2 SVD and PCA

2.1 Singular Value Decomposition

The singular value decomposition (SVD) of a matrix $A \in \mathbb{C}^{n\times m}$ is arguably the most useful of all the available matrix factorizations. The SVD reveals the complexity of A, and can be used both when A is used as a data storage and as a linear mapping. Almost all modern model reduction techniques use the SVD in one way or another. In this section, the basic properties of the SVD are reviewed.

First, we introduce some notation. We use the induced 2-norm of a matrix and the Frobenius norm,

$$\begin{aligned}
\|u\| &:= \|u\|_2 = \sqrt{u^* u}, \quad u \in \mathbb{C}^m \\
\|A\| &:= \sup_x \frac{\|Au\|}{\|u\|} = \sqrt{\lambda_{\max}(A^* A)}, \quad A \in \mathbb{C}^{n\times m} \\
\|A\|_F &:= \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |A^{ij}|^2 \right)^{1/2} = \sqrt{\text{Trace}(A^* A)},
\end{aligned}$$

where * is the complex conjugate transpose. The induced norm has the property $\|AB\| \leq \|A\|\|B\|$, for all matrices B such that the product AB is defined. For unitary matrices U, V, we have $\|UAV\| = \|A\|$ and $\|UAV\|_F = \|A\|_F$.

The SVD of A is defined as follows. For all matrices $A \in \mathbb{C}^{n\times m}$ there exist unitary matrices

$$\begin{aligned}
U &= \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \in \mathbb{C}^{n\times n}, \\
V &= \begin{bmatrix} v_1 & \dots & v_m \end{bmatrix} \in \mathbb{C}^{m\times m},
\end{aligned}$$

($U^*U = UU^* = I$ and $V^*V = VV^* = I$) such that

$$A = U\Sigma V^*, \tag{2.1}$$

where

$$\Sigma_1 = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{pmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times m},$$

$p = \min\{n, m\}$, with the *singular values*

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0.$$

The proof of this statement is given in most modern books on linear algebra.

The following properties are useful:

- $\|A\| = \sigma_1 =: \overline{\sigma}$;
- $\|A\|_F = \left(\sum_{i=1}^{p} \sigma_i^2\right)^{1/2}$;
- if $\sigma_1 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_p = 0$, then Rank(A) $= k$;
- $N(A) =$ Span$\{v_{k+1}, \dots, v_m\}$ (orthonormal basis of the nullspace of A);
- $R(A) =$ Span$\{u_1, \dots, u_k\}$ (orthonormal basis of the range space of A);
- $AA^* = U\Sigma^2 U^*$, and $A^*A = V\Sigma^2 V^*$.

An alternative way to write (2.1) is the *dyadic expansion*

$$A = \sum_{i=1}^{k} \sigma_i u_i v_i^*. \tag{2.2}$$

One can view (2.2) as a series expansion of A, with the terms in decreasing order of importance. This follows from the *Schmidt-Mirsky* approximation theorem which is stated next. Assume we want to approximate the matrix A with a matrix $B \in \mathbb{C}^{n\times m}$ that has a rank smaller or equal to r. Then it holds that

$$\min_{\text{Rank}(B)\leq r} \|A - B\| = \|A - A_r\| = \sigma_{r+1},$$
$$\min_{\text{Rank}(B)\leq r} \|A - B\|_F = \|A - A_r\|_F = \left(\sum_{i=r+1}^{p} \sigma_i^2\right)^{1/2},$$

where A_r is a truncated dyadic expansion of A, $A_r := \sum_{i=1}^{r} \sigma_i u_i v_i^*$, retaining the r dominant terms. In the induced norm $\|\cdot\|$, A_r is not the unique minimizer (a fact that is used in optimal Hankel norm approximation). In the Frobenius norm $\|\cdot\|_F$, A_r is the unique minimizer if the singular values are distinct.

2.2 Principal Component Analysis (Proper Orthogonal Decomposition [POD])

Principal Component Analysis (PCA) can be viewed as an SVD of a function. PCA is extensively used for model reduction since it is a flexible tool for choosing good coordinate transformations T, either for truncation or for singular perturbation, see Section 1.3.

Consider a signal $x \in L_2^n[0, T]$, where

$$L_2^n[0,T] = \{x : x(t) \in \mathbb{C}^n,\ \|x\| < \infty\},$$
$$(x, y) = \int_0^T x(t)^* y(t) dt, \quad \|x\| = \sqrt{(x,x)} = \left(\int_0^T x(t)^* x(t) dt\right)^{1/2}.$$

We define the *Gramian* of x by

$$W = \int_0^T x(t)x(t)^* dt \in \mathbb{C}^{n\times n}.$$

The Gramian is a Hermitian positive semidefinite matrix ($W = W^*$). We define the n singular values of x by

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0, \quad \sigma_i = \sqrt{\lambda_i(W)},$$

with $\lambda_i(W)$ being the eigenvalues of W. We can now expand x as

$$x(t) = \sum_{i=1}^{n} \sigma_i u_i v_i(t)^*, \tag{2.3}$$

where

- $v_i \in L_2^1[0,T]$, $(v_i, v_j) = 0$ if $i \neq j$, and $\|v_i\| = 1$;
- $u_i \in \mathbb{C}^n$, $u_i^* u_j = 0$, if $i \neq j$, and $\|u_i\| = 1$;
- $v_i(t)^* = u_i^* x(t)/\sigma_i$, and $\sigma_i^2 u_i = W u_i$.

The expansion (2.3) should be compared to the dyadic expansion (2.2). As seen, it can be computed from the eigenvalues and the eigenvectors of the Gramian W (or the SVD). The i-th *principal component* of x is defined by $\sigma_i u_i v_i(t)^*$, the i-th *component vector* by u_i, and the i-th *component function* by $v_i(t)$. The component vectors describe in decreasing order of importance where the energy of the signal x is found in $\mathbb{C}^n$, as seen next. The total energy of the signal x is given by the singular values,

$$\|x\| = \left(\sum_{i=1}^{n} \sigma_i^2 \right)^{1/2}.$$

Define the subspace in $\mathbb{C}^n$ that is spanned by the component vectors u_i that correspond to the $k \leq n$ strictly positive singular values of x ($\sigma_1 \geq \ldots \geq \sigma_k > \sigma_{k+1} = \ldots = \sigma_n = 0$) as

$$S_x = \text{Span}\{u_1, \ldots, u_k\}.$$

If x evolves over the entire space $\mathbb{C}^n$, then $\dim S_x = n$. Often, however, x falls approximately on a subspace of lower dimension. We want to find a subspace of dimension $r < k \leq n$ that captures as much as possible of the energy of x. This can be done by truncating the expansion (2.3). We have

$$\min_{\dim S_y \leq r} \|x - y\| = \|x - x_r\| = \left(\sum_{i=r+1}^{n} \sigma_i^2 \right)^{1/2}, \tag{2.4}$$

where $x_r(t) := \sum_{i=1}^{r} \sigma_i u_i v_i(t)^*$. If there is a significant drop in the magnitude of the singular values after σ_r, then typically only r dimensions, spanned by $u_1, \ldots, u_r$, are needed to model x accurately.

2.3 Controllability Analysis

Let us apply PCA to analyze how the state-space of a model G is excited by an input u. The input-to-state mapping is given by

$$x(t) = \int_0^t e^{A(t-\tau)} B u(\tau) d\tau.$$

To get a unique signal $x \in L_2^n[0, T]$ to analyze, let us apply an impulse $u(t) = \delta(t)$. We then have the impulse response signal

$$x(t) = e^{At} B,$$

which lies in $L_2^n[0, \infty)$, if A is Hurwitz. The corresponding Gramian is called the *reachability Gramian*, and is given by

$$P(T) = \int_0^T e^{At} B B^T e^{A^T t} dt \in \mathbb{R}^{n \times n}. \tag{2.5}$$

By analyzing the principal components of x through $P(T)$, we can find new coordinate systems $T\bar{x} = x$ that capture the subspace of $\mathbb{R}^n$ that is most excited by the input u.

In practice, it is often convenient to compute the reachability Gramian through the Lyapunov differential equation

$$\dot{P} = AP + PA^T + BB^T, \quad P(0) = 0. \tag{2.6}$$

When $T \to \infty$ and A is Hurwitz, P can be computed from the algebraic Lyapunov equation

$$AP + PA^T + BB^T = 0. \tag{2.7}$$

As we shall see in the coming lectures, $P(T)$ contains a lot of information. For example, the system is controllable if, and only if, $P(T)$ is nonsingular, and the controllable subspace is spanned by $u_1, \ldots, u_r$, when $\text{Rank}(P(T)) = r$.

2.4 Recommended Reading

The SVD is treated in most modern text books on linear algebra. In Section 2.2 of *Linear Robust Control*, many basic properties of the SVD are listed.

The paper *Principal Component Analysis in Linear Systems* by Bruce Moore pioneered the use of PCA for model reduction. Sections I–IV in the paper is recommended reading. Balanced coordinates, which were introduced for model reduction by Moore, will be treated in the next lecture.

2.5 Exercises

EXERCISE 2.1

a) Assume that $A = U\Sigma V^T$ is invertible. What is the SVD of A^{-1}?

b) Compute the SVD of $A = \begin{pmatrix} 1 & 2 \\ 4 & 1 \end{pmatrix}$. Use the SVD to illustrate the mapping $y = Au$. What direction in $\mathbb{R}^2$ is amplified most by A?

EXERCISE 2.2

a) Compute the principal components of the impulse response $x(t) = e^{At}B$, $0 \leq t \leq \infty$, where

$$A = \begin{pmatrix} -a & 0 \\ 0 & -a-\epsilon \end{pmatrix}, \; a > 0, \; a > |\epsilon| \geq 0, \quad B = \begin{pmatrix} b \\ b \end{pmatrix}.$$

Sketch the phase portrait $(x^1(t), x^2(t))$ and the component vectors as a function of ϵ. (You may assume ϵ is small and use suitable approximations.)

b) Compute the principal components of the impulse response $x(t) = e^{At}B$, $0 \leq t \leq 10$, where

$$A = \begin{pmatrix} -0.5 & 2 \\ -2 & -0.5 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Sketch the phase portrait $(x^1(t), x^2(t))$ and the component vectors.

c) Suggest coordinate transformations $x = T\bar{x}$ for a)–b) above, and peform model truncations, see Section 1.3. You can use

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

EXERCISE 2.3

a) Prove that the reachability Gramian $P(t)$ as defined in (2.5) satisfies (2.6).

b) How is P transformed when the coordinates are changed using a linear coordinate transformation T, $T\bar{x} = x$?

c) Show that $P(t_2) \geq P(t_1)$, $t_2 \geq t_1$. ($P(t_2) \geq P(t_1)$ means that $P(t_2) - P(t_1)$ is positive semidefinite.)