Outline

1 Recap
   • Vector and Matrix Norm Inequalities
   • Missing Proof

2 Existence and Uniqueness of the SVD
   • The Theorem
   • Proof

3 The SVD
   • “Every Matrix is Diagonal”
   • Singular Values and Eigenvalues
   • The SVD $\rightsquigarrow$ Matrix Properties
Last Time

- Hölder and Cauchy-[Bunyakovsky]-Schwarz inequalities:
  \[ \left| \bar{v}^* \bar{w} \right|_H \leq \left\| \bar{v} \right\|_p \left\| \bar{w} \right\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1, \quad \left| \bar{v}^* \bar{w} \right|_{CS} \leq \left\| \bar{v}_2 \right\| \left\| \bar{w} \right\|_2 \]

- Bounds on the norms of matrix products
  \[ \left\| AB \right\| \leq \left\| A \right\| \left\| B \right\| \]

- General matrix norms: The Frobenius norm \( \left\| A \right\|_F^2 = \sum_{ij} |a_{ij}|^2 \).
- A geometrical introduction to the SVD.
- The reduced vs. the full SVD.
We ended last lecture with: \textbf{If we can show that every matrix }A\textbf{ has a SVD, then it follows that the image of the unit sphere under any linear map is a hyper-ellipse...}"

We now turn our attention to showing that this indeed is the case...

\textbf{Figure:} The unit-sphere $S^2$, and the image $AS^2$, where $A = \begin{bmatrix} 1.3127 & 0.6831 & 0.6124 \\ 0.0129 & 1.0928 & 0.6085 \\ 0.3840 & 0.0353 & 1.0158 \end{bmatrix}$. 

Peter Blomgren, $\langle$blomgren.peter@gmail.com$\rangle$
Theorem (Existence and Uniqueness of the SVD)

Every matrix $A \in \mathbb{C}^{m \times n}$ has a singular value decomposition $A = U \Sigma V^*$, where

- $U \in \mathbb{C}^{m \times m}$ is unitary
- $V \in \mathbb{C}^{n \times n}$ is unitary
- $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal, non-negative.

Furthermore, the singular values $\{\sigma_j\}$ are uniquely determined, and if $A$ is square and the $\sigma_j$ are distinct, the left $\{\tilde{u}_j\}$ and right $\{\tilde{v}_j\}$ singular vectors are uniquely determined up to complex scalar factors $s \in \mathbb{C} : |s| = 1$. 
The proof is by induction. Let $\sigma_1 = \|A\|_2$. There must exist $\bar{u}_1 \in \mathbb{C}^m$, $\|\bar{u}_1\|_2 = 1$, and $\bar{v}_1 \in \mathbb{C}^n$, $\|\bar{v}_1\|_2 = 1$, such that $A\bar{v}_1 = \sigma_1 \bar{u}_1$:

$$\sigma_1 = \frac{\|A\bar{x}^*\|_2}{\|\bar{x}^*\|_2}, \text{ for some } \bar{x}^*. \text{ Let } \bar{v}_1 = \frac{\bar{x}^*}{\|\bar{x}^*\|_2}.$$ 

Clearly, $A\bar{v}_1 = \bar{p}$, for some $\bar{p}$. Let $\bar{u}_1 = \frac{\bar{p}}{\|\bar{p}\|_2}$, and $\sigma_1 = \|\bar{p}\|_2$.

Now, consider any extensions (exists Movie) of $\bar{v}_1$ to an orthonormal basis $\{\bar{v}_j\}$ of $\mathbb{C}^n$ and of $\bar{u}_1$ to an orthonormal basis $\{\bar{u}_j\}$ of $\mathbb{C}^m$. Let $U_1$ and $V_1$ denote the matrices with columns $u_j$ and $v_j$, respectively.
We have (by construction)

\[ U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & \bar{w}^* \\ \bar{0} & B \end{bmatrix}, \]

where \( \bar{0} \) is a column-vector of size \( m - 1 \), and \( \bar{w}^* \) is a row vector of size \( n - 1 \), and the matrix \( B \in \mathbb{C}^{(m-1)\times(n-1)} \).

Now,

\[
\left\| \begin{bmatrix} \sigma_1 & \bar{w}^* \\ \bar{0} & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \bar{w} \end{bmatrix} \right\|_2 \geq \sigma_1^2 + \bar{w}^* \bar{w} = \sqrt{\sigma_1^2 + \bar{w}^* \bar{w}} \left\| \begin{bmatrix} \sigma_1 \\ \bar{w} \end{bmatrix} \right\|_2,
\]

Hence,

\[
\|S\|_2 \geq \sqrt{\sigma_1^2 + \bar{w}^* \bar{w}}.
\]
Theorem: $A = U\Sigma V^*$

We have $\|S\|_2 \geq \sqrt{\sigma_1^2 + \bar{w}^*\bar{w}}$, and $S = U_1^*AV_1$. Since $U_1$ and $V_1$ are unitary, we must have $\|S\|_2 = \|A\|_2 = \sigma_1$.

Therefore $\|\bar{w}\|_2^2 = \bar{w}^*\bar{w} = 0$, which means $\bar{w} = 0$, hence

$$U_1^*AV_1 = S = \begin{bmatrix} \sigma_1 & \bar{0}^* \\ 0 & B \end{bmatrix}, \quad \iff \quad A = U_1 \begin{bmatrix} \sigma_1 & \bar{0}^* \\ 0 & B \end{bmatrix} V_1^*$$

If $m = 1$, or $n = 1$, we are done. Otherwise, the sub-matrix $B$ describes the action of $A$ on the subspace orthogonal to $\bar{v}_1$.

We can now recursively (inductively) apply the same process to $B$, and establish existence of the SVD of $A$:

$$A = U_1 \begin{bmatrix} 1 & \bar{0}^* \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \bar{0}^* \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \bar{0}^* \\ 0 & V_2 \end{bmatrix}^* V_1^* = U\Sigma V^*.$$
**The uniqueness proof remains —**

[Geometric version] If the singular values $\sigma_j$ are distinct, then the lengths of the semi-axes of the hyper-ellipse $A_S^{(n-1)}$ must be distinct.

The semi-axes themselves are determined by the geometry, up to a complex sign. $\square_{\text{geometric}}$.

[Algebraic version] $\sigma_1 = \|A\|_2$ is uniquely determined. Now, suppose that in addition to $\tilde{v}_1$, there is another linearly independent vector $\tilde{w}_1$ with $\|\tilde{w}_1\| = 1$, and $\|A\tilde{w}_1\| = \sigma_1$.

We define a unit vector $\tilde{v}_2$, orthogonal to $\tilde{v}_1$, as a linear combination of $\tilde{v}_1$ and $\tilde{w}_1$:

$$\tilde{v}_2 = \frac{\tilde{w}_1 - (\tilde{v}_1^*\tilde{w}_1)\tilde{v}_1}{\|\tilde{w}_1 - (\tilde{v}_1^*\tilde{w}_1)\tilde{v}_1\|_2}.$$
Theorem: \( A = U \Sigma V^* \)

Since \( \|A\|_2 = \sigma_1 \), \( \|A\bar{v}_2\|_2 \leq \sigma_1 \); but this must be an equality, otherwise since

\[
\bar{w}_1 = \cos(\theta)\bar{v}_1 + \sin(\theta)\bar{v}_2, \quad \bar{v}_1 \perp \bar{v}_2, \quad |\cos(\theta)|^2 + |\sin(\theta)|^2 = 1
\]

we would have \( \|A\bar{w}_1\|_2 < \sigma_1 \).

This vector \( \bar{v}_2 \) is a second right singular vector corresponding to the singular value \( \sigma_1 \); it will lead to the appearance of a \( \bar{y} \) (the last \( n - 1 \) elements of \( V_1^*\bar{v}_2 \)) with \( \|\bar{y}\|_2 = 1 \), and \( \|B\bar{y}\|_2 = \sigma_1 \).

Hence, if the singular vector \( \bar{v}_1 \) is not unique, then the corresponding singular value \( \sigma_1 \) is not simple (\( \sigma_1 \not> \sigma_2 \)). Therefore there cannot exist a vector \( \bar{w}_1 \) as above.

Now, the uniqueness of the remaining singular vectors follows by induction. \( \square_{\text{algebraic}} \)
The SVD: \( A = U\Sigma V^* \)

**Bold Statement**

SVD enables us to say that **every matrix is diagonal** — as long as we use the proper bases for the domain \( \in \mathbb{C}^n \), and range \( \in \mathbb{C}^m \) spaces.

**Changing Bases — Rotating the Map!**

Any \( \vec{b} \in \mathbb{C}^m \) can be expanded in the basis of the left singular vectors of \( A \) (i.e. the columns of \( U \)), and any \( \vec{x} \in \mathbb{C}^n \) in the basis of the right singular vectors of \( A \) (i.e. the columns of \( V \))...

The coordinates for these expansions are

\[
\vec{b}' = U^*\vec{b}, \quad \vec{x}' = V^*\vec{x}.
\]

Now, the relation \( \vec{b} = A\vec{x} \) can be written in terms of \( \vec{b}' \) and \( \vec{x}' \):

\[
\vec{b} = A\vec{x} \Leftrightarrow \quad U^*\vec{b} = U^*A\vec{x} = U^* \underbrace{U\Sigma V^*}_{A} \vec{x} \Leftrightarrow \quad \vec{b}' = \Sigma \vec{x}'
\]
The idea of **diagonalizing** a matrix by a change of basis is the foundation for the study of eigenvalues.

A non-defective square matrix $A$ can be expressed as a diagonal matrix of eigenvalues $\Lambda$, if the range and domain are expressed in a basis of the eigenvectors. The **eigenvalue decomposition** of $A \in \mathbb{C}^{m \times m}$ is

$$A = X\Lambda X^{-1}$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$, and the columns of $X \in \mathbb{C}^{m \times m}$ contain linearly independent eigenvectors of $A$.

We can change basis for the expression $\mathbf{b} = A\mathbf{x}$:

$$\mathbf{b}' = X^{-1}\mathbf{b}, \quad \mathbf{x}' = X^{-1}\mathbf{x}.$$
## Fundamental differences between the SVD and Eigenvalue Decomposition

<table>
<thead>
<tr>
<th>The SVD</th>
<th>Eigenvalue Decomposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Properties</strong></td>
<td></td>
</tr>
<tr>
<td>Uses two different bases — the set of right and left singular vectors.</td>
<td>Uses one basis — the eigenvectors.</td>
</tr>
<tr>
<td>Uses orthonormal bases</td>
<td>Uses a basis which is generally not orthogonal.</td>
</tr>
<tr>
<td>All matrices (even rectangular ones) have a singular value decomposition.</td>
<td>Not all matrices (even square ones) have an eigenvalue decomposition.</td>
</tr>
<tr>
<td><strong>(Typical) Application Relevance</strong></td>
<td></td>
</tr>
<tr>
<td>Behavior of $A$ itself, or $A^{-1}$.</td>
<td>Behavior of $A^k$, $e^{tA}$.</td>
</tr>
</tbody>
</table>
The SVD has many connections with other fundamental topics in linear algebra...

In the following slides, assume that $A \in \mathbb{C}^{m \times n}$, let $p = \min(m, n)$, and let $r \leq p$ denote the number of non-zero singular values of $A$; finally let $\langle \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m \rangle$ denote the space spanned by the vectors $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$.

Theorem (Rank of a Matrix)

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

Proof: The rank of a diagonal matrix is the number of non-zero entries. In the decomposition $A = U \Sigma V^*$, both $U$ and $V$ are full rank. Therefore $\text{rank}(A) = \text{rank}(\Sigma) = r$. □
Theorem (Range and Nullspace of a Matrix)

\[ \text{range}(A) = \langle \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r \rangle, \text{ and } \text{null}(A) = \langle \mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n \rangle. \]

Proof: This follows directly from the fact that

\[ \text{range}(\Sigma) = \langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_r \rangle \subseteq \mathbb{C}^m, \]
and \[ \text{null}(\Sigma) = \langle \mathbf{e}_{r+1}, \mathbf{e}_{r+2}, \ldots, \mathbf{e}_n \rangle \subseteq \mathbb{C}^n. \]
Theorem (Euclidean and Frobenius Matrix Norms)

\[ \|A\|_2 = \sigma_1, \text{ and } \|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}. \]

**Proof:** We already established that \( \sigma_1 = \|A\|_2 \) in the existence proof, since \( A = U\Sigma V^* \) with unitary \( U \) and \( V \),

\[ \|A\|_2 = \|\Sigma\|_2 = \max\{|\sigma_i|\} = \sigma_1. \]

Now, since the Frobenius norm is invariant under unitary transformations, \( \|A\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}. \]
Theorem

The non-zero singular values of \( A \) are the square roots of the non-zero eigenvalues of \( A^*A \) or \( AA^* \) (these two matrices have the same non-zero eigenvalues).

Proof: From

\[
A^* A = (U \Sigma V^*)^* (U \Sigma V^*) = V \Sigma^* U^* U \Sigma V^* = V (\Sigma^* \Sigma) V^* = V (\Sigma^* \Sigma) V^{-1}
\]

we see that \( A^* A \) and \( \Sigma^* \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2) \) have the same eigenvalues, \( \lambda_i = \sigma_i^2 \), \( i = 1, 2, \ldots, p \).

If \( n > p \), we have an additional \( n - p \) zero eigenvalues.

The same argument works for \( AA^* \) (just substitute \( m \) for \( n \))...
Theorem

If $A = A^*$, then the singular values of $A$ are the absolute values of the eigenvalues of $A$.

Proof: The eigenvalues of a Hermitian matrix are real since if $(\lambda, \vec{v})$ is an eigenvalue-eigenvector pair ($\lambda \neq 0$), then

$$\langle \vec{v}, A\vec{v} \rangle = \vec{v}^* A\vec{v} = (A^*\vec{v})^* \vec{v} = \langle A^*\vec{v}, \vec{v} \rangle$$
$$\langle \vec{v}, A\vec{v} \rangle = \langle \vec{v}, \lambda \vec{v} \rangle = \lambda \langle \vec{v}, \vec{v} \rangle$$
$$\langle \vec{v}, A\vec{v} \rangle = \langle A^*\vec{v}, \vec{v} \rangle = \langle A\vec{v}, \vec{v} \rangle = \langle \lambda \vec{v}, \vec{v} \rangle = \lambda^* \langle \vec{v}, \vec{v} \rangle$$

Hence, $\lambda = \lambda^* \Rightarrow \lambda \in \mathbb{R}$. Further, a Hermitian matrix has a complete set of orthogonal eigenvectors. This means that we can diagonalize $A$

$$A = Q\Lambda Q^* = Q(|\Lambda| \text{sign}(\Lambda))Q^*$$

for some unitary matrix $Q$ and $\Lambda$ a real diagonal matrix...
Since \( \text{sign}(\Lambda)Q^* \) is unitary, we have

\[
A = Q \left[ \begin{array}{c|c}
|\Lambda| & (\text{sign}(\Lambda)Q^*) \\
\hline
U & \Sigma & V^*
\end{array} \right]
\]

a SVD of \( A \), where \( \sigma_i = |\lambda_i|, \ i = 1, 2, \ldots, p \). (An appropriate ordering of the columns of \( U \) guarantees that the singular values are ordered in decreasing order.) □
Theorem

For $A \in \mathbb{C}^{m \times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.

Proof:

$$|\det(A)| = |\det(U\Sigma V^*)| = |\det(U)||\det(\Sigma)||\det(V^*)|$$

$$= 1 \cdot |\det(\Sigma)| \cdot 1 = |\det(\Sigma)| = \prod_{i=1}^{m} \sigma_i$$

where we have used the fact that $\det(AB) = \det(A)\det(B)$ and that the magnitude of the determinant of a unitary matrix is one.

$\square$
Given the SVD of $A$, $A = U\Sigma V^*$, we can represent $A$ as a sum of $r$ rank-one matrices

$$A = \sum_{j=1}^{r} \sigma_j \bar{u}_j \bar{v}_j^*$$

This is certainly not the only way to write $A$ as a sum of rank-one matrices: it could be written as a sum of its $m$ rows, $n$ columns, or even its $mn$ entries...

The decomposition above has the special property that if we truncate the sum at $\nu < r$, then that partial sum captures as much energy of $A$ as possible for a rank-$\nu$ sub-matrix of $A$.

We formalize this in a theorem...
Theorem (Optimal Low-Rank Approximation)

For any \( \nu \) with \( 0 \leq \nu < r \), define

\[
A_\nu = \sum_{j=1}^{\nu} \sigma_j \bar{u}_j \bar{v}_j^*
\]

if \( \nu = p = \min(m, n) \), define \( \sigma_{\nu+1} = 0 \). Then

\[
\|A - A_\nu\|_2 = \inf_{B \in \mathbb{C}^{m \times n} \text{ with } \text{rank}(B) \leq \nu} \|A - B\|_2 = \sigma_{\nu+1}
\]
Proof: Suppose that there is some $B$ with $\text{rank}(B) \leq \nu$ such that

$$\|A - B\|_2 < \|A - A_\nu\|_2 = \sigma_{\nu+1}.$$ 

Then there is an 

$(n - \nu)$-dimensional subspace $\mathbb{W} \subseteq \mathbb{C}^n$ such that

$\bar{w} \in \mathbb{W} \Rightarrow B\bar{w} = 0$. Thus $\forall \bar{w} \in \mathbb{W}$:

$$\|A\bar{w}\|_2 = \|(A - B)\bar{w}\|_2 \leq \|A - B\|_2 \|\bar{w}\|_2 < \sigma_{\nu+1} \|\bar{w}\|_2.$$ 

Now, $\mathbb{W}$ is an $(n - \nu)$-dimensional subspace where

$$\|A\bar{w}\|_2 < \sigma_{\nu+1} \|\bar{w}\|_2.$$ 

But there is a $(\nu + 1)$-dimensional subspace

where $\|A\bar{w}\|_2 \geq \sigma_{\nu+1} \|\bar{w}\|_2$ — the space spanned by the first $(\nu + 1)$ right singular vectors of $A$.

Since the sum of the two subspaces $(\nu + 1) + (n - \nu) = n + 1$ exceeds $n$, there must be a non-zero vector lying in both. This is a contradiction. □
The preceding theorem has a nice geometrical interpretation.

Ponder the issue of finding the best approximation of an $n$-dimensional hyper-ellipsoid.

The best approximation by a 2-dimensional ellipse must be the ellipse spanned by the largest and second largest axis.

We get the best 3-dimensional approximation by adding span of the 3rd largest axis, etc...

This is useful in many applications, e.g. signal compression (images, audio, etc.), analysis of large data sets, etc.
We state the following theorem, and leave the proof as an “exercise.”

Theorem

*For the matrix* $A_\nu$ *as defined in the previous theorem*

$$
\|A - A_\nu\|_F = \inf_{B \in \mathbb{C}^{m \times n}} \|A - B\|_F = \sqrt{\sigma_{\nu+1}^2 + \sigma_{\nu+2}^2 + \cdots + \sigma_r^2}
$$

We will get back to how to compute the SVD later. For now, we note that it is a powerful tool which can be used to

- find the numerical rank of a matrix;
- find the orthonormal basis for the range and null-space;
- computing $\|A\|_2$;
- computing low-rank approximations.

The SVD shows up in least squares fitting, regularization, intersection of subspaces (video games?), and many, many other problems.