Notes on the Singular Value Decomposition

Robert A. van de Geijn
The University of Texas at Austin
Austin, TX 78712

September 11, 2014

NOTE: I have not thoroughly proof-read these notes!!

We recommend the reader review Weeks 9-11 of Linear Algebra: Foundations to Frontiers - Notes to LAFF With.

1 Orthogonality and Unitary Matrices

Definition 1. Let $u, v \in \mathbb{C}^m$. These vectors are orthogonal (perpendicular) if $u^H v = 0$.

Definition 2. Let $q_0, q_1, \ldots, q_{n-1} \in \mathbb{C}^m$. These vectors are said to be mutually orthonormal if for all $0 \leq i, j < n$

$$q_i^H q_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$  

Notice that for $n$ vectors of length $m$ to be mutually orthonormal, $n$ must be less than or equal to $m$. This is because $n$ mutually orthonormal vectors are linearly independent and there can be at most $m$ linearly independent vectors of length $m$.

Definition 3. Let $Q \in \mathbb{C}^{m \times n}$ (with $n \leq m$). Then $Q$ is said to be an orthonormal matrix if $Q^H Q = I$ (the identity).

Exercise 4. Let $Q \in \mathbb{C}^{m \times n}$ (with $n \leq m$). Partition $Q = \left( \begin{array}{c|c|c|c} q_0 & q_1 & \cdots & q_{n-1} \end{array} \right)$. Show that $Q$ is an orthonormal matrix if and only if $q_0, q_1, \ldots, q_{n-1}$ are mutually orthonormal.

Definition 5. Let $Q \in \mathbb{C}^{m \times m}$. Then $Q$ is said to be a unitary matrix if $Q^H Q = I$ (the identity).

Exercise 6. Let $Q \in \mathbb{C}^{m \times m}$. Show that if $Q$ is unitary then $Q^{-1} = Q^H$ and $QQ^H = I$.

Exercise 7. Let $Q_0, Q_1 \in \mathbb{C}^{m \times m}$ both be unitary. Show that their product, $Q_0 Q_1$, is unitary.

Exercise 8. Let $Q_0, Q_1, \ldots, Q_{k-1} \in \mathbb{C}^{m \times m}$ all be unitary. Show that their product, $Q_0 Q_1 \cdots Q_{k-1}$, is unitary.

The following is a very important observation: Let $Q$ be a unitary matrix with

$$Q = \left( \begin{array}{c|c|c|c} q_0 & q_1 & \cdots & q_{m-1} \end{array} \right).$$
Let $x \in \mathbb{C}^m$. Then

$$x = QQ^H x = \left( \begin{array}{c|c|c|c|c} q_0 & q_1 & \cdots & q_{m-1} \end{array} \right) \left( \begin{array}{c|c|c|c|c} q_0^H & \cdots & q_{m-1}^H \end{array} \right) x$$

$$= \left( \begin{array}{c|c|c|c|c} q_0 & q_1 & \cdots & q_{m-1} \end{array} \right) \left( \begin{array}{c} q_0 x \\ q_1 x \\ \vdots \\ q_{m-1} x \end{array} \right)$$

$$= \left( q_0 x \right) q_0 + \left( q_1 x \right) q_1 + \cdots + \left( q_{m-1} x \right) q_{m-1}.$$ 

What does this mean?

- The vector $x = \left( \begin{array}{c} \chi_0 \\ \chi_1 \\ \vdots \\ \chi_{m-1} \end{array} \right)$ gives the coefficients when the vector $x$ is written as a linear combination of the unit basis vectors:

$$x = \chi_0 e_0 + \chi_1 e_1 + \cdots + \chi_{m-1} e_{m-1}.$$ 

- The vector

$$Q^H x = \left( \begin{array}{c} q_0^H x \\ q_1^H x \\ \vdots \\ q_{m-1}^H x \end{array} \right)$$

gives the coefficients when the vector $x$ is written as a linear combination of the orthonormal vectors $q_0, q_1, \ldots, q_{m-1}$:

$$x = \left( q_0^H x \right) q_0 + \left( q_1^H x \right) q_1 + \cdots + \left( q_{m-1}^H x \right) q_{m-1}.$$ 

- The vector $\left( q_i^H x \right) q_i$ equals the component of $x$ in the direction of vector $q_i$.

Another way of looking at this is that if $q_0, q_1, \ldots, q_{m-1}$ is an orthonormal basis for $\mathbb{C}^m$, then any $x \in \mathbb{C}^m$ can be written as a linear combination of these vectors:

$$x = \alpha_0 q_0 + \alpha_1 q_1 + \cdots + \alpha_{m-1} q_{m-1}.$$ 

Now,

$$q_i^H x = \left( q_i^H\alpha_0 q_0 + \alpha_1 q_1 + \cdots + \alpha_{i-1} q_{i-1} + \alpha_i q_i + \alpha_{i+1} q_{i+1} + \cdots + \alpha_{m-1} q_{m-1} \right)$$

$$= \alpha_0 q_i^H q_0 + \alpha_1 q_i^H q_1 + \cdots + \alpha_{i-1} q_i^H q_{i-1} + \alpha_i q_i^H q_i + \alpha_{i+1} q_i^H q_{i+1} + \cdots + \alpha_{m-1} q_i^H q_{m-1}$$

$$= \alpha_i.$$ 

Thus $q_i^H x = \alpha_i$, the coefficient that multiplies $q_i$. 

2
Exercise 10. Let \( U \in \mathbb{C}^{m \times m} \) and \( V \in \mathbb{C}^{n \times n} \) be unitary matrices and \( A \in \mathbb{C}^{m \times n} \). Then \( \|UA\|_2 = \|AV\|_2 = \|A\|_2 \).

Exercise 11. Let \( U \in \mathbb{C}^{m \times m} \) and \( V \in \mathbb{C}^{n \times n} \) be unitary matrices and \( A \in \mathbb{C}^{m \times n} \). Then \( \|UA\|_F = \|AV\|_F = \|A\|_F \).

2 Toward the SVD

Lemma 12. Given \( A \in \mathbb{C}^{m \times n} \) there exists unitary \( U \in \mathbb{C}^{m \times m} \), unitary \( V \in \mathbb{C}^{n \times n} \), and diagonal \( D \in \mathbb{R}^{m \times n} \) such that \( A = UDV^H \) where \( D = \left( \begin{smallmatrix} D_{TL} & 0 \\ 0 & 0 \end{smallmatrix} \right) \) with \( D_{TL} \) is diagonal and \( \delta_i > 0 \) for \( 0 \leq i < r \).

Proof: First, let us observe that if \( A = 0 \) (the zero matrix) then the theorem trivially holds: \( A = UDV^H \) where \( U = I_{m \times m}, V = I_{n \times n} \), and \( D = \left( \begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix} \right) \), so that \( D_{TL} = 0 \times 0 \). Thus, w.l.o.g. assume that \( A \neq 0 \).

We will prove this for \( m \geq n \), leaving the case where \( m \leq n \) as an exercise, employing a proof by induction on \( n \).

- **Base case:** \( n = 1 \). In this case \( A = \left( \begin{array}{c} a_0 \\ 0 \end{array} \right) \) where \( a_0 \in \mathbb{R}^m \) is its only column. By assumption, \( a_0 \neq 0 \). Then

\[
A = \left( \begin{array}{c} a_0 \\ 0 \end{array} \right) = \left( \begin{array}{c} u_0 \\ 0 \end{array} \right) (\|a_0\|_2) \left( \begin{array}{c} 1 \\ 0 \end{array} \right)^H
\]

where \( u_0 = a_0/\|a_0\|_2 \). Choose \( U_1 \in \mathbb{C}^{m \times (m-1)} \) so that \( U = \left( \begin{array}{c} u_0 \\ U_1 \end{array} \right) \) is unitary. Then

\[
A = \left( \begin{array}{c} a_0 \\ 0 \end{array} \right) (\|a_0\|_2) \left( \begin{array}{c} 1 \\ 0 \end{array} \right)^H = \left( \begin{array}{c} u_0 \\ U_1 \end{array} \right) \left( \begin{array}{c} \|a_0\|_2 \\ 0 \end{array} \right) \left( \begin{array}{c} 1 \\ 0 \end{array} \right)^H = UDV^H
\]

where \( D_{TL} = \left( \begin{smallmatrix} \delta_0 \\ \|a_0\|_2 \end{smallmatrix} \right) \) and \( V = \left( \begin{array}{c} 1 \end{array} \right) \).

- **Inductive step:** Assume the result is true for all matrices with \( 1 \leq k < n \) columns. Show that it is true for matrices with \( n \) columns.

Let \( A \in \mathbb{C}^{m \times n} \) with \( n \geq 2 \). W.l.o.g., \( A \neq 0 \) so that \( \|A\|_2 \neq 0 \). Let \( \delta_0 \) and \( v_0 \in \mathbb{C}^n \) have the property that \( \|v_0\|_2 = 1 \) and \( \delta_0 = \|Av_0\|_2 = \|A\|_2 \). (In other words, \( v_0 \) is the vector that maximizes \( \max_{\|x\|_2 = 1} \|Ax\|_2 \).) Let \( u_0 = Av_0/\delta_0 \). Note that \( \|u_0\|_2 = 1 \). Choose \( U_1 \in \mathbb{C}^{m \times (m-1)} \) and \( V_1 \in \mathbb{C}^{n \times (n-1)} \) so that \( \hat{U} = \left( \begin{array}{c} u_0 \\ U_1 \end{array} \right) \) and \( \hat{V} = \left( \begin{array}{c} v_0 \\ V_1 \end{array} \right) \) are unitary. Then

\[
\hat{U}^H A \hat{V} = \left( \begin{array}{c} u_0 \\ U_1 \end{array} \right)^H A \left( \begin{array}{c} v_0 \\ V_1 \end{array} \right)
\]

\[
= \left( \begin{array}{c} u_0^H Av_0 \\ U_1^H Av_0 \end{array} \right) \left( \begin{array}{c} u_0^H AV_1 \\ U_1^H AV_1 \end{array} \right) = \left( \frac{\delta_0 u_0^H u_0}{\delta_0 U_1^H u_0} \right) \left( \begin{array}{c} u_0^H AV_1 \\ U_1^H AV_1 \end{array} \right) = \left( \begin{array}{c} \delta_0 \\ 0 \end{array} \right) \left( \begin{array}{c} w^H \\ B \end{array} \right),
\]

where \( w = V_1^H A^H u_0 \) and \( B = U_1^H AV_1 \). Now, we will argue that \( w = 0 \), the zero vector of appropriate size:

\[
\delta_0^2 = \|A\|_2^2 = \|U^H AV\|_2^2 = \max_{x \neq 0} \frac{\|U^H AVx\|_2^2}{\|x\|_2^2} = \max_{x \neq 0} \left\| \left( \frac{\delta_0}{0 \ B} \right) x \right\|_2^2
\]
\[
\begin{align*}
\geq & \left\| \begin{pmatrix} \delta_0 & w^H \\ 0 & B \end{pmatrix} \begin{pmatrix} \delta_0 \\ w \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} \delta_0^2 + w^H w \\ Bw \end{pmatrix} \right\|_2^2 \\
\geq & \frac{(\delta_0^2 + w^H w)^2}{\delta_0^2 + w^H w} = \delta_0^2 + w^H w.
\end{align*}
\]

Thus \( \delta_0^2 \geq \delta_0^2 + w^H w \) which means that \( w = 0 \) and \( \tilde{U}^H A \tilde{V} = \begin{pmatrix} \delta_0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ B \end{pmatrix} \).

By the induction hypothesis, there exists unitary \( \tilde{U} \in \mathbb{C}^{(m-1)\times(m-1)} \), unitary \( \tilde{V} \in \mathbb{C}^{(n-1)\times(n-1)} \), and \( \tilde{D} \in \mathbb{R}^{(m-1)\times(n-1)} \) such that \( B = \tilde{U} \tilde{D} \tilde{V}^H \) where \( \tilde{D} = \begin{pmatrix} \bar{D}_{TL} & 0 \\ 0 & 0 \end{pmatrix} \) with \( \bar{D}_{TL} = \text{diag}(\delta_1, \cdots, \delta_{r-1}) \).

Now, let \( U = \tilde{U} \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U} \end{pmatrix} \), \( V = \tilde{V} \begin{pmatrix} 1 & 0 \\ 0 & \tilde{V} \end{pmatrix} \), and \( D = \begin{pmatrix} \delta_0 & 0 \\ 0 & 0 \end{pmatrix} \).

(There are some really tough to see "checks" in the definition of \( U \), \( V \), and \( D \)!!) Then \( A = UDV^H \) where \( U \), \( V \), and \( D \) have the desired properties.

\begin{itemize}
  \item **By the Principle of Mathematical Induction** the result holds for all matrices \( A \in \mathbb{C}^{m\times n} \) with \( m \geq n \).
\end{itemize}

\[ \square \]

**Exercise 13.** Let \( D = \text{diag}(\delta_0, \ldots, \delta_{n-1}) \). Show that \( \|D\|_2 = \max_{i=0}^{n-1} |\delta_i| \).

**Exercise 14.** Let \( A = \begin{pmatrix} A_T \\ 0 \end{pmatrix} \). Use the SVD of \( A \) to show that \( \|A\|_2 = \|A_T\|_2 \).

**Exercise 15.** Assume that \( U \in \mathbb{C}^{m\times m} \) and \( V \in \mathbb{C}^{n\times n} \) be unitary matrices. Let \( A, B \in \mathbb{C}^{m\times n} \) with \( B = UAV^H \).

Show that the singular values of \( A \) equal the singular values of \( B \).

**Exercise 16.** Let \( A \in \mathbb{C}^{m\times n} \) with \( A = \begin{pmatrix} \sigma_0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ B \end{pmatrix} \) and assume that \( \|A\|_2 = \sigma_0 \). Show that \( \|B\|_2 \leq \|A\|_2 \).

(Hint: Use the SVD of \( B \).)

**Exercise 17.** Prove Lemma 12 for \( m \leq n \).

You can use the following as an outline for your proof: **Proof:** First, let us observe that if \( A = 0 \) (the zero matrix) then the theorem trivially holds: \( A = UDV^H \) where \( U = I_{m\times m} \), \( V = I_{n\times n} \), and \( D = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \), so that \( \bar{D}_{TL} \) is \( 0 \times 0 \). Thus, w.l.o.g. assume that \( A \neq 0 \).

We will employ a proof by induction on \( m \).

\begin{itemize}
  \item **Base case:** \( m = 1 \). In this case \( A = \begin{pmatrix} \tilde{a}_0^T \end{pmatrix} \) where \( \tilde{a}_0^T \in \mathbb{R}^{1\times n} \) is its only row. By assumption, \( \tilde{a}_0^T \neq 0 \). Then

\[
A = \begin{pmatrix} \tilde{a}_0^T \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} \left( \|\tilde{a}_0^T\|_2 \right) \begin{pmatrix} v_0 \end{pmatrix}^H
\]

\end{itemize}
where \( v_0 = (\hat{a}_0^T)^H / \|\hat{a}_0^T\|_2 \). Choose \( V_1 \in \mathbb{C}^{n \times (n-1)} \) so that \( V = \left( \begin{array}{c} v_0 \\ V_1 \end{array} \right) \) is unitary. Then
\[
A = \left( \begin{array}{c} \hat{a}_0^T \\ 1 \end{array} \right) \left( \begin{array}{c} \|\hat{a}_0^T\|_2 \\ 0 \end{array} \right) (v_0 \mid V_1)^H = UDV^H
\]
where \( D_{TL} = \left( \begin{array}{c} \delta_0 \\ \|\hat{a}_0^T\|_2 \end{array} \right) \) and \( U = \left( \begin{array}{c} 1 \end{array} \right) \).

- **Inductive step:** Similarly modify the inductive step of the proof of the theorem.
- **By the Principle of Mathematical Induction** the result holds for all matrices \( A \in \mathbb{C}^{m \times n} \) with \( m \geq n \).

### 3 The Theorem

**Theorem 18** (Singular Value Decomposition). Given \( A \in \mathbb{C}^{m \times n} \) there exists unitary \( U \in \mathbb{C}^{m \times m} \), unitary \( V \in \mathbb{C}^{n \times n} \), and \( \Sigma \in \mathbb{R}^{m \times n} \) such that \( A = U \Sigma V^H \) where \( \Sigma = \left( \begin{array}{c} \Sigma_{TL} \\ 0 \end{array} \right) \) with \( \Sigma_{TL} = \text{diag}(\sigma_0, \ldots, \sigma_{r-1}) \) and \( \sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{r-1} > 0 \). The \( \sigma_0, \ldots, \sigma_{r-1} \) are known as the singular values of \( A \).

**Proof:** Notice that the proof of the above theorem is identical to that of Lemma 12. However, thanks to the above exercises, we can conclude that \( \|B\|_2 \leq \sigma_0 \) in the proof, which then can be used to show that the singular values are found in order. □

**Proof:** (Alternative) An alternative proof uses Lemma 12 to conclude that \( A = UDV^H \). If the entries on the diagonal of \( D \) are not ordered from largest to smallest, then this can be fixed by permuting the rows and columns of \( D \), and correspondingly permuting the columns of \( U \) and \( V \). □

### 4 Geometric Interpretation (Again)

We will now quickly illustrate what the SVD Theorem tells us about matrix-vector multiplication (linear transformations) by examining the case where \( A \in \mathbb{R}^{2 \times 2} \). Let \( A = U \Sigma V^T \) be its SVD decomposition. (Notice that all matrices are now real valued, and hence \( V^H = V^T \).) Partition
\[
A = \left( \begin{array}{c|c} u_0 & u_1 \\ \hline \sigma_0 & 0 \\ 0 & \sigma_1 \end{array} \right) \left( \begin{array}{c|c} v_0 & v_1 \end{array} \right)^T .
\]

Since \( U \) and \( V \) are unitary matrices, \( \{u_0, u_1\} \) and \( \{v_0, v_1\} \) form orthonormal bases for the range and domain of \( A \), respectively:
Let us manipulate the decomposition a little:

\[
A = \begin{pmatrix} u_0 & u_1 \end{pmatrix} \begin{pmatrix} \sigma_0 & 0 \\ 0 & \sigma_1 \end{pmatrix} \begin{pmatrix} v_0 & v_1 \end{pmatrix}^T
= \begin{pmatrix} \sigma_0 u_0 & \sigma_1 u_1 \end{pmatrix} \begin{pmatrix} v_0 & v_1 \end{pmatrix}^T.
\]

Now let us look at how \( A \) transforms \( v_0 \) and \( v_1 \):

\[
A v_0 = \begin{pmatrix} \sigma_0 u_0 & \sigma_1 u_1 \end{pmatrix} \begin{pmatrix} v_0 & v_1 \end{pmatrix}^T v_0 = \begin{pmatrix} \sigma_0 u_0 & \sigma_1 u_1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \sigma_0 u_0
\]

and similarly \( A v_1 = \sigma_1 u_1 \). This motivates the pictures
Now let us look at how $A$ transforms any vector with (Euclidean) unit length. Notice that $x = \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix}$ means that

$$x = \chi_0 e_0 + \chi_1 e_1,$$

where $e_0$ and $e_1$ are the unit basis vectors. Thus, $\chi_0$ and $\chi_1$ are the coefficients when $x$ is expressed using $e_0$ and $e_1$ as basis. However, we can also express $x$ in the basis given by $v_0$ and $v_1$:

$$x = \underbrace{V V^T}_{I} x = \begin{pmatrix} v_0 & v_1 \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \end{pmatrix}^T x = \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \begin{pmatrix} v_0^T x \\ v_1^T x \end{pmatrix} = \frac{v_0^T x}{\alpha_0} v_0 + \frac{v_1^T x}{\alpha_1} v_1 = \alpha_0 v_0 + \alpha_1 v_1 = \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}.$$

Thus, in the basis formed by $v_0$ and $v_1$, its coefficients are $\alpha_0$ and $\alpha_1$. Now,

$$Ax = \begin{pmatrix} \sigma_0 u_0 & \sigma_1 u_1 \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \end{pmatrix}^T x = \begin{pmatrix} \sigma_0 u_0 & \sigma_1 u_1 \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \end{pmatrix}^T \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \alpha_0 \sigma_0 u_0 + \alpha_1 \sigma_1 u_1.$$

This is illustrated by the following picture, which also captures the fact that the unit ball is mapped to an “ellipse” with major axis equal to $\sigma_0 = \|A\|_2$ and minor axis equal to $\sigma_1$:

Finally, we show the same insights for general vector $x$ (not necessarily of unit length).

---

1 It is not clear that it is actually an ellipse and this is not important to our observations.
Another observation is that if one picks the right basis for the domain and codomain, then the computation $Ax$ simplifies to a matrix multiplication with a diagonal matrix. Let us again illustrate this for nonsingular $A \in \mathbb{R}^{2 \times 2}$ with

$$A = \begin{pmatrix} u_0 & u_1 \\ \Sigma & \end{pmatrix} \begin{pmatrix} \sigma_0 & 0 \\ 0 & \sigma_1 \end{pmatrix} \begin{pmatrix} v_0 & v_1 \\ \end{pmatrix}^T.$$

Now, if we chose to express $y$ using $u_0$ and $u_1$ as the basis and express $x$ using $v_0$ and $v_1$ as the basis, then

$$\hat{y} = \begin{pmatrix} \sigma_0 \hat{\psi}_0 \\ \sigma_1 \hat{\psi}_1 \end{pmatrix}, \quad \hat{x} = \begin{pmatrix} \hat{\chi}_0 \\ \hat{\chi}_1 \end{pmatrix}.$$

If $y = Ax$ then

$$U U^T y = U \Sigma V^T x = U \Sigma \hat{x},$$

so that $\hat{y} = \Sigma \hat{x}$ and

$$\begin{pmatrix} \hat{\psi}_0 \\ \psi_1 \end{pmatrix} = \begin{pmatrix} \sigma_0 \hat{\chi}_0 \\ \sigma_1 \hat{\chi}_1 \end{pmatrix}.$$

These observation generalize to $A \in \mathbb{C}^{m \times m}$.

### 5 Consequences of the SVD Theorem

Throughout this section we will assume that
Corollary 19. \( A = U_L \Sigma TL V_L^H \). This is called the reduced SVD of \( A \).

**Proof:**
\[
A = U \Sigma V^H = \left( \begin{array}{c|c} U_L & U_R \end{array} \right) \left( \begin{array}{c|c} \Sigma TL & 0 \\ \hline 0 & 0 \end{array} \right) \left( \begin{array}{c|c} V_L & V_R \end{array} \right)^H = U_L \Sigma TL V_L^H.
\]

\[\blacksquare\]

Corollary 20. Let \( A = U_L \Sigma TL V_L^H \) be the reduced SVD with \( U_L = \left( \begin{array}{c|c} u_0 & \cdots & u_{r-1} \end{array} \right) \), \( \Sigma TL = \text{diag}(\sigma_0, \ldots, \sigma_{r-1}) \), and \( V_L = \left( \begin{array}{c|c} v_0 & \cdots & v_{r-1} \end{array} \right) \). Then \( A = \sigma_0 u_0 v_0^H + \sigma_1 u_1 v_1^H + \cdots + \sigma_{r-1} u_{r-1} v_{r-1}^H \) (each term nonzero and \( \text{an outer product, and hence a rank-1 matrix} \)).

**Proof:** We leave the proof as an exercise.

\[\blacksquare\]

Corollary 21. \( \mathcal{C}(A) = \mathcal{C}(U_L) \).

**Proof:**
Let \( y \in C(A) \). Then there exists \( x \in \mathbb{C}^n \) such that \( y = Ax \) (by the definition of \( y \in C(A) \)). But then

\[
y = Ax = U_L \frac{\Sigma TL V_L^H}{z} x = U_L z,
\]

i.e., there exists \( z \in \mathbb{C}^r \) such that \( y = U_L z \). This means \( y \in C(U_L) \).

Let \( y \in C(U_L) \). Then there exists \( z \in \mathbb{C}^r \) such that \( y = U_L z \). But then

\[
y = U_L z = U_L \frac{\Sigma TL \Sigma_{TL}^{-1}}{I} z = U_L \Sigma TL \frac{V_L^H V_L}{I} \Sigma_{TL}^{-1} z = A \frac{V_L \Sigma_{TL}^{-1}}{x} z = Ax
\]

so that there exists \( x \in \mathbb{C}^n \) such that \( y = Ax \), i.e., \( y \in C(A) \).

---

**Corollary 22.** The rank of \( A \) is \( r \).

**Proof:** The rank of \( A \) equals the dimension of \( C(A) = C(U_L) \). But the dimension of \( C(U_L) \) is clearly \( r \). \( \square \)

**Corollary 23.** \( N(A) = C(V_R) \).

**Proof:**

- Let \( x \in N(A) \). Then

\[
x = VV^H \frac{x}{I}(V_L \mid V_R) = (V_L \mid V_R) \left( \begin{array}{c} V_L^H x \\ V_R^H x \end{array} \right) = V_L V_L^H x + V_R V_R^H x.
\]

If we can show that \( V_R^H x = 0 \) then \( x = V_R z \) where \( z = V_R^H x \). Assume that \( V_R^H x \neq 0 \). Then \( \Sigma_{TL}(V_R^H x) \neq 0 \) (since \( \Sigma_{TL} \) is nonsingular) and \( U_L(\Sigma_{TL}(V_R^H x)) \neq 0 \) (since \( U_L \) has linearly independent columns). But that contradicts the fact that \( Ax = U_L \Sigma_{TL} V_L^H x = 0 \).

- Let \( x \in C(V_R) \). Then \( x = V_R z \) for some \( z \in \mathbb{C}^r \) and \( Ax = U_L \Sigma_{TL} \frac{V_L^H V_R}{z} z = 0 \).

\( \square \)

**Corollary 24.** For all \( x \in \mathbb{C}^n \) there exists \( z \in C(V_L) \) such that \( Ax = A z \).

**Proof:**

\[
Ax = A \frac{VV^H}{I} x = A (V_L \mid V_R) \left( \begin{array}{c} V_L^H x \\ V_R^H x \end{array} \right) x
\]

\[
= A (V_L V_L^H x + V_R V_R^H x) = AV_L V_L^H x + AV_R V_R^H x
\]

\[
= AV_L V_L^H x + U_L \Sigma_{TL} \frac{V_L^H V_R}{z} V_R^H x = A \frac{V_L V_L^H}{z} x.
\]
\( \mathbb{C}^n \)

\[ \dim r \]

\( \mathcal{C}(A^H) = \mathcal{C}(\mathcal{V}_L) \)

\[ x = z + x_n \]

\( y = Az \)

\[ y = Ax = U_L \Sigma_T L V_L^H x \]

\( \mathcal{N}(A) = \mathcal{C}(\mathcal{V}_R) \)

\[ \dim n - r \]

\( \mathcal{N}(A^H) = \mathcal{C}(\mathcal{U}_R) \)

\[ \dim m - r \]

\[ \mathbb{C}^m \]

\( \mathcal{C}(A) = \mathcal{C}(\mathcal{U}_L) \)

Figure 1: A pictorial description of how \( x = z + x_n \) is transformed by \( A \in \mathbb{C}^{m \times n} \) into \( y = Ax = A(z + x_n) \). We see that \( \mathcal{C}(\mathcal{V}_L) \) and \( \mathcal{C}(\mathcal{V}_R) \) are orthogonal complements of each other within \( \mathbb{C}^n \). Similarly, \( \mathcal{C}(\mathcal{U}_L) \) and \( \mathcal{C}(\mathcal{U}_R) \) are orthogonal complements of each other within \( \mathbb{C}^m \). Any vector \( x \) can be written as the sum of a vector \( z \in \mathcal{C}(\mathcal{V}_L) \) and \( x_n \in \mathcal{N}(A) = \mathcal{C}(\mathcal{V}_R) \).

Alternative proof (which uses the last corollary):

\[
A x = A \left( V_L V_L^H x + V_R V_R^H x \right) = A V_L V_L^H x + A \left( V_R V_R^H x \right) = A V_L V_L^H x \in \mathcal{N}(A).
\]

The proof of the last corollary also shows that

**Corollary 25.** Any vector \( x \in \mathbb{C}^n \) can be written as \( x = z + x_n \) where \( z \in \mathcal{C}(\mathcal{V}_L) \) and \( x_n \in \mathcal{N}(A) = \mathcal{C}(\mathcal{V}_R) \).

**Corollary 26.** \( A^H = V_L \Sigma_T L U_L^H \) so that \( \mathcal{C}(A^H) = \mathcal{C}(\mathcal{V}_L) \) and \( \mathcal{N}(A^H) = \mathcal{C}(\mathcal{U}_R) \).

The above corollaries are summarized in Figure 1.

**Theorem 27.** Let \( A \in \mathbb{C}^{n \times n} \) be nonsingular. Let \( A = U \Sigma V^H \) be its SVD. Then

1. The SVD is the reduced SVD.
2. \( \sigma_{n-1} \neq 0 \).
3. If

\[
U = \left( \begin{array}{c} u_0 \\ \vdots \\ u_{n-1} \end{array} \right), \Sigma = \text{diag}(\sigma_0, \ldots, \sigma_{n-1}), \text{ and } V = \left( \begin{array}{c} v_0 \\ \vdots \\ v_{n-1} \end{array} \right),
\]

then

\[
A^{-1} = (V P T)(P \Sigma^{-1} P^T)(U P T)^H = \left( \begin{array}{c} v_{n-1} \\ \vdots \\ v_0 \end{array} \right) \text{diag} \left( \frac{1}{\sigma_{n-1}}, \ldots, \frac{1}{\sigma_0} \right) \left( \begin{array}{c} u_{n-1} \\ \vdots \\ u_0 \end{array} \right),
\]
where \( P = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{pmatrix} \) is the permutation matrix such that \( Px \) reverses the order of the entries in \( x \). (Note: for this permutation matrix, \( P^T = P \). In general, this is not the case. What is the case for all permutation matrices \( P \) is that \( P^T P = PP^T = I \).

4. \( \|A^{-1}\|_2 = 1/\sigma_{n-1} \).

**Proof:** The only item that is less than totally obvious is (3). Clearly \( A^{-1} = V \Sigma^{-1} U^H \). The problem is that in \( \Sigma^{-1} \) the diagonal entries are not ordered from largest to smallest. The permutation fixes this. \( \square \)

Corollary 28. If \( A \in \mathbb{C}^{m \times n} \) has linearly independent columns then \( A^H A \) is invertible (nonsingular) and \( (A^H A)^{-1} = V_L (\Sigma_{TL}^2)^{-1} V_L^H \).

**Proof:** Since \( A \) has linearly independent columns, \( A = U_L \Sigma_{TL} V_L^H \) is the reduced SVD where \( U_L \) has \( n \) columns and \( V_L \) is unitary. Hence

\[
A^H A = (U_L \Sigma_{TL} V_L^H)^H U_L \Sigma_{TL} V_L^H = V_L \Sigma_{TL}^H U_L^H U_L \Sigma_{TL} V_L^H = V_L \Sigma_{TL} \Sigma_{TL} V_L^H = V_L \Sigma_{TL}^2 V_L^H.
\]

Since \( V_L \) is unitary and \( \Sigma_{TL} \) is diagonal with nonzero diagonal entries, they are both nonsingular. Thus

\[
(V_L \Sigma_{TL}^2 V_L^H) \left( V_L \left( \Sigma_{TL}^2 \right)^{-1} V_L^H \right) = I.
\]

This means \( A^T A \) is invertible and \( (A^T A)^{-1} \) is as given. \( \square \)

### 6 Projection onto the Column Space

**Definition 29.** Let \( U_L \in \mathbb{C}^{m \times k} \) have orthonormal columns. The projection of a vector \( y \in \mathbb{C}^m \) onto \( \mathcal{C}(U_L) \) is the vector \( U_L x \) that minimizes \( \|y - U_L x\|_2 \), where \( x \in \mathbb{C}^k \). We will also call this vector \( y \) the component of \( x \) in \( \mathcal{C}(U_L) \).

**Theorem 30.** Let \( U_L \in \mathbb{C}^{m \times k} \) have orthonormal columns. The projection of \( y \) onto \( \mathcal{C}(U_L) \) is given by \( U_L U_L^H y \).

**Proof:** The vector \( U_L x \) that we want must satisfy

\[
\|U_L x - y\|_2 = \min_{w \in \mathbb{C}^k} \|U_L w - y\|_2.
\]

Now, the 2-norm is invariant under multiplication by the unitary matrix \( U^H = \begin{pmatrix} U_L & U_R \end{pmatrix}^H \)

\[
\|U_L x - y\|_2^2 = \min_{w \in \mathbb{C}^k} \|U_L w - y\|_2^2
\]

\[
= \min_{w \in \mathbb{C}^k} \|U^H (U_L w - y)\|_2^2 \quad \text{(since the two norm is preserved)}
\]

\[
= \min_{w \in \mathbb{C}^k} \left\| \begin{pmatrix} U_L & U_R \end{pmatrix}^H (U_L w - y) \right\|_2^2
\]

\[
= \min_{w \in \mathbb{C}^k} \left\| \begin{pmatrix} U_L^H \\ U_R^H \end{pmatrix} (U_L w - y) \right\|_2^2
\]

12
Theorem 34. Let \( A \in \mathbb{C}^{m \times n} \) have SVD \( A = U \Sigma V^H \) and assume \( A \) has rank \( r \). Partition 
\[
U = \begin{pmatrix} U_L \\ U_R \end{pmatrix}, \quad V = \begin{pmatrix} V_L \\ V_R \end{pmatrix}, \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{TL} & 0 \\ 0 & \Sigma_{BR} \end{pmatrix},
\]

\[
= \min_{w \in \mathbb{C}^k} \left\| \begin{pmatrix} U^H_L \\ U^H_R \end{pmatrix} U_L w - \begin{pmatrix} U^H_L \\ U^H_R \end{pmatrix} y \right\|_2^2
\]
\[
= \min_{w \in \mathbb{C}^k} \left\| \begin{pmatrix} U^H_L U_L w \end{pmatrix} - \begin{pmatrix} U^H_L y \\ U^H_R y \end{pmatrix} \right\|_2^2
\]
\[
= \min_{w \in \mathbb{C}^k} \left\| \begin{pmatrix} w \\ 0 \end{pmatrix} - \begin{pmatrix} U^H_L y \\ U^H_R y \end{pmatrix} \right\|_2^2
\]
\[
= \min_{w \in \mathbb{C}^k} \left\| \begin{pmatrix} w - U^H_L y \\ -U^H_R y \end{pmatrix} \right\|_2^2
\]
\[
= \min_{w \in \mathbb{C}^k} \left( \| w - U^H_L y \|_2^2 + \| -U^H_R y \|_2^2 \right) \quad \text{(since \( \left\| \begin{pmatrix} u \\ v \end{pmatrix} \right\|_2^2 = \| u \|_2^2 + \| v \|_2^2 \})
\]
\[
= \left( \min_{w \in \mathbb{C}^k} \| w - U^H_L y \|_2^2 \right) + \| U^H_R y \|_2^2.
\]

This is minimized when \( w = U^H_L y \). Thus, the vector that is closest to \( y \) in the space spanned by \( U_L \) is given by \( x = U_L U^H_L y \).

Corollary 31. Let \( A \in \mathbb{C}^{m \times n} \) and \( \bar{A} = U_L \Sigma_{TL} V^H_L \) be its reduced SVD. Then the projection of \( y \in \mathbb{C}^m \) onto \( \mathcal{C}(A) \) is given by \( U_L U^H_L y \).

Proof: This follows immediately from the fact that \( \mathcal{C}(A) = \mathcal{C}(U_L) \).

Corollary 32. Let \( A \in \mathbb{C}^{m \times n} \) have linearly independent columns. Then the projection of \( y \in \mathbb{C}^m \) onto \( \mathcal{C}(A) \) is given by \( A(A^H A)^{-1} A^H y \).

Proof: From Corollary 28 we know that \( A^H A \) is nonsingular and that \( (A^H A)^{-1} = V_L (\Sigma_{TL}^2)^{-1} V_L^H \). Now,
\[
A(A^H A)^{-1} A^H y = (U_L \Sigma_{TL} V_L^H)(V_L (\Sigma_{TL}^2)^{-1} V_L^H)(U_L \Sigma_{TL} V_L^H)^H y
\]
\[
= U_L \Sigma_{TL} V_L^H V_L \Sigma_{TL}^{-1} \Sigma_{TL}^{-1} V_L^H V_L \Sigma_{TL} U_L^H y = U_L U^H_L y.
\]

Hence the projection of \( y \) onto \( \mathcal{C}(A) \) is given by \( A(A^H A)^{-1} A^H y \).

Definition 33. Let \( A \) have linearly independent columns. Then \( (A^H A)^{-1} A^H \) is called the pseudo-inverse or Moore-Penrose generalized inverse of matrix \( A \).

7 Low-rank Approximation of a Matrix

Theorem 34. Let \( A \in \mathbb{C}^{m \times n} \) have SVD \( A = U \Sigma V^H \) and assume \( A \) has rank \( r \). Partition 
\[
U = \begin{pmatrix} U_L \\ U_R \end{pmatrix}, \quad V = \begin{pmatrix} V_L \\ V_R \end{pmatrix}, \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{TL} & 0 \\ 0 & \Sigma_{BR} \end{pmatrix},
\]
where \( U_L \in \mathbb{C}^{m \times k}, V_L \in \mathbb{C}^{n \times k}, \) and \( \Sigma_{TL} \in \mathbb{R}^{k \times k} \) with \( k \leq r \). Then \( B = U_L \Sigma_{TL} V_L^H \) is the matrix in \( \mathbb{C}^{m \times n} \) closest to \( A \) in the following sense:

\[
\| A - B \|_2 = \min_{C \in \mathbb{C}^{m \times n}} \| A - C \|_2 = \sigma_k.
\]

\[
\text{rank}(C) \leq k
\]

**Proof:** First, if \( B \) is as defined, then clearly \( \| A - B \|_2 = \sigma_k \):

\[
\| A - B \|_2 = \| U^H (A - B) V \|_2 = \| U^H AV - U^H BV \|_2
\]

\[
= \left\| \Sigma - \begin{pmatrix} U_L & U_R \end{pmatrix}^H B \begin{pmatrix} V_L & V_R \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} \Sigma_{TL} & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} \Sigma_{TL} & 0 \\ 0 & 0 \end{pmatrix} \right\|_2
\]

\[
= \| \Sigma_{BR} \|_2 = \sigma_k
\]

Next, assume that \( C \) has rank \( t \leq k \) and \( \| A - C \|_2 < \| A - B \|_2 \). We will show that this leads to a contradiction.

- The null space of \( C \) has dimension at least \( n - k \) since \( \dim(N(C)) + \text{rank}(C) = n \).

- If \( x \in N(C) \) then
  \[
  \| Ax \|_2 = \| (A - C)x \|_2 \leq \| A - C \|_2 \| x \|_2 < \sigma_k \| x \|_2.
  \]

- Partition \( U = \begin{pmatrix} u_0 & \cdots & u_{m-1} \end{pmatrix} \) and \( V = \begin{pmatrix} v_0 & \cdots & v_{n-1} \end{pmatrix} \). Then \( \| Av_j \|_2 = \| \sigma_j u_j \|_2 = \sigma_j \geq \sigma_s \) for \( j = 0, \ldots, k \). Now, let \( x \) be any linear combination of \( v_0, \ldots, v_k \): \( x = \alpha_0 v_0 + \cdots + \alpha_k v_k \). Notice that
  \[
  \| x \|_2^2 = \| \alpha_0 v_0 + \cdots + \alpha_k v_k \|_2^2 \leq |\alpha_0|^2 + \cdots + |\alpha_k|^2.
  \]

Then

\[
\| Ax \|_2^2 = \| A(\alpha_0 v_0 + \cdots + \alpha_k v_k) \|_2^2 = \| \alpha_0 A v_0 + \cdots + \alpha_k A v_k \|_2^2
\]

\[
= \| \alpha_0 \sigma_0 u_0 + \cdots + \alpha_k \sigma_k u_k \|_2^2 = \| \alpha_0 \sigma_0 u_0 \|_2^2 + \cdots + \| \alpha_k \sigma_k u_k \|_2^2
\]

\[
= |\alpha_0|^2 \sigma_0^2 + \cdots + |\alpha_k|^2 \sigma_k^2 \geq (|\alpha_0|^2 + \cdots + |\alpha_k|^2) \sigma_k^2
\]

so that \( \| Ax \|_2 \geq \sigma_k \| x \|_2 \). In other words, vectors in the subspace of all linear combinations of \( \{ v_0, \ldots, v_k \} \) satisfy \( \| Ax \|_2 \geq \sigma_k \| x \|_2 \). The dimension of this subspace is \( k + 1 \) (since \( \{ v_0, \cdots, v_k \} \) form an orthonormal basis).

- Both these subspaces are subspaces of \( \mathbb{C}^n \). Since their dimensions add up to more than \( n \) there must be at least one nonzero vector \( z \) that satisfies both \( \| Az \|_2 < \sigma_k \| z \|_2 \) and \( \| Az \|_2 \geq \sigma_k \| z \|_2 \), which is a contradiction.

The above theorem tells us how to pick the best approximation with given rank to a given matrix.

8 An Application

Let \( Y \in \mathbb{R}^{m \times n} \) be a matrix that, for example, stores a picture. In this case, the \((i, j)\) entry in \( Y \) is, for example, a number that represents the grayscale value of pixel \((i, j)\). The following instructions, executed in octave or matlab, generate the picture of Mexican artist Frida Kahlo in Figure 2 (top-left). The file FridaPNG.png can be found at http://www.cs.utexas.edu/users/flame/Notes/FridaPNG.png.
Figure 2: Multiple pictures as generated by the code
octave> IMG = imread( 'FridaPNG.png' ); % this reads the image
octave> Y = IMG(:,1:1);
octave> imshow( Y ) % this dispays the image

Although the picture is black and white, it was read as if it is a color image, which means a \( m \times n \times 3 \) array of pixel information is stored. Setting \( Y = \text{IMG}(::,1) \) extracts a single matrix of pixel information. (If you start with a color picture, you will want to approximate \( \text{IMG}(::,1), \text{IMG}(::,2), \) and \( \text{IMG}(::,3) \) separately.)

Now, let \( Y = U \Sigma V^T \) be the SVD of matrix \( Y \). Partition, conformally,

\[
U = \begin{pmatrix} U_L & U_R \end{pmatrix}, \quad V = \begin{pmatrix} V_L & V_R \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{TL} & 0 \\ 0 & \Sigma_{BR} \end{pmatrix},
\]

where \( U_L \) and \( V_L \) have \( k \) columns and \( \Sigma_{TL} \) is \( k \times k \). so that

\[
Y = \begin{pmatrix} U_L & U_R \end{pmatrix} \begin{pmatrix} \Sigma_{TL} & 0 \\ 0 & \Sigma_{BR} \end{pmatrix} \begin{pmatrix} V_L & V_R \end{pmatrix}^T = U_L \Sigma_{TL} V_L^T + U_R \Sigma_{BR} V_R^T.
\]

Recall that then \( U_L \Sigma_{TL} V_L^T \) is the best rank-\( k \) approximation to \( Y \).

Let us approximate the matrix that stores the picture with \( U_L \Sigma_{TL} V_L^T \):

\[
\begin{align*}
\gg & \text{IMG = imread( 'FridaPNG.png' )}; \% \text{read the picture} \\
\gg & Y = \text{IMG}(::,1); \\
\gg & \text{imshow( Y )}; \% \text{this dispays the image} \\
\gg & k = 1; \\
\gg & [ \text{U, Sigma, V} ] = \text{svd( Y )}; \\
\gg & \text{UL} = \text{U}(::,1:k); \% \text{first k columns} \\
\gg & \text{VL} = \text{V}(::,1:k); \% \text{first k columns} \\
\gg & \text{SigmaTL} = \text{Sigma}(1:k,1:k); \% \text{TL submatrix of Sigma} \\
\gg & \text{Yapprox} = \text{uint8} ( \text{UL} * \text{SigmaTL} * \text{VL}' ); \\
\gg & \text{imshow( Yapprox )};
\end{align*}
\]

As one increases \( k \), the approximation gets better, as illustrated in Figure 2. The graph in Figure 3 helps explain. The original matrix \( Y \) is \( 387 \times 469 \), with 181,503 entries. When \( k = 10 \), matrices \( U, V, \) and \( \Sigma \) are \( 387 \times 10, 469 \times 10 \) and \( 10 \times 10 \), respectively, requiring only 8,660 entries to be stores.

9 SVD and the Condition Number of a Matrix

In “Notes on Norms” we saw that if \( Ax = b \) and \( A(x + \delta x) = b + \delta b \), then

\[
\frac{\|\delta x\|_2}{\|x\|_2} \leq \kappa_2(A) \frac{\|\delta b\|_2}{\|b\|_2},
\]

where \( \kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 \) is the condition number of \( A \), using the 2-norm.

**Exercise 35.** Show that if \( A \in \mathbb{C}^{m \times m} \) is nonsingular, then
• $\|A\|_2 = \sigma_0$, the largest singular value;
• $\|A^{-1}\|_2 = 1/\sigma_{m-1}$, the inverse of the smallest singular value; and
• $\kappa_2(A) = \sigma_0/\sigma_{m-1}$.

If we go back to the example of $A \in \mathbb{R}^{2 \times 2}$, recall the following pictures that shows how $A$ transforms the unit circle:
In this case, the ratio $\sigma_0/\sigma_{n-1}$ represents the ratio between the major and minor axes of the “ellipse” on the right.

10 An Algorithm for Computing the SVD?

It would seem that the proof of the existence of the SVD is constructive in the sense that it provides an algorithm for computing the SVD of a given matrix $A \in \mathbb{C}^{m \times m}$. Not so fast! Observe that

- Computing $\|A\|_2$ is nontrivial.
- Computing the vector that maximizes $\max_{\|x\|_2 = 1} \|Ax\|_2$ is nontrivial.
- Given a vector $q_0$ computing vectors $q_0, \ldots, q_{m-1}$ is expensive (as we will see when we discuss the QR factorization).

Towards the end of the course we will discuss algorithms for computing the eigenvalues and eigenvectors of a matrix, and related algorithms for computing the SVD.