Dynamic Mode Decomposition: Theory and Data Reconstruction

Tim Krake*, Daniel Weiskopf†, Bernhard Eberhardt‡

February 15, 2022

Abstract

Dynamic Mode Decomposition (DMD) is a data-driven decomposition technique extracting spatio-temporal patterns of time-dependent phenomena. In this paper, we perform a comprehensive theoretical analysis of various variants of DMD. We provide a systematic advancement of these and examine the interrelations. In addition, several results of each variant are proven. Our main result is the exact reconstruction property. To this end, a new modification of scaling factors is presented and a new concept of an error scaling is introduced to guarantee an error-free reconstruction of the data.

Keywords. Dynamic Mode Decomposition, Data Reconstruction, Numerical Analysis, Matrix Decomposition

1 Introduction

The analysis of time-dependent phenomena is at the heart of investigation in a broad range of scientific research. Within these studies, the integration of data in the form of time-series has increased considerably. Therefore, the application of innovative algorithms is necessary to gain deep insights into the characteristics of data. In this paper, we address time-series analysis by Dynamic Mode Decomposition (DMD), which was first introduced by Schmid and Sesterhenn in 2008.

DMD is a data-driven and model-free algorithm extracting spatio-temporal patterns in the form of so-called DMD modes and DMD eigenvalues. As an efficient tool in fluid mechanics, DMD has gained much attention. DMD has been investigated on both practical and theoretical grounds. Nonetheless, the focus of these analyses was mainly a practical one. For example, various types of flow were considered, such as airflow around an airfoil, fuel flow in a combustion chamber, or heat conduction in various cases. Completely different fields of application comprise financial trading, video processing, epidemiology, neuroscience, and control theory.

In contrast, we focus on theoretical investigations. The paper is thus structured as follows: After discussing related work, we introduce the theoretical framework of DMD dealing with the background mechanisms. In this process, we define the so-called system matrix,

*University of Stuttgart and Hochschule der Medien, Germany (tim.krake@visus.uni-stuttgart.de)
†University of Stuttgart, Germany (weiskopf@visus.uni-stuttgart.de)
‡Hochschule der Medien, Germany (eberhardt@hdm-stuttgart.de)
which is pioneering for DMD and prove the following results: a characterization for the exactness and diagonalizability of the system matrix as well as the resulting reconstruction of data with its spectral components. These theorems are central for the following sections introducing the three common variants of DMD: The original formulation \cite{9}, the modification by a singular value decomposition \cite{8}, and Exact-Dynamic Mode Decomposition \cite{10}. In this context, a systematic advancement will be presented that clarifies precisely the interrelation of these algorithms. This especially includes algebraic identities as well as spectral-theoretic results leading, e.g., to a new approach for the extension to the most recent variant of DMD. In addition, the exact reconstruction property of DMD will be proven for each DMD variant that guarantees an error-free reconstruction of the data. To this end, a new variant of scaling factors is introduced involving a new concept of an error scaling for the reconstruction of the first snapshot. Some concluding remarks will be given in the last section.

2 Related Work

Rowley et al. provided a first theoretical investigation \cite{6} for the fundamental version of DMD, here denoted as Companion Dynamic Mode Decomposition (CDMD). They dealt with the reconstruction property of CDMD, however, they did not take appropriate scaling factors into account. An algorithmic improvement through the singular value decomposition was achieved by Schmid resulting in another variant of DMD \cite{7}. We refer to this algorithm as Singular Value Decomposition Dynamic Mode Decomposition (SDMD). The reconstruction property of SDMD was mentioned by Chen et al. as well as further properties of CDMD \cite{1}. Tu et al. introduced the advancement of SDMD to Exact-Dynamic Mode Decomposition (EXDMD) \cite{10}, which is the most recent version of DMD. They prove basic algebraic identities and show primarily spectral-theoretic connections between these two algorithms. We generalize and extend all results or derive them as a corollary. In addition, we present a new approach for the extension of SDMD to EXDMD characterizing precisely the connection between these two algorithms. Despite these theoretical investigations, the problem of an exact reconstruction is left open. However, Jovanovic et al. \cite{3} as well as Drmač et al. \cite{2} discussed efficient techniques for finding appropriate coefficients by solving certain (convex) minimization problems. We introduce a new variant of scaling factors that lead to an error-free reconstruction of the snapshots under appropriate conditions.

3 Theoretical Framework

This section is dedicated to the basic theoretical background of DMD. In this context, the general setting will be presented as well as an intuitive interpretation of the principles of DMD. These are crucial for the precise understanding of DMD, forming the basis for the subsequent sections. In addition, basic notation will be formalized and consistently used in this paper.

The application of DMD starts with the availability of data that may stem either from empirical experiments or numerical simulations alike. The objective of DMD is to extract spatio-temporal patterns out of the data in the form of DMD modes, eigenvalues, and amplitudes. As the modes are related to spatial structures, the corresponding eigenvalues determine the temporal behavior of these. The amplitudes characterize the impact of individual modes on the whole system, i.e. the dominance structure.
Now, consider data (snapshots) \( x_0, x_1, \ldots, x_m \in \mathbb{C}^n \) with the following two quantities:

\[
\begin{align*}
n &= \text{“size/dimension of the data points”}, \\
m &= \text{“number of data points”}.
\end{align*}
\]

In the application areas of DMD such as fluid dynamics or non-linear dynamics, the connection between these variables is typically given by \( n \gg m \), which means that the size of the data points is considerably larger than the number of snapshots. In this context, typical values are \( n \approx 10^6 \text{–} 10^{12} \) (depending on whether we address 2D or 3D scenarios) and \( m \approx 100 \text{–} 1000 \). This basic setting is crucial for understanding the principles of DMD and will be assumed in the following derivation. However, DMD can also be mathematically formulated and applied without this assumption, as we will see later.

In short, DMD calculates the relevant dynamic information of a high-dimensional linear operator that connects the given data points \( x_0, x_1, \ldots, x_m \) in a least square sense, without explicitly computing it. This is achieved by an eigenvalue decomposition of a low-dimensional representation. The corresponding eigenvectors will be embedded as DMD modes into the high-dimensional space endowed with appropriate scaling factors, the DMD amplitudes.

In order to obtain the high-dimensional matrix \( A \in \mathbb{C}^{n \times n} \) connecting the data points, we consider the following (least-squares) minimization problem:

\[
\min_{A \in \mathbb{C}^{n \times n}} \sum_{j=0}^{m-1} \| Ax_j - x_{j+1} \|^2_2.
\]

Note that the high dimensionality stems from the fact \( n \gg m \). An explicit solution of \( A \) is necessary to formulate an algorithmic approach. To this end, we rewrite the data into the matrices

\[
X = \begin{bmatrix} x_0 & x_1 & \ldots & x_{m-1} \\ 1 & 1 & \ldots & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x_1 & x_2 & \ldots & x_m \\ 1 & 1 & \ldots & 1 \end{bmatrix} \in \mathbb{C}^{n \times m}
\]

obtaining the following equivalent minimization problem:

\[
\min_{A \in \mathbb{C}^{n \times n}} \| AX - Y \|_F^2,
\]

where \( \| \cdot \|_F \) denotes the Frobenius norm of a matrix. An explicit solution is now given by

\[
A = Y X^+ \in \mathbb{C}^{n \times n},
\]

where \( X^+ \) denotes the Moore-Penrose pseudoinverse [4] of \( X \). Since the pseudoinverse always exists, the solution \( A \) can be used for an algorithmic formulation.

Now, assuming the diagonalizability of the matrix \( A \), i.e., \( A = V \Lambda V^{-1} \) with the matrices \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \) and \( V = [v_1, v_2, \ldots, v_n] \) containing the eigenvalues and eigenvectors respectively, we obtain the characteristic reconstruction property of the matrix \( A \) by

\[
x_k \approx A^k x_0 = (V \Lambda V^{-1})^k x_0 = V \Lambda^k V^{-1} x_0 = V \Lambda^k b = \sum_{j=1}^{n} \lambda_j^k b_j v_j
\]

for \( k = 0, 1, \ldots, m \), where \( b = (b_1, \ldots, b_n)^T \) are the coefficients of the linear combination of \( x_0 \) in the eigenvector basis, i.e. \( b = V^{-1} x_0 \). Since the rank of \( A \) is at most
\[
\min\{\text{rank}(X), \text{rank}(Y)\}\text{ and consequently not more than } m, \text{ there are at least } n - m \text{ eigenvalues of } A \text{ that are equal to zero. The dynamic behavior will be thus captured by at most } m \text{ components, which are considerable fewer components. Consequently, we obtain the following reconstruction of the data:}
\]
\[
x_0 \approx \sum_{j=1}^{m} b_j v_j + q_0 \quad x_k \approx \sum_{j=1}^{m} \lambda_j b_j v_j,
\]
for \( k = 1, \ldots, m \), where \( q_0 \) is the resulting error arising from the missing \( m - n \) components. In sum, we gain a reasonable low-dimensional decomposition of the data into the triples \((\lambda_j, v_j, b_j) \in \mathbb{C} \times \mathbb{C}^n \times \mathbb{C}\), providing an instrument for diagnostic approaches as well as a tool for prediction, long-term analysis, and stability analysis.

The different versions of DMD presented in the subsequent sections are based on various techniques to produce a low-dimensional representation of the matrix \( A \) in order to (approximately) compute its eigenvalues and eigenvectors as well as new appropriate scaling factors. These procedures yield similar triples that will be denoted by \((\lambda_j, \vartheta_j, a_j) \in \mathbb{C} \times \mathbb{C}^n \times \mathbb{C}\) throughout the paper. These triples consist of the so-called DMD eigenvalues, DMD modes, and DMD amplitudes corresponding to a particular algorithm (see each section).

Before studying the variants of DMD, we first concentrate on an analysis of the high-dimensional structures involving the matrix \( A \). Through a deeper understanding of the matrix \( A \) representing the starting point of DMD, we obtain insights into the desired action of DMD. In particular, the success of an error-free reconstruction of DMD depends on the following two aspects:

1. The exactness of the matrix \( A \), i.e., \( AX = Y \).
2. The diagonalizability of the matrix \( A \).

These two aspects will be examined throughout this section, however, before, the matrix \( A \) will be captured in the following definition.

**Definition 3.1** For data \( x_0, \ldots, x_m \in \mathbb{C}^n \) with associated matrices \( X = [x_0 \ldots x_{m-1}] \) and \( Y = [x_1 \ldots x_m] \), we call the matrix \( A = YX^* \in \mathbb{C}^{n \times n} \) the system matrix to the data \( x_0, \ldots, x_m \).

First, we recall some well-known facts. The definition is well-defined, as the pseudoinverse always exists and is unique. Furthermore, the system matrix is the unique solution to the the minimization problem \( \min_{A \in \mathbb{C}^{n \times n}} \|AX - Y\|_F^2 \), if the rows of the matrix \( X \) are linearly independent (or equivalently, if the matrix \( X \) is surjective).

An important condition is the exactness of the system matrix, i.e., the equality \( AX = Y \) or equivalently \( Ax_j = x_{j+1} \) for \( j = 0, \ldots, m - 1 \). The following proposition characterizes this property on linear functionals.

**Proposition 3.2** Let the system matrix \( A \) be given to the data \( x_0, \ldots, x_m \in \mathbb{C}^n \) as well as an arbitrary vector \( w \in \mathbb{C}^n \). Then the following statements are equivalent:

(i) \( \ker(X) \subseteq \ker(w^*Y) \), i.e., \( Xz = 0 \implies w^*Yz = 0 \) for all \( z \in \mathbb{C}^n \).

(ii) The system matrix is exact on \( w \), i.e. \( w^*AX = w^*Y \).
Proof. “(i) \(\implies\) (ii)”. Consider the following equation
\[ w^*Y - w^*AX = w^*Y - w^*YX^+X = w^*Y(I - P_X), \]
where \(P_X\) is the orthogonal projection onto the image of \(X^*\). Therefore \(I - P_X\) is the orthogonal projection onto the kernel of \(X\) and consequently the assertion follows by the assumption \(\ker(X) \subseteq \ker(w^*Y)\).

“(ii) \(\implies\) (i)”. Let \(z \in \ker(X)\), i.e., \(Xz = 0\). This implies \(w^*AXz = 0\), which is by assumption equivalent to \(w^*Yz = 0\). Hence \(z \in \ker(w^*Y)\).

A simple consequence of this proposition is the following corollary [10, Theorem 2].

**Corollary 3.3** Let the system matrix \(A\) be given to the data \(x_0, \ldots, x_m \in \mathbb{C}^n\). Then the following assertions are equivalent:

(i) \(\ker(X) \subseteq \ker(Y)\).

(ii) The system matrix is exact, i.e. \(AX = Y\).

In the case of linear independent data points \(x_0, \ldots, x_{m-1}\), the system matrix is exact, since \(\ker(X) = \{0\}\) and hence condition (i) of [Corollary 3.3] is trivially satisfied. In this context, the condition \(n \gg m\) (which is typical for the application areas of DMD) suggests the linear independence of the data. Consequently, the first aspect (exactness of the system matrix) is characterized. For a full reconstruction of the data, however, we still need the diagonalizability of the system matrix. To this end, we examine the inner structure of the system matrix, i.e., its kernel and image.

**Lemma 3.4** Let the system matrix \(A\) be given to the data \(x_0, \ldots, x_m \in \mathbb{C}^n\), where \(x_0, \ldots, x_{m-1}\) and \(x_1, \ldots, x_m\) are linear independent, respectively. Then the dimension of the image and the kernel of \(A\) is given by
\[ \dim \text{im}(A) = m, \quad \dim \ker(A) = n - m. \]

**Proof.** For the first assertion, we use the following rank inequality:
\[ m = \text{rank}(Y) + \text{rank}(X^+) - m \leq \text{rank}(YX^+) \leq \min\{\text{rank}(Y), \text{rank}(X^+)\} = m, \]
which implies \(\text{rank}(A) = m\) or equivalently \(\dim \text{im}(A) = m\). Consequently, we obtain that \(\dim \ker(A) = n - \dim \text{im}(A) = n - m\).

**Corollary 3.5** Let the system matrix \(A\) to the data \(x_0, \ldots, x_m \in \mathbb{C}^n\) be given, where \(x_0, \ldots, x_{m-1}\) and \(x_1, \ldots, x_m\) are linear independent, respectively. If \(A\) has \(m\) non-zero distinct eigenvalues, then it is diagonalizable.

Regarding the reconstruction property of the system matrix, the final result will be stated in the following theorem, which deals with both sufficient and necessary conditions. For a simple notational handling of the proof, we define the Vandermonde matrix.

**Definition 3.6** For \(\lambda_1, \ldots, \lambda_m \in \mathbb{C}\) we define the \(k\)-\(K\)-Vandermonde matrix by
\[ \text{Vand}(\lambda_1, \ldots, \lambda_m; k, K) = \begin{pmatrix} \lambda_{1}^{k-1} & \lambda_{1}^{k} & \cdots & \lambda_{1}^{K-1} \\ \lambda_{2}^{k-1} & \lambda_{2}^{k} & \cdots & \lambda_{2}^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m}^{k-1} & \lambda_{m}^{k} & \cdots & \lambda_{m}^{K-1} \end{pmatrix} \in \mathbb{C}^{m \times (K - (k - 1))}, \]
with \(\text{Vand}(\lambda_1, \ldots, \lambda_m) = \text{Vand}(\lambda_1, \ldots, \lambda_m; 1, m)\) for the usual Vandermonde matrix.
Theorem 3.7 (Reconstruction-property system matrix) Let the system matrix \( A \) be given to the data \( x_0, \ldots, x_m \in \mathbb{C}^n \). Then the two following assertions are equivalent:

(i) The system matrix \( A \) has the following properties:

a) \( \ker(X) \subseteq \ker(Y) \).

b) \( A \) is diagonalizable, where the non-zero eigenvalues are distinct.

c) \( \text{rank}(Y) = r_1 \).

(ii) There are distinct numbers \( 0 \neq \lambda_1, \ldots, \lambda_r \in \mathbb{C} \), coefficients \( 0 \neq b_1, \ldots, b_r \in \mathbb{C} \) and linearly independent vectors \( v_1, \ldots, v_r \) with \( r_1 \leq r_2 \leq n \) and \( r_1 \leq m \) such that the following identities hold for \( k = 1, \ldots, m \):

\[
x_0 = \sum_{j=1}^{r_2} b_j v_j, \quad x_k = \sum_{j=1}^{r_1} \lambda_j^k b_j v_j.
\]

In this case, the non-zero distinct numbers \( \lambda_1, \ldots, \lambda_r \) are the eigenvalues of the system matrix and the related (scaled) vectors \( v_1, \ldots, v_r \) are the corresponding eigenvectors. The remaining vectors \( v_{r_1+1}, \ldots, v_r \) are eigenvectors of the system matrix to the eigenvalue zero.

Proof. “(i) \( \implies \) (ii)” By condition c), the system matrix has at most \( r_1 \) non-zero eigenvalues, because \( \text{rank}(A) \leq \min\{\text{rank}(X^+), \text{rank}(Y)\} \leq r_1 \). As the system matrix is diagonalizable by assumption b), there exist non-zero distinct eigenvalues \( \lambda_1, \ldots, \lambda_r \) with \( r \leq r_1 \) and zero eigenvalues \( \lambda_{r+1}, \ldots, \lambda_n \) with corresponding eigenvectors \( v_1, \ldots, v_r, v_{r+1}, \ldots, v_n \). Rewriting into matrices leads to \( W^{-1}AW = \Lambda \) for \( W = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \). By Corollary 3.3, the first assumption a) is equivalent to \( AX = Y \), which implies the identity

\[
x_k = A^k x_0 = W A^k W^{-1} Wb = WA^k b = \sum_{j=1}^{\tilde{r}} \lambda_j^k b_j v_j,
\]

for \( k = 0, 1, \ldots, m \), where \( b = (b_1, \ldots, b_n)^T \) contains the coefficients of the linear combination of \( x_0 \) in the eigenvector basis, i.e., \( b = W^{-1} x_0 \). Some of the coefficients \( b_j \) may be zero, such that we obtain, after reordering, the identity

\[
x_k = \sum_{j=1}^{\tilde{r}} \lambda_j^k b_j v_j,
\]

for \( k = 1, \ldots, m \) and \( b_1, \ldots, b_{\tilde{r}} \neq 0 \) with \( \tilde{r} \leq r \leq r_1 \). Since the rank of \( Y \) is \( r_1 \) by condition c), the data points \( x_1, \ldots, x_m \) span an \( r_1 \)-dimensional vector subspace. As the vectors \( v_1, \ldots, v_{\tilde{r}} \) are linearly independent, the sum have to has at least \( r_1 \) terms. Hence, \( \tilde{r} = r = r_1 \) and \( b_1, \ldots, b_{r_1} \neq 0 \) as well as \( \lambda_1, \ldots, \lambda_{r_1} \neq 0 \). Finally, the first data point can be expressed by the remaining non-zero coefficients \( b_1, \ldots, b_{r_1}, b_{r_1+1}, \ldots, b_{r_2} \) through

\[
x_0 = \sum_{j=1}^{r_2} b_j v_j.
\]
“(ii) $\implies (i)$”. First, we define the following matrices

$\begin{align*}
W_1 &= [v_1 \ldots v_r_1] \in \mathbb{C}^{n \times r_1}, & \quad W_2 &= [v_1 \ldots v_r_2] \in \mathbb{C}^{n \times r_2} \\
K_{r_1} &= \text{diag}(b_1, \ldots, b_{r_1}) \in \mathbb{C}^{r_1 \times r_1}, & \quad K_{r_2} &= \text{diag}(b_1, \ldots, b_{r_2}) \in \mathbb{C}^{r_2 \times r_2} \\
M_{r_1} &= \text{Vand}(\lambda_1, \ldots, \lambda_{r_1}; 1, m) \in \mathbb{C}^{r_1 \times m}, & \quad M_{r_2} &= \text{Vand}(\lambda_1, \ldots, \lambda_{r_1}, 0, \ldots, 0; 1, m) \in \mathbb{C}^{r_2 \times m} \\
\Lambda_{r_1} &= \text{diag}(\lambda_1, \ldots, \lambda_{r_1}) \in \mathbb{C}^{r_1 \times r_1}, & \quad \Lambda_{r_1 \times r_2} &= [\Lambda_{r_1} \ 0] \in \mathbb{C}^{r_1 \times r_2}.
\end{align*}$

Now, we can rewrite the data matrices $X$ and $Y$ with the above introduced notation by

$X = W_{r_2} K_{r_2} M_{r_2}, \quad Y = W_{r_1} K_{r_1} \Lambda_{r_1} M_{r_1} = W_{r_1} K_{r_1} \Lambda_{r_1 \times r_2} M_{r_2}.$

Consider the linear operator $A_* = W_{r_1} \Lambda_{r_1 \times r_2} W_{r_2}^+ \in \mathbb{C}^{n \times n}$. This operator exactly connects the data, since

$A_* X = W_{r_1} \Lambda_{r_1 \times r_2} W_{r_2}^+ W_{r_2} K_{r_2} M_{r_2} = W_{r_1} \Lambda_{r_1 \times r_2} K_{r_2} M_{r_2} = W_{r_1} K_{r_1} \Lambda_{r_1 \times r_2} M_{r_2} = Y.$

As a result, the system matrix $A = Y X^+$ is exact, too, because it minimizes the problem $\|AX - Y\|_F$, i.e., $AX = Y$. By Corollary 3.3, the first assertion $\ker(X) \subseteq \ker(Y)$ is proven.

The second claim, $\text{rank}(Y) = r_1$, follows from the application of simple rank inequalities onto $Y = W_{r_1} K_{r_1} \Lambda_{r_1} M_{r_1}$ together with the conditions in (ii).

As a consequence, we have $\text{rank}(A) \leq r_1$, which implies that $A$ has at most $r_1$ non-zero eigenvalues. Now, if we prove that $\lambda_1, \ldots, \lambda_{r_1}$ are these eigenvalues, then the proof is complete, since this implies the diagonalizability of the system matrix with distinct non-zero eigenvalues (Corollary 3.3). More precisely, we have to show that

$AW_{r_1} K_{r_1} = W_{r_1} K_{r_1} \Lambda_{r_1}.$

Note that the column vectors of $W_{r_1} K_{r_1}$ (representing the eigenvectors) are already non-zero by assumption. To prove the equality we consider two cases:

1. case: $r_1 < m$. Defining the matrix $\hat{M}_{r_1} = \text{Vand}(\lambda_1, \ldots, \lambda_{r_1}; 2, m) \in \mathbb{C}^{r_1 \times m-1}$ (i.e., without the first column), then we obtain by the exactness of the system matrix

$AW_{r_1} K_{r_1} \hat{M}_{r_1} = W_{r_1} K_{r_1} \Lambda_{r_1} \hat{M}_{r_1}.$

Since the eigenvalues are distinct and $r_1 < m$, the matrix $\hat{M}_{r_1}$ has a right inverse (which is given by the pseudoinverse), hence, we arrive at the desired assertion.

2. case: $r_1 = m$. From $\text{rank}(Y) = m$ follows $\ker(Y) = \{0\}$ and, consequently, $\ker(X) = \{0\}$ by the already proven condition a). Hence, $\text{rank}(X) = m$, too and by Lemma 3.4 the system matrix has $\text{rank}(A) = m$ and $\dim \ker(A) = n - m$. As the second identity geometrically implies $\text{im}(Y) \subseteq \langle v_1, \ldots, v_m \rangle$, we obtain the following relationship

$\text{im}(A) \subseteq \text{im}(Y) \subseteq \langle v_1, \ldots, v_m \rangle.$

Since $\text{rank}(A) = m$ and $v_1, \ldots, v_m$ are linearly independent, it follows $\text{im}(A) = \langle v_1, \ldots, v_m \rangle$. Consequently, the remaining linearly independent vectors $v_{m+1}, \ldots, v_{r_2}$ belong to the kernel of the system matrix $A$. Rewriting the matrix $X$ into $X = W_{r_1} K_{r_1} M_{r_1} + q e_1^T$ with $q = \sum_{j=m+1}^{r_2} b_j v_j$ with the first standard basis vector $e_1$, we obtain

$W_{r_1} K_{r_1} \Lambda_{r_1} M_{r_1} = Y = AX = AW_{r_1} K_{r_1} M_{r_1} + A q e_1^T = AW_{r_1} K_{r_1} M_{r_1},$

because $q \in \ker(A)$. Multiplying this equation by $M_{r_1}^+$ from the right side, we obtain the desired algebraic identity, since $M_{r_1}^+$ is a right inverse. \hfill \square
4 Companion Dynamic Mode Decomposition (CDMD)

In 2008, Schmid and Sesterhenn presented the first version of DMD [9]. This variant will be referred to as Companion Dynamic Mode Decomposition (CDMD). CDMD has been investigated by experimental and numerical data by Schmid [7, 8]. The first approaches of a theoretic analysis were performed by Rowley et al. [6]. Before we discuss the derivation of CDMD, the companion matrix will be defined, which is eponymous for this variant of DMD.

**Definition 4.1** For a vector \( c = (c_0, \ldots, c_{m-1})^T \in \mathbb{C}^m \), we define the matrix \( C_c \in \mathbb{C}^{m \times m} \) of the form

\[
\begin{pmatrix}
0 & 0 & \cdots & c_0 \\
1 & 0 & \cdots & c_1 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 1 & c_{m-1}
\end{pmatrix}
\]

as the companion matrix to the vector \( c \).

For the derivation of CDMD, we consider a given data set \( x_0, \ldots, x_m \in \mathbb{C}^n \). Again, we rewrite the data points into the matrices \( X = [x_0 \ldots x_{m-1}] \) and \( Y = [x_1 \ldots x_m] \). Now, instead of using the system matrix \( A \), which connects the matrices \( X \) and \( Y \) (from the left), i.e., \( AX \approx Y \), another approach is to look for a right-hand multiplied matrix \( C \in \mathbb{C}^{m \times m} \) such that \( Y \approx X C \). By construction, we can choose \( C \) as a companion matrix \( C_c \) and, hence, the problem reduces to find a vector \( c \) such that

\[
Y = XC_c + qe_m^T
\]

with a minimal error \( q \in \mathbb{C}^n \), where \( e_m \) represent the last standard basis vector. Under the assumption \( n \gg m \), the companion matrix \( C_c \in \mathbb{C}^{m \times m} \) is substantially lower dimensional than the system matrix \( A \in \mathbb{C}^{n \times n} \). In addition, with decreasing error \( q \) the companion matrix approximates the spectral-theoretic properties of \( A \). In fact, for an eigenvector \( v \) of \( C_c \) to the eigenvalue \( \lambda \) the transformed eigenvector \( \vartheta = Xv \) satisfies

\[
A\vartheta = AXv \approx XC_Cv = X\lambda v = \lambda \vartheta.
\]

In sum, the computation of \( C_c \) is reduced to the calculation of the associated vector \( c = (c_0, \ldots, c_{m-1})^T \). This vector minimizes the error \( q = x_m - Xc \) and hence only need to solve the following minimization problem

\[
\min_{c \in \mathbb{C}^m} \|Xc - x_m\|_2^2,
\]

which will be solved by \( c = X^+x_m \). Consequently, we can formulate CDMD as an algorithm, however, before, we define the companion matrix to a given data set.

**Definition 4.2** For data \( x_0, \ldots, x_m \in \mathbb{C}^n \) with associated matrix \( X = [x_0 \ldots x_{m-1}] \), we call the companion matrix \( C_c \) to the vector \( c = X^+x_m \in \mathbb{C}^m \) the companion matrix to the data \( x_0, \ldots, x_m \).

**Algorithm 4.3 (Companion Dynamic Mode Decomposition (CDMD))**

Input: Data \( x_0, \ldots, x_m \in \mathbb{C}^n \).
Output: DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \in \mathbb{C} \), DMD modes \( \vartheta_1, \ldots, \vartheta_m \in \mathbb{C}^n \), and DMD amplitudes \( a_1, \ldots, a_m \in \mathbb{C} \).
1. Define the matrices $X := [x_0 \ldots x_{m-1}], Y := [x_1 \ldots x_m] \in \mathbb{C}^{n \times m}$.

2. Compute the $c = X^+x_m \in \mathbb{C}^m$.

3. Construct the companion matrix $C_c \in \mathbb{C}^{m \times m}$ to the vector $c$.

4. Compute the DMD eigenvalues $\lambda_1, \ldots, \lambda_m$ and eigenvectors $W^{-1} = [v_1 \ldots v_m]$ of $C_c$ to the vector $c$.

5. Calculate DMD modes $\Theta = [\vartheta_1 \ldots \vartheta_m] = XW^{-1} \in \mathbb{C}^{n \times m}$.

6. Compute the DMD amplitudes by $K = \text{diag}(a_1, \ldots, a_m) = (V \text{and}(\lambda_1, \ldots, \lambda_m)W^{-1})^+$.

The algorithm described here differs from the standard literature as we introduce the necessary concept of DMD amplitudes. Note, that the matrix $K$ in step 6, which defines the DMD amplitudes, may not be diagonal. However, if the eigenvalues are distinct, then $K$ is a diagonal matrix, as we will observe later. The current definition of the DMD amplitudes seems to be obscure and not very intuitive. Later on, we will show that these DMD amplitudes are the right scaling factors for an exact reconstruction. In addition, we will prove that this choice equals $(a_1, \ldots, a_m)^T = \Theta x_0$ under some further assumption.

For a theoretical investigation, we need a more compact representation of the companion matrix $C_c$ from Algorithm 4.3. To this end, consider the following minimization problem

$$\min_{C \in \mathbb{C}^{m \times m}} \|XC - Y\|_F^2,$$

which is closely related to the construction of the companion matrix. An explicit solution of the minimization problem is given by $C = X^+Y \in \mathbb{C}^{m \times m}$. This matrix will be captured in the following definition.

**Definition 4.4** For data $x_0, \ldots, x_m \in \mathbb{C}^n$ with associated matrices $X = [x_0 \ldots x_{m-1}]$ and $Y = [x_1 \ldots x_m]$, we call the matrix $C = X^+Y \in \mathbb{C}^{n \times n}$ the twisted system matrix to the data $x_0, \ldots, x_m$.

**Remark 4.5** At first glance, the two approaches seems to be equivalent. However, note that the twisted system matrix $C$ and the companion matrix $C_c$ are not necessarily equal or similar. This fact can be observed by a rank truncation of the matrices $X$ and $Y$. In particular, the companion matrix $C_c$ has at least rank $m - 1$, because the first $m - 1$ columns are linearly independent. However, the rank of the matrix $C = X^+Y$ depends only on the matrices $X$ and $Y$ and therefore it may be less than $m - 1$, which implies a non-similarity in general.

The following lemma and corollary characterize the relations between the three objects $A, C,$ and $C_c$.

**Lemma 4.6** Let the system matrix $A$, the twisted system matrix $C$, and the companion matrix $C_c$ be given to the data $x_0, \ldots, x_m \in \mathbb{C}^n$ as well as the projections $P_X = XX^+$ and $P_{X^*} = X^+X$ onto $\text{im}(X) \subseteq \mathbb{C}^n$ and $\text{im}(X^*) \subseteq \mathbb{C}^m$, respectively. Then the following assertions hold:

1. $X^+q = 0$, especially $Aq = 0$.
2. $P_{X^*}C_cP_{X^*} = CP_{X^*} = X^+AX$. 

9
(3) \( X_c X^+ = X C X^+ = P X A \).

**Proof.** The identities are proven by simple algebraic manipulations. \( \square \)

A direct consequence of this lemma is that the twisted system matrix and the companion matrix are equal, if the first \( m \) snapshots are linearly independent. In addition, in the case of \( n \gg m \), it illustrates that the companion matrix \( C_c \in C^{m \times m} \) is a low-dimensional representation of the system matrix \( A \in C^{n \times n} \).

**Corollary 4.7** Let the system matrix \( A \), the twisted system matrix \( C \) and the companion matrix \( C_c \) be given to the data \( x_0, \ldots, x_m \in C^n \). If \( x_0, \ldots, x_m-1 \) are linearly independent, then the following assertions hold:

1. \( C_c = C \).
2. \( C_c = X^+ AX \).

**Proof.** Since \( x_0, \ldots, x_{m-1} \) are linearly independent, the identity \( X^+ X = I \) holds, which implies the assertions by Lemma 4.6 \( \square \)

The property of a low-dimensional representation suggests that CDMD inherits the characteristic reconstruction of the system matrix. Indeed, the following theorem proves this fact using the new concept of DMD amplitudes from Algorithm 4.3. In the literature a similar theorem is known [6, Theorem 1] that does not account for amplitudes. Hence, the choice of the corresponding modes is not appropriate for the reconstruction as these are computed up to a scaling (since they stem from eigenvectors).

**Theorem 4.8 (Reconstruction-property of CDMD)** Let DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \), DMD amplitudes \( a_1, \ldots, a_m \), and DMD modes \( \vartheta_1, \ldots, \vartheta_m \) be given by Algorithm 4.3 to the data \( x_0, \ldots, x_m \in C^n \). If the DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \) are distinct, then the following identities hold:

\[
x_k = \sum_{j=0}^{m} \lambda_k^j a_j \vartheta_j, \quad x_m = \sum_{j=0}^{m} \lambda_m^j a_j \vartheta_j + q,
\]

for \( k = 0, \ldots, m - 1 \) and \( q = x_m - X_c \).

**Proof.** Since the eigenvalues are distinct, the companion matrix \( C_c \) will be diagonalized by the Vandermonde matrix:

\[
\text{Vand}(\lambda_1, \ldots, \lambda_m) C_c \text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1} = \Lambda.
\]

However, the eigenvectors \( [z_1 \ \ldots \ \ z_m] = \text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1} \) may not coincide with the eigenvectors \( v_1, \ldots, v_m \) of the companion matrix produced by Algorithm 4.3. As the eigenvalues are distinct, there exist scaling factors \( \alpha_1, \ldots, \alpha_m \in C \) such that for \( j = 1, \ldots, m \):

\[
\alpha_j v_j = z_j.
\]

For the eigenvectors \( v_1, \ldots, v_m \) and scaling factors \( \alpha_1, \ldots, \alpha_m \) we define the matrices \( W^{-1} = [v_1 \ \ldots \ \ v_m] \) and \( K = \text{diag}(\alpha_1, \ldots, \alpha_m) \), respectively. As a result, we get the following connection

\[
W^{-1} K = \text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1}.
\]
Consequently, \( K = (V \text{and}(\lambda_1, \ldots, \lambda_m)W^{-1})^{-1} \), which shows that \( K \) equals the matrix in step 6 of Algorithm 4.3 consisting of the DMD amplitudes, i.e., \( a_j = a_j \) for \( j = 1, \ldots, m \). Multiplying the above equation by the matrix \( X \), we obtain

\[
XW^{-1}K = XV\text{and}(\lambda_1, \ldots, \lambda_m)^{-1}.
\]

By this notation the DMD modes \( \Theta = [\vartheta_1 \ldots \vartheta_m] \) are given by \( \Theta = XW^{-1} \) and, therefore, the first assertions follows by rearranging the above equation to \( X = \Theta KV\text{and}(\lambda_1, \ldots, \lambda_m) \). The second identity is a result of the following calculation

\[
x_m = Xc + q = XC\vartheta_m^T + q = \Theta K V\text{and}(\lambda_1, \ldots, \lambda_m) \vartheta_m^T + q = \Theta K \Lambda V\text{and}(\lambda_1, \ldots, \lambda_m) \vartheta_m^T + q = \sum_{j=0}^{m} \lambda_j^m a_j \vartheta_j + q.
\]

\[\square\]

**Corollary 4.9** Let DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \), DMD amplitudes \( a_1, \ldots, a_m \), and DMD modes \( \vartheta_1, \ldots, \vartheta_m \) be given by Algorithm 4.3 to the data \( x_0, \ldots, x_m \in \mathbb{C}^n \). If the DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \) are distinct and \( x_0, \ldots, x_{m-1} \) are linearly independent, then the DMD amplitudes \( a = (a_1, \ldots, a_m)^T \) can be calculated by

\[
a = \Theta^+ x_0.
\]

**Proof.** By Theorem 4.8 we obtain the relation \( x_0 = \Theta a \), where \( \Theta = [\vartheta_1 \ldots \vartheta_m] = XW^{-1} \) with \( W^{-1} = [v_1 \ldots v_m] \). Since the matrix \( X \) and \( W^{-1} \) have full rank, the DMD modes will be linear independent and, consequently, there exist a left-inverse of \( \Theta \), which is given by its pseudoinverse \( \Theta^+ \). Hence, \( a = \Theta^+ x_0 \). \[\square\]

Even though, the method of CDMD is mathematically correct, a practical implementation leads to an ill-conditioned algorithm \[7\]. The reason for this is the external computation of the vector \( c \) (which define companion matrix \( C_c \)) that leads to unsatisfied approximation properties of the system matrix \( A \). This problem can be tackled by using the robust singular value decomposition (SVD), and will be discussed in the next section.

## 5 SVD-Dynamic Mode Decomposition (SDMD)

In 2010, the algorithm of DMD was modified radically by Schmid. He published a variant of DMD based on a reduced singular value decomposition \[7\]. For this reason, we will denote this method by SVD-Dynamic Mode Decomposition (SDMD). The result was a robust and stable algorithm, which serves as a basis for the most modern version of DMD. We start this section by deducing this algorithm.

Let the system matrix \( A \) to the data \( x_0, \ldots, x_m \in \mathbb{C}^n \) be given as well as the reduced singular value decomposition of \( X \) with \( r = \text{rank}(X) \) and \( X = U \Sigma V^* \in \mathbb{C}^{n \times m} \), where \( U \in \mathbb{C}^{n \times r} \), \( V \in \mathbb{C}^{m \times r} \) and \( \Sigma \in \mathbb{R}^{r \times r} \). Utilizing the transformation matrix \( U \), we construct the low-dimensional representation \( S \) of the system matrix \( A \) by

\[
S := U^* AU \in \mathbb{C}^{r \times r}.
\]
For an explicit calculation of $S$, it is necessary to avoid the computation of the system matrix $A$. To this end, we calculate

$$S = U^*AU = U^*AUΣV^*VΣ^{-1} = U^* AXVΣ^{-1} = U^* YVΣ^{-1}.$$ 

Now, we compute the eigenvalues $λ_i$ and eigenvectors $v_i$ of the matrix $S$ and finally transform them by the matrix $U$ into

$$θ_i := Uv_i ∈ C^n,$$

in order to obtain an approximation of the eigenvalues and eigenvectors of the system matrix:

$$Aθ_i = AUv_i ≈ UU^*AUv_i = USv_i = Uλv_i = λθ_i.$$

Consequently, the algorithm can be formulated as follows.

**Algorithm 5.1 (SVD-Dynamic Mode Decomposition)**

**Input:** Data $x_0, . . . , x_m ∈ C^n$.

**Output:** DMD eigenvalues $λ_1, . . . , λ_r ∈ C$, DMD modes $θ_1, . . . , θ_r ∈ C^n$, and DMD amplitudes $a_1, . . . , a_r ∈ C$.

1. Define the matrices $X := (x_0, . . . , x_{m-1}), Y := (x_1, . . . , x_m) ∈ C^{n×m}$.

2. Compute the reduced singular value decomposition $X = UΣV^*$ with $r = \text{rank}(X)$.

3. Define the DMD matrix $S := U^* YVΣ^{-1} ∈ C^{r×r}$.

4. Compute the DMD eigenvalues $λ_i$ and eigenvectors $v_i$ of $S$.

5. Calculate the DMD modes $θ_i = Uv_i ∈ C^n$ and define $Θ = (θ_1, . . . , θ_r) ∈ C^{n×r}$.

6. Compute the DMD amplitudes $a = Θ^+x_0 ∈ C^r$ with $a = (a_1, . . . , a_r)^T$.

In **Algorithm 5.1** the DMD amplitudes will be defined intuitively by a best-fit linear combination of the first snapshots $x_0$ in the modes selection (compare **Corollary 4.9**). For a deeper understanding of SDMD, we first characterize the connection to CDMD, which will be examined in the following lemma.

**Lemma 5.2** Let the companion matrix $C_c$ and the DMD matrix $S$ be given by Algorithm 5.1 to the data $x_0, . . . , x_m ∈ C^n$. Then the following identity holds:

$$S = (ΣV^*)C_c(VΣ^{-1}).$$

In particular, if $x_0, . . . , x_{m-1}$ are linearly independent, then the matrices $C_c$ and $S$ are similar.

**Proof.** By **Lemma 4.6** and the reduced singular value decomposition $X = UΣV^*$, we obtain

$$S = U^* YVΣ^{-1} = U^* (XC_c + qe_m^T)VΣ^{-1} = \Sigma V^*X^+(XC_c + qe_m^T)VΣ^{-1} = \Sigma V^*C_cVΣ^{-1} □.$$

**Corollary 5.3** Let the data $x_0, . . . , x_m ∈ C^n$ be given. If $x_0, . . . , x_{m-1}$ are linearly independent, then CDMD and SDMD produce the same DMD eigenvalues.
The corollary suggests that the reconstruction property of CDMD from Theorem 4.8 will be also transferred onto SDMD. In the case of distinct DMD eigenvalues, the DMD modes associated to equal eigenvalues only differ by a scaling factor. Therefore, the DMD modes of SDMD have to be rescaled for an exact reconstruction. The following theorem shows the reconstruction property of SDMD, where the scaling factors are given by the DMD amplitudes.

**Theorem 5.4 (Reconstruction-property SDMD)** Let DMD eigenvalues $\lambda_1, \ldots, \lambda_m$, DMD amplitudes $a_1, \ldots, a_m$, and DMD modes $\vartheta_1, \ldots, \vartheta_m$ be given by Algorithm 5.1 to the data $x_0, \ldots, x_m \in \mathbb{C}^n$, where $x_0, \ldots, x_{m-1}$ are linearly independent. If the DMD eigenvalues are distinct, then the following identities hold:

$$x_k = \sum_{j=1}^{m} \lambda_j^k a_j \vartheta_j, \quad x_m = \sum_{j=1}^{m} \lambda_j^m a_j \vartheta_j + q$$

for $k = 0, \ldots, m-1$ and $q = x_m - Xc$ with $c = X^+x_m$.

**Proof.** By the notation of the proof of Theorem 4.8 and Lemma 5.2 we obtain

$$WSW^{-1} = \Lambda = \text{Vand}(\lambda_1, \ldots, \lambda_m)C_\vartheta \text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1} = \text{Vand}(\lambda_1, \ldots, \lambda_m)(\Sigma V^*)^{-1}S(\Sigma V^*)\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1},$$

where the matrix $W^{-1}$ contains the eigenvectors of $S$. Consider a scaling matrix $K = \text{diag}(\alpha_1, \ldots, \alpha_m)$ that satisfies the equation

$$W^{-1}K = (\Sigma V^*)\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1}.$$

Since the scaling matrix $K$ adjusts eigenvectors (in the same one-dimensional eigenspace), the scaling factors are $\alpha_1, \ldots, \alpha_m \neq 0$ and, therefore, the matrix $K$ is invertible. By multiplying the above equation with $U$ from the left, we obtain

$$\Theta K = UW^{-1} = U(\Sigma V^*)\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1} = XV\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1},$$

where $\Theta = UW^{-1}$ are the DMD modes (by Algorithm 5.1). In sum, we obtain

$$X = \Theta K\text{Vand}(\lambda_1, \ldots, \lambda_m),$$

which shows the first identity concerning the scalings $\alpha_1, \ldots, \alpha_m$. The second statement follows analogously by

$$x_m = XC e_m^T + q = \Theta K\text{Vand}(\lambda_1, \ldots, \lambda_m)C e_m^T + q = \Theta K\Lambda\text{Vand}(\lambda_1, \ldots, \lambda_m)e_m^T + q = \sum_{j=1}^{m} \lambda_j^m a_j \vartheta_j + q.$$

It misses to show that the scaling factors $\alpha_i$ are given by the DMD amplitudes $a_i$. Rewriting $\alpha = (\alpha_1, \ldots, \alpha_m)^T$ and using the identity $X\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1}e = x_0$ (from Theorem 4.8), we obtain:

$$\alpha = Ke = W\Sigma V^*\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1}e = WU^*X\text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1}e = WU^*x_0 = \Theta^{-1}x_0 = a. \quad \Box$$
Remark 5.5 SDMD is characterized by the robust singular value decomposition and the reconstruction property. However, the spectral-theoretical connection of SDMD to the system matrix is not clear. Therefore, let the system matrix \( A \) as well as DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \) and DMD modes \( \vartheta_1, \ldots, \vartheta_m \) be given by Algorithm 5.1 to the data \( x_0, \ldots, x_m \in \mathbb{C}^n \). Then the following equation hold:

\[
\lambda_i \vartheta_i = \lambda_i U v_i = U S v_i = U^* Y V \Sigma^{-1} v_i = U^* Y V \Sigma^{-1} U^* U v_i = U^* Y X^+ \vartheta_i = U^* A \vartheta_i.
\]

We observe that the eigenvalue equation is correct up to the projection \( P_X = X X^+ = U U^* \) onto the image of \( X \).

The following proposition presents characterizations for the equality of the eigenvector equation of Remark 5.5.

Proposition 5.6 Let the system matrix \( A \) as well as DMD eigenvalues \( \lambda_1, \ldots, \lambda_m \) and DMD modes \( \vartheta_1, \ldots, \vartheta_m \) be given by Algorithm 5.1 to the data \( x_0, \ldots, x_m \in \mathbb{C}^n \). In addition, let the error \( q = x_m - X c \) with \( c = X^+ x_m \) be given. Then the following assertions are equivalent:

(i) The DMD mode \( \vartheta_i \) is an eigenvector of the system matrix \( A \) to the eigenvalue \( \lambda_i \).

(ii) \( A \vartheta_i \in \langle x_0, \ldots, x_{m-1} \rangle \).

(iii) \( \langle e_m, X^+ \vartheta_i \rangle \cdot q = 0 \).

(iv) \( x_m \in \langle x_0, \ldots, x_{m-1} \rangle \) or \( \langle e_m, V \Sigma^{-1} v_i \rangle = 0 \).

Proof. Let the companion matrix \( C_c \) to the data \( x_0, \ldots, x_m \) be given.

(iii) \( \implies \) (ii) By Lemma 4.6 we obtain the equality

\[
P_X A \vartheta_i = X C_c X^+ \vartheta_i + q e_m^T q X^+ \vartheta_i = [X C_c + q e_m^T] \vartheta_i = Y \vartheta_i = A \vartheta_i,
\]

which implies \( A \vartheta_i \in \im(X) \), i.e. \( A \vartheta_i \in \langle x_0, \ldots, x_{m-1} \rangle \).

(iii) \( \implies \) (i)" Let \( A \vartheta_i \in \langle x_0, \ldots, x_{m-1} \rangle \). Then the (algebraic) eigenvalue equation is trivially satisfied, since the projection \( P_X \) can be ignored in the equation of Remark 5.5. Furthermore, since the columns of \( U \) are orthogonal and \( v_i \neq 0 \), the vector \( \vartheta_i = U v_i \) is non-zero and, therefore, an eigenvector of \( A \).

("i) \( \implies \) (iii)" By the preliminary Remark 5.5, Lemma 4.6 and assumption (i), we obtain

\[
0 = A \vartheta_i - \lambda_i \vartheta_i = Y X^+ \vartheta_i - P_X A \vartheta_i = [X C_c + q e_m^T] X^+ \vartheta_i - X C_c X^+ \vartheta_i = q e_m^T X^+ \vartheta_i.
\]

("iii) \( \iff \) (iv)" We reformulate the conditions by

\[
\langle e_m, X^+ \vartheta_i \rangle \cdot q = 0 \iff q = 0 \text{ or } \langle e_m, X^+ \vartheta_i \rangle = 0
\]

\[
\iff q = 0 \text{ or } \langle e_m, V \Sigma^{-1} U v_i \rangle = 0
\]

\[
\iff x_m \in \langle x_0, \ldots, x_{m-1} \rangle \text{ or } \langle e_m, V \Sigma^{-1} v_i \rangle = 0. \quad \Box
\]

A trivial consequence of the above proposition is the following corollary, which was first proven by Tu et al. [10].
Corollary 5.7 Let the system matrix $A$ as well as DMD eigenvalues $\lambda_1, \ldots, \lambda_m$ and DMD modes $\vartheta_1, \ldots, \vartheta_m$ be given by Algorithm 5.1 to the data $x_0, \ldots, x_m \in \mathbb{C}^n$. If $x_m \in \langle x_0, \ldots, x_{m-1} \rangle$, then the DMD modes and DMD eigenvalues are eigenvectors and eigenvalues of the system matrix $A$, respectively.

The previous statement implies $\sigma(S) \subseteq \sigma(A)$. However, the other inclusion is also true for all non-zero eigenvalues of the system matrix $A$.

Proposition 5.8 Let the system matrix $A$ as well as DMD eigenvalues $\lambda_1, \ldots, \lambda_m$ and DMD modes $\vartheta_1, \ldots, \vartheta_m$ be given by Algorithm 5.1 to the data $x_0, \ldots, x_m \in \mathbb{C}^n$. Then all non-zero eigenvalues $\lambda \neq 0$ of $A$ will be calculated by DMD, i.e., it holds

$$\sigma(A) \setminus \{0\} \subseteq \sigma(S).$$

Proof. For an arbitrary eigenvector $z$ of $A$ to the eigenvalue $\lambda \neq 0$, we obtain by defining the vector $v := U^*z$:

$$Sv = U^*Y\Sigma^{-1}U^*z = U^*X^+z = U^*\vartheta z = U^*\lambda z = \lambda v.$$

Assume $v = 0$. Then $U^*z = 0$ and we obtain

$$0 = Y\Sigma^{-1}U^*z = YX^+z = Az = \lambda z.$$

As $z \neq 0$, it follows $\lambda = 0$, which contradicts the assumptions. $\square$

Consequently, under appropriate assumptions of the data, we obtain the spectral-theoretic relation

$$\sigma(A) \setminus \{0\} \subseteq \sigma(S) \subseteq \sigma(A).$$

However, the condition $x_m \in \langle x_0, \ldots, x_{m-1} \rangle$ in Corollary 5.7 is actually never satisfied in the case of $n \gg m$ and, hence, not practically applicable.

This raises the question, whether the DMD modes can be modified such that we obtain eigenvectors of the system matrix without any assumptions. A solution to this problem is presented in the next theorem, which is inspired by the assertion (iii) in Proposition 5.6.

Theorem 5.9 Let the system matrix $A$ as well as DMD eigenvalues $\lambda_1, \ldots, \lambda_m$ and DMD modes $\vartheta_1, \ldots, \vartheta_m$ be given by Algorithm 5.1 to the data $x_0, \ldots, x_m \in \mathbb{C}^n$. In addition, let the error $q = x_m - Xc$ with $c = X^+x_m$ be given. For $\lambda_i \neq 0$

$$\zeta_i = \vartheta_i + \frac{1}{\lambda_i}(e_m, X^+\vartheta_i) \cdot q$$

is an eigenvector of the system matrix $A$ to the eigenvalue $\lambda_i$.

Proof. By Lemma 4.6 and Remark 5.5 we obtain the eigenvalue equation by

$$A\zeta_i = A\vartheta_i + \frac{1}{\lambda_i}(e_m, X^+\vartheta_i) \cdot q = A\vartheta_i = YX^+\vartheta_i = [XC + qe_m^T]X^+\vartheta_i = XcX^+\vartheta_i + qe_m^T X^+\vartheta_i = P_XA\vartheta_i + (e_m, X^+\vartheta_i) \cdot q = \lambda_i\vartheta_i + (e_m, X^+\vartheta_i) \cdot q = \lambda_i\zeta_i.$$

Assume $\zeta = 0$. Then we obtain

$$0 = X^+\zeta = X^+(\vartheta_i + \frac{1}{\lambda_i}(e_m, X^+\vartheta_i) \cdot q) = X^+\vartheta_i,$$
and therefore

\[ 0 = U^*YX^+\vartheta_i = U^*YV\Sigma^{-1}U^*Uv_i = Sv_i = \lambda v_i. \]

Since \( v_i \neq 0 \), it follows \( \lambda = 0 \), which contradicts the assumption. \( \square \)

The previous proposition characterizes exactly the spectral-theoretic relation between the DMD matrix \( S \) and the system matrix \( A \). In fact, it holds \( \sigma(A) \setminus \{0\} \subseteq \sigma(S) \subseteq \sigma(A) \) (without any assumption) and thus

\[ \sigma(A) \setminus \{0\} = \sigma(S) \setminus \{0\}. \]

Consequently, the dynamic behavior of the system matrix \( A \) will be completely captured by the low-dimensional DMD matrix \( S \). However, the DMD modes have to be modified according to Theorem 5.9 in order to get eigenvectors of the system matrix. This motivates the formulation of a new variant of DMD with modified DMD modes and possibly new DMD amplitudes such that the reconstruction property is preserved. In particular, as Theorem 5.9 states, only the non-zero eigenvalues can be used. Nevertheless, these eigenvalues are sufficient to capture the temporal evolution and consequently no dynamical information is lost.

**Remark 5.10** For the formulation of an efficient algorithm, we need a more compact representation of the modified DMD modes \( \zeta_i \), which arises directly from transformations of the eigenvectors of \( S \). By Remark 5.5 and Lemma 4.6, we rearrange the modified DMD modes by

\[
\zeta_i = \vartheta_i + \frac{1}{\lambda_i} \langle e_m, X^+\vartheta_i \rangle \cdot q = \frac{1}{\lambda_i} \lambda_i \vartheta_i + \frac{1}{\lambda_i} q e_m^T X^+\vartheta_i = \frac{1}{\lambda_i} P X A \vartheta_i + \frac{1}{\lambda_i} q e_m^T X^+\vartheta_i = \frac{1}{\lambda_i} X C_c X^+\vartheta_i + \frac{1}{\lambda_i} q e_m^T X^+\vartheta_i = \frac{1}{\lambda_i} (X C_c + q e_m^T) X^+\vartheta_i = \frac{1}{\lambda_i} Y X^+\vartheta_i = \frac{1}{\lambda_i} Y V \Sigma^{-1} v_i.
\]

Defining the DMD modes in this way, we obtain the most modern version of DMD, called Exact-Dynamic Mode Decomposition (EXDMD). In the following section, we will introduce this variant of DMD.

### 6 Exact-Dynamic Mode Decomposition (EXDMD)

In 2014, Tu et al. [10] presented the most modern version of DMD, called Exact-Dynamic Mode Decomposition (EXDMD). However, the algorithm presented here differs from the standard literature as we use a different definition of the DMD amplitudes. In addition, we introduce a novel relevant variable: The error scaling.

**Algorithm 6.1 (Exact-Dynamic Mode Decomposition)**

**Input:** Data \( x_0, \ldots, x_m \in \mathbb{C}^n \).

**Output:** DMD eigenvalues \( \lambda_1, \ldots, \lambda_{r_0} \in \mathbb{C} \), DMD modes \( \vartheta_1, \ldots, \vartheta_{r_0} \in \mathbb{C}^n \), DMD amplitudes \( a_1, \ldots, a_{r_0} \in \mathbb{C} \), and the error scaling \( a_0 \).

1. Define the matrices \( X := (x_0, \ldots, x_{m-1}), Y := (x_1, \ldots, x_m) \in \mathbb{C}^{n \times m} \).
2. Compute the reduced singular value decomposition \( X = U \Sigma V^* \) with \( r = \text{rank}(X) \).
3. Define the DMD matrix \( S := U^* Y V \Sigma^{-1} \in \mathbb{C}^{r \times r} \).
4. Compute the DMD eigenvalues $\lambda_i$ and eigenvectors $v_i$ of $S$.

5. Calculate the DMD modes $\vartheta_1 = \frac{1}{\lambda_1} Y \Sigma^{-1} v_1 \in \mathbb{C}^n$ and define $\Theta = (\vartheta_1, \ldots, \vartheta_m) \mathbb{C}^{n \times \tau_0}$.

6. Compute the DMD amplitudes $a = \Lambda^{-1} X \vartheta_1 \in \mathbb{C}^r$ with $a = (a_1, \ldots, a_r)^T$.

7. Calculate the error scaling $a_0 = -\sum_{j=1}^{\tau_0} \frac{1}{\lambda_j} \prod_{k=p}^{\tau_0} \frac{1}{\lambda_j - \lambda_k} \in \mathbb{C}$, if it exists.

The introduction of the error scaling $a_0$ and the fundamental change of the definition of the DMD amplitude will be justified by the subsequent theorem, which proves the reconstruction property of EXDMD. The reconstruction property of EXDMD, however, differs from the previous ones of CDMD (see Theorem 4.8) and SDMD (see Theorem 5.4). The reason for this is the spectral-theoretic relation of EXDMD to the system matrix, as we will examine in the proof.

**Lemma 6.2** Let data $x_0, \ldots, x_m \in \mathbb{C}^n$ with associated matrices $X = [x_0 \ldots x_{m-1}]$ and $Y = [x_1 \ldots x_m]$ be given. If $\ker(X) \subseteq \ker(Y)$ and $\ker(X) \neq \{0\}$, then $x_m \in \text{im}(X)$. In particular, it holds $\text{im}(X) \subseteq \text{im}(Y)$.

**Proof.** Since the kernel of $X$ is non-trivial, there exist a vector $v = (v_1, \ldots, v_m)^T \neq 0 \in \mathbb{C}^m$ with $Xv = 0$. Since $\ker(X) \subseteq \ker(Y)$, we obtain $Yv = 0$.

1. case: $v_m \neq 0$. Then the equation $v_1 x_1 + \ldots + v_m x_m = 0$ implies

$$x_m = -\frac{1}{v_m} (v_1 x_1 + \ldots + v_{m-1} x_{m-1}) = -\frac{v_1}{v_m} x_1 - \ldots - \frac{v_{m-1}}{v_m} x_{m-1},$$

and hence the assertion is proven.

2. case: $v_m = 0$. Then we obtain

$$0 = Y \begin{pmatrix} v_1 \\ \vdots \\ v_{m-1} \\ 0 \end{pmatrix} = X \begin{pmatrix} 0 \\ v_1 \\ \vdots \\ v_{m-1} \end{pmatrix}. $$

By assumption $\ker(X) \subseteq \ker(Y)$, we obtain

$$X \begin{pmatrix} 0 \\ v_1 \\ \vdots \\ v_{m-1} \end{pmatrix} = 0 \implies Y \begin{pmatrix} 0 \\ v_1 \\ \vdots \\ v_{m-1} \end{pmatrix} = 0.$$

If $v_{m-1} \neq 0$, then we get a representation of $x_m$ analogous to the first case. Otherwise, we repeat this steps as long as an entry $v_j \neq 0$ appears. As $v$ is non-zero, there exists such an entry $v_j \neq 0$. \hfill $\square$

**Theorem 6.3 (Reconstruction property EXDMD)** Let DMD eigenvalues $\lambda_1, \ldots, \lambda_r$, DMD amplitudes $a_1, \ldots, a_r$, DMD modes $\vartheta_1, \ldots, \vartheta_r$, and the error scaling $a_0$ be given by [Algorithm 6.1] to the data $x_0, \ldots, x_m \in \mathbb{C}^n$. If $\ker(X) \subseteq \ker(Y)$ and the DMD amplitudes are distinct, then the following identities hold:

$$x_0 = \sum_{j=1}^{r} a_j \vartheta_j + q_0, \quad x_k = \sum_{j=1}^{r} \lambda_j^k a_j \vartheta_j,$$
for \( k = 1, \ldots, m \) and a vector \( q_0 \) with

\[
\begin{cases}
q_0 = a_0 \cdot q & r = m \\
q_0 = 0 & r \neq m
\end{cases}
\]

where \( q = x_m - Xc \) and \( c = X^+x_m \).

**Proof.** By **Theorem 5.9** and **Remark 5.10**, we obtain that the DMD eigenvalues \( \lambda_i \) and modes \( \vartheta_i \) are eigenvalues and eigenvectors of the system matrix \( A \) for \( i = 1, \ldots, m \). Since \( \ker(X) \subseteq \ker(Y) \), we obtain the relation

\[
\text{rank}(Y) = m - \ker(Y) \leq m - \ker(X) = \text{rank}(X) = r.
\]

In addition, the condition \( \text{im}(A) \subseteq \text{im}(Y) \) implies \( \text{rank}(A) \leq \text{rank}(Y) \) and consequently \( r \leq \text{rank}(Y) \), because \( A \) has \( r \) non-zeros distinct eigenvalues. Hence, \( \text{rank}(Y) = r \) and the system matrix \( A \) is diagonalizable by \( r \) non-zero distinct eigenvalues. Therefore, the assumptions of **Theorem 3.7** are satisfied such that there exist scaling factors \( \alpha_1, \ldots, \alpha_r \in \mathbb{C} \) and an error \( q_0 \in \mathbb{C}^n \) such that

\[
x_0 = \sum_{j=1}^{r} \alpha_j v_j + q_0, \quad x_k = \sum_{j=1}^{r} \lambda_j^k \alpha_j v_j,
\]

for \( k = 1, \ldots, m \). Using the notation of **Theorem 3.7**, we obtain the equation

\[
Y = \Theta K \Lambda \text{Vand}(\lambda_1, \ldots, \lambda_r; 1, m),
\]

where \( K = \text{diag}(\alpha_1, \ldots, \alpha_m) \) is the scaling matrix. Since the DMD modes (which are eigenvectors of the system matrix) are related to distinct eigenvalues, they are linearly independent and, consequently, the following calculation

\[
a = \Lambda^{-1} \Theta^+ x_1 = \Lambda^{-1} \Theta^+ Y e_1 \\
\quad = \Lambda^{-1} \Theta^+ \Theta K \Lambda \text{Vand}(\lambda_1, \ldots, \lambda_r; 1, m) e_1 \\
\quad = K \text{Vand}(\lambda_1, \ldots, \lambda_r; 1, m) e_1 \\
\quad = (\alpha_1, \ldots, \alpha_m)^T
\]

shows that the scaling factors \( \alpha_k \) coincide with the DMD amplitude \( a_k \). Hence, the two identities are proven. For the additional statement, we consider two cases:

1. case: \( r = m \). Consider \( Y = X C_m + q e_m^T \) and rearrange this equation by using the inverse of \( C_m \) (which exists since \( r = m \)):

\[
X = (Y - q e_m^T) C_m^{-1} = Y C_m^{-1} - q e_m^T C_m^{-1}.
\]

By **Lemma 5.2** and relation \( C_m = (\Sigma V^*) S (\Sigma V^*)^{-1} \), we obtain the identity

\[
YC_m^{-1} = Y V \Sigma^{-1} S^{-1} \Sigma V^* = \Theta A W S^{-1} \Sigma V^* = \Theta W \Sigma V^* = \Theta W \Sigma V^* M^{-1} M = \Theta K M,
\]

and, therefore, it follows

\[
x_0 = X e_1 = (Y C^{-1} - q e_m^T C_{c}^{-1}) e_1 = \Theta K M e_1 - q e_m^T C_{c}^{-1} e_1 = \sum_{j=1}^{m} a_j \vartheta_j - \langle e_m, C_{c}^{-1} e_1 \rangle \cdot q.
\]
Now, we examine the remaining term \( \langle e_m, C^{-1}e_1 \rangle \cdot q \). To this end, we represent the companion matrix \( C \) by its eigenvectors and eigenvalues

\[
e_m^T C^{-1} e_1 = e_m^T \text{Vand}(\lambda_1, \ldots, \lambda_m)^{-1} \Lambda^{-1} \text{Vand}(\lambda_1, \ldots, \lambda_m) e_1
\]

\[
= ((\text{Vand}(\lambda_1, \ldots, \lambda_m)^T)^{-1} e_m)^T \Lambda^{-1} \text{Vand}(\lambda_1, \ldots, \lambda_m) e_1.
\]

The inverse of the transposed Vandermonde matrix \( \text{Vand}(\lambda_1, \ldots, \lambda_m)^T \) is given by a LU-decomposition [5], i.e., there exist a lower triangle matrix \( L \) and an upper triangle matrix \( U \) with

\[
(\text{Vand}(\lambda_1, \ldots, \lambda_m)^T)^{-1} = (LU)^{-1} = U^{-1}L^{-1}.
\]

More precisely, these matrices are given by

\[
U^{-1}_{i,j} = \begin{cases} 
0 & i > j \\
\prod_{k=1}^{j} \frac{1}{\lambda_i - \lambda_k} & i \leq j
\end{cases}
\]

and

\[
L^{-1}_{i,j} = \begin{cases} 
0 & i < j \\
1 & i = j \\
L_{i-1,j-1} - L_{i-1,j} \cdot \lambda_{i-1} & i = 2, \ldots, m, \quad j = 2, \ldots, i - 1
\end{cases}
\]

This yields

\[
(\text{Vand}(\lambda_1, \ldots, \lambda_m)^T)^{-1} e_m = U^{-1}L^{-1} e_m = U^{-1}e_m = \begin{pmatrix}
\prod_{k=1}^{m} \frac{1}{\lambda_1 - \lambda_k} \\
\vdots \\
\prod_{k=1}^{m} \frac{1}{\lambda_m - \lambda_k}
\end{pmatrix}
\]

and finally

\[
e_m^T C^{-1} e_1 = ((\text{Vand}(\lambda_1, \ldots, \lambda_m^T)^{-1} e_m)^T \Lambda^{-1} \text{Vand}(\lambda_1, \ldots, \lambda_m) e_1
\]

\[
= \frac{1}{\lambda_1} \prod_{k=1, k \neq 1}^{m} \frac{1}{\lambda_1 - \lambda_k} + \ldots + \frac{1}{\lambda_m} \prod_{k=1, k \neq m}^{m} \frac{1}{\lambda_m - \lambda_k}.
\]

This factor equals the error scaling \( a_0 \), which completes the proof for the first case.

2. case: \( r \neq m \). Hence, \( \ker(X) \) is non-trivial, i.e., \( \ker(X) \neq \{0\} \). By Lemma 6.2, we obtain \( \text{im}(X) \subseteq \text{im}(Y) \) and, therefore, it holds

\[
\text{im}(X) = \text{im}(Y) = \text{im}(A) = \langle \vartheta_1, \ldots, \vartheta_r \rangle,
\]

because \( \text{im}(A) \subseteq \text{im}(Y) \). Hence, \( x_0 \) is in the span of \( \vartheta_1, \ldots, \vartheta_r \) and therefore \( q_0 \in \text{im}(A) \), especially. However, by the proof of Theorem 3.7, we obtain that the vector \( q_0 \in \ker(A) \). Consequently, we obtain \( q_0 = 0 \). \( \Box \)

Theorem 6.3 clearly demonstrates the functionality of EXDMD. Via a reduced singular value decomposition, the dynamically relevant properties of the system matrix were extracted, i.e., the eigenvectors to non-zero eigenvalues. As a result, the eigenvectors (or DMD modes) do not generate a basis anymore and consequently the first snapshot \( x_0 \) will be reconstructed with an error. However, Theorem 6.3 gives us an exact representation of the error. The subsequent corollary illustrates the connection between the DMD amplitudes and the coefficients of the first snapshot in the eigenvector basis of the system matrix.
Corollary 6.4 Let the system matrix $A$ as well as DMD eigenvalues $\lambda_1, \ldots, \lambda_r$, DMD modes $\vartheta_1, \ldots, \vartheta_r$, DMD amplitudes $a_1, \ldots, a_r$ be given by Algorithm 6.1 to the data $x_0, \ldots, x_m \in \mathbb{C}^n$. In addition, let the eigenvalues $\lambda_1, \ldots, \lambda_r, \lambda_{r+1}, \ldots, \lambda_n$ and eigenvectors $\vartheta_1, \ldots, \vartheta_r, v_{r+1}, \ldots, v_m$ of the system matrix as well as the coefficient vector $b = (b_1, \ldots, b_n)^T = W^{-1}x_0$ with $W = (\vartheta_1, \ldots, \vartheta_r, v_{r+1}, \ldots, v_m)$ be given. If $\ker(X) \subseteq \ker(Y)$, rank($Y$) = $r$, and the (DMD) eigenvalues $\lambda_1, \ldots, \lambda_r$ are distinct, then the DMD amplitudes coincide with the coefficients:

$$a_j = b_j$$

for $j = 1, \ldots, r$.

Proof. The assumptions of Theorem 6.3 and Theorem 3.7 are satisfied, because the system matrix $A$ is diagonalizable by the $r$ distinct eigenvalues. Hence we receive

$$\sum_{j=1}^{r} \lambda_j a_j \vartheta_j = x_1 = \sum_{j=1}^{r} \lambda_j b_j \vartheta_j.$$

Since the eigenvectors (to distinct eigenvalues) are linear independent and the eigenvalues are non-zero, the coefficients have to match. □

7 Conclusion

A comprehensive theoretical analysis of Dynamic Mode Decomposition has been developed that clarify the connection between different variants of DMD (CDMD, SDMD, and EXDMD) and demonstrates several features of them. One of these features is the reconstruction property, which was proven for all variants and the system matrix as well. To this end, different scaling factors were used and new ones introduced to ensure this property. Especially for EXDMD, it was shown that under appropriate conditions the algorithm calculates the dynamically relevant, high-dimensional structures of the system matrix with the help of low-dimensional, spectral-theoretical techniques. The new findings facilitate the application with DMD since precise reconstructions are obtained which lead to a clearer decomposition of the data into DMD eigenvalues, modes and amplitudes.

Acknowledgment

The first author would like to thank Rainer Nagel for ideas, helpful suggestions and inspiring discussions. Moreover, he expresses his gratitude to the entire working group functional analysis (AGFA) from the University of Tübingen for the numerous discussions and the support.

This work is partly supported by “Kooperatives Promotionskolleg Digital Media” at Hochschule der Medien and the University of Stuttgart.

References

[1] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses. Journal of Nonlinear Science, 22(6):887–915, 2012.
[2] Zlatko Drmač, Igor Mezić, and Ryan Mohr. Data driven Koopman spectral analysis in Vandermonde-Cauchy form via the DFT: numerical method and theoretical insights. arXiv preprint arXiv:1808.09557, 2018.

[3] Mihailo R. Jovanović, Peter J. Schmid, and Joseph W. Nichols. Sparsity-promoting dynamic mode decomposition. Physics of Fluids, 26(2):024103, 2014.

[4] Eliakim H. Moore. On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society, 26:394–395, 1920.

[5] Hector Moya-Cessa and Francisco Soto-Eguiabar. Inverse of the Vandermonde and Vandermonde confluent matrices. Applied Mathematics & Information Sciences, 5:361, 2011.

[6] Clarence W. Rowley, Igor Mezic, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. Journal of Fluid Mechanics, 641:115–127, 2009.

[7] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. Journal of Fluid Mechanics, 656:5–28, 2010.

[8] Peter J. Schmid, Larry Li, Matthew P. Juniper, and Oliver Pust. Applications of the dynamic mode decomposition. Theoretical and Computational Fluid Dynamics, 25:249–259, 2011.

[9] Peter J Schmid and Jörn Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In APS Division of Fluid Dynamics Meeting Abstracts, page MR.007, November 2008.

[10] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. Journal of Computational Dynamics, 1:391–421, 2014.