Sparsity-promoting dynamic mode decomposition

Mihailo R. Jovanovic, Peter J. Schmid, Joseph W. Nichols

Sparsity-promoting dynamic mode decomposition

Mihailo R. Jovanović,[1,a)] Peter J. Schmid,[2,b)] and Joseph W. Nichols[3,c)]

[1]*Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA*

[2]*Laboratoire d'Hydrodynamique (LadHyX), Ecole Polytechnique, 91128 Palaiseau cedex, France*

[3]*Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, Minnesota 55455, USA*

Dynamic mode decomposition (DMD) represents an effective means for capturing the essential features of numerically or experimentally generated flow fields. In order to achieve a desirable tradeoff between the quality of approximation and the number of modes that are used to approximate the given fields, we develop a sparsity-promoting variant of the standard DMD algorithm. Sparsity is induced by regularizing the least-squares deviation between the matrix of snapshots and the linear combination of DMD modes with an additional term that penalizes the ℓ_1-norm of the vector of DMD amplitudes. The globally optimal solution of the resulting regularized convex optimization problem is computed using the alternating direction method of multipliers, an algorithm well-suited for large problems. Several examples of flow fields resulting from numerical simulations and physical experiments are used to illustrate the effectiveness of the developed method. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4863670]

I. INTRODUCTION

Even though fluid flows are infinite-dimensional systems governed by nonlinear partial differential equations, the essential features of their dynamical responses can often be approximated reliably by models of low complexity. This observation has given rise to the notion of coherent structures – organized fluid elements that, together with dynamic processes, are responsible for the bulk of momentum and energy transfer in the flow. Recent decades have witnessed significant advances in the extraction of coherent structures from data collected in experiments and numerical simulations. Proper orthogonal decomposition (POD) modes,[1,2] global eigenmodes, frequential modes,[3] and balanced modes[4,5] have provided useful insight into the dynamics of fluid flows. Recently, Koopman modes[6–8] and dynamic mode decomposition (DMD)[9] have joined the group of feature extraction techniques. Both POD and DMD are snapshot-based post-processing algorithms which may be applied equally well to data obtained in simulations or in experiments. While POD modes are characterized by spatial orthogonality and multi-frequential temporal content, DMD modes may be non-orthogonal but each of them possesses a single temporal frequency. This lack of non-orthogonality of DMD modes may be essential to capturing important dynamical effects in systems with non-normal dynamical generators.[10–12] For an in-depth discussion of the connection between DMD and other data decomposition methods, we refer the reader to Refs. 8, 9, and 13.

The importance of POD and DMD modes goes beyond identification of coherent structures in fluid flows. In particular, they may be used to obtain models of low complexity; by projecting the full system onto the subspace spanned by the extracted modes, the governing equations may

a)URL: http://umn.edu/home/mihailo/. mihailo@umn.edu
b)URL: http://yakari.polytechnique.fr/people/peter/. peter.schmid@ladhyx.polytechnique.fr
c)URL: http://www.aem.umn.edu/people/faculty/bio/nichols.shtml. jwn@umn.edu

be approximated by a dynamical system with fewer number of degrees of freedom. This facilitates computationally tractable study of flow stability and receptivity as well as a model-based control design. In many situations, however, it is challenging to identify a subset of modes that have the strongest impact on the flow dynamics. For example, spatial non-orthogonality of the DMD modes may introduce a poor quality of approximation of experimentally or numerically generated snapshots when only a subset of modes with the largest amplitude is retained. Recent attempts at extracting only a subset of desired frequencies and spatial profiles rely on formulation of non-convex optimization problems. In Ref. 14, a variant of DMD, referred to as the Optimized DMD, was introduced; determining a solution to this problem in general requires an intractable combinatorial search. In Refs. 15 and 16, a gradient-based algorithm was employed to simultaneously search for the low-rank basis and the matrix that governs temporal evolution on a lower-dimensional subspace in order to reduce the least-squares residual achieved by DMD.

In this paper, we develop a sparsity-promoting variant of the standard DMD algorithm. This algorithm is aimed at achieving a desirable tradeoff between the quality of approximation (in the least-squares sense) and the number of modes that are used to approximate numerical or experimental snapshot sequences. To achieve this objective, we combine tools and ideas from convex optimization[17] with the emerging area of compressive sensing.[18–20] Our approach to inducing sparsity relies on regularization of the least-squares deviation (between the matrix of snapshots and the linear combination of DMD modes) with an additional term that penalizes the ℓ_1-norm of the vector of DMD amplitudes. The ℓ_1-norm can be interpreted as a convex relaxation of the non-convex cardinality function,[17] and it has been effectively used as a proxy for promoting sparsity in a number of applications.[20–25] The alternating direction method of multipliers (ADMM) – a state-of-the-art algorithm for solving large-scale and distributed optimization problems[24] – is then employed to solve the resulting convex optimization problem and to efficiently compute the globally optimal solution. Since this algorithm alternates between promoting sparsity and minimizing the least-squares residual, we exploit the respective structures of the underlying penalty functions to decompose the optimization problem into easily solvable modules. In particular, we show that the least-squares minimization step amounts to solving an unconstrained regularized quadratic program and that sparsity is promoted through the application of a convenient soft-thresholding operator. After a desirable tradeoff between the quality of approximation and the number of DMD modes has been achieved, for the fixed sparsity structure resulting from the ADMM algorithm we compute the optimal amplitudes of the retained dynamic modes as the solution to the constrained quadratic program. This "polishing step" improves the quality of approximation of numerically or experimentally generated snapshots and provides optimal amplitudes of the dominant DMD modes.

Our presentation is organized as follows. In Sec. II, we formulate the problem, provide a brief overview of the dynamic mode decomposition, and address the optimal selection of amplitudes of extracted DMD modes. In Sec. III, we apply a sparsity-promoting framework to select a subset of DMD modes which strikes a user-defined balance between the approximation error (with respect to the full data sequence) and the number of extracted modes. In Sec. IV, we use three databases resulting from the two-dimensional linearized Navier–Stokes equations for plane Poiseuille flow, the unstructured large-eddy simulation (LES) of a supersonic jet, and the time-resolved particle image velocimetry (TR-PIV) experiment of a flow through a cylinder bundle to illustrate the utility of the developed method. We conclude our presentation in Sec. V with a summary of our contributions and an outlook for future research directions, and relegate algorithmic developments to the appendixes.

II. PROBLEM FORMULATION

A. Dynamic mode decomposition

The DMD is a data processing algorithm that extracts coherent structures with a single temporal frequency from a numerical or experimental data-sequence.[9] In what follows, we briefly outline the key steps of DMD.

We begin by collecting a sequence of snapshots from numerical simulations or physical experiments and form a data matrix whose columns represent the individual data samples. Even though

we confine our attention to temporal evolution processes, the DMD-framework can accommodate a variety of "evolution coordinates" (e.g., spatial directions, or curved base-flow streamlines).[9] Furthermore, we assume that the data are equispaced in time, with a time step Δt,

$$\{\psi_0, \psi_1, \ldots, \psi_N\},$$

where each $\psi_i := \psi(i\,\Delta t)$ is, in general, a complex vector with M components (measurement points), i.e., $\psi_i \in \mathbb{C}^M$.

Next, we form two data matrices from the snapshot sequence

$$\Psi_0 := \begin{bmatrix} \psi_0\ \psi_1\ \cdots\ \psi_{N-1} \end{bmatrix} \in \mathbb{C}^{M \times N},$$

$$\Psi_1 := \begin{bmatrix} \psi_1\ \psi_2\ \cdots\ \psi_N \end{bmatrix} \in \mathbb{C}^{M \times N},$$

and postulate that the snapshots have been generated by a discrete-time linear time-invariant system

$$\psi_{t+1} = A\,\psi_t, \quad t = \{0, \ldots, N-1\}. \tag{1}$$

For fluid flows, the matrix A typically contains a large number of entries (which are in general complex numbers). The dynamic mode decomposition furnishes a procedure for determining a low-order representation of the matrix $A \in \mathbb{C}^{M \times M}$ that captures the dynamics inherent in the data sequence. In fluid problems, the number of components (measurement points) in each snapshot ψ_i is typically much larger than the number of snapshots, $M \gg N$, thereby implying that Ψ_0 and Ψ_1 are tall rectangular matrices. Using the linear relation (1) between the snapshots at two consecutive time steps, we can link the two data matrices Ψ_0 and Ψ_1 via the matrix A and express Ψ_1 as

$$\begin{aligned} \Psi_1 &= \begin{bmatrix} \psi_1\ \psi_2\ \cdots\ \psi_N \end{bmatrix} \\ &= \begin{bmatrix} A\,\psi_0 & A\,\psi_1 & \cdots & A\,\psi_{N-1} \end{bmatrix} \\ &= A\,\Psi_0. \end{aligned} \tag{2}$$

For a rank-r matrix of snapshots Ψ_0, the DMD algorithm provides an optimal representation $F \in \mathbb{C}^{r \times r}$ of the matrix A in the basis spanned by the POD modes of Ψ_0,

$$A \approx U\,F\,U^*.$$

Here, U^* denotes the complex-conjugate-transpose of the matrix of POD modes U which is obtained from an economy-size singular value decomposition (SVD) of $\Psi_0 \in \mathbb{C}^{M \times N}$,

$$\Psi_0 = U\,\Sigma\,V^*,$$

where Σ is an $r \times r$ diagonal matrix with non-zero singular values $\{\sigma_1, \ldots, \sigma_r\}$ on its main diagonal, and

$$U \in \mathbb{C}^{M \times r} \text{ with } U^* U = I,$$

$$V \in \mathbb{C}^{r \times N} \text{ with } V^* V = I.$$

The matrix F can be determined from the matrices of snapshots Ψ_0 and Ψ_1 by minimizing the Frobenius norm of the difference between Ψ_1 and $A\,\Psi_0$ with $A = U F U^*$ and $\Psi_0 = U\,\Sigma\,V^*$,

$$\underset{F}{\text{minimize}} \quad \|\Psi_1 - U\,F\,\Sigma\,V^*\|_F^2, \tag{3}$$

where the Frobenius norm of the matrix Q is determined by

$$\|Q\|_F^2 = \text{trace}\left(Q^* Q\right) = \text{trace}\left(Q\,Q^*\right).$$

It is straightforward to show that the optimal solution to (3) is determined by

$$F_{\text{dmd}} = U^*\,\Psi_1\,V\,\Sigma^{-1}.$$

This expression is identical to the expression provided in Ref. 9 and it concludes the implementation of the DMD algorithm, starting from matrices of data snapshots Ψ_0 and Ψ_1. For a discussion about

the relation between F_{dmd} and the companion form matrix A_c – which provides a representation of the matrix A on an N-dimensional subspace of $\mathbb{C}^M$ that is spanned by the columns of Ψ_0 – we refer the reader to Ref. 9.

B. Optimal amplitudes of DMD modes

The matrix $F_{dmd} \in \mathbb{C}^{r \times r}$ determines an optimal low-dimensional representation of the inter-snapshot mapping $A \in \mathbb{C}^{M \times M}$ on the subspace spanned by the POD modes of Ψ_0. The dynamics on this r-dimensional subspace are governed by

$$x_{t+1} = F_{dmd}\, x_t, \tag{4}$$

and the matrix of POD modes U can be used to map x_t into a higher dimensional space $\mathbb{C}^M$,

$$\psi_t \approx U\, x_t.$$

If F_{dmd} has a full set of linearly independent eigenvectors $\{y_1, \ldots, y_r\}$, with corresponding eigenvalues $\{\mu_1, \ldots, \mu_r\}$, then it can be brought into a diagonal coordinate form,

$$F_{dmd} = \underbrace{\begin{bmatrix} y_1 & \cdots & y_r \end{bmatrix}}_{Y} \underbrace{\begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_r \end{bmatrix}}_{D_\mu} \underbrace{\begin{bmatrix} z_1^* \\ \vdots \\ z_r^* \end{bmatrix}}_{Z^*}.$$

Here, each y_j is of unit length, $y_j^* y_j = 1$, and $\{z_1, \ldots, z_r\}$ are the eigenvectors of F_{dmd}^*, corresponding to the eigenvalues $\{\bar{\mu}_1, \ldots, \bar{\mu}_r\}$, which are suitably scaled so that the following bi-orthogonality condition holds:

$$z_i^* y_j = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Now, the solution to (4) is determined by

$$x_t = Y\, D_\mu^t\, Z^* x_0 = \sum_{i=1}^{r} y_i\, \mu_i^t\, z_i^* x_0 = \sum_{i=1}^{r} y_i\, \mu_i^t\, \alpha_i,$$

where $\alpha_i := z_i^* x_0$ represents the ith modal contribution of the initial condition x_0. We can thus approximate experimental or numerical snapshots using a linear combination of the DMD modes, $\phi_i := U y_i$,

$$\psi_t \approx U\, x_t = \sum_{i=1}^{r} \phi_i\, \mu_i^t\, \alpha_i, \quad t \in \{0, \ldots, N-1\}, \tag{5}$$

and each α_i can be interpreted as the "amplitude" of the corresponding DMD mode.[9] It is worth noting that the selection of the amplitudes α_i can be interpreted as the selection of the Koopman modes that have the strongest influence on the system's response resulting from the use of the particular initial condition and the time interval on which the snapshots are collected; cf. Eq. (5) with Eq. (2.5) in Ref. 7. Equivalently, in matrix form, we have

$$\underbrace{\begin{bmatrix} \psi_0 & \psi_1 & \cdots & \psi_{N-1} \end{bmatrix}}_{\Psi_0} \approx \underbrace{\begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_r \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_r \end{bmatrix}}_{D_\alpha \; := \; \mathrm{diag}\{\alpha\}} \underbrace{\begin{bmatrix} 1 & \mu_1 & \cdots & \mu_1^{N-1} \\ 1 & \mu_2 & \cdots & \mu_2^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mu_r & \cdots & \mu_r^{N-1} \end{bmatrix}}_{V_{and}},$$

which demonstrates that the temporal evolution of the dynamic modes is governed by the Vandermonde matrix $V_{and} \in \mathbb{C}^{r \times N}$. This matrix is determined by the r complex eigenvalues μ_i of

F_{dmd} which contain information about the underlying temporal frequencies and growth/decay rates.

Determination of the unknown vector of amplitudes $\alpha := \begin{bmatrix} \alpha_1 & \cdots & \alpha_r \end{bmatrix}^T$ then amounts to finding the solution to the following optimization problem:

$$\underset{\alpha}{\mathrm{minimize}} \quad \|\Psi_0 \, - \, \Phi \, D_\alpha \, V_{\mathrm{and}}\|_F^2 .$$

Using the economy-size SVD of $\Psi_0 = U \, \Sigma \, V^*$ and the definition of the matrix $\Phi := U \, Y$, we bring this problem into the following form:

$$\underset{\alpha}{\mathrm{minimize}} \quad J(\alpha) := \|\Sigma \, V^* \, - \, Y \, D_\alpha \, V_{\mathrm{and}}\|_F^2 , \tag{6}$$

which is a convex optimization problem that can be solved using standard methods.[17,26] We note that this optimization problem does not require access to the POD modes of the matrix of snapshots Ψ_0; the problem data in (6) are the matrices Σ and V, which are obtained from the economy-size SVD of Ψ_0, and the matrices Y and V_{and}, which result from the eigenvalue decomposition of F_{dmd}.

In Appendix A, we show that the objective function $J(\alpha)$ in (6) can be equivalently represented as

$$J(\alpha) = \alpha^* P \, \alpha \, - \, q^* \alpha \, - \, \alpha^* q \, + \, s , \tag{7}$$

where

$$P := \left(Y^* \, Y\right) \circ \left(\overline{V_{\mathrm{and}} \, V_{\mathrm{and}}^*}\right), \quad q := \overline{\mathrm{diag}\,(V_{\mathrm{and}} \, V \, \Sigma^* \, Y)}, \quad s := \mathrm{trace}\,(\Sigma^* \Sigma) .$$

Here, an asterisk denotes the complex-conjugate-transpose of a vector (matrix), an overline signifies the complex-conjugate of a vector (matrix), diag of a vector is a diagonal matrix with its main diagonal determined by the elements of a given vector, diag of a matrix is a vector determined by the main diagonal of a given matrix, and $\circ$ is the elementwise multiplication of two matrices. The optimal vector of DMD amplitudes that solves the optimization problem (6) can thus be obtained by minimizing the quadratic function (7) with respect to α,

$$\alpha_{\mathrm{dmd}} = P^{-1} q = \left(\left(Y^* \, Y\right) \circ \left(\overline{V_{\mathrm{and}} \, V_{\mathrm{and}}^*}\right)\right)^{-1} \overline{\mathrm{diag}\,(V_{\mathrm{and}} \, V \, \Sigma^* \, Y)} .$$

A superposition of all DMD modes, properly weighted by their amplitudes and advanced in time according to their temporal growth/decay rate, optimally approximates the entire data sequence. The key challenge that this paper addresses is the identification of a low-dimensional representation in order to capture the most important dynamic structures (by eliminating features that contribute weakly to the data sequence).

III. SPARSITY-PROMOTING DYNAMIC MODE DECOMPOSITION

In this section, we direct our attention to the problem of selecting the subset of DMD modes that has the most profound influence on the quality of approximation of a given sequence of snapshots. In other words, we are interested in a hierarchical description of the data sequence in terms of a set of dynamic modes. Our approach to sparsity-promoting dynamic mode decomposition consists of two steps:

- In the first step, we seek a *sparsity structure* that achieves a user-defined tradeoff between the number of extracted modes and the approximation error (with respect to the full data sequence); see (9) below;
- In the second step, we *fix* the sparsity structure of the vector of amplitudes (identified in the first step) and determine the optimal values of the non-zero amplitudes; see (10) below.

We approach the problem of inducing sparsity by augmenting the objective function $J(\alpha)$ in (6) with an additional term, **card**(α), that penalizes the number of non-zero elements in the vector of unknown amplitudes α,

$$\underset{\alpha}{\mathrm{minimize}} \quad J(\alpha) \, + \, \gamma \, \mathbf{card}\,(\alpha) . \tag{8}$$

In the modified optimization problem (8), γ is a positive regularization parameter that reflects our emphasis on sparsity of the vector $\alpha \in \mathbb{C}^r$. Larger values of γ place stronger emphasis on the number of non-zero elements in the vector α (relative to the quality of the least-squares approximation, $J(\alpha)$), thereby encouraging sparser solutions to (8).

In general, finding a solution to the problem (8) amounts to a combinatorial search that quickly becomes intractable for any problem of interest. To circumvent this issue, we introduce a relaxed version of (8) by replacing the cardinality function with the ℓ_1-norm of the vector α,

$$\underset{\alpha}{\text{minimize}} \quad J(\alpha) \,+\, \gamma \sum_{i=1}^{r} |\alpha_i|. \tag{9}$$

The sparsity-promoting DMD problem (9) is a *convex optimization problem* whose global solution, for small and medium sizes, can be obtained using standard optimization solvers.[17,26] In Sec. III A, we develop an efficient algorithm for solving (9). This algorithm utilizes the ADMM, a state-of-the-art method for solving large-scale and distributed optimization problems.[24]

After a desired balance between the quality of approximation of experimental or numerical snapshots and the number of DMD modes is achieved, we *fix the sparsity structure* of the unknown vector of amplitudes and determine only the non-zero amplitudes as the solution to the following constrained convex optimization problem:

$$\begin{aligned} \underset{\alpha}{\text{minimize}} \quad & J(\alpha) \\ \text{subject to} \quad & E^T \alpha \,=\, 0. \end{aligned} \tag{10}$$

In this expression, the matrix $E \in \mathbb{R}^{r \times m}$ encodes information about the sparsity structure of the vector α. The columns of E are the unit vectors in $\mathbb{R}^r$ whose non-zero elements correspond to zero components of α. For example, for $\alpha \in \mathbb{C}^4$ with

$$\alpha \,=\, \begin{bmatrix} \alpha_1 & 0 & \alpha_3 & 0 \end{bmatrix}^T,$$

the matrix E is given as

$$E \,=\, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

An efficient algorithm for solving (10) is provided in Appendix C.

In summary, optimization problem (9) identifies location of non-zero entries in the vector of amplitudes and optimization problem (10) adjusts the values of these non-zero entries in order to optimally approximate the entire data sequence.

A. Alternating direction method of multipliers

We next use the alternating direction method of multipliers algorithm to find the globally optimal solution to the sparsity-promoting optimization problem (9),

$$\underset{\alpha}{\text{minimize}} \quad J(\alpha) \,+\, \gamma \, g(\alpha), \tag{11}$$

with $J(\alpha)$ determined by (7) and

$$g(\alpha) \,:=\, \sum_{i=1}^{r} |\alpha_i|.$$

In order to bring the problem into the form that is convenient for the application of ADMM, we need the following two steps:

- **Step 1:** Replace the vector of amplitudes α in the sparsity-promoting term g with a new variable $\beta \in \mathbb{C}^r$,

$$
\begin{aligned}
&\underset{\alpha,\,\beta}{\text{minimize}} \quad J(\alpha) \,+\, \gamma\, g(\beta) \\
&\text{subject to} \quad \alpha \,-\, \beta \,=\, 0.
\end{aligned}
\tag{12}
$$

For any feasible α and β, the optimization problems (11) and (12) are equivalent. Even though the number of optimization variables in (12) is twice as big as in (11), formulation (12) allows us to exploit the respective structures of the quadratic function $J(\alpha)$ and the sparsity-promoting function $g(\beta)$ in the ADMM algorithm, as outlined below.

- **Step 2:** Introduce the augmented Lagrangian,

$$
\mathcal{L}_\rho\,(\alpha,\beta,\lambda) \,:=\, J(\alpha) \,+\, \gamma\, g(\beta) \,+\, \frac{1}{2}\left(\lambda^*(\alpha \,-\, \beta) \,+\, (\alpha \,-\, \beta)^*\lambda \,+\, \rho\,\|\alpha \,-\, \beta\|_2^2\right).
$$

Here, $\lambda \in \mathbb{C}^r$ is the vector of Lagrange multipliers, ρ is a positive parameter that introduces a quadratic penalty on the deviation between α and β, and $\|\cdot\|_2$ is the Euclidean norm of a given vector. For $\rho = 0$, $\mathcal{L}_\rho$ simplifies to the standard Lagrangian associated with the optimization problem (12). For $\rho \neq 0$, the additional quadratic term in $\mathcal{L}_\rho$ increases curvature of the objective function and extends applicability of the augmented Lagrangian methods to the problems that the standard Lagrangian methods cannot handle. In particular, the augmented Lagrangian methods are guaranteed to converge for convex optimization problems with non-differentiable objective functions that are even allowed to take infinite values.[24]

ADMM is an iterative algorithm for minimization of the augmented Lagrangian that combines desirable features of the dual ascent method with the method of multipliers.[24] It consists of an α-minimization step, a β-minimization step, and a Lagrange multiplier update step

$$
\alpha^{k+1} \,:=\, \underset{\alpha}{\arg\min}\; \mathcal{L}_\rho\left(\alpha, \beta^k, \lambda^k\right),
\tag{13a}
$$

$$
\beta^{k+1} \,:=\, \underset{\beta}{\arg\min}\; \mathcal{L}_\rho\left(\alpha^{k+1}, \beta, \lambda^k\right),
\tag{13b}
$$

$$
\lambda^{k+1} \,:=\, \lambda^k \,+\, \rho\left(\alpha^{k+1} \,-\, \beta^{k+1}\right).
\tag{13c}
$$

Starting with an initial point (β^0, λ^0), the iterations are conducted until the desired feasibility tolerances, ϵ_{prim} and ϵ_{dual}, are met

$$
\|\alpha^{k+1} \,-\, \beta^{k+1}\|_2 \,\leq\, \epsilon_{\text{prim}} \text{ and } \|\beta^{k+1} \,-\, \beta^k\|_2 \,\leq\, \epsilon_{\text{dual}}.
$$

We note that the selection of the positive parameter ρ determines the rate of convergence of the ADMM algorithm. The optimal selection of ρ is a topic of an active research interest; for a class of convex optimization problems, recent progress has been reported in Refs. 27 and 28.

In Appendix B, we utilize the respective structures of the functions J and g in (11) and show that the α-minimization step amounts to solving an unconstrained regularized quadratic program and that the solution to the β-minimization step is obtained through the application of a soft-thresholding operator.

IV. EXAMPLES

In this section, we apply sparsity-promoting DMD to three databases of snapshots. The first database is obtained using a numerically generated sequence of snapshots of the two-dimensional linearized Navier–Stokes equations for plane Poiseuille flow with $Re = 10\,000$, the second database results from an LES of a screeching supersonic rectangular jet, and the third database contains the time-resolved particle image velocimetry data of a flow through a cylinder bundle. In all three examples, the value of the parameter ρ in the ADMM algorithm is set to one, $\rho = 1$.

A. Two-dimensional Poiseuille flow with $Re = 10\,000$

We begin illustration of the sparsity-promoting DMD algorithm by considering the two-dimensional linearized Navier–Stokes equations for plane Poiseuille flow with $Re = 10\,000$ (based on the centerline velocity and the channel half-width); see Figure 1 for geometry. The dynamics of the wall-normal velocity fluctuations with streamwise wavenumber k_x are governed by the Orr-Sommerfeld equation.[29] The discretized version of the linearized dynamics is obtained using a pseudo-spectral scheme with $M = 150$ Chebyshev collocation points in the wall-normal direction,[30] and the flow fluctuations with a single wavenumber $k_x = 1$ are advanced in time using the matrix exponential with time step $\Delta t = 1$ and a randomly generated initial profile (that satisfies both homogeneous Dirichlet and Neumann boundary conditions in the wall-normal direction). After a transient period of ten time-steps, $N = 100$ snapshots are taken to form the snapshot matrices, and we apply the standard DMD algorithm along with its sparsity-promoting variant.

The spectrum of the Orr-Sommerfeld operator (circles) along with the eigenvalues resulting from the standard DMD algorithm (crosses) are shown in Figure 2(a). The rank of the matrix of snapshots Ψ_0 is $r = 26$, and the dependence of the absolute value of the DMD amplitudes α_i on the frequency and the real part of the corresponding DMD eigenvalues μ_i is displayed in Figures 2(b) and 2(c), respectively. As expected, the largest amplitude of DMD modes corresponds to an unstable eigenvalue that generates an exponentially growing Tollmien-Schlichting (TS) wave.

Figure 3 shows the residual $\|\Psi_0 - \Phi \, D_\alpha \, V_{\mathrm{and}}\|_F$ of the optimal vector of amplitudes α, resulting from the sparsity-promoting DMD algorithm, in fraction of $\|\Psi_0\|_F$,

$$\% \, \Pi_{\mathrm{loss}} := 100 \sqrt{\frac{J(\alpha)}{J(0)}} = 100 \frac{\|\Psi_0 - \Phi \, D_\alpha \, V_{\mathrm{and}}\|_F}{\|\Psi_0\|_F},$$

as a function of the number of the retained DMD modes, $N_z := \mathbf{card}\,(\alpha)$. As our emphasis on sparsity increases, a progressively smaller number of non-zero elements in the vector α are obtained and the quality of the least-squares approximation deteriorates. The sparsity-promoting DMD algorithm with 18 modes provides nearly identical performance as the full-rank DMD algorithm. By reducing the number of modes to 13, the least-squares residual gets compromised by only 1.3%. While a reduction in the number of modes from 13 to 12 introduces performance degradation of approximately 3%, the reduction in the number of modes from 12 to 6 degrades performance by only additional 2.5%. Further reduction in the number of DMD modes has a much more profound influence on the quality of approximation of numerically generated snapshots; for example, performance deterioration of almost 10% takes place when the number of DMD modes is reduced from 3 to 2. As we further discuss below, this sharp performance drop may be attributed to the fact that at least one fast, one slow, and one unstable mode has to be selected in order to capture the essential dynamical features of the original data sequence. The results displayed in Figure 3 suggest that a reasonable compromise between the quality of approximation (in the least-squares sense) and the number of modes in the Poiseuille flow example may be achieved with $N_z = 6$ DMD modes.

It is also instructive to assess how the sparsity-promoting procedure selects the eigenvalues identified by the standard DMD algorithm. For decreasing numbers of retained modes, the first row in Figure 4 illustrates this process in the complex plane. The DMD-spectrum for plane Poiseuille flow consists of branches that describe distinct features of the perturbation dynamics: the fast perturbation dynamics in the center of the channel is captured by eigenvalues with phase velocities
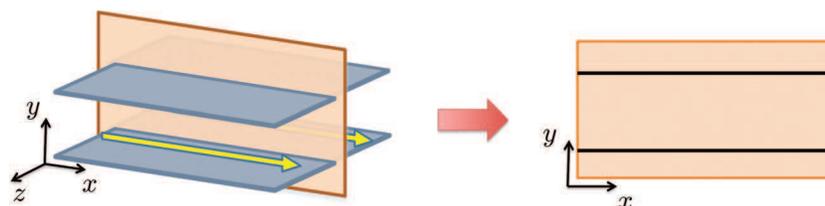

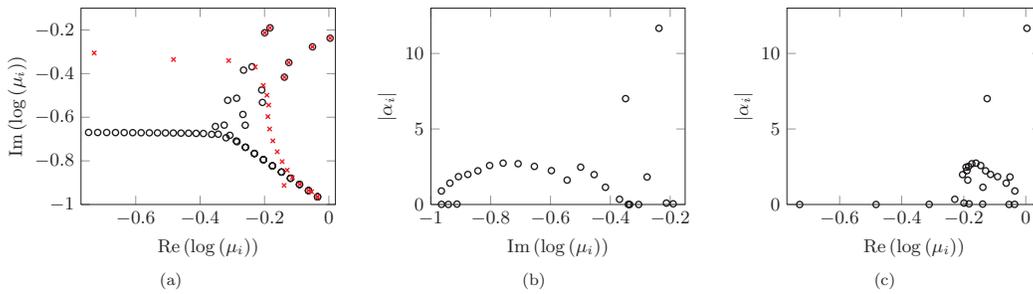FIG. 1. Geometry of a two-dimensional channel flow.

FIG. 2. (a) Spectrum of the Orr-Sommerfeld operator (circles) along with the eigenvalues resulting from the standard DMD algorithm (crosses) for the two-dimensional Poiseuille flow with $Re = 10\,000$ and $k_x = 1$. Dependence of the absolute value of the DMD amplitudes α_i on (b) the frequency and (c) the real part of the corresponding DMD eigenvalues μ_i.

$\text{Re}\left(\log(\mu_i)\right)/k_x$ that are larger than the average base velocity ($\bar{U} = 2/3$), and the slower dynamics which takes place near the channel walls is described by eigenvalues with smaller phase velocities. As our emphasis on sparsity increases, we observe a selection of modes from both the slow and fast branches. While for $N_z = 13$ modes nearly the entire slow and fast branches (equivalent to the A- and P-branches of plane Poiseuille flow[29]) are selected, a progressive coarsening on each branch is seen for a smaller number of modes. This process continues until, for $N_z = 3$, only one fast, one slow, and one unstable mode is chosen to represent the dynamics contained in the original data sequence. Finally, in the limit of only a single mode, for $N_z = 1$, the unstable TS wave is selected by the sparsity-promoting DMD algorithm.

The second row in Figure 4 shows the frequency-dependence of the absolute values of the DMD amplitudes α_i for both the standard DMD algorithm (circles) and its sparsity-promoting variant (crosses). We observe discrepancy between the amplitudes resulting from the two algorithms. This is because the sparsity-promoting algorithm adjusts amplitudes of the retained DMD modes in order to provide optimal approximation of the original data sequence.

B. A screeching supersonic jet

Screech is a component of supersonic jet noise that is connected to the presence of a train of shock cells within the jet column.[31] For a turbulent jet, the unsteady shear layers interact with the shocks to create sound. While this process is generally broadbanded, screech is a special case of shock-noise that arises from the creation of a feedback loop between the upstream-propagating part of the
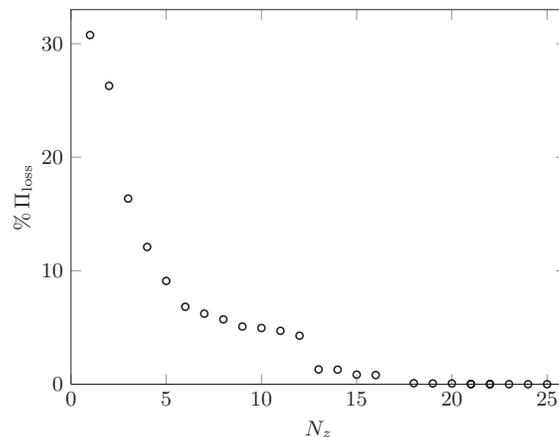

FIG. 3. Performance loss $\%\,\Pi_{\text{loss}} := 100\,\sqrt{J(\alpha)/J(0)}$ of the optimal vector of amplitudes α resulting from the sparsity-promoting DMD algorithm (with N_z DMD modes) for the Poiseuille flow example.

(a) $N_z = 13$ (b) $N_z = 3$ (c) $N_z = 1$
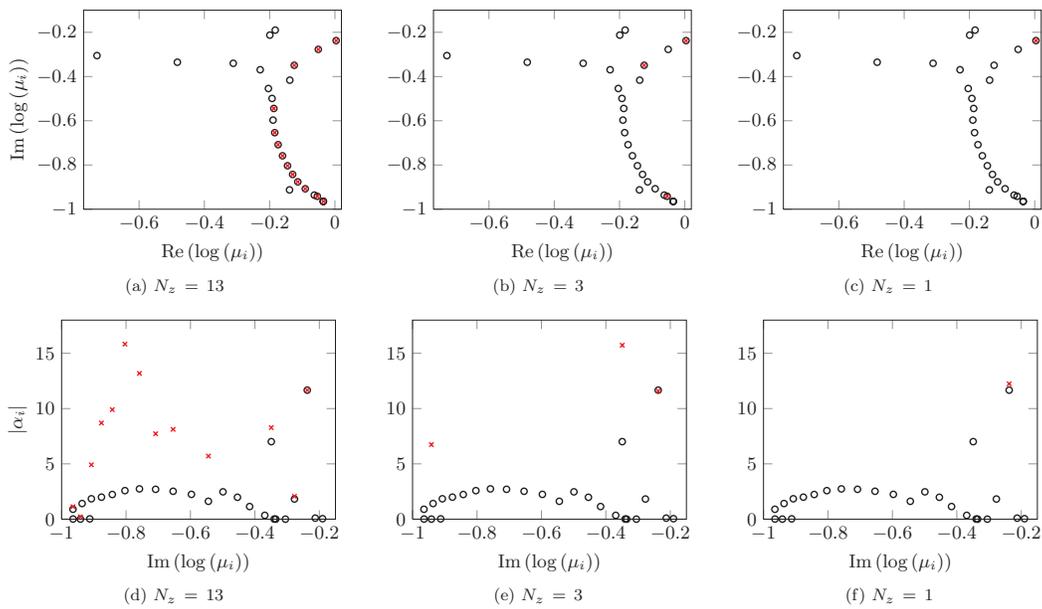
(d) $N_z = 13$ (e) $N_z = 3$ (f) $N_z = 1$

FIG. 4. Eigenvalues of F_{dmd} (a)–(c) and the absolute values of the DMD amplitudes α_i (d)–(f) for the Poiseuille flow example. The results are obtained using the standard DMD algorithm (circles) and the sparsity-promoting DMD algorithm (crosses) with N_z DMD modes.

acoustic field and the generation of new disturbances at the nozzle lip. This self-sustaining feedback loop leads to an extremely loud (narrow-banded) screech tone at a specific fundamental frequency. The presence of a tonal process embedded in an otherwise broadbanded turbulent flow makes the screeching jet an excellent test case for the sparsity-promoting DMD method. The objective of the developed method is to extract the entire coherent screech feedback loop from the turbulent data and to describe the screech mechanism with as few modes as possible.

The supersonic jet used in this example was produced by a convergent rectangular nozzle of aspect ratio 4, precisely matching the geometry of an experimental nozzle;[32] see Figure 5(a) for geometry. The entire flow inside, outside, and downstream of the nozzle was simulated using the low-dissipation low-dispersion LES solver `charles` on an unstructured mesh with about 45×10^6 control volumes. This simulation was a part of a sequence involving different mesh resolutions, and it was validated against the experimental measurements.[33]
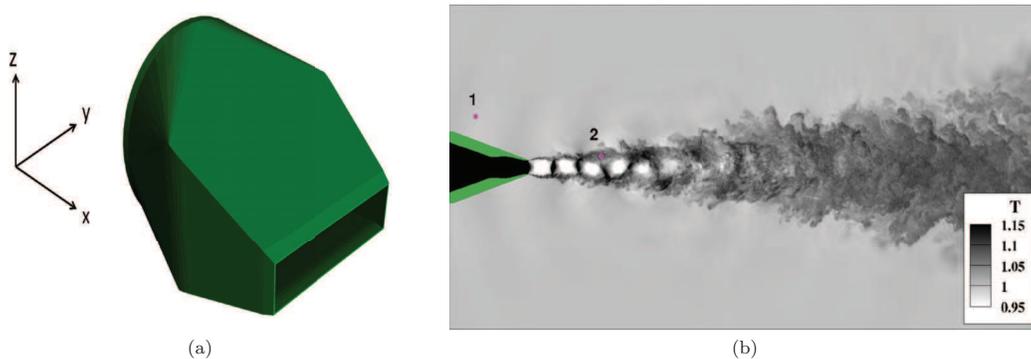

(a)   (b)

FIG. 5. (a) Rectangular nozzle geometry of aspect ratio 4. (b) Temperature contours in the (x, z)-centerplane taken from a snapshot of the screeching jet. Time histories of pressure were recorded at probe locations indicated by circles 1 and 2 (Multimedia view). [URL: http://dx.doi.org/10.1063/1.4863670.1]
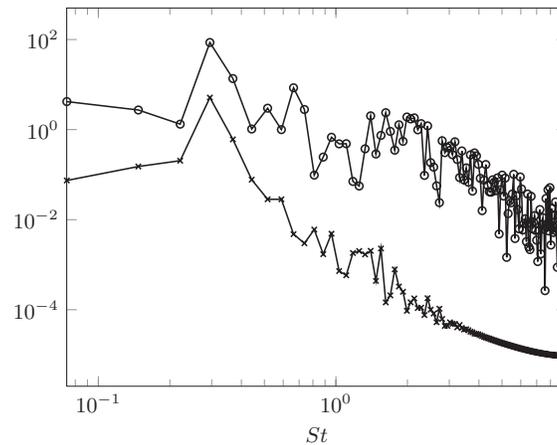
FIG. 6. Power spectra of pressure corresponding to the measurement locations indicated in Figure 5(b). While both spectra peak at $St \approx 0.3$, the spectrum at location 2 (circles) is more broadbanded than at location 1 (crosses).

The stagnation pressure and temperature inside the nozzle were set so that the jet Mach number $M_j = 1.4$ and the fully expanded jet temperature matched the ambient temperature. A rectangular nozzle whose interior cross section area was decreasing monotonically from inlet to exit was used. Since the nozzle did not have a diverging section before its exit, the flow left the nozzle in an under-expanded (sonic) state and continued to expand downstream to reach the supersonic fully expanded condition. This induced a train of diamond-shaped shock cells as shown in Figure 5(b) (Multimedia view). The figure shows contours of temperature on a centerplane cross section taken through the narrow dimension of the nozzle. From an animation of the jet, it can be observed that the first two shock cells are almost stationary, but the third and fourth shock cells undergo transverse oscillations along the narrow (vertical) dimension of the jet. The transverse oscillation of the shock cells occurs precisely at the screech frequency of the jet, and is connected to a strong upstream-oriented acoustic tone.

Figure 6 shows the spectra of the pressure recorded at the locations indicated by the circles in Figure 5(b). At location 1, which coincides with the center of the upstream-directed acoustic beam associated with the screech tone, the spectrum is relatively narrow-banded. Since the sample window contains approximately four oscillations of the screech tone, the fourth Fourier coefficient has the largest amplitude. In contrast to location 1, location 2 was taken in the center of one of the turbulent shear layers. The turbulent fluctuations induce a broad range of frequencies in the corresponding spectrum (circles in Figure 6). At both locations, spectra show a strong peak at Strouhal number $St \approx 0.3$ (based on the fully expanded jet velocity and the nozzle equivalent diameter). This peak corresponds to the screech tone frequency predicted for a Mach 1.4 jet from a 4:1 aspect ratio rectangular nozzle.[31,33]

The database used for the DMD analysis consisted of 257 snapshots (so that $N = 256$) of the full three-dimensional pressure and velocity fields. The snapshots were equispaced in time with an interval of $\Delta t = 0.0528 D_e/u_j$, where D_e is the nozzle equivalent diameter (the diameter of the circle of same area as the nozzle exit) and u_j is the fully expanded jet velocity. Although the computational domain for the LES extended approximately $32 D_e$ downstream of the nozzle exit, DMD was applied to a subdomain focusing on the shock cells within the jet's potential core and the surrounding acoustic field (this domain extends to $10 D_e$). This restriction reduced the number of cells from 45×10^6 to 8×10^6. In spite of this reduction, each snapshot required 256 Mb of storage in double precision format. To handle such large matrices, DMD was implemented using a MapReduce framework so that the matrix could be stored and processed across several storage discs. In particular, the algorithm relied upon a MapReduce QR-factorization of tall-and-skinny matrices developed in Ref. 34.

Figure 7(a) illustrates the frequency dependence of the absolute value of the amplitudes of the DMD modes obtained by solving the optimization problem (6). We note that it is not trivial to identify, by mere inspection, a subset of DMD modes that has the strongest influence on the quality
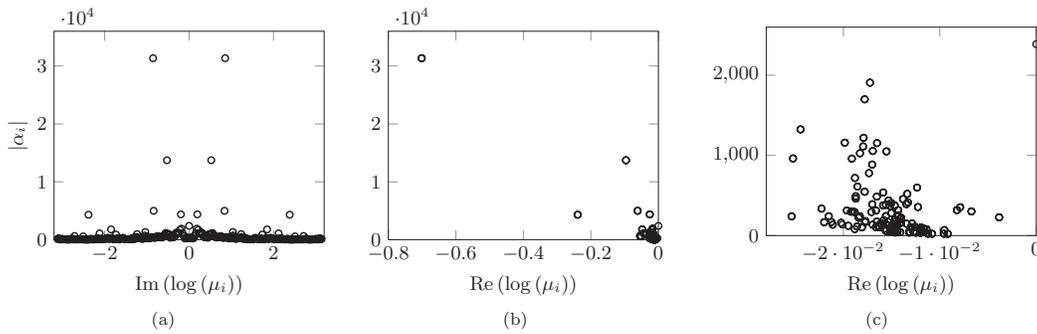
FIG. 7. Dependence of the absolute value of the DMD amplitudes α_i on (a) the frequency and (b) and (c) the real part of the corresponding DMD eigenvalues μ_i for the screeching jet example. Subfigure (c) represents a zoomed version of subfigure (b) and it focuses on the amplitudes that correspond to lightly damped eigenvalues.

of the least-squares approximation. As shown in Figure 7(b), the largest amplitudes originate from eigenvalues that are strongly damped. In what follows, we demonstrate that keeping only a subset of modes with largest amplitudes can lead to poor quality of approximation of numerically generated snapshots.

The sparsity level **card** (α) and the performance loss % $\Pi_{\text{loss}} := 100\sqrt{J(\alpha)/J(0)}$ for the optimal vector of amplitudes α resulting from the sparsity-promoting DMD algorithm are shown in Figure 8 as a function of the user-specified parameter γ (a measure of preference between approximation quality and solution sparsity). As expected, larger values of γ encourage sparser solutions, at the expense of compromising quality of the least-squares approximation. The values of γ in Figure 8 are selected in such a way that $\gamma_{\min}$ induces a dense vector α (with 256 non-zero elements), and $\gamma_{\max}$ induces α with a single non-zero element.

Eigenvalues resulting from the standard DMD algorithm (circles) along with the subset of N_z eigenvalues selected by the sparsity-promoting DMD algorithm (crosses) are shown in Figure 9. Eigenvalues in the interior of the unit circle are strongly damped. Since a strongly damped mode influences only early stages of time evolution, the associated amplitude $|\alpha_i|$ can be large; see Figure 7(b) for an illustration. Rather than focusing only on the modes with largest amplitudes, the sparsity-promoting DMD identifies modes that have the strongest influence on the entire time history of available snapshots. While the selection of the retained eigenvalues is non-trivial, it increasingly concentrates on the low-frequency modes as N_z decreases. For $N_z = 3$, only the mean flow and one dominant frequency pair (that corresponds to the fundamental frequency of the screech tone) remain, while for $N_z = 5$ a second, lower frequency is identified. In this low-N_z limit, we do see that the sparsity-promoting DMD approximates the data sequence using the most prevalent structures.
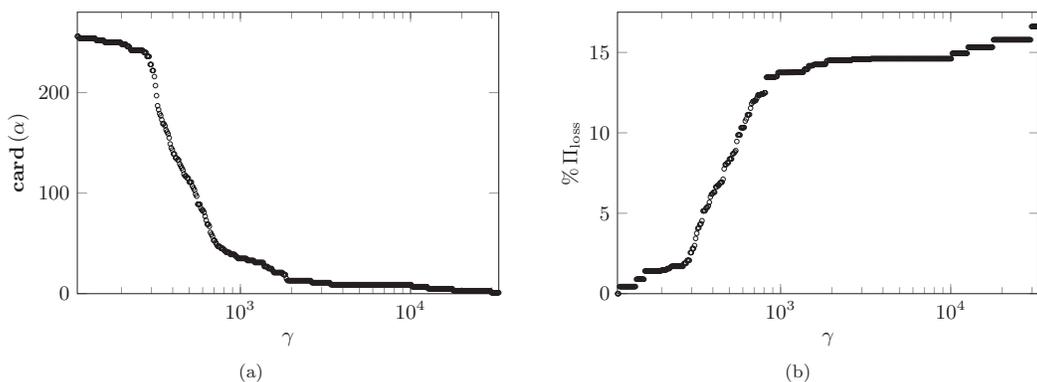

FIG. 8. (a) The sparsity level **card** (α) and (b) the performance loss % $\Pi_{\text{loss}} := 100\sqrt{J(\alpha)/J(0)}$ of the optimal vector of amplitudes α resulting from the sparsity-promoting DMD algorithm for the screeching jet example.
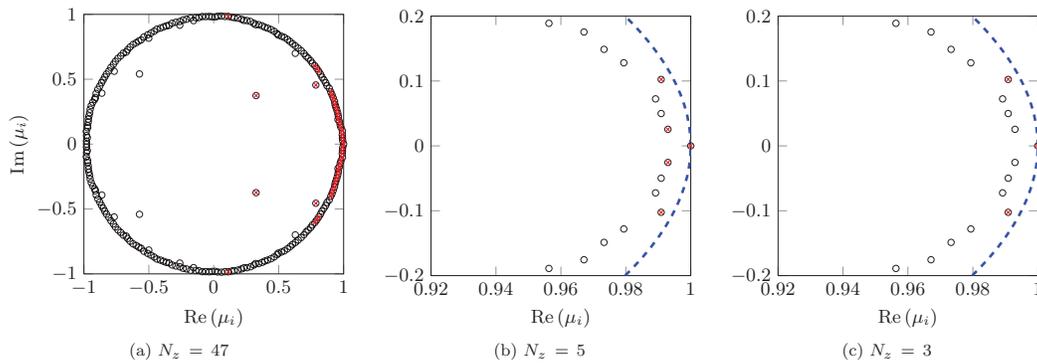
FIG. 9. Eigenvalues resulting from the standard DMD algorithm (circles) along with the subset of N_z eigenvalues selected by the sparsity-promoting DMD algorithm (crosses) for the screeching jet example. In subfigures (b) and (c), the dashed curves identify the unit circle.

The corresponding amplitudes for the various truncations N_z are displayed in Figure 10. Again, as N_z decreases a concentration on low frequencies is observed. We also note that the amplitudes of the original DMD modes do not provide sufficient guidance in reducing the full set of DMD modes to a few relevant ones.

Figure 11 illustrates performance of the sparsity-promoting DMD algorithm developed in this paper in terms of the number of identified structures. The performance is quantified by the Frobenius norm of the approximation error between a low-dimensional representation and the full data sequence in fraction of the Frobenius norm of the full data sequence. As expected, the performance of our algorithm improves as the number of modes is increased. For the screeching jet example, the sparsity-promoting DMD with $N_z = 3$ gives $St = 0.3104$ which agrees well with the frequency $St = 0.3$ measured from a sequence containing 10 times the number of snapshots.

Finally, Figure 12 visualizes the three-dimensional mode associated with the dominant frequency pair resulting from the sparsity-promoting DMD algorithm (Multimedia view). Isosurfaces of pressure and dilatation fluctuations are identified by darker gray and lighter gray colors, respectively. An animation of this mode reveals that the dilatation corresponds to a flapping mode along the narrow dimension of the jet. As the shock cells oscillate transversely, they encroach into the jet shear layers, and compress the oncoming jet fluid in a region where previously there was no compression. This time-periodic compression is connected to the upstream propagating acoustic wave at the exterior of the jet. Furthermore, the animation of the DMD mode reveals that – close to the jet – the acoustic wave does not propagate smoothly in the upstream direction. Instead, the acoustic wave hops from one spot of negative dilatation to the next as it proceeds upstream. As the oscillating shock cells penetrate into the jet shear layers, they are also rotated by the shear. When this
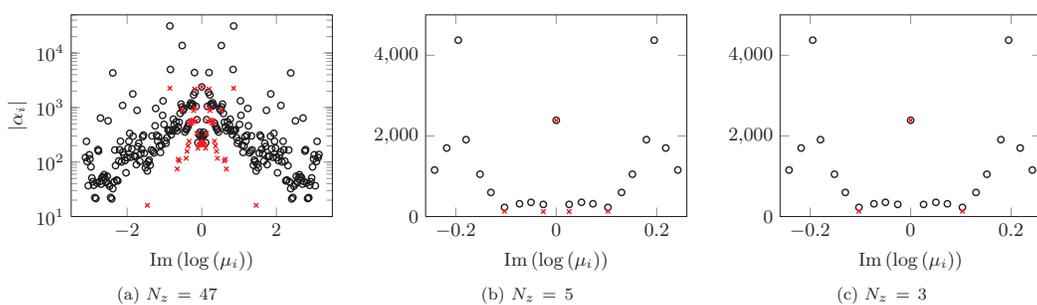

FIG. 10. (a) $N_z = 47$, (b) $N_z = 5$, (c) $N_z = 3$: Dependence of the absolute value of the amplitudes α_i on the frequency (imaginary part) of the corresponding eigenvalues μ_i for the screeching jet example. The results are obtained using the standard DMD algorithm (circles) and the sparsity-promoting DMD algorithm (crosses) with N_z DMD modes.
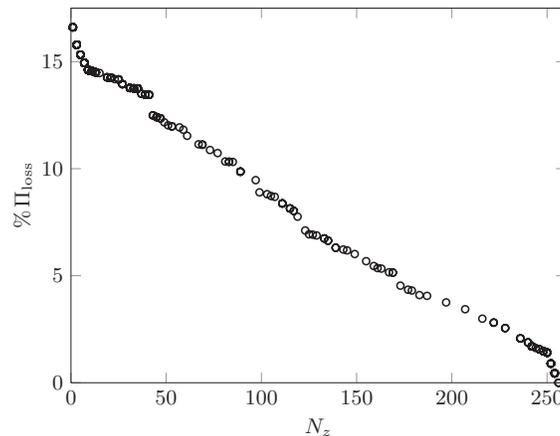
FIG. 11. The optimal performance loss, $\% \, \Pi_{\text{loss}} := 100 \, \sqrt{J(\alpha)/J(0)}$, of the sparsity-promoting DMD algorithm for the screeching jet example as a function of the number of retained modes.

happens, the outer tip of the shock cell sweeps from downstream to upstream, providing an impulse to the exterior acoustic perturbation which "kicks" it upstream to the next shock cell.

C. Flow through a cylinder bundle

Flow through cylinders in a bundle configuration is often encountered in the utility and energy conversion industry. Owing to their versatility and efficiency,[35] cross-flow heat exchangers account for the vast majority of heat exchangers in oil refining, process engineering, petroleum extraction, and power generation sectors. Consequently, even a modest improvement in their effectiveness and operational margins would have a significant impact on production efficiency. In spite of their widespread use, the details of the flow (and heat transfer) through cylinder bundles are far from fully understood, thereby making this type of flow, along with its simplified variants, the subject of active research.[36–39]

Vortex shedding and complex wake interactions can induce vibrations and structural resonances which, in turn, may result in fretting wear, collisional damage, material fatigue, creep, and ultimately in cracking. Even though the flow through a cylinder bundle is very complex, it is characterized by well-defined shedding frequencies.[40] For single-row cylinders, only a few distinct frequencies are detected, while for more complex array configurations, a multitude of precise shedding frequencies are observed. The presence of distinct shedding frequencies makes this type of flow well-suited for a decomposition of the flow fields into single-frequency modes.
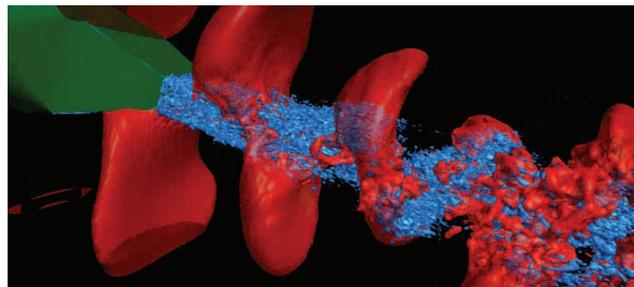

FIG. 12. The three-dimensional DMD mode associated with the dominant frequency pair (that corresponds to the fundamental frequency of the screech tone, $St \approx 0.3$). A darker gray (red online) isosurface of perturbation pressure is shown together with a lighter gray (blue online) isosurface of the perturbation dilatation (Multimedia view). [URL: http://dx.doi.org/10.1063/1.4863670.2]
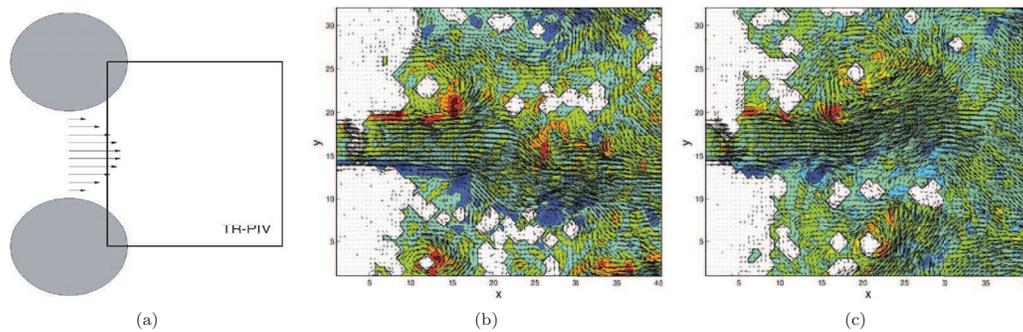
FIG. 13. (a) Geometry of a flow between two cylinders along with the interrogation window where PIV measurements are taken. (b) and (c) Two representative snapshots of velocity vectors and vorticity contours, resulting from time-resolved PIV measurements, illustrate the oscillatory motion of the exiting jet. See text for details on the experimental parameters.

In addition to the multiple-frequency behavior, a characteristic oscillatory pattern – labelled as a "flip-flop phenomenon" – is typically observed in the exit stream of the flow through cylinder bundles.[41] The observed oscillatory pattern consists of a meta-stable deflection of the exiting jet off the centerline location. A statistical analysis of the velocity field in a cross-stream section shows a strongly bimodal distribution. Over a wide range of Reynolds numbers, the Strouhal number associated with the oscillation between these two states is approximately Reynolds-number-independent. This flow feature can also be represented by a dynamic mode decomposition of the flow fields.

In this paper, we use a simplified geometric configuration that nonetheless inherits the main features from more complex settings, including the presence of multiple distinct frequencies and the flip-flop oscillations of the exit stream. In particular, we consider the flow passing between two cylinders. The PIV interrogation window shown in Figure 13(a) measures 40.36 mm in the streamwise and 32.08 mm in the cross-stream direction. The flow field is resolved on a 63×79 measurement grid, and two in-plane velocity components are recorded. The flow fields are represented in a fully time-resolved manner with 4 ms between two consecutive PIV measurements. The inter-cylinder gap is 10.7 mm and the jet passing between the two cylinders (of diameter 12 mm) has a mean velocity of 0.663 m/s. The resulting Reynolds number, based on the volume flux velocity ($\dot{Q} = 18 \, \mathrm{m}^3/\mathrm{h}$) and the cylinder diameter, is $Re = 3000$.

Two representative snapshots resulting from a time-resolved flow field sequence display velocity vectors and contours of the spanwise vorticity in Figure 13. The two flow fields are 0.28 s apart and show a downward (Figure 13(b)) and upward (Figure 13(c)) deflection of the jet exiting between the two cylinders. A time-sequence of flow fields substantiates the existence of a low-frequency oscillation of the jet, which forms the basis of the afore-mentioned flip-flop phenomenon. A streamwise velocity signal has been extracted from the center of the experimental domain, and a subsequent spectral analysis of this signal reveals a distinct frequency of 7.81 Hz which corresponds to a Strouhal number $St = 0.126$ (based on the jet mean velocity and cylinder gap).

Figure 14 shows the amplitudes of the DMD modes versus the identified frequencies and corresponding growth/decay rates. These results were obtained using a database with $N = 100$ snapshots. As in the screeching jet example, strongly damped modes are characterized by large amplitudes, and a clear indication about the relative importance of individual DMD modes cannot be inferred from these two plots. Instead, we apply our sparsity-promoting framework to aid in the delineation of modes that contribute significantly to the data sequence and modes that capture only transient effects. By adjusting the user-specified parameter γ, we encourage solutions that consist of only a limited number of modes but still optimally represent the original data-sequence.

Figure 15 superimposes the eigenvalues resulting from the standard DMD algorithm (circles) and its sparsity-promoting variant (crosses). Even though the eigenvalues in the vicinity of the unit circle are associated with much smaller amplitudes than the strongly damped eigenvalues (that reside well within the unit circle), they are selected by our algorithm. As the sparsity parameter is
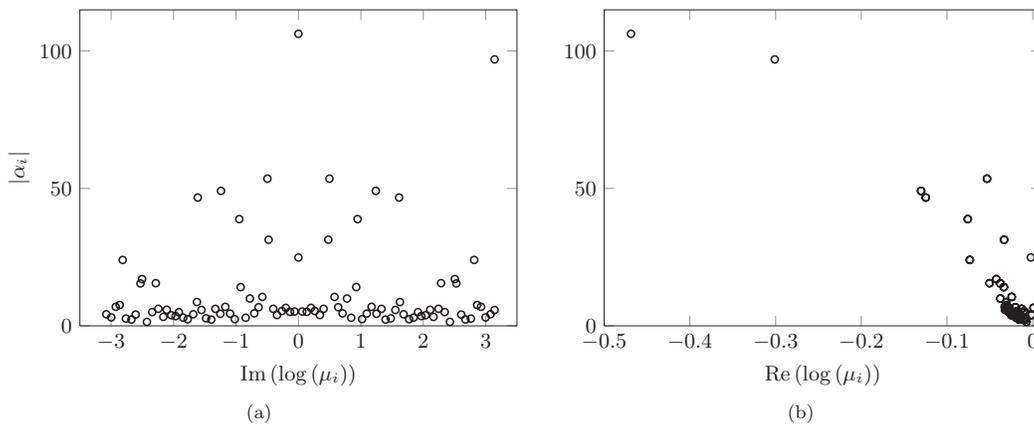
FIG. 14. Dependence of the absolute value of the DMD amplitudes α_i on (a) the frequency and (b) the real part of the corresponding DMD eigenvalues μ_i for the flow through a cylinder bundle.

increased and fewer eigenvalues are selected, a clustering of eigenvalues at and near the point $(1, 0)$ is observed; see Figure 15(c).

A similar picture emerges from displaying the selected modes in the frequency-amplitude plane; see Figure 16. The sparsity-promoting DMD algorithm provides a rapid concentration on low-frequency modes, thereby eliminating structures with substantially larger amplitudes identified by the standard DMD algorithm. As the limit of only $N_z = 3$ DMD modes is reached, the sparsity-promoting algorithm concentrates on frequencies that still optimally describe the principal oscillatory components of the full data set. As evident from Figure 16(c), for $N_z = 3$ the last oscillatory representation is characterized by $\mathrm{Im}\,(\log(\mu_i)) \approx 0.201$ (which corresponds to a frequency of 7.99 Hz and a Strouhal number of $St = 0.129$). This frequency is in good agreement with the frequency identified by the aforementioned point measurements of the streamwise velocity. The flow field associated with the three-term ($N_z = 3$) representation of the data sequence is displayed in Figure 17; it consists of a deflected jet and strong vortical components off the symmetry axis (Multimedia view). An animation of this flow field has been obtained using the identified and selected frequency (7.99 Hz) and it reproduces the main features of the full data sequence, i.e., the lateral swaying of the jet under the influence of the wake vortices of the two cylinders.

The sparse approximation of the full data-sequence focuses on the most relevant structures by removing large-amplitude but transient flow features and it results in progressively larger residuals as the sparsity is enhanced. This inherent trade-off between a more compact data representation and the quality of approximation compared to the original data sequence is depicted in Figure 18.
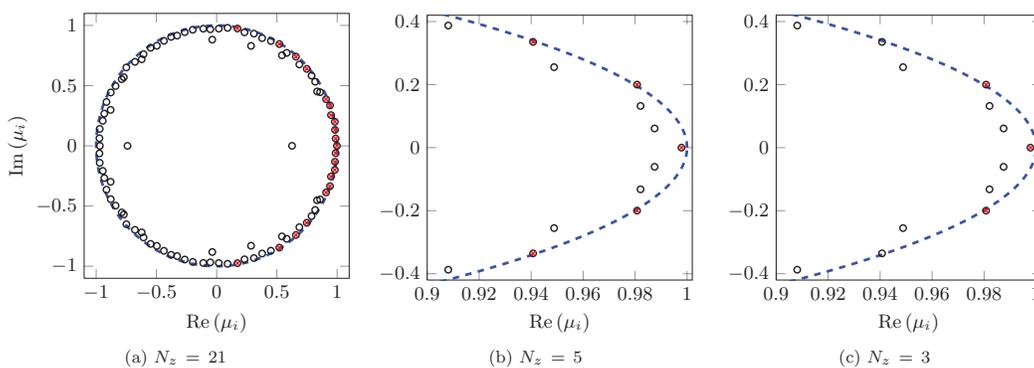

FIG. 15. (a) $N_z = 21$, (b) $N_z = 5$, (c) $N_z = 3$: Eigenvalues resulting from the standard DMD algorithm (circles) along with the subset of N_z eigenvalues selected by its sparsity-promoting variant (crosses) for the flow through a cylinder bundle. The dashed curves identify the unit circle.
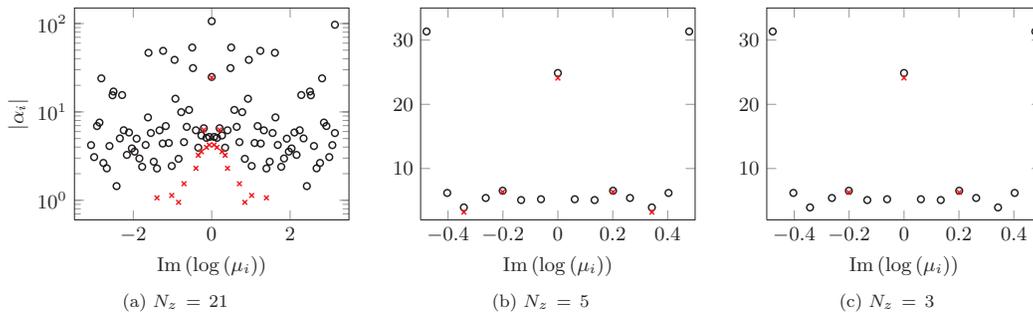
(a) $N_z = 21$ (b) $N_z = 5$ (c) $N_z = 3$

FIG. 16. (a) $N_z = 21$, (b) $N_z = 5$, (c) $N_z = 3$: Dependence of the absolute value of the amplitudes α_i on the frequency (imaginary part) of the corresponding eigenvalues μ_i for the flow through a cylinder bundle. The results are obtained using the standard DMD algorithm (circles) and the sparsity-promoting DMD algorithm (crosses) with N_z DMD modes.
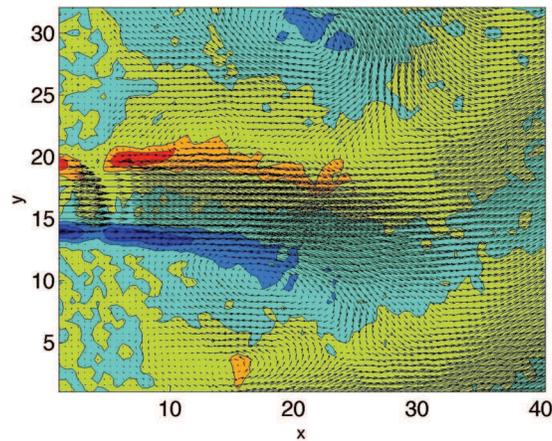

FIG. 17. Snapshot resulting from a sparse three-component ($N_z = 3$) representation of the full data sequence, visualized by velocity vectors and vorticity contours (Multimedia view). [URL: http://dx.doi.org/10.1063/1.4863670.3]

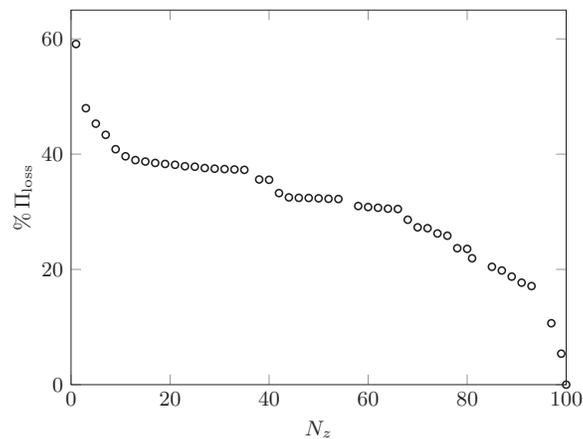

FIG. 18. The optimal performance loss, % $\Pi_{\text{loss}} := 100\sqrt{J(\alpha)/J(0)}$, of the sparsity-promoting DMD algorithm for the flow through a cylinder bundle as a function of the number of retained modes.

In contrast to the channel flow and screeching jet examples, the sparsity-promoting version of the DMD-algorithm displays higher performance degradation. This behavior may perhaps be attributed to the presence of unstructured measurement noise in the experimental dataset for a jet passing between two cylinders.

V. CONCLUDING REMARKS

We have introduced an extension of the standard DMD algorithm that addresses the reduction in dimensionality of the full rank decomposition. This reduction is accomplished by a sparsity-promoting procedure which augments the standard least-squares optimization problem by a term that penalizes the ℓ_1-norm of the vector of unknown amplitudes. A user-specified regularization parameter balances the trade-off between the quality of approximation and the number of retained DMD modes. The sparsity-promoting DMD algorithm thus selects specific dynamic modes (along with their associated temporal frequencies and growth/decay rates) which exhibit the strongest contribution to a representation of the original data sequence over the considered time interval.

The ℓ_1-regularized least-squares problem can be formulated as a convex optimization problem for which the ADMM provides an efficient tool for computing the globally optimal solution. ADMM is an iterative method that alternates between minimization of the least-squares residual and sparsity enhancement. We have shown that the least-squares minimization step amounts to solving an unconstrained regularized quadratic program and that sparsity is promoted through the application of a soft-thresholding operator. After a desirable tradeoff between the quality of approximation and the number of DMD modes has been accomplished, we fix the sparsity structure and compute the optimal amplitudes of the retained dynamic modes.

The sparsity-promoting dynamic mode decomposition has been applied to three examples. The linearized plane Poiseuille flow at a supercritical Reynolds number was used to showcase the algorithm on a canonical flow configuration; the sparsity-promoting DMD extracted modal contributions from each of the familiar eigenvalue branches until only the exponentially growing TS wave was retained in the limit of maximal sparsity. For the unstructured large-eddy simulation of a supersonic screeching jet, the sparsity-promoting DMD algorithm successfully and efficiently identified the screech frequency and the associated flow fields; dynamic modes with large amplitudes and substantial decay rates, representing transient flow features, have been eliminated as our emphasis on sparsity increased. Finally, for two-dimensional time-resolved PIV measurements of a jet passing between two cylinders, the prevailing Strouhal number has been identified by the dynamic mode decomposition, and the sparsity-promoting algorithm effectively differentiated the dominant flow structures from less relevant flow phenomena embedded in the full data sequence.

The developed method builds on recent attempts at optimizing the standard dynamic mode decomposition[14–16] and at providing a framework for the automated detection of a few pertinent modal flow features. While these efforts resulted in complex or nearly intractable non-convex optimization problems, the sparsity-promoting DMD algorithm provides an efficient paradigm for detection and extraction of a limited subset of flow features that optimally approximate the original data sequence. Furthermore, in contrast to the aforementioned attempts, our approach *regularizes* the least-squares deviation between the matrix of snapshots and the linear combination of DMD modes with a term that penalizes the ℓ_1-norm of the vector of unknown DMD amplitudes. While the former term depends both on the problem data and on the optimization variable, the latter term is problem-data-independent; its primary purpose is to enhance desirable features in the solution to the resulting regularized optimization problem. We note that regularization may be of essence in the situations where the problem data are corrupted by noise or contain numerical or experimental outliers. In the absence of regularization, the predictive capability of low-dimensional models resulting from corrupted or incomplete snapshots may be significantly diminished. Exploration of different types of regularization penalties that are well-suited for the problems in fluid mechanics is an effort worth pursuing.

The developed algorithm has shown its value on the numerical and experimental snapshot sequences considered in this paper. It is expected that the sparsity-promoting dynamic mode decomposition will become a valuable tool in the quantitative analysis of high-dimensional datasets, in the

interpretation of identified dynamic modes, and in the eduction of relevant physical mechanisms. For additional information about the examples considered in this paper, including Matlab source codes and problem data, see the supplementary material.[42]

ACKNOWLEDGMENTS

APPENDIX A: AN ALTERNATIVE FORMULATION OF (6)

We show that the objective function $J(\alpha)$ in (6),

$$J(\alpha) \;=\; \|G \,-\, L\,D_\alpha\,R\|_F^2, \tag{A1}$$

with

$$G \;:=\; \Sigma\,V^*,\, L \;:=\; Y,\, D_\alpha \;:=\; \mathrm{diag}\,\{\alpha\},\, R \;:=\; V_{\mathrm{and}},$$

can be equivalently represented as (7). The equivalence between (7) and (A1) can be established through a sequence of straightforward algebraic manipulations in conjunction with the repeated use of the following properties of the matrix trace:

P1: Commutativity invariance,

$$\mathrm{trace}\,(A\,B) \;=\; \mathrm{trace}\,(B\,A)\,;$$

P2: A product between a matrix Q and a diagonal matrix $D_\alpha := \mathrm{diag}\,\{\alpha\}$,

$$\mathrm{trace}\,(Q\,D_\alpha) \;=\; (\mathrm{diag}\,\{Q\})^T\alpha \;=\; \big(\overline{\mathrm{diag}\,\{Q\}}\big)^*\alpha\,;$$

P3: For vectors $\alpha \in \mathbb{C}^n$ and $\beta \in \mathbb{C}^m$ and matrices $A,\, B \in \mathbb{C}^{m\times n}$,

$$\mathrm{trace}\,\big(D_\beta^*\,A\,D_\alpha\,B^T\big) \;=\; \beta^*(A \circ B)\,\alpha\,.$$

Now, a repeated use of the matrix trace commutativity invariance property **P1** leads to

$$\begin{aligned}
J(\alpha) &= \|G \,-\, L\,D_\alpha\,R\|_F^2 \\
&= \mathrm{trace}\,\big((G \,-\, L\,D_\alpha\,R)^*\,(G \,-\, L\,D_\alpha\,R)\big) \\
&= \mathrm{trace}\,\big(D_\alpha^*\,(L^*L)\,D_\alpha\,(R\,R^*) \,-\, Q\,D_\alpha \,-\, Q^*D_\alpha^* \,+\, G^*G\big),
\end{aligned}$$

and the equivalence between (7) and (A1) follows from **P2** and **P3**.

APPENDIX B: α- AND β-MINIMIZATION STEPS IN ADMM

We exploit the respective structures of the functions J and g in (11) and show that the α-minimization step amounts to solving an unconstrained regularized quadratic program and that the β-minimization step amounts to an opportune use of the soft thresholding operator.

- **α-minimization step:** Completion of squares with respect to α in the augmented Lagrangian $\mathcal{L}_\rho$ can be used to show that the α-minimization step (13a) is equivalent to

$$\underset{\alpha}{\mathrm{minimize}} \quad J(\alpha) \,+\, \frac{\rho}{2}\,\|\alpha \,-\, u^k\|_2^2,$$

where

$$u^k = \beta^k - (1/\rho)\lambda^k.$$

Now, the definition (7) of J leads to the unconstrained quadratic programming problem

$$\underset{\alpha}{\text{minimize}} \quad \alpha^*(P + (\rho/2)I)\alpha - (q + (\rho/2)u^k)^*\alpha - \alpha^*(q + (\rho/2)u^k) + s + \rho\|u^k\|_2^2,$$

whose solution is determined by

$$\alpha^{k+1} = (P + (\rho/2)I)^{-1}(q + (\rho/2)u^k).$$

- **β-minimization step:** The completion of squares with respect to β in the augmented Lagrangian $\mathcal{L}_\rho$ can be used to show that the β-minimization step (13b) is equivalent to

$$\underset{\beta}{\text{minimize}} \quad \gamma g(\beta) + \frac{\rho}{2}\|\beta - v^k\|_2^2,$$

where

$$v^k = \alpha^{k+1} + (1/\rho)\lambda^k.$$

It is a standard fact that the solution to this optimization problem is determined by

$$\beta_i^{k+1} = S_\kappa(v_i^k), \quad \kappa = \gamma/\rho,$$

where $S_\kappa(\,\cdot\,)$ denotes the *soft thresholding* operator

$$S_\kappa(v_i^k) = \begin{cases} v_i^k - \kappa, & v_i^k > \kappa, \\ 0, & v_i^k \in [-\kappa, \kappa], \\ v_i^k + \kappa, & v_i^k < -\kappa. \end{cases}$$

APPENDIX C: AN EFFICIENT ALGORITHM FOR SOLVING (10)

From Appendix A it follows that finding the solution to (10) amounts to solving the following equality-constrained quadratic programming problem

$$\underset{\alpha}{\text{minimize}} \quad J(\alpha) = \alpha^*P\alpha - q^*\alpha - \alpha^*q + s$$
$$\text{subject to} \quad E^T\alpha = 0.$$

It is a standard fact that the variation of the Lagrangian

$$\mathcal{L}(\alpha, \nu) = J(\alpha) + \nu^*E^T\alpha + (E^T\alpha)^*\nu,$$

with respect to α and the vector of Lagrange multipliers ν can be used to obtain the conditions for optimality of $\mathcal{L}$,

$$\begin{bmatrix} P & E \\ E^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \nu \end{bmatrix} = \begin{bmatrix} q \\ 0 \end{bmatrix}.$$

The optimal sparse vector of amplitudes, α_{sp}, is then determined by

$$\alpha_{\text{sp}} = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} P & E \\ E^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} q \\ 0 \end{bmatrix}.$$

[1] J. Lumley, *Stochastic Tools in Turbulence* (Dover Publications, Mineola, NY, 2007).
[2] L. Sirovich, "Turbulence and the dynamics of coherent structures. Part I: Coherent structures," Q. Appl. Math. **45**(3), 561–571 (1987).
[3] D. Sipp, O. Marquet, P. Meliga, and A. Barbagallo, "Dynamics and control of global instabilities in open flows: A linearized approach," Appl. Mech. Rev. **63**, 030801 (2010).

[4] B. Moore, "Principal component analysis in linear systems: Controllability, observability and model reduction," IEEE Trans. Autom. Control **AC-26**(1), 17–32 (1981).

[5] C. Rowley, "Model reduction for fluids using balanced proper orthogonal decomposition," Int. J. Bifurcation Chaos, **15**, 997–1013 (2005).

[6] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," Nonlinear Dyn. **41**(1), 309–325 (2005).

[7] C. Rowley, I. Mezic, S. Bagheri, P. Schlatter, and D. Henningson, "Spectral analysis of nonlinear flows," J. Fluid Mech. **641**, 115–127 (2009).

[8] I. Mezić, "Analysis of fluid flows via spectral properties of Koopman operator," Annu. Rev. Fluid Mech. **45**(1), 357–378 (2013).

[9] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," J. Fluid Mech. **656**, 5–28 (2010).

[10] L. N. Trefethen, A. E. Trefethen, S. C. Reddy, and T. A. Driscoll, "Hydrodynamic stability without eigenvalues," Science **261**, 578–584 (1993).

[11] M. R. Jovanović and B. Bamieh, "Componentwise energy amplification in channel flows," J. Fluid Mech. **534**, 145–183 (2005).

[12] P. J. Schmid, "Nonmodal stability theory," Annu. Rev. Fluid Mech. **39**, 129–162 (2007).

[13] S. Bagheri, "Koopman-mode decomposition of the cylinder wake," J. Fluid Mech. **726**, 596–623 (2013).

[14] K. K. Chen, J. H. Tu, and C. W. Rowley, "Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses," J. Nonlinear Sci. **22**(6), 887–915 (2012).

[15] P. J. Goulart, A. Wynn, and D. Pearson, "Optimal mode decomposition for high dimensional systems," in *Proceedings of the 51st IEEE Conference on Decision and Control*, 2012 (IEEE, 2012), pp. 4965–4970.

[16] A. Wynn, D. Pearson, B. Ganapathisubramani, and P. J. Goulart, "Optimal mode decomposition for unsteady flows," J. Fluid Mech. **733**, 473–503 (2013).

[17] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, NY, 2004).

[18] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**(2), 489–509 (2006).

[19] D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).

[20] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," IEEE Trans. Inf. Theory **52**(12), 5406–5425 (2006).

[21] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," J. Fourier Anal. Appl. **14**, 877–905 (2008).

[22] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, New York, NY, 2009).

[23] T. Goldstein and S. Osher, "The split Bregman method for ℓ_1 regularized problems," SIAM J. Imaging Sci. **2**(2), 323–343 (2009).

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Found. Trends Mach. Learning **3**(1), 1–124 (2011).

[25] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," IEEE Trans. Automat. Control **58**(9), 2426–2431 (2013).

[26] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," 2012, see http://cvxr.com/cvx.

[27] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "On the optimal step-size selection for the alternating direction method of multipliers," in *Preprints of the 3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems, Santa Barbara*, CA, 2012 (International Federation of Automatic Control, 2012), pp. 139–144.

[28] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems" IEEE Trans. Automat. Control (submitted); e-print: arXiv:1306.2454.

[29] P. J. Schmid and D. S. Henningson, *Stability and Transition in Shear Flows* (Springer-Verlag, New York, NY, 2001).

[30] J. A. C. Weideman and S. C. Reddy, "A MATLAB differentiation matrix suite," ACM Trans. Math. Softw. **26**(4), 465–519 (2000).

[31] C. K. W. Tam, "Supersonic jet noise," Annu. Rev. Fluid Mech. **27**, 17–43 (1995).

[32] F. C. Frate and J. E. Bridges, "Extensible rectangular nozzle model system," AIAA Paper 2011–975, 2011.

[33] J. W. Nichols, F. E. Ham, and S. K. Lele, "High-fidelity large-eddy simulation for supersonic rectangular jet noise prediction," AIAA Paper 2011–2919, 2011.

[34] P. G. Constantine and D. F. Gleich, "Tall and skinny QR factorizations in MapReduce architectures," in *Proceedings of the 2nd International Workshop on MapReduce and its Applications*, 2011 (ACM, New York, NY, 2011), pp. 43–50.

[35] S. Kakac and H. Liu, *Heat Exchangers: Selection, Rating and Thermal Design* (CRC Press, Boca Raton, FL, 1997).

[36] P. Rollet-Miet, D. Laurence, and J. Ferziger, "LES and RANS of turbulent flow in tube bundles," Int. J. Heat Fluid Flow **20**, 241–254 (1999).

[37] D. Sumner, S. Price, and M. Paidoussis, "Flow-pattern identification for two staggered circular cylinders in cross-flow," J. Fluid Mech. **411**, 263–303 (2000).

[38] S. Benhamadouche and D. Laurence, "LES, coarse LES, and transient RANS comparisons on the flow across a tube bundle," Int. J. Heat Fluid Flow **24**, 470–479 (2003).

[39] C. Moulinec, M. Pourquié, B. Boersma, T. Buchal, and F. Nieuwstadt, "Direct numerical simulation on a Cartesian mesh of the flow through a tube bundle," Int. J. Comput. Fluid Dyn. **18**, 1–14 (2004).

[40] C. Liang and G. Papadakis, "Large eddy simulation of cross-flow through a staggered tube bundle at subcritical Reynolds number," J. Fluids Struct. **23**, 1215–1230 (2007).

[41] Y. Hassan and H. Barsamian, "Turbulence simulation in tube bundle geometries using the dynamic subgrid-scale model," Nucl. Tech. J. **128**, 58–74 (1999).

[42] See supplementary material at http://dx.doi.org/10.1063/1.4863670 for a brief description of MATLAB implementation of the Sparsity-Promoting Dynamic Mode Decomposition (DMDSP) algorithm and for additional information about the examples considered in this paper, including Matlab source codes and problem data.