# Notes on Dynamic Mode Decomposition
## (with some code)



Kamila Zdybał

*Université libre de Bruxelles, kamila.zdybal@ulb.ac.be*
*camillejr.github.io/science-docs, kamila.zdybal@gmail.com*

## Preface

*Dynamic Mode Decomposition* (DMD) is a data-driven method for approximating a generally nonlinear dynamical system with a linear one. The technique can be used for finding low-rank structures in multi-dimensional data sets, as well as predicting the future state of the system with inference made solely from the collected data. As a data decomposition technique, DMD offers the possibility of performing a low-rank approximation of the original data set.
This document combines the knowledge from various sources that shaped my journey through understanding DMD. Finally, the notes are accompanied by a set of codes that aim to show DMD in action.

This document is still in preparation. Please feel free to contact me with any suggestions, corrections or comments.

## Keywords

*Dynamic Mode Decomposition (DMD), linear algebra, matrix decomposition, matrix approximation, linear dynamical systems, complex plane*

## Contents



# 1 Setting the stage

This small opening section is here to give you a bit of intuition for DMD which hopefully will be handy in understanding the mathematical formulation in the next chapters. It is assumed that you have some general idea on what it means to perform data decomposition[1].
The main purpose of DMD is to find *dynamic modes* underlying the data set in a purely data-driven way. The modes are *dynamic* in a sense that they are spatial structures, each associated with a certain time evolution given by frequency of oscillations. In other words, in DMD we not only know how the system can be decomposed spatially (what spatial structures it is composed of), but also how each found spatial structure behaves in time.
DMD is also a data reduction/approximation technique. The original data set can be reconstructed by summing up the dynamic modes, or approximated by summing up a *significant* portion of the dynamic modes.

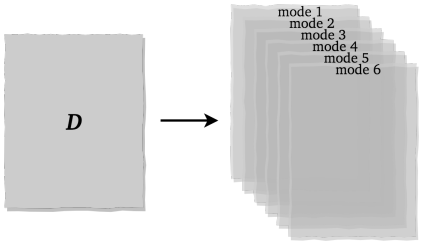


Figure 1: Modal decomposition of a data matrix **D**.

In many data reduction techqniues, in addition to reconstructing data from its modes, different data decomposition techniques can give a different meaning to the modes themselves. In general, for physical data sets such as ones coming from fluid dynamics, we would like the modes to have a physical interpretation (which is not *a priori* guaranteed by some decomposition techniques). That would allow to extract additional information about the physics of the original data set. We will see later how DMD performs with respect to this idea.

## 1.1 DMD of a flipbook

To see a pictorial example before we jump into the mathematical formalism, we may think of DMD as analyzing a flipbook[2].
Let's say that each slide is composed of a background which is the same in every slide (it is constant in time). There might also be an animation of moving objects happening in the foreground where some parts are animated as moving faster and some as moving slowly.

[1] For this you can check out my other note [8].

[2] If you are unfamiliar with what a flipbook is, take a look at this YouTube video [4].

Notice then, that every page of a flipbook is not a DMD mode. Rather, a mode can be such element of all the flipbook pages which has got the same behaviour in time and is also *spatially significant*.
We associate a certain number, say $\lambda$, to describe the evolution in time of every separate element (spatial structure) of the animation. This number will have a meaning of frequency of appearance.
And hence the background might be thought of as one spatial mode that has got a a constant time behaviour - it is present on every flipbook page. A ball bouncing on the foreground might be another mode, associated with its $\lambda$.
The analogy presented in this section is based on [1].
In order to attach a physical meaning to the spatial modes extracted, the DMD technique can distinguish a particular physical object as a separate dynamic mode (for instance: a jumping ball in the foreground, or a moving cloud in the background).

# 2 Dynamical systems

Dynamic Mode Decomposition connects strongly with the arguments made for *dynamical systems* - systems that change in time. There is an interesting intuition carried from one-dimensional linear dynamical systems to multi-dimensional linear dynamical systems and we will briefly present it here. For a more thorough explanation the reader is encouraged to look into [5], which is an excellent introductory position for the first encounter with dynamical systems.

## 2.1 One-dimensional linear dynamical system

A discrete one-dimensional linear dynamical system is characterized by the following evolution:

$$x(i+1) = ax(i) + b \tag{1}$$

where $x(i)$ and $x(i+1)$ are two consecutive measurements of the system and $a$ and $b$ are the system parameters.
If we know the initial condition of this system $x(0) = x_0$, we may also write that in the general case for $i$ equal to some $k$:

$$x(k) = a^k x_0 + \frac{a^k - 1}{a - 1} b \tag{2}$$

for $a \neq 1$, or:

$$x(k) = x_0 + kb \tag{3}$$

when $a = 1$.
Let's think now what could a linear dynamical system behaviour be? What are the possible ways in which the measured variable propagates? Surely it can stay constant in time, or if not, it can either grow or decay as we keep on measuring the system. The constant behaviour is captured by the solution described by eq.(3). The other two behaviours are somehow hidden in eq.(2). At this point, I would like to draw your attention to an important observation: the character of the system's evolution in time is dependent on the parameter $a$.
For a deeper explanation of various cases I encourage you to take a look at Chapter 2 in [5].

## 2.2 Multi-dimensional linear dynamical system

A discrete multi-dimensional linear dynamical system is, on the other hand, characterized by the following evolution:

$$\mathbf{x}(i+1) = \mathbf{A}\mathbf{x}(i) + \mathbf{b} \tag{4}$$

The measurement $\mathbf{x}$ is no longer a single variable (a scalar) but an entire vector. It could come from measurements taken at different points on the mesh at a particular moment in time. Let's assume the vector size is $n$. We may call again $\mathbf{A}$ and $\mathbf{b}$ the parameters of the system but this time $\mathbf{A}$ is a matrix of size $(n \times n)$ and $\mathbf{b}$ is a vector of size $n$.
Can we find the analogous parameter to $a$ from the 1D dynamical systems that would define the time behaviour of a multi-dimensional one? Turns out this parameter are the eigenvalues of a matrix $\mathbf{A}$! If you are new to dynamical systems this probably seems like a very strange but also fortunate result. Again, for a thorough proof of that I encourage you to turn to Chapter 2 in [5].
The most important thing to remember from this section is the following: **the behaviour of a multi-dimensional linear dynamical system is dependent on the eigenvalues of the system**. This is the piece of knowledge that will soon show prove important in the explanation of the steps of Dynamic Mode Decomposition.

## 2.3 Solution to a dynamical system

We have a system described by a differential equation:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), t, \dots) \tag{5}$$

where the function $f(\mathbf{x}(t), t, \dots)$ is a way of *modeling* that system.
We also have collected *measurements* of the system at different points in space at time $k$, in the form of a vector(s) $\mathbf{y}_k$:

$$\mathbf{y}_k = g(\mathbf{x}_k) \tag{6}$$

where $\mathbf{x}_k$ is the quantity of interest that we are aiming at measuring. The fact that we might not be able to measure it directly is accounted for by some function $g()$ (although it might happen that $\mathbf{y}_k = \mathbf{x}_k$, meaning that we are able to measure $\mathbf{x}_k$ directly).
Notice, that for measurements at many moments in time, we may stack all the collected vectors $\mathbf{y}_i$ for different times $i$ to create a matrix whose columns represent time snapshots and whose rows represent position in space - see Figure 2. This is a matrix that we are going to work with.

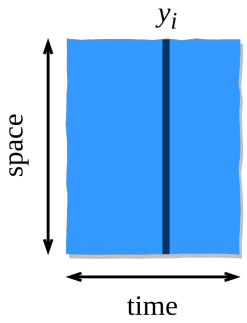

Figure 2: Data matrix with measurements of the system.

We are from now interested in systems where the governing equation from eq.(5) is not known (in other words, the function $f$ is unknown) and we solely rely on measurements of the system which, in general, form a high-dimensional data set.
In the Dynamic Mode Decomposition we approximate that data set by a linear dynamical system of the form:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) \tag{7}$$

This is in fact a very handy approximation since we are able to write down exact solutions to linear systems.
Once we assume that the general solution is of the form:

$$\mathbf{x}(t) = \mathbf{v}e^{\lambda t} \tag{8}$$

to obtain the parameters we effectively solve the eigenvalue problem:

$$\mathbf{A}\mathbf{v} = \check{}\mathbf{v} \tag{9}$$

The exact solution to the linear system from eq.(7) is:

$$x = \sum_{j=1}^{n} b_j \phi_j e^{\lambda_j t} \tag{10}$$

For a reader interested in how this solution was derived, more can be found in appendix A.

# 3 Dynamic Mode Decomposition theory

## 3.1 Exact DMD

For the moment, we assume that we can measure the system directly, that is we measure $\mathbf{y}_i = \mathbf{x}_i$. Moreover, we assume that our data is collected in equal[3] time steps $\Delta t$. The measurements are combined inside a large matrix $\mathbf{X}$ where each of its columns represents one time snapshot:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \dots & \mathbf{x}_m \end{bmatrix} \tag{11}$$

We split the large matrix $\mathbf{X}$ into two matrices $\mathbf{X_1}$ and $\mathbf{X_2}$ such that:

$$\mathbf{X_1} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \dots & \mathbf{x}_{m-1} \end{bmatrix} \tag{12}$$

$$\mathbf{X_2} = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \dots & \mathbf{x}_m \end{bmatrix} \tag{13}$$

If we now assume that a linear operator will map the first element of $\mathbf{X_1}$ with the first element of $\mathbf{X_2}$, second with the second, third with the third, and so on, matrix $\mathbf{X_2}$ can be thought of as a matrix representing the *future state* of the matrix $\mathbf{X_1}$. That linear operator is assumed to be a matrix $\mathbf{A}$.

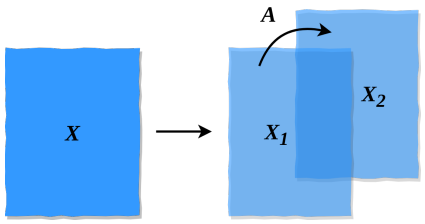


Figure 3: Splitting the data matrix into *past* and *future* matrices $\mathbf{X_1}$ and $\mathbf{X_2}$, linked by the linear operator $\mathbf{A}$.

Note here, that for nonlinear systems, a matrix that transforms $\mathbf{x}_1$ to $\mathbf{x}_2$ is different from a matrix that transforms $\mathbf{x}_2$ to $\mathbf{x}_3$ and so on. DMD assumes, however, that there is one matrix $\mathbf{A}$ that does all these transformations at once, with the least amount of error. It finds the *best-fit* linear dynamical system for the non-linear data set. In mathematical terms, we are looking for such $\mathbf{A}$ that:

$$\mathbf{X_2} = \mathbf{A}\mathbf{X_1} \tag{14}$$

At this point, we should pause and appreciate what the above equation is representing. It describes a linear system of equations which (in the most practical case) will be underdetermined. The unknown in this equation is the matrix $\mathbf{A}$, hence the number of unknowns is equal to the number of elements of this matrix: $n_s \cdot n_s$. The number of linear equations however is $n_s \cdot n_t$ and in many dynamical systems of interest $n_s >> n_t$. Such system might have infinitely many solutions and therefore some additional constraint on $\mathbf{A}$ is needed in order to select one solution out of infinitely many. This boils down to the problem of regression - what regression condition we chose is up to us. In particular, we might restrict ourselves to such $\mathbf{A}$ that will minimize the L2 norm. It is worth noting that there might be different *best* $\mathbf{A}$ that gets us from snapshot 5 to 6 then the one that gets us from snapshot 100 to 101. What our regression will aim at optimizing is to find one matrix $\mathbf{A}$ that *sort of* does all the past-future links in the *best* way it can.
If we chose to proceed with the regression that minimizes the L2 norm (as is typically done in DMD algorithm) the solution to the system of eq.(14) is obtained by multiplying both sides by the *pseudo-inverse* of matrix $\mathbf{X_1}$ which we denote by $\mathbf{X_1}^+$:

$$\mathbf{A} = \mathbf{X_2}\mathbf{X_1}^+ \tag{15}$$

The pseudo-inverse described here, also known as the Moore-Penrose inverse[4], is computed using the least squares method. There is therefore certain information lost when going from eq.(14) to eq.(15).
Once we have solved for matrix $\mathbf{A}$, we can go back to eq.(9) and solve for eigenvalues and eigenvectors.
Up to this point, this is what the **exact DMD** computes. There is however a problem that the eq.(15) may pose when numerics are involved and this will be addressed in the next section.

## 3.2 Going low-rank

Matrices $\mathbf{X_1}^+$ and $\mathbf{X_2}$ typically represent huge spacial dimensionality[5] which in turn means that the matrix $\mathbf{A}$ can become a square matrix of a massive size.
We are hence reluctant to perform the multiplication of matrices as is stated in eq.(15).
The hope comes from the *Singular Value Decomposition* (SVD). We belive that there are low-rank structures hidden in the data set and we are able to reduce the dimensionality of matrix $\mathbf{A}$ without significant loss of information[6].
We perform the SVD on matrix $\mathbf{X_1}$ and decompose it to the component matrices: $\mathbf{U}$, $\mathbf{\Sigma}$ and $\mathbf{V}$.

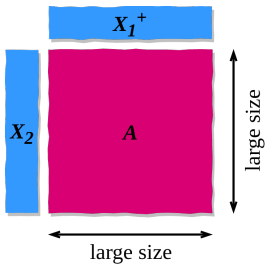


Figure 4: Building the linear operator $\mathbf{A}$ in exact DMD.

$$\mathbf{X_1} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{16}$$

Based on the rank structure of the matrix $\mathbf{X_1}$ (one way to get information about the rank structure is to plot the elements from the diagonal of the matrix $\mathbf{\Sigma}$) we perform a rank-$r$ truncation on the SVD decomposition and approximate the matrix $\mathbf{X_1}$ by its low-rank (rank-$r$) representation:

$$\mathbf{X_1} \approx \mathbf{X_{1r}} = \mathbf{U_r}\mathbf{\Sigma_r}\mathbf{V_r}^T \tag{17}$$

The pseudo-inverse of the truncated matrix is:

[3] Which is indeed a special case for real life measurements. Check section 4 for more information.

[4] Check appendix B for more information.

[5] This is often the case for data sets where we have very few snapshots in time but a large number of spacial points where the measurements were taken. Graphically, we might think of those matrices as being "tall" and this is illustrated in Figure 2.

[6] Professor Kutz said a very interesting sentence here, that the multiplication presented in Figure 4 completely ignores the fact that there might be low-rank structures in our data set.

$$\mathbf{X_{1r}}^{+} = \mathbf{V_r}\mathbf{\Sigma_r}^{-1}\mathbf{U_r}^{T} \tag{18}$$

The usefulness of this decomposition might not yet be evident, since the matrix $\mathbf{X_{1r}}$ is of the same size as matrix $\mathbf{X_1}$, they only differ by rank. The idea is to nevertheless use the SVD decomposition but also, to generate a matrix similar to the matrix $\mathbf{A}$ (since similar matrices share eigenvalues and eigenvectors, among some other properties) but one that will have a smaller size (in fact, it will be size $(r \times r)$). This similar matrix will be denoted $\underline{\mathbf{A}}$. Since it has a lower size than the original matrix $\mathbf{A}$, we will only retrieve $r$ eigenvectors and eigenvalues.
What will now follow are clever mathematical steps performed to avoid computation of the large matrix $\mathbf{A}$.
We come back to the eq.(15) and
We perform a *similarity transform* of the matrix $\mathbf{A}$:

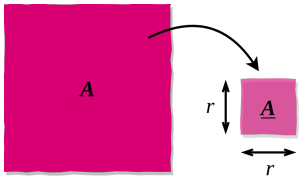


Figure 5: Similarity transform of matrix $\mathbf{A}$ to reduce its size.

$$\underline{\mathbf{A}} = \mathbf{U_r}^{T}\mathbf{A}\mathbf{U_r} \tag{19}$$

Matrix $\mathbf{A}$ can be written as:

$$\mathbf{A} = \mathbf{X_2}\mathbf{V_r}\mathbf{\Sigma_r}^{-1}\mathbf{U_r}^{T} \tag{20}$$

The similar matrix $\underline{\mathbf{A}}$ can be written as:

$$\underline{\mathbf{A}} = \mathbf{U_r}^{T}\mathbf{X_2}\mathbf{V_r}\mathbf{\Sigma_r}^{-1} \tag{21}$$

taking into account that $\mathbf{U_r}^{T}\mathbf{U_r} = \mathbf{I}$.
We have thus chosen a low-dimensional subspace by performing rank-$r$ truncation in which we now find the solution to the linear dynamical system presented initially. The solution is built in this low-dimensional subspace.

## 3.3 Eigendecomposition

Now that we have computed the similar matrix $\underline{\mathbf{A}}$, we move on to perform the eigendecomposition:

$$[\mathbf{W}, \mathbf{\Lambda}] = \text{eig}(\underline{\mathbf{A}}) \tag{22}$$

The matrix $\mathbf{W}$ is a matrix whose columns are the eigenvectors of $\underline{\mathbf{A}}$. The matrix $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues $\lambda_i$ which, in the most general case, are complex numbers. We have therefore obtained "coupled" quantities - each eigenvector has got its corresponding eigenvalue - and that link is exploited by the Dynamic Mode Decomposition to give a physical meaning to eigenvectors and eigenvalues.
The eigenvectors will form the spacial modes of DMD.
We can write the complex eigenvalue as:

$$\lambda_i = e^{(\sigma_i + i\omega_i)\Delta t} \tag{23}$$

The real part of the eigenvalue corresponds to the growth rate of that particular mode and the imaginary part represents the oscillatory frequency of a mode.

## 3.4 Going back to the original dimensions

In the previous section we have retrieved $r$ eigenvalues and eigenvectors but our original dimension for the propagator matrix was $n_t$. We have reduced the dimensionality of the problem but our final solution will only make sense to us if it is expressed in $n_t$ dimensions. Once the model has been built in the low-dimensional subspace, we want to move to the original dimensions.
The DMD modes are obtained from:

$$\mathbf{\Phi} = \mathbf{X_2}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W} \tag{24}$$

The above equation can be interpreted as transforming the eigenvectors matrix $\mathbf{W}$ to a new *basis vectors* matrix $\mathbf{\Phi}$ where the linear transformation is given by the composition: $\mathbf{X_2}\mathbf{V}\mathbf{\Sigma}^{-1}$.
What is worth mentioning here, is that DMD modes are not guaranteed to be orthogonal after such transformation. This creates a great capacity of DMD to be applicable to systems where data structure does not exhibit orthogonality.

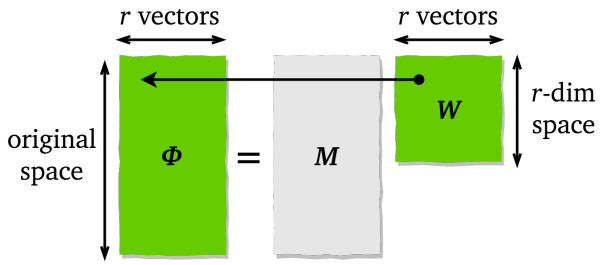


Figure 6: Every of the $r$ vectors in the eigenvectors matrix $\mathbf{W}$ gets transformed into one of the $r$ vectors in the DMD modes matrix $\mathbf{\Phi}$. $\mathbf{M}$ is the matrix associated with this transformation. We have moved from the $r$-dimensional space onto the original space.

## 3.5 DMD solution

The solution to the original dynamical system is finally computed:

$$\mathbf{x}(t) = \mathbf{\Phi}e^{\mathbf{\Omega}t}\mathbf{b} \tag{25}$$

the above equation is equivalent to:

$$\mathbf{x}(t) = \sum_{k=1}^{r} \phi_k e^{\omega_k t} b_k \tag{26}$$

The vector $\mathbf{b}$ is a vector of initial amplitudes for each mode. It is computed by projecting the initial state of the data matrix (given by $\mathbf{x}_1$) onto the DMD modes. This is done in order for our initial condition to be formulated in the obtained DMD basis.

# 4 A broader view

What can go different with our data sets?

## 4.1 Optimized DMD

- varying time steps
We mentioned earlier, that

## 4.2 Robust DMD

Sparse Identification

### 4.3 Multi-diagnostics DMD

## 5 Python example

## A Solution to linear dynamical systems

We first recall the general solution to the differential equation:

$$\frac{df(x)}{dt} = f(x) \tag{27}$$

to be the exponential function: $f(x) = a \cdot e^x$.

In an analogous way, the general solution to the linear dynamical system of the form:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} \tag{28}$$

is:

$$\mathbf{x} = \mathbf{v}e^{\check{}t} \tag{29}$$

Computing the time derivative of the eq. 29 we get:

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}\check{}e^{\check{}t} \tag{30}$$

And substituting the eq. 29 to eq. 28 we get:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{v}e^{\check{}t} \tag{31}$$

The nontrivial solution for the equality of these two above equations is obtained when:

$$\mathbf{A}\mathbf{v} = \check{}\mathbf{v} \tag{32}$$

which is the statement of eigenvalue problem.

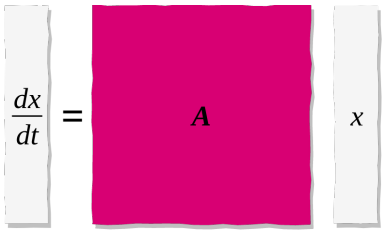


Figure 7: Linear dynamical system.

## B Moore-Penrose inverse

## C Singular Value Decomposition

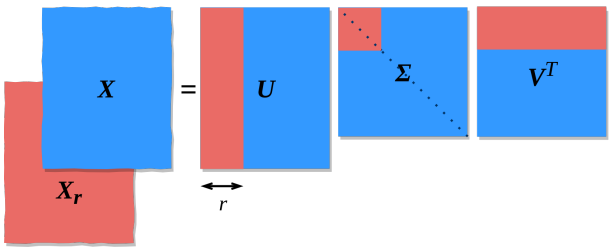


Figure 8: Sizes of component matrices in the Singular Value Decomposition and after rank truncation.

## References


[1] J. Grosek, N. Kutz, *Dynamic Mode Decomposition for Real-Time Background/Foreground Separation in Video*, 2014

[2] N. Kutz, *Dynamic Mode Decomposition Theory*, an online lecture: `https://youtu.be/bYfGVQ1Sg98`

[3] N. Kutz, *Dynamic Mode Decomposition Code* , an online lecture: `https://youtu.be/KAau5TBUOSc`

[4] `https://www.youtube.com/watch?v=ZCCETV-8950`

[5] E. R. Scheinerman, *Invitation to Dynamical Systems*

[6] G. Strang, *Introduction to Linear Algebra*, 5th edition

[7] K. Zdybal, *POD and DMD decomposition of numerical and experimental data*, von Karman Institute for Fluid Dynamics, stagiaire report

[8] K. Zdybal, *Exactly what is data decomposition?*