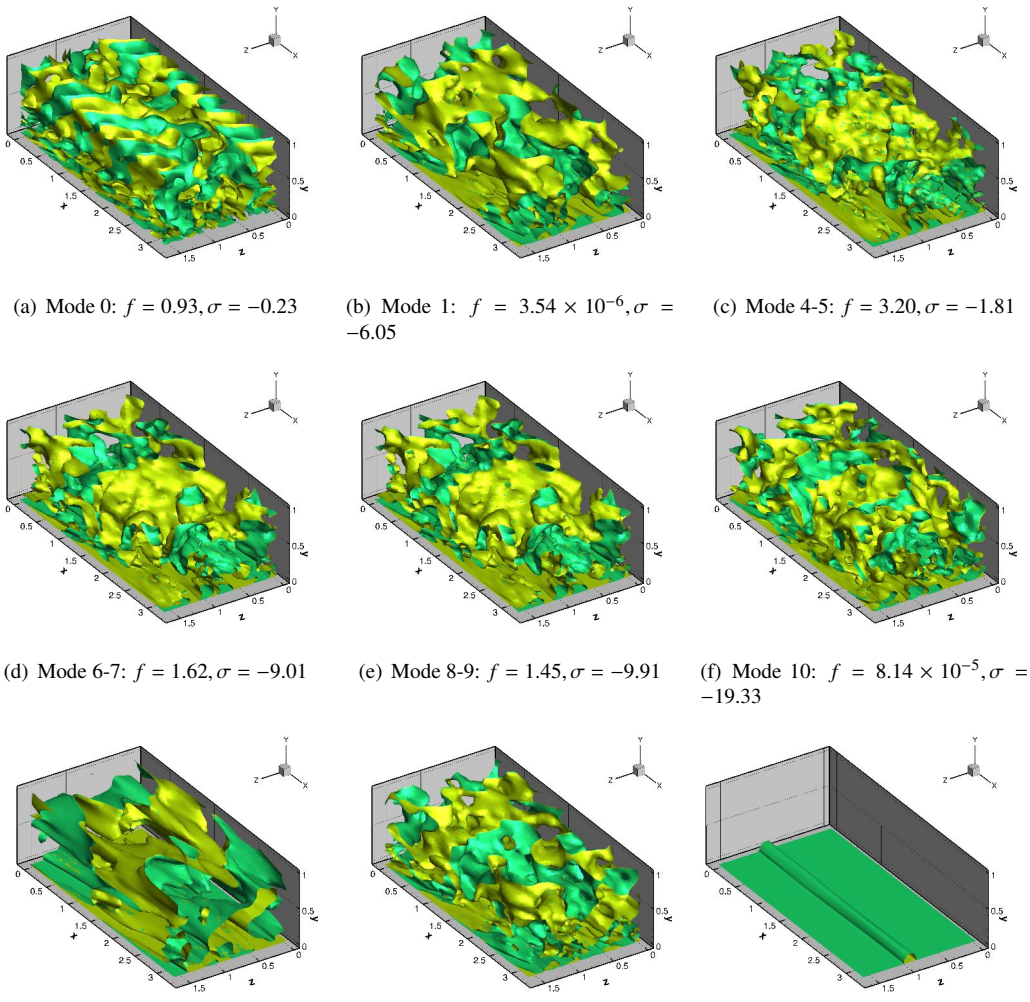


(a) The eigenvalues by DMD (b) The accumulation of amplitude α by DMD (c) The eigenvalues by NU-DMD

Figure 14: The 31 main DMD modes modes of channel flow

3.3.2 NU-DMD analysis

A similar analysis based on NU-DMD has also been conducted. In this case, the algorithm has been set to extract 100 modes. Fig. 15 shows the most relevant modes, together with their associated frequencies and temporal growth/decrease rates. Also for the eigenvalues distribution in Fig. 14, which show similarity for top 100 modes and also the 31 modes marked.



(a) Mode 0: $f = 0.93, \sigma = -0.23$ (b) Mode 1: $f = 3.54 \times 10^{-6}, \sigma = -6.05$ (c) Mode 4-5: $f = 3.20, \sigma = -1.81$

(d) Mode 6-7: $f = 1.62, \sigma = -9.01$ (e) Mode 8-9: $f = 1.45, \sigma = -9.91$ (f) Mode 10: $f = 8.14 \times 10^{-5}, \sigma = -19.33$

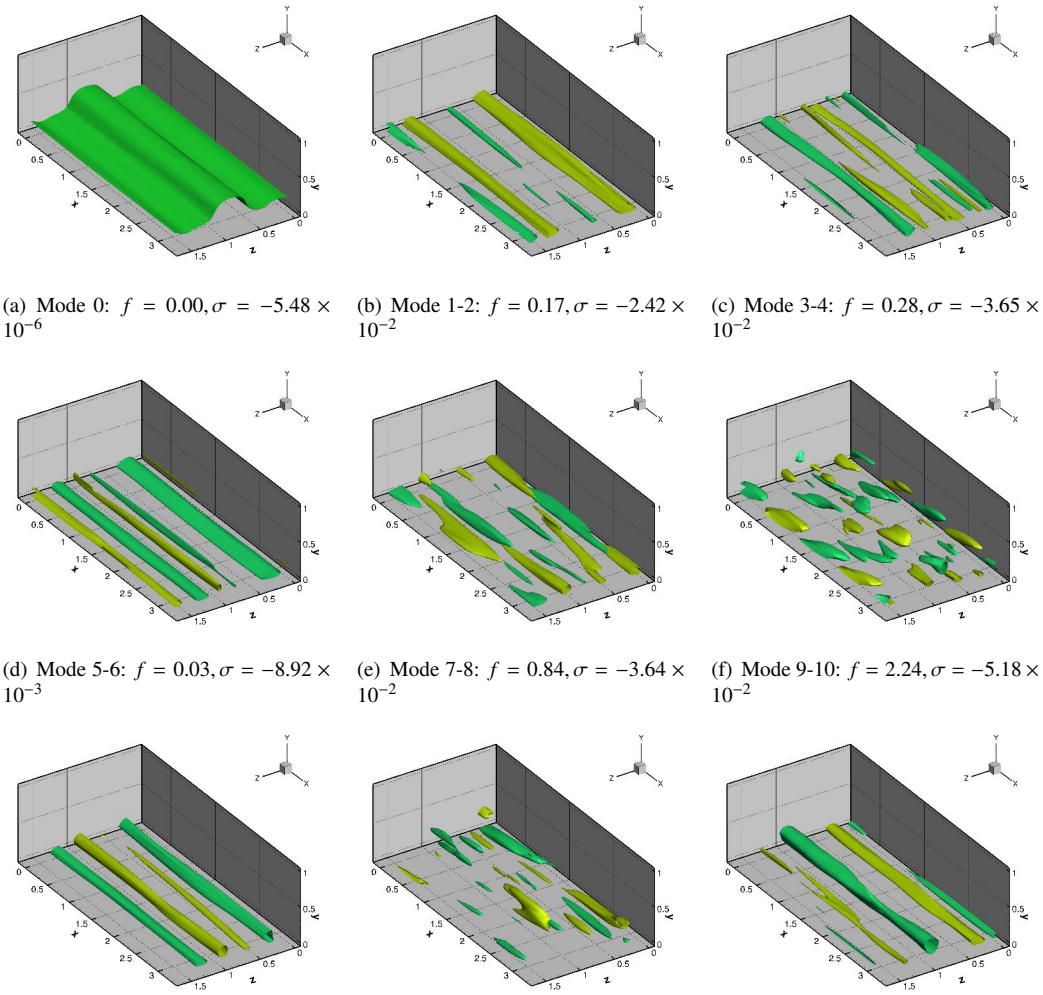
(g) Mode 12-13: $f = 0.16, \sigma = -8.08 \times 10^{-5}$ (h) Mode 14-15: $f = 0.45, \sigma = -6.45$ (i) Mode 16: $f = 3.30 \times 10^{-6}, \sigma = 2.54 \times 10^{-3}$

Figure 15: The 14 main DMD modes and their corresponding frequencies and growth rates

DOI: 10.13009/EUCASS2019-111

3.3.3 θ-DMD analysis

Finally, a θ-DMD-based analysis is also conducted for the turbulent channel flow. Fig. 16 presents the 17 most relevant modes. The corresponding λ spectrum (compared with DMD) and the cumulative moduli sum are shown in Fig. 17.



(a) Mode 0: $f = 0.00, \sigma = -5.48 \times 10^{-6}$

(b) Mode 1-2: $f = 0.17, \sigma = -2.42 \times 10^{-2}$

(c) Mode 3-4: $f = 0.28, \sigma = -3.65 \times 10^{-2}$

(d) Mode 5-6: $f = 0.03, \sigma = -8.92 \times 10^{-3}$

(e) Mode 7-8: $f = 0.84, \sigma = -3.64 \times 10^{-2}$

(f) Mode 9-10: $f = 2.24, \sigma = -5.18 \times 10^{-2}$

(g) Mode 11-12: $f = 0.20, \sigma = -1.01 \times 10^{-2}$

(h) Mode 13-14: $f = 1.54, \sigma = -3.25 \times 10^{-2}$

(i) Mode 15-16: $f = 0.26, \sigma = -8.63 \times 10^{-3}$

Figure 16: The 17 main modes and their corresponding frequencies and growth rates by θ-DMD.

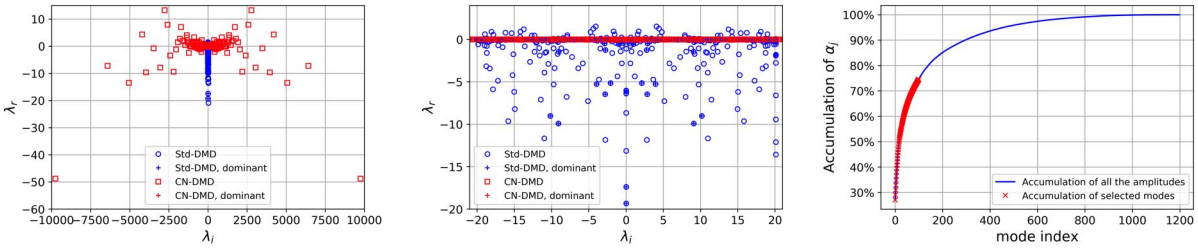
From the modes considered, the average flow is ranked the first. Main frequencies retrieved are $f = 0.03$, 0.17, 0.20, 0.26, 0.28, 0.84, 1.54 and 2.24. Compared to the results by DMD and NU-DMD, θ-DMD identifies more modes with lower frequencies. The modes present well defined spatial structures, and two of them ($f_{mode16-17} = 1.62$ and $f_{mode19-20} = 0.45$) appear at frequencies also identified by DMD and NU-DMD. We observe how modes at lower frequencies present spatial structures elongated and remain close to the wall (see *e.g.*, mode 11-12). Modes at relatively high frequency extend further into the interior of the channel.

The distribution of λ again shows a wider range of the frequency of the modes, instead of limited around the origin point in the complex plane λ. What's more, the accumulation of the amplitudes α showed uniform distribution after the average mode, which give a signal that there are many modes with range of frequencies giving the similar contribution to this flow, and results to the common knowledge of the multi-scale of the turbulent boundary layer flow.

4. Conclusions

In this contribution we have focused on strategies that allow to perform a Dynamic Mode Decomposition analysis of data at a reduced computational cost. We have departed from the the classical DMD method of Schmid[21] and the Non-Uniform DMD method of Guéniat *et al.* [10] Specifically, we have investigated how both *temporal compression*

DOI: 10.13009/EUCASS2019-111



(a) The λ distribution of 31 dominant DMD modes

(b) Partial view of (a)

(c) The accumulation of amplitude α of 31 modes

Figure 17: The spectrum of 31 main modes of turbulent channel flow by θ-DMD.

(*i.e.*, $N_r/N < 1$) and *spatial agglomeration* (*i.e.*, $\widetilde{n}_p/n_p < 1$) –as proposed in[10]– can be leveraged to attain the desired objective of reducing both the computational and memory efforts.
Additionally we have presented what, to the best of our knowledge, is a novel DMD based strategy, the θ-DMD method. This method, as NU-DMD, can process databases acquired at non-uniform temporal sampling rates.
The three DMD based methods considered have been compared with respect to their numerical performance and their capabilities of handle shortened (*i.e.*, compressed), possibly non-uniformly sampled databases and to accommodate spatial agglomeration strategies. The comparison has been made in terms of robustness and ability to identify physical features.
The aforementioned comparison has been conducted on three different test cases:

- In the first synthetic system case, we compared the three decomposition methods in combination with the compressed and agglomerated settings. Sensitivity to noise is also considered. In general, θ-DMD outperforms the other two methods in almost all the conditions considered.
- In the two-dimensional cylinder flow case, the θ-DMD identifies four dominant modes (steady component and three harmonics). When considering temporal compression, θ-DMD works better on the capture process with error in every target frequency. However, in the spatially agglomerated situation, the standard DMD performs relatively better for almost all the reduction levels. In this case situation, θ-DMD shows slightly worse, but comparable, results to DMD.
- Regarding the turbulent channel flow, θ-DMD captures practically all the frequencies identified by DMD and NU-DMD methods and some additional ones. Moreover, θ-DMD provides modes with better defined spatial structures.

In conclusion, DMD methods arise as a useful tool to identify relevant features of complicated flows.

5. Acknowledgments

The support provided by China Scholarship Council (CSC, No.201806320222) during a visit of Mr. Binghua Li to UPM-ETSIAE is acknowledged. And this work is funded by Projects: Drag Reduction via Turbulent Boundary Layer Flow Control (DRAGY, GA-690623), 2016-2019, China-EU Aeronautical Cooperation project, co-funded by Ministry of Industry and Information Technology (MIIT), China, and Directorate -General for Research and Innovation (DG RTD), European Commission.

References

[1] D.A. Bistrian and I.M. Navon. Randomized dynamic mode decomposition for nonintrusive reduced order modelling. *International Journal for Numerical Methods in Engineering*, 112(1):3–25, 2017.

[2] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer, 2013.

[3] A. Cassinelli, M. de Giovanetti, and Y. Hwang. Streak instability in near-wall turbulence revisited. *Journal of Turbulence*, 18(5):443–464, 2017.

[4] K.K. Chen, J.H. Tu, and C.W. Rowley. Variants of Dynamic Mode Decomposition: boundary condition, Koopman, and Fourier analyses. *Journal of Nonlinear Science*, 22(6):887–915, 2012.

DOI: 10.13009/EUCASS2019-111

[5] S. Le Clainche, J.M. Vega, and J. Soria. Higher Order Dynamic Mode Decomposition of noisy experimental data: The flow structure of a zero-net-mass-flux jet. *Experimental Thermal and Fluid Science*, 88(Supplement C):336 – 353, 2017.

[6] D. Duke, J. Soria, and D. Honnery. An error analysis of the Dynamic Mode Decomposition. *Exp. Fluids*, 52(2):529–542, 2012.

[7] N.B. Erichson, L. Mathelin, S.L. Brunton, and J.N. Kutz. Randomized Dynamic Mode Decomposition. *ArXiv e-prints*, Sep 2018.

[8] J. Garicano-Mena, B. Li, E. Ferrer, and E. Valero. A composite dynamic mode decomposition method for the data-driven analysis of turbulent channel flows. *submitted to Comput.& Fluids*, 2019.

[9] T. Grenga, J.F. MacArt, and M.E. Mueller. Dynamic Mode Decomposition of a direct numerical simulation of a turbulent premixed planar jet flame: convergence of the modes. *Combustion Theory and Modelling*, 0(0):1–17, 2018.

[10] F. Guéniat, L. Mathelin, and L. R. Pastur. A Dynamic Mode Decomposition approach for large and arbitrarily sampled systems. *Phys. Fluids*, 27(2), 2015.

[11] N. Halko, P. G. Martinsson, and J.A. Tropp. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Rev.*, 53(2):217 – 288, 2011.

[12] M. R. Jovanović, P. J. Schmid, and J. W. Nichols. Sparsity-promoting Dynamic Mode Decomposition. *Phys. Fluids*, 26(2), 2014.

[13] J. Kou and W. Zhang. An improved criterion to select dominant modes from Dynamic Mode Decomposition. *European Journal of Mechanics - B/Fluids*, 62:109 – 129, 2017.

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

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[16] C.W. Rowley and S.T.M. Dawson. Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics*, 49(1):387–417, 2017.

[17] C.W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D.S. Henningson. Spectral analysis of nonlinear flows. *J. Fluid Mech.*, 641:115–127, 12 2009.

[18] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, Manchester, 1992.

[19] T. Sayadi and P.J. Schmid. Parallel data-driven decomposition algorithm for large-scale datasets: with application to transitional boundary layers. *Theor. Comp. Fluid Dyn.*, 30(5):415–428, Oct 2016.

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

[21] P.J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5–28, 2010.

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