18 Lecture 18: Introduction to Singular Value Decompositions

Outline

1. Geometric Motivation: $AV = U\Sigma$
   (a) Matrix Form
   (b) Comparison with Eigendecomposition
2. Properties of the SVD
3. Computing the SVD - 1st Attempt
   (a) Example

This lecture introduces the final decomposition called the **singular value decomposition**. Lecture 4 of Trefethen & Bau provides more detail, see https://people.maths.ox.ac.uk/trefethen/text.html.

18.1 Geometric Motivation: $AV = U\Sigma$

The **image** of the unit hypersphere S in $\mathbb{R}^n$ under any $m \times n$ matrix transformation A is a **hyperellipse** in $\mathbb{R}^m$. Figure 18.31 shows the geometric interpretation of this transformation. Both the hypersphere and hyperellipse are in $\mathbb{R}^2$ in this example. However, the dimensions can be any n and m, not necessarily $n = m$.

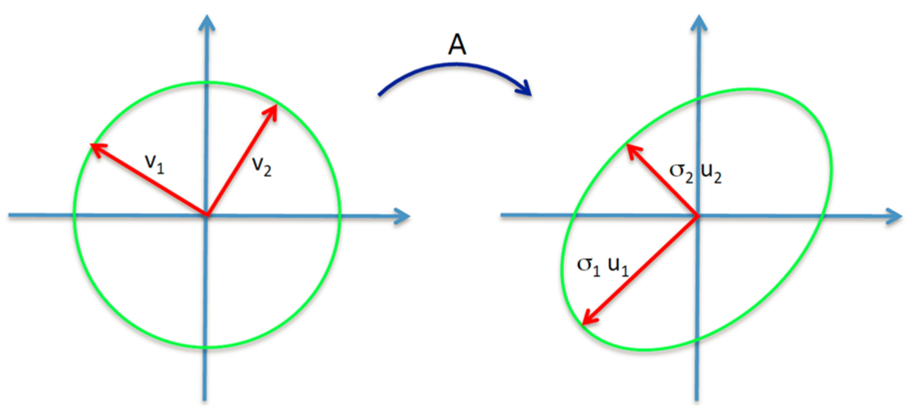


Figure 18.31: Transformation of unit hypersphere S (left) by matrix A into hyperellipse AS (right).

The factors by which the hypersphere is scaled in each of the principal semi-axes of the hyperellipse are called the **singular values** of A. The n singular values are denoted σ_i. By convention we will order them such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

Notice that all the singular values are non-negative.

The n **left singular vectors**, u_i, of A are the unit vectors in the directions of the principal semi-axes of the ellipse. The n **right singular vectors**, v_i, are the unit vectors in S such that

$$Av_j = \sigma_j u_j.$$

In other words, v_i's are the **pre-image** of u_i's under the transformation A.

18.1.1 Matrix Form

We can write the above equation

$$Av_j = \sigma_j u_j, \qquad \text{for } j = 1, 2, \ldots, n,$$

in matrix form to define the **reduced SVD**. Pictorially, we have

$$\underbrace{\begin{bmatrix} \\ & A & \\ \\ \end{bmatrix}}_{A,\ m\times n} \underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}}_{V,\ n\times n}$$

$$= \underbrace{\begin{bmatrix} \\ u_1 & u_2 & \cdots & u_n \\ \\ \end{bmatrix}}_{\hat{U},\ m\times n} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}}_{\hat{\Sigma},\ n\times n}$$

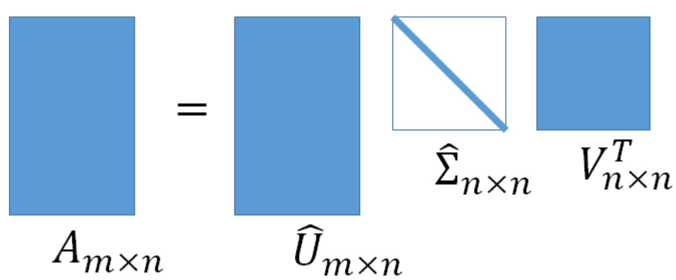
The matrix $\hat{\Sigma}$ is a diagonal matrix, with the singular values of A on its diagonal. The matrices $\hat{U}$ and V have orthonormal columns. Note that the hat notation indicates **reduced** or **economy-sized** SVD

$$AV = \hat{U}\hat{\Sigma}.$$

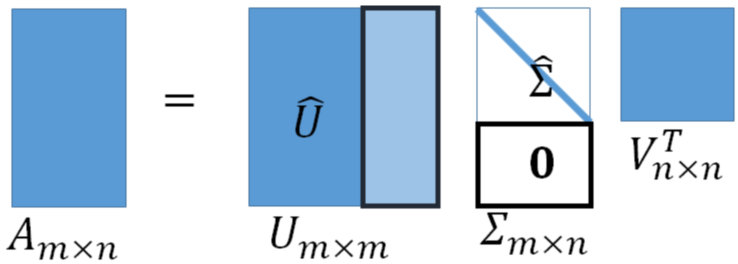
Since V is orthogonal, if we multiply by V^T on the right, we can equivalently write it as

$$A = \hat{U}\hat{\Sigma}V^T.$$

The figure below shows this reduced SVD of A pictorially.



The **full SVD** is constructed in a similar way to how the full QR factorization was created from the reduced QR factorization. We can define a full SVD by adding $m - n$ more orthonormal columns to $\hat{U}$ to give a square, **orthogonal** U. Then we must also add extra empty rows to $\hat{\Sigma}$ to construct Σ. That is, replace $\hat{U} \to U$ and $\hat{\Sigma} \to \Sigma$ as shown in the figure below.



- Every matrix $A \in \mathbb{R}^{m\times n}$ has a singular value decomposition.
  - We will prove this in the next lecture.
  - Furthermore, the singular values are uniquely determined.
- Also, if A is square and σ_j are distinct, then the left and right singular vectors are unique **(up to signs)**.

18.1.2 Comparison with Eigendecomposition

The SVD is similar to the eigendecomposition we have seen previously. Consider the SVD vs the eigendecomposition

$$A = U\Sigma V^T \quad \text{vs} \quad A = X\Lambda X^{-1}.$$

- Both decompositions act to diagonalize a matrix.
- The SVD uses two bases: U and V, the left and right singular vectors.
- The eigendecomposition uses only one basis, the set of eigenvectors.
- The SVD always uses orthonormal vectors.
- The eigenvectors are not orthonormal **in general** (though for the real symmetric matrices we considered, they are).
- Finally, not all matrices have an eigendecomposition, but **all matrices** have an SVD, even rectangular matrices.

18.2 Properties of the SVD

Next we will discuss some properties of the SVD. For the following theorems let $A \in \mathbb{R}^{m\times n}$ and $r = \#$ of non-zero singular values.

Theorem 18.1.

$$\text{rank}(A) = r.$$

Proof. Rank of a diagonal matrix is the number of non-zero diagonal entries. U and V are both of full rank, by definition. Hence $\text{rank}(A) = \text{rank}(\Sigma) = r$. □

Theorem 18.2.

$$\begin{aligned}\text{range}(A) &= \text{span}\{u_1, u_2, \ldots, u_r\},\\ \text{null}(A) &= \text{span}\{v_{r+1}, \ldots, v_n\}.\end{aligned}$$

Proof. We will <u>not</u> give a full proof of this theorem. Instead for the second property, we show that a vector in $\text{span}\{v_{r+1},\ldots,v_n\}$ is in $\text{null}(A)$ (in other words, $\text{span}\{v_{r+1},\ldots,v_n\} \subseteq \text{null}(A)$).

Let $x \in \text{span}\{v_{r+1},\ldots,v_n\}$ be arbitrary. Then

$$x = \sum_{i=r+1}^{n} w_i v_i \qquad \text{and so} \qquad Ax = \sum_{i=r+1}^{n} w_i (A v_i).$$

Observe that

$$\begin{aligned} Av_i &= U\Sigma V^T v_i \\ &= U\Sigma e_i, \end{aligned}$$

with the last equality holding because V is an orthogonal matrix. But $\Sigma e_i = 0$ for $i \in [r+1, n]$ since the corresponding entries of Σ are zero. Therefore $Ax = 0$, so $x \in \text{null}(A)$. $\square$

I claim that

$$\|A\|_2^2 = \lambda_{\max}\left(A^T A\right).$$

Proof.

- Recall that

$$\begin{aligned} \|A\|_2^2 &= \max_{\|x\|_2=1} \|Ax\|_2^2 \\ &= \max_{\lambda \in \Lambda(A)} |\lambda|^2 \\ &= \max_{\lambda \in \Lambda(A)} \lambda^2. \end{aligned}$$

- The SVD of A gives

$$\begin{aligned} A &= U\Sigma V^T, \text{ so that} \\ A^T A &= \left(U\Sigma V^T\right)^T U\Sigma V^T \\ &= V\Sigma^T \underbrace{U^T U}_{=I} \Sigma V^T \\ &\underbrace{=}_{\Sigma \text{ is diagonal, thus symmetric}} V\Sigma^2 V^T. \end{aligned}$$

- This is a similarity transformation, hence the eigenvalues of $A^T A$ equal the eigenvalues of Σ^2.
- But Σ is a diagonal matrix with A's singular values on its diagonal.
- Hence the eigenvalues of $A^T A$ equal the squares of the singular values of A.

$\square$

Notation:

$$\|A\|_F = \sum_{i,j} a_{ij}^2 = \text{tr}\left(A^T A\right) \text{ is the } \textbf{Frobenius norm}.$$

Recall that

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii},$$

the sum of the diagonal entries of A.

Theorem 18.3. $\|A\|_2 = \sigma_1$ *and* $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.

Proof. We have $\lambda_{\max}(A^TA) = \lambda_{\max}(\Sigma^2) = \sigma_1^2 \ \Rightarrow \ \|A\|_2 = \sigma_1$.

Now for the Frobenius norm we have

$$\begin{aligned}
& \|A\|_F^2 \\
= \quad & \mathrm{tr}(A^TA) \\
\underbrace{=}_{A^TA=V\Sigma^2V^T} \quad & \mathrm{tr}\left(V\Sigma^2V^T\right) \\
= \quad & \mathrm{tr}\left((V\Sigma)(V\Sigma)^T\right), \\
= \quad & \mathrm{tr}\left((V\Sigma)^T(V\Sigma)\right), \text{ trace identity } \mathrm{tr}(X^TY) = \mathrm{tr}(XY^T), \\
= \quad & \mathrm{tr}\left(\Sigma V^T V\Sigma\right), \\
= \quad & \mathrm{tr}(\Sigma^2), \text{ by the orthogonality of } V, \\
= \quad & \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2.
\end{aligned}$$

□

Theorem 18.4. *Non-zero singular values of A are the square roots of non-zero eigenvalues of AA^T or A^TA.*

Proof. A^TA and AA^T are similar to Σ^2. We proved this for A^TA above; the proof for AA^T is similar.

1. We showed above that $A^TA = V\Sigma^2V^T$.
2. Similarly,

$$\begin{aligned}
AA^T &= U\Sigma V^T\left(U\Sigma V^T\right)^T \\
&= U\Sigma\overset{I}{\cancel{(V^TV)}}\Sigma^TU^T \\
&= U\Sigma^2U^T, \text{ since } \Sigma \text{ is diagonal.}
\end{aligned}$$

□

Recall Notation: $\Lambda(A)$ is the set of eigenvalues of A.

New Notation: $\sigma(A)$ is the set of the singular values of A.

Theorem 18.5. *If $A = A^T$, then $\sigma(A) = \{|\lambda| : \lambda \in \Lambda(A)\}$. In particular, if A is SPD then $\sigma(A) = \Lambda(A)$.*

Proof. Real symmetric matrices have orthogonal eigenvectors and real eigenvalues, so

$$A = Q\Lambda Q^T, \quad \text{with } Q \text{ orthogonal.}$$

Construct the SVD as

$$A = \underbrace{Q}_{U}\,\underbrace{|\Lambda|}_{\Sigma}\,\underbrace{\mathrm{sign}(\Lambda)Q^T}_{V^T},$$

where $|\Lambda|$ and $\mathrm{sign}(\Lambda)$ are diagonal matrices with entries $|\lambda_j|$ and $\mathrm{sign}(\lambda_j)$, respectively. If desired one can also insert orthogonal permutation matrices to sort the σ's. $\square$

Theorem 18.6. *The* ***condition number*** *for $A \in \mathbb{R}^{n\times n}$ is $\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$.*

Proof. By the definition of κ and by Theorem 18.3, we have

$$\kappa_2(A) = \|A\|_2\|A^{-1}\|_2 = \sigma_1\|A^{-1}\|_2.$$

Since $A = U\Sigma V^T$, therefore $A^{-1} = V\Sigma^{-1}U^T$ is the SVD of A^{-1}. Therefore

$$\begin{aligned} \|A^{-1}\|_2 &= \frac{1}{\sigma_n} \\ \Rightarrow \kappa_2(A) &= \frac{\sigma_1}{\sigma_n}. \end{aligned}$$

$\square$

18.3 Computing the SVD - 1st Attempt

We first consider a naïve approach to computing the SVD. Since $A = U\Sigma V^T$ we showed above that $A^TA = V\Sigma^2V^T$, which is an eigendecomposition of A^TA! Therefore, the eigen**values** of A^TA are **squares** of the singular values of A. The eigen**vectors** of A^TA are the **right singular** vectors of A.

This suggests a (naïve) method for computing the SVD:

1. Form A^TA (it's symmetric and positive semi-definite, so its eigenvalues are real and non-negative),
2. Compute eigendecomposition of $A^TA = V\Lambda V^T$,
3. Compute $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$, where $\sigma_i = \sqrt{\lambda_i}$ and $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$,
4. Solve $U\Sigma = AV$ for orthogonal U (e.g., by QR factorization).

Recovering U from the above algorithm involves (note, we already have Σ, A, V):

- Multiply AV to get A',
- QR factor $A' = QR$,
- Identify $U = Q, \Sigma = R$.

This ensures that $U = Q$ is properly orthogonal. Conveniently, $R = \Sigma$ will be diagonal.

Unfortunately, this naïve method is inaccurate; the error satisfies

$$|\tilde{\sigma}_k - \sigma_k| = O\left(\frac{\epsilon \|A\|^2}{\sigma_k}\right),$$

which can be very bad for small singular values! (**Conceptually**, this is similar to how solving least squares by normal equations used $A^T A$. Effectively this "squares the condition number", therefore making it less accurate than QR factorization). In the next lecture we will discuss a better alternative for computing the SVD.

18.3.1 Example

We can find the SVD of $A = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}$ in a few different ways.

1. **Method 1:**

$$\begin{aligned} A^T A &= \begin{bmatrix} 0 & 3 & 0 \\ -\frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \\ &= U\Sigma^2 V^T \\ &= Q\Lambda Q^T \end{aligned}$$

Therefore $\lambda_1 = 9, \lambda_2 = \frac{1}{4}, v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ since $Q = I$. Therefore $\sigma_1 = 3, \sigma_2 = \frac{1}{2}$, so

$$\hat{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then find U from $U\Sigma = AV$

$$\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}$$

Hence

$$\begin{aligned} 3u_1 &= \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \text{ therefore } u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ \frac{1}{2}u_2 &= \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 0 \end{bmatrix} \text{ therefore } u_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

Thus $\hat{U} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$.

2. **Method 2:** Use AA^T instead, same idea.
3. **Method 3:** Let's exploit intuition about SVD and the simple structure of this matrix. By inspection, range(A) = span$\{u_1, u_2\}$ for $u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, u_1 and u_2 are orthonormal. The lengths of the principal axes are 3 and $\frac{1}{2}$.
   Then by the definition of SVD

$$\begin{aligned}
Av_1 &= \sigma_1 u_1 \\
\begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} v_1 &= 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
\Rightarrow v_1 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
Av_2 &= \sigma_2 u_2 \\
\begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} v_2 &= \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
\Rightarrow v_2 &= \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\end{aligned}$$

   The details of solving both systems follow.

$$
\begin{aligned}
& \left[\begin{array}{cc|c} 0 & -\frac{1}{2} & 0 \\ 3 & 0 & 3 \\ 0 & 0 & 0 \end{array}\right] \\
\sim\ & \left[\begin{array}{cc|c} 3 & 0 & 3 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{array}\right] \begin{array}{l} {\scriptstyle R_1 \leftarrow R_2} \\ {\scriptstyle R_2 \leftarrow R_1} \\ \end{array} \\
\sim\ & \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \begin{array}{l} {\scriptstyle R_1 \leftarrow \frac{1}{3} R_1} \\ {\scriptstyle R_2 \leftarrow -2 R_2} \\ \end{array} \\
\Rightarrow v_1 = & \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
& \left[\begin{array}{cc|c} 0 & -\frac{1}{2} & \frac{1}{2} \\ 3 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \\
\sim\ & \left[\begin{array}{cc|c} 3 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{array}\right] \begin{array}{l} {\scriptstyle R_1 \leftarrow R_2} \\ {\scriptstyle R_2 \leftarrow R_1} \\ \end{array} \\
\sim\ & \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right] \begin{array}{l} {\scriptstyle R_1 \leftarrow \frac{1}{3} R_1} \\ {\scriptstyle R_2 \leftarrow -2 R_2} \\ \end{array} \\
\Rightarrow v_2 = & \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\end{aligned}
$$

So $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ i.e. same solution, up to signs in U and V.

19 Lecture 19: Singular Value Decompositions Versus Eigendecomposition

Outline

1. Alternative Formulation
   (a) Alternate Approach Example
2. Proof of Existence of SVD
3. Stability Comparison
4. Golub-Kahan Bidiagonalization

Recall that SVD is the decomposition of any matrix A into $U\Sigma V^T$, where Σ is diagonal with non-negative entries, and U, V are orthogonal. In the previous lecture we seen that the SVD can be found from the eigendecomposition of A^TA or AA^T.

In this lecture we will see a more stable method using the eigendecomposition of

$$H = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}.$$

We will also prove the existence of the SVD, discuss stability, and discuss how to compute the SVD efficiently.

19.1 Alternative Formulation

Assume A is square, i.e. $A \in \mathbb{R}^{n\times n}$. Consider the $2n \times 2n$ symmetric matrix

$$H = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}.$$

By computing the eigendecomposition of $H = Q\Lambda Q^T$ we can extract the singular values and vectors. We have that $\sigma_A = |\lambda_H|$, and U, V can be recovered from the eigenvectors. Let us see why all of this holds.

Write $A = U\Sigma V^T$, then we have $AV = U\Sigma$ because V is orthogonal. Likewise,

$$\begin{aligned} A^T &= (U\Sigma V^T)^T \\ &= V\Sigma^T U^T \\ &= V\Sigma U^T, \text{ because } \Sigma \text{ is diagonal.} \end{aligned}$$

Thus $A^TU = V\Sigma$ because U is orthogonal. Hence we have

$$\begin{aligned} \underbrace{\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} V & V \\ U & -U \end{bmatrix}}_{Q} &= \begin{bmatrix} A^TU & -A^T{\color{red}U} \\ {\color{red}AV} & {\color{red}AV} \end{bmatrix} \\ &= \begin{bmatrix} V\Sigma & -V\Sigma \\ U\Sigma & U\Sigma \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} V & V \\ U & -U \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}}_{\Lambda} \end{aligned}$$

Therefore, $HQ = Q\Lambda$, equivalently $H = Q\Lambda Q^T$, gives an **eigendecomposition** of H. Note, we need to normalize the columns of Q, to make Q an orthogonal matrix.

Explanation Of Why Q's Columns Are Orthogonal, Given U, V Are Orthogonal Matrices

- Let $1 \le i < j \le n$ be arbitrary.
- Then we have

$$\begin{aligned}
& \begin{bmatrix} v_i \\ u_i \end{bmatrix}^T \begin{bmatrix} v_j \\ \pm u_j \end{bmatrix} \\
= \quad & \begin{bmatrix} v_i^T & u_i^T \end{bmatrix} \begin{bmatrix} v_j \\ \pm u_j \end{bmatrix} \\
= \quad & v_i^T v_j \pm u_i^T u_j \\
\underbrace{=}_{U,V \text{ are orthogonal}} \quad & 0 \pm 0 \\
= \quad & 0.
\end{aligned}$$

To summarize we have the following steps:

1. Form $H = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}$,
2. Compute eigendecomposition $HQ = Q\Lambda$,
3. Set $\sigma_A = |\lambda_H|$,
4. Extract U, V from Q (normalizing for orthogonality).

This algorithm is preferable with respect to stability (see the more detailed section below). The error in the singular values satisfies $|\tilde{\sigma}_k - \sigma_k| = O\left(\epsilon \|A\|\right)$, compared to $O\left(\epsilon \|A\|^2 / \sigma_k\right)$ for the algorithm using $A^T A$. This approach can be extended to non-square matrices too. Practical algorithms are based on this premise, but without explicitly forming the (large) matrix H.

19.1.1 Alternate Approach Example

$$A = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \end{bmatrix}$$

Therefore

$$H = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix}$$

MATLAB eigendecomposition gives

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \Lambda = \begin{bmatrix} -3 & & & \\ & -\frac{1}{2} & & \\ & & \boxed{\dfrac{1}{2}} & \\ & & & \boxed{3} \end{bmatrix}$$

Order may be different (of cols) so read off desired cols, for positive Σ entries.

One can verify that the following eigendecomposition (permuting the columns of the previous one, to change the order of the eigenvalues) is also correct, and perfectly fits the shape required of our setup:

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{bmatrix}, \Lambda = \begin{bmatrix} \boxed{3} & & & \\ & \boxed{\dfrac{1}{2}} & & \\ & & -3 & \\ & & & -\frac{1}{2} \end{bmatrix}$$

Even better, this modified eigendecomposition fits in perfectly with the remainder of the computation.

Therefore

$$\begin{aligned} \sigma_1 &= 3 \\ v_1 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ u_1 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ \sigma_2 &= \frac{1}{2} \\ v_2 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ u_2 &= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \end{aligned}$$

So

$$\begin{aligned} U &= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\ V &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \Sigma &= \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \end{aligned}$$

Check:

$$
\begin{aligned}
U\Sigma V^T &= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \end{bmatrix} \\
&= A
\end{aligned}
$$

19.2 Proof of Existence of SVD

We claimed in Lecture 18 that every matrix $A \in \mathbb{R}^{m \times n}$ has a singular value decomposition. We will now prove this result.

Proof. Let A be an arbitrary $m \times n$ matrix. The proof is by induction on $n \geq 1$.

Recall that the induced matrix norm is defined as

$$\|A\| := \max_{\|x\|=1} \|Ax\| .$$

Let $\sigma_1 = \|A\|_2$. Let v_1 have $\|v_1\|_2 = 1$ and a direction such that $\|Av_1\|_2 = \|A\|_2 = \sigma_1$. Also, let $u_1 = \frac{Av_1}{\sigma_1}$, so that $Av_1 = \sigma_1 u_1$.

Consider any extensions of vectors u_1 and v_1 to orthonormal bases U_1 and V_1:

$$
\begin{aligned}
U_1 &= \begin{bmatrix} u_1 | \cdots \end{bmatrix}, \\
V_1 &= \begin{bmatrix} v_1 | \cdots \end{bmatrix}.
\end{aligned}
$$

Then we have

$$U_1^T A V_1 \underbrace{=}_{\text{definition}} S = \begin{bmatrix} \sigma_1 & w^T \\ 0 & B \end{bmatrix},$$

where 0 is the $m-1$ column vector, w^T is a $n-1$ row vector, and B has dimensions $(m-1) \times (n-1)$.

Note, the top-left comes from

$$
\begin{aligned}
Av_1 &= \sigma_1 u_1 \\
u_1^T A v_1 &= \sigma_1 \underbrace{u_1^T u_1}_{=1} \\
&= \sigma_1.
\end{aligned}
$$

The bottom-left is zero because

$$
\begin{aligned}
u_i^T A v_1 &= u_i^T (\sigma_1 u_1), \\
&= \sigma_1 u_i^T u_1, \\
&= 0, \ \forall i > 1.
\end{aligned}
$$

Now, we can show $w = 0$ as follows: Consider

$$\begin{aligned}
& \left\| \begin{bmatrix} \sigma_1 & w^T \\ 0 & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 \\
= & \left\| \begin{bmatrix} \sigma_1^2 + w^T w \\ Bw \end{bmatrix} \right\|_2 \\
= & \sqrt{\left(\sigma_1^2 + w^T w\right)^2 + (Bw)^T Bw}, \\
= & \sqrt{\left(\sigma_1^2 + w^T w\right)^2 + w^T \underbrace{B^T B}_{\text{symm, + semi-def}} w}, \\
\geq & \ \sigma_1^2 + w^T w \\
= & \sqrt{\sigma_1^2 + w^T w} \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 .
\end{aligned}$$

By the induced matrix norm definition, this implies $\|S\|_2 \geq \sqrt{\sigma_1^2 + w^T w}$. However, since U_1 and V_1 are orthogonal, therefore $\|S\|_2 = \|A\|_2 = \sigma_1$. Thus we must have $w = 0$, and so

$$U_1^T A V_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}.$$

For $n = 1$ (base case) this completes the proof. For $n > 1$ (induction case), by the inductive hypothesis, the SVD of B exists. Write $B = U_2 \Sigma_2 V_2^T$. Then if we let

$$A = \underbrace{U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix} V_1^T}_{V^T},$$

it is easy to verify that this is an SVD of A. (Some explanation is given below.)

Therefore, in either case, the SVD of A always exists. □

Additional Explanation: Earlier, we had

$$U_1^T A V_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix}.$$

Recall that U_1 and V_1 are orthogonal. Therefore left multiplying by U_1 and right multiplying by V_1 yields

$$A = U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix} V_1^T.$$

Thus it suffices to prove that

$$\begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix},$$

where $B = U_2 \Sigma_2 V_2^T$.