Throughout these notes we assume $V, W$ are finite dimensional inner product spaces over $\mathbb{C}$.

1. Upper Triangular Representation

**Proposition:** Let $T \in \mathcal{L}(V)$. There exists an orthonormal basis $\mathcal{B} = \{u_1, u_2, ..., u_n\}$ of $V$ such that $[T]_\mathcal{B}$ is upper triangular. In particular, if $A \in M_{nn}(\mathbb{C})$ then there exists an upper triangular matrix $\Delta$ and a unitary matrix $U$ such that $A = U \Delta U^*$.

**Proof:** Let $T \in \mathcal{L}(V)$. We prove this proposition by induction on dim $V$. The base case, is trivial because the matrix of $T$ with respect to any basis is upper triangular if $V$ is 1-dimensional.

Let $n > 1$ be an integer. Assume that if dim $W = n - 1$ then for any $S \in \mathcal{L}(W)$ there exists an orthonormal basis $\mathcal{C}$ of $W$ such that $[S]_\mathcal{C}$ is upper triangular.

Let $n = \text{dim} \ V$. Let $\lambda$ be an eigenvalue of $T$ and $v$ be a corresponding eigenvector such that $\|v\| = 1$. Let $W = \text{span}(v)^\perp$. Then dim $W = n - 1$.

Consider that for any orthonormal basis $\mathcal{C}$ of $W$, $\mathcal{B} = \{v\} \cup \mathcal{C}$ is an orthonormal basis of $V$. In such a basis $\mathcal{B}$ we have

$$[T]_\mathcal{B} = \begin{pmatrix} \lambda & * & * & ... & * \\ 0 & b_{11} & b_{12} & ... & b_{1(n-1)} \\ 0 & b_{21} & b_{22} & ... & b_{2(n-1)} \\ . & . & . & ... & . \\ . & . & . & ... & . \\ 0 & b_{(n-1)1} & b_{(n-1)2} & ... & b_{(n-1)(n-1)} \end{pmatrix}.$$

In particular, there exists $S \in \mathcal{L}(W)$ such that

$$[S]_\mathcal{C} = \begin{pmatrix} b_{11} & b_{12} & ... & b_{1(n-1)} \\ b_{21} & b_{22} & ... & b_{2(n-1)} \\ . & . & ... & . \\ . & . & ... & . \\ b_{(n-1)1} & b_{(n-1)2} & ... & b_{(n-1)(n-1)} \end{pmatrix}.$$

By induction we can pick an orthonormal basis $\mathcal{C}$ of $W$ such that this matrix above is upper triangular. Then $\mathcal{B}$ is an orthonormal basis of $V$ and $[T]_\mathcal{B}$ would be upper triangular.
Now let \( A \in M_{nn}(\mathbb{C}) \). Then there exists \( T \in \mathcal{L}(\mathbb{C}^n) \) such that its standard matrix is \( A \). Then there exists an orthonormal basis \( B = \{v_1, v_2, \ldots, v_n\} \) of \( \mathbb{C}^n \) such that \([T]_B\) is upper triangular.

In particular, using \( S \) to denote the standard basis, \( \Delta = [T]_B = [I_n]_BA[I_n]_B = [I_n]_B^{-1}A[I_n]_B \) is upper triangular.

Let \( U = [I_n]_S \). Clearly, \( U = [v_1 \ v_2 \ \cdots \ v_n] \) is unitary, hence \( U^{-1} = U^* \) and \( A = U \Delta U^* \). □

**Remark:** This argument does not hold for \( \mathbb{R} \), since a real matrix may have complex eigenvalues and eigenvectors. Indeed we know that the entries on the diagonal of \( \Delta \) are the eigenvalues!

**Corollary:** Let \( X \) be a real inner product space. Let \( T \in \mathcal{L}(X) \) have only real eigenvalues. Then there exists an orthonormal basis \( B = \{u_1, u_2, \ldots, u_n\} \) of \( X \) such that \([T]_B\) is upper triangular. In particular, if \( A \in M_{nn}(\mathbb{R}) \) only has real eigenvalues then there exists an upper triangular matrix \( \Delta \) and a unitary (orthogonal) matrix \( U \) such that \( A = U \Delta U^T \).

**Proof:** Begin by copying the proof of the previous proposition with \( \mathbb{R} \) in place of \( \mathbb{C} \). It only remains to show that the eigenvalues of \( S \in \mathcal{L}(W) \) are real.

Notice that \( \det([T]_B - zI_n) = (\lambda - z)\det([S]_C - zI_{n-1}) \). This means that eigenvalues of \( S \) are eigenvalues of \( T \), and hence are real, which completes the proof by induction □
2. Spectral Theorem for Self-Adjoint and Normal Operators

$T \in \mathcal{L}(V)$ is called self-adjoint if $T^* = T$.

A matrix $A \in M_{nn}(\mathbb{C})$ is called self-adjoint or Hermitian if $A^* = A$.

**Proposition:** Let $T \in \mathcal{L}(V)$ be self-adjoint. Then the eigenvalues of $T$ are real and there exists an orthonormal basis of eigenvectors. In particular, if $A \in M_{nn}(\mathbb{C})$ is Hermitian, then there exists a unitary matrix $U$ and a real diagonal matrix $D$ such that $A = UDU^*$.

**Proof:** Let $T \in \mathcal{L}(V)$ be self-adjoint. Pick an orthonormal basis $\mathcal{B}$ of $V$ such that $A = [T]_{\mathcal{B}}$ is upper triangular. $A$ is Hermitian or self-adjoint (as $T$ is). Then the entries above the main diagonal must be zero as the entries below the main diagonal are zero. Moreover, the entries on the diagonal are real since they satisfy the equation $\bar{z} = z$, which requires that their imaginary parts are 0. Thus $A$ is a real diagonal matrix, so the eigenvalues of $T$ are real. Furthermore, the basis $\mathcal{B}$ consists of orthonormal eigenvectors.

Now let $A \in M_{nn}(\mathbb{C})$ be Hermitian. There exists a unitary matrix $U$ such that $A = UDU^*$ and $D$ is upper triangular. Then $D = U^*AU$ and $D^* = (U^*AU)^* = U^*A^*(U^*)^* = U^*AU = D$. Thus $D$ is real and diagonal $\square$

As a consequence of this proposition, eigenvectors of a self-adjoint operator corresponding to different eigenvalues must be orthogonal:

**Corollary:** Let $T \in \mathcal{L}(V)$ be self-adjoint and $\lambda, \mu \in \sigma(T)$ (eigenvalues) corresponding to eigenvectors $u, v$ respectively. If $\lambda \neq \mu$ then $u \perp v$. 
An operator \( T \in \mathcal{L}(V) \) (or matrix \( T \in M_{nn}(\mathbb{C}) \)) is called normal if \( T^*T = TT^* \).

If \( T \in \mathcal{L}(V) \) is self-adjoint (\( T^* = T \)) or unitary (\( T^*T = I = TT^* \)) then \( T \) is normal.

**Proposition:** Let \( T \in \mathcal{L}(V) \) be normal. Then \( T \) has an orthonormal basis of eigenvectors. In particular, if \( A \in M_{nn}(\mathbb{C}) \) is normal, then there exists a unitary matrix \( U \) and a diagonal matrix \( D \) such that \( A = UDU^* \).

**Proof:** Let \( T \in \mathcal{L}(V) \) be normal. Pick an orthonormal basis \( \mathcal{B} \) of \( V \) such that \( A = [T]_\mathcal{B} \) is upper triangular. \( A \) is normal. It suffices to show that \( A \) is diagonal. We will do this by induction on the size of \( A \).

Clearly, if \( A \) is a \( 1 \times 1 \) matrix, then it is diagonal and \( T \) has an orthonormal basis of eigenvectors.

Let \( n > 1 \) be an integer. Assume any \( (n-1) \times (n-1) \) matrix that is normal and upper-triangular must be diagonal.

Let \( A \in M_{nn} \) be normal and upper triangular. Then

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\
0 & & & & \\
\vdots & & B & & \\
0 & & & &
\end{pmatrix},
\]

where \( B \in M_{(n-1)(n-1)} \) is upper triangular. By direct calculation \( (A^*A)_{11} = \overline{a_{11}}a_{11} = |a_{11}|^2 \) and

\[
(A^*A)_{11} = \sum_{j=1}^{n} |a_{1j}|^2.
\]

Since \( A \) is normal, this implies \( a_{12} = \cdots = a_{1n} = 0 \). Hence

\[
A = \begin{pmatrix}
a_{11} & 0 & 0 & \ldots & 0 \\
0 & & & & \\
\vdots & & B & & \\
0 & & & &
\end{pmatrix}.
\]
\[
A^*A = \begin{pmatrix}
|a_{11}|^2 & 0 & 0 & \ldots & 0 \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & B^*B & \ddots & 0 \\
0 & \ldots & \ldots & \ddots & 0 \\
0 & \ldots & \ldots & \ldots & 0
\end{pmatrix} = AA^* = \begin{pmatrix}
|a_{11}|^2 & 0 & 0 & \ldots & 0 \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & BB^* & \ddots & 0 \\
0 & \ldots & \ldots & \ddots & 0 \\
0 & \ldots & \ldots & \ldots & 0
\end{pmatrix}
\]
implies \( B^*B = BB^* \).

So \( B \) is normal and by induction diagonal. This means \( A \) is diagonal \( \square \)

**Proposition:** \( T \in \mathcal{L}(V) \) is normal iff for all \( x \in V \), \( \|Tx\| = \|T^*x\| \).

**Proof:** Let \( T \in \mathcal{L}(V) \). Assume \( T \) is normal. Let \( x \in V \).

\[
\|Tx\|^2 = (Tx, Tx) = (x, T^*Tx) = (x, TT^*x) = (T^*x, T^*x) = \|T^*x\|^2.
\]

So \( \|Tx\| = \|T^*x\| \).

Now assume for all \( x \in V \), \( \|Tx\| = \|T^*x\| \).

Let \( x, y \in V \). Using polarization identities,

\[
(T^*Tx, y) = (Tx, Ty) = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|Tx + \alpha Ty\|^2 = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|T(x + \alpha y)\|^2 = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|T^*(x + \alpha y)\|^2 = (T^*x, T^*y) = (T^*x, y).
\]

Hence \( T^*T = TT^* \), so \( T \) is normal \( \square \)

**Suggested Homework:** 2.1, 2.3, 2.4, and 2.10.
3. Polar and Singular Value Decompositions

A self-adjoint operator $T \in \mathcal{L}(V)$ is positive definite if for all non-zero $x \in V$, $(Tx, x) > 0$, and positive semidefinite if for all $x \in V$, $(Tx, x) \geq 0$.

We use the notations $T > 0$ and $T \geq 0$ for positive definite and semidefinite respectively.

**Proposition:** Let $T \in \mathcal{L}(V)$ be self-adjoint. Then

(1) $T > 0$ iff the eigenvalues of $T$ are positive.

(2) $T \geq 0$ iff the eigenvalues of $T$ are non-negative.

**Proof:** Let $T \in \mathcal{L}(V)$ be self-adjoint. There exists an orthonormal basis $\mathcal{B} = \{v_1, v_2, ..., v_n\}$ of $V$ of eigenvectors of $T$ such that $D = [T]_\mathcal{B} = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}$ is diagonal, where $\lambda_1, \lambda_2, ..., \lambda_n$ are the corresponding real eigenvalues of $T$ (counting multiplicity).

Assume $T > 0$. Then $(Tv_i, v_i) = \lambda_i > 0$ for $i = 1, 2, ..., n$.

Now assume $\lambda_i > 0$ for $i = 1, 2, ..., n$. Let $v \in V$ be non-zero. There $c_1, c_2, ..., c_n \in \mathbb{C}$, not all zero, such that $v = \sum_{i=1}^n c_i v_i$. Then $(Tv, v) = \sum_{i=1}^n |c_i|^2 \lambda_i > 0$. Thus $T > 0$.

This proves (1). We get a proof of (2) by replacing all $>$ with $\geq$.

**Proposition:** Let $T \in \mathcal{L}(V)$ be a positive semidefinite, self-adjoint operator. Then there exists a unique positive semidefinite, self-adjoint operator $S \in \mathcal{L}(V)$ such that $S^2 = T$.

$S$ is denoted $\sqrt{T}$ and is called the (positive) square root of $T$.

**Proof:** Let $T \in \mathcal{L}(V)$ be a positive semidefinite, self-adjoint operator. Pick a basis $\mathcal{B} = \{v_1, v_2, ..., v_n\}$ of $V$ of eigenvectors of $T$ corresponding to eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ (counting multiplicity). Since $T \geq 0$, the eigenvalues are non-negative.

Let $S \in \mathcal{L}(V)$ be given by $[S]_\mathcal{B} = \text{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_n}\}$.

The eigenvalues of $S$ are $\sqrt{\lambda_i} \geq 0$ for $i = 1, 2, ..., n$. So $S$ is a positive semidefinite, self-adjoint operator. Clearly $S^2 = T$ since $[S]^2 = [T]_\mathcal{B}$.
It only remains to show that $S$ is unique. Suppose $R \in \mathcal{L}(V)$ is a positive semidefinite, self-adjoint operator satisfying $R^2 = T$. Let $\mu_1, \mu_2, \ldots, \mu_n$ be the eigenvalues of $R$ corresponding to an orthonormal basis $\mathcal{C} = \{u_1, u_2, \ldots, u_n\}$ of $V$ consisting of eigenvectors of $R$. Then $[T]_C = [R^2]_C = \text{diag}\{\mu_1^2, \mu_2^2, \ldots, \mu_n^2\}$. So (up to a re-ordering) the eigenvalues $\lambda_i$ of $T$ satisfy $\lambda_i = \mu_i^2$ for $i = 1, 2, \ldots, n$. Hence, since $R$ is positive semidefinite, $\mu_i = \sqrt{\lambda_i}$ for $i = 1, 2, \ldots, n$. So $R = S$. Thus $S$ is unique \(\square\)

Consider an operator $T \in \mathcal{L}(V, W)$. Its **Hermitian square** is the operator $T^*T \in \mathcal{L}(V)$.

The Hermitian square is a positive semidefinite, self-adjoint operator:

$$(T^*T)^* = T^*(T^*)^* = T^*T$$

and for all $x \in V$, $(T^*Tx, x) = (Tx, Tx) = \|Tx\|^2 \geq 0$.

Let $T \in \mathcal{L}(V, W)$. Let $R = \sqrt{T^*T} \in \mathcal{L}(V)$, the square root of the Hermitian square. This (positive semidefinite, self-adjoint) operator $R$ is called the **modulus** of the operator $T$, and is denoted $|T|$.

**Proposition**: Let $T \in \mathcal{L}(V, W)$. For all $x \in V$, $\|T|x\| = \|Tx\|$.

**Proof**: Let $T \in \mathcal{L}(V, W)$. Let $x \in V$. Then

$$\|T|x\|^2 = (|T|x, |T|x) = (|T|^2|x, x) = (T^*Tx, x) = (Tx, Tx) = \|Tx\|^2.$$  

Thus $\|T|x\| = \|Tx\|$ \(\square\)

As a consequence...

**Corollary**: Let $T \in \mathcal{L}(V, W)$. $\text{Ker } T = \text{Ker } |T| = (\text{range } |T|)\perp$.

The first equality follows immediately from the previous proposition. The second equality makes use of the fact that $\text{Ker } S = (\text{range } S^*)\perp$ for $S \in \mathcal{L}(V)$, and $|T|^* = |T|$.
Proposition (Polar Decomposition): Let \( T \in \mathcal{L}(V) \). Then \( T = U|T| \) for some unitary operator \( U \in \mathcal{L}(V) \).

**Proof:** Let \( T \in \mathcal{L}(V) \). Let \( x \in \text{range } |T| \). Then there exists \( v \in V \) such that \( x = |T|v \). Define \( U_0 : \text{range } |T| \to V \) by \( U_0x = T v \).

We must make sure \( U_0 \) is well-defined (has a unique output). Suppose \( v_1 \in V \) satisfies \( x = |T|v_1 \). But then \( v - v_1 \in \text{Ker } |T| = \text{Ker } T \), so \( Tv = Tv_1 \) as required.

It is easy to check that \( U_0 \) is a linear transformation.

By construction \( T = U_0|T| \) and \( \|U_0x\| = \|Tv\| = \||T|v\| = \|x\| \). So \( U_0 \) is an isometry.

\((\text{range } |T|)^\perp = \text{Ker } |T|^* = \text{Ker } |T| = \text{Ker } T.\)

Pick a unitary operator \( U_1 : \text{Ker } T \to (\text{range } T)^\perp = \text{Ker } T^* \), which can be done since \( \dim(\text{Ker } T) = \dim(\text{Ker } T^*) \) (Rank Theorem for square matrices!)

Define \( U \in \mathcal{L}(V) \) by \( U = U_0P_{\text{range } |T|} + U_1P_{\text{Ker } T} \), where \( P_X \) denotes orthogonal projection onto the subspace \( X \).

(ELFY) \( U \) is a unitary transformation and \( T = U|T| \)

Let \( T \in \mathcal{L}(V,W) \). The eigenvalues of \( |T| = \sqrt{T^*T} \) are called the singular values of \( T \).

In particular, if \( \lambda \) is an eigenvalue of \( T^*T \) then \( \sqrt{\lambda} \) is a singular value of \( T \).

Let \( T \in \mathcal{L}(V,W) \). Let \( \sigma_1, \sigma_2, ..., \sigma_n \) be the singular values of \( T \) (counting multiplicity).

Assume \( \sigma_1, \sigma_2, ..., \sigma_r \) are the non-zero singular values. So \( \sigma_k = 0 \) for \( k = r + 1, ..., n \).

By definition, \( \sigma_1^2, \sigma_2^2, ..., \sigma_n^2 \) are the eigenvalues of \( T^*T \in \mathcal{L}(V) \).

Let \( v_1, v_2, ..., v_n \) be an orthonormal basis of \( V \) consisting of eigenvectors of \( T^*T \) corresponding to the eigenvalues \( \sigma_1^2, \sigma_2^2, ..., \sigma_n^2 \). Such a basis can be selected since \( T^*T \) is self-adjoint.

So \( T^*Tv_i = \sigma_i^2v_i \) for \( i = 1, 2, ..., n \).
Let \( w_i = \frac{1}{\sigma_i} T v_i \) for \( i = 1, 2, \ldots, r \). Then the vectors \( w_1, w_2, \ldots, w_r \) are orthonormal:

Let \( i, j \in \{ 1, 2, \ldots, r \} \). Then

\[
(w_i, w_j) = \left( \frac{1}{\sigma_i} T v_i, \frac{1}{\sigma_j} T v_j \right) = \frac{1}{\sigma_i \sigma_j} (T^* T v_i, v_j) = \frac{1}{\sigma_i \sigma_j} (\sigma_i^2 v_i, v_j) = \frac{\sigma_i}{\sigma_j} (v_i, v_j) = \delta_{ij},
\]

where \( \delta_{ij} \) is 1 when \( i = j \) and 0 otherwise (Kronecker delta function).

Claim: The operator \( T \) can be represented by

\[
T = \sum_{k=1}^{r} \sigma_k w_k v_k^*.
\]

To justify the claim, it suffices to check that these operators coincide on a basis of \( V \).

For \( i = 1, 2, \ldots, r \),

\[
\sum_{k=1}^{r} \sigma_k w_k v_k^* v_i = \sum_{k=1}^{r} T v_k (v_i, v_k) = T v_i.
\]

For \( i = r + 1, \ldots, n \), \( \sigma_i = 0 \) both

\[
\sum_{k=1}^{r} \sigma_k w_k v_k^* v_i = 0 \quad \text{and} \quad T v_i = 0 \quad \text{since} \quad \sigma_i = 0 \implies \parallel T v_i \parallel^2 = (T v_i, T v_i) = (T^* T v_i, v_i) = (0, v_i) = 0.
\]

So since the operators agree on a (orthonormal) basis of \( V \) they must be equal.

The representation

\[
T = \sum_{k=1}^{r} \sigma_k w_k v_k^*
\]

is called a singular value decomposition of the operator \( T \).
Proposition: Let $T \in \mathcal{L}(V, W)$. Suppose $T$ can be represented by

$$T = \sum_{k=1}^{r} \sigma_k w_k v_k^*,$$

where $\sigma_i > 0$ for $i = 1, 2, ..., r$, $v_1, v_2, ..., v_r$ are orthonormal in $V$, and $w_1, w_2, ..., w_r$ orthonormal in $W$.

Then this representation is a singular value decomposition of $T$.

Proof: Assume $T \in \mathcal{L}(V, W)$ has the representation described in the proposition. It suffices to show that $T^* T v_i = \sigma_i^2 v_i$ for $i = 1, 2, ..., r$ and that this accounts for all non-zero eigenvalues of $T^* T$. Let $i \in \{1, 2, ..., r\}$.

First notice that

$$T^* = \sum_{k=1}^{r} \sigma_k v_k w_k^*$$

and

$$T v_i = \sum_{k=1}^{r} \sigma_k w_k v_k^* v_i = \sigma_i w_i.$$

Thus

$$T^* T v_i = \sigma_i \sum_{k=1}^{r} \sigma_k v_k w_k^* w_i = \sigma_i^2 v_i.$$

Extend $v_1, v_2, ..., v_r$ to an orthonormal basis $v_1, v_2, ..., v_n$ of eigenvectors of $T^* T$. Let $\lambda$ be an eigenvalue of $T^* T$ not in $\{\sigma_i^2 | i = 1, 2, ..., r\}$. Then for some $i = r + 1, ..., n$, $T^* T v_i = \lambda v_i$.

However, $T v_i = 0$ since

$$\sum_{k=1}^{r} \sigma_k w_k v_k^* v_i = 0.$$

So $T^* T v_i = 0$ and thus $\lambda = 0$.
Matrix representation of a singular value decomposition:

By fixing orthonormal bases for $V$ and $W$ and working in these coordinates, we may assume $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$.

Let

$$T = \sum_{k=1}^{r} \sigma_k w_k v_k^*$$

be a singular value decomposition of $T$.

This can be rewritten in matrix form as

$$T = \tilde{B} \tilde{D} \tilde{A}^*,$$

where $\tilde{D} = \text{diag}\{\sigma_1, \sigma_2, ..., \sigma_r\} \in M_{rr}$, $\tilde{B} \in M_{mr}$ has columns $w_k \ (k = 1, 2, ..., r)$ and $\tilde{A} \in M_{nr}$ has columns $v_k \ (k = 1, 2, ..., r)$.

In particular, $\tilde{A}$ and $\tilde{B}$ are isometries as their columns are orthonormal. In the case when $m = n = r$, $\tilde{A}$ and $\tilde{B}$ are unitary.

By extending $v_1, v_2, ..., v_r$ and $w_1, w_2, ..., w_r$ to orthonormal bases $v_1, v_2, ..., v_n$ and $w_1, w_2, ..., w_m$ of $V, W$ respectively, we get unitary matrices $A = [v_1 \ v_2 \ \cdots \ v_n] \in M_{nn}$ and $B = [w_1 \ w_2 \ \cdots \ w_m] \in M_{mm}$ such that

$$T = BDA^*,$$

where $D \in M_{mn}$ has $\tilde{D}$ in the upper left $r \times r$ block and zeroes elsewhere.

This matrix form is also often called a singular value decomposition of $T$.

(ELFY) Write down a random $3 \times 2$ matrix and perform the singular value decomposition (in matrix form).
From a singular value decomposition to a polar decomposition:

Let $T = BDA^*$ be a singular value decomposition of $T \in \mathcal{L}(V)$ in matrix form.

Then set $U = BA^*$. $U$ is unitary since $U^{-1} = (BA^*)^{-1} = (A^*)^{-1}B^{-1} = AB^* = (BA^*)^* = U^*$.

Notice that $T^*T = AD^*B^*BDA^* = AD^2A^* = (ADA^*)^2 \rightarrow |T| = ADA^*$.

Then $U|T| = (BA^*)(ADA^*) = BDA^* = T$ is a polar decomposition.

**Suggested Homework:** 3.1, 3.3 (first and third matrices only), 3.5(a), 3.6(a)(b), and 3.9.
4. What do Singular Values Describe?

Essentially a singular value decomposition is a “diagonalization” of $T \in \mathcal{L}(V, W)$ with respect to a pair of orthonormal bases (one for $V$, one for $W$).

So what does it tell us?

To gain the appropriate intuition, consider the case where $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Let $B = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ (the closed unit ball).

Let $\sigma_1, \sigma_2, ..., \sigma_n$ be the singular values (counting multiplicity, 0’s at the end of the list) of $T$. Let $\mathcal{B}, \mathcal{C}$ be the corresponding orthonormal bases from the singular value decomposition of $V, W$ respectively.

Then the matrix of $T$ with respect to these bases, $[T]_{\mathcal{B}\mathcal{C}}$, has the diagonal matrix $D = \text{diag}\{\sigma_1, \sigma_2, ..., \sigma_r\}$ (corresponding to the non-zero singular values $\sigma_1, \sigma_2, ..., \sigma_r$) as its upper left $r \times r$ block and zeros everywhere else.

Let $v \in B$. Set $x = [v]_B = [I_V]_B v$, where $B$ denotes the standard basis. Since the columns of $[I_V]_B = [I_V]_B^{-1}$ are the orthonormal basis vectors of $B$, this matrix unitary (an isometry), and hence $x \in B$.

Then $x = (x_1 \ x_2 \ · · · \ x_n)^T$ satisfies $\sum_{i=1}^{n} x_i^2 \leq 1$. Let $y = [T]_{\mathcal{B}\mathcal{C}} x = (y_1 \ y_2 \ · · · \ y_n)^T$.

Clearly $y_i = \sigma_i x_i$ for $i = 1, 2, ..., r$, and $y_j = 0$ for $j = r + 1, ..., n$.

Thus $y$ satisfies

$$\sum_{i=1}^{r} \frac{y_i^2}{\sigma_i^2} \leq 1.$$ 

So in the first $r$ orthonormal basis directions in basis $\mathcal{C}$ of $\mathbb{R}^m$ (corresponding to non-zero singular values) $y$ is on or within the ellipsoid given by

$$\sum_{i=1}^{r} \frac{y_i^2}{\sigma_i^2} = 1.$$ 

The last $n - r$ orthonormal basis directions in $y$ are “killed” (sent to 0).

Furthermore, when $x$ is on the unit sphere (boundary of the unit ball), then $y$ will be on the ellipsoid above, and $T(B)$ is the closed ellipsoidal ball with axes of length $2\sigma_1, 2\sigma_2, ..., 2\sigma_r$ with respect to the first $r$ coordinates given by $\mathcal{C}$. 
(ELFY) Let $\mathcal{B} = \{v_i|i = 1, 2, ..., n\}$ be an orthonormal basis of $V$. Then $\|x\| = ||[x]_{\mathcal{B}}||$.

Now let’s discuss the notion of operator norm. Let $T \in \mathcal{L}(V,W)$.

The operator norm of $T$ is given by $\|T\| = \max\{\|Tx\| : x \in V, \|x\| \leq 1\}$.

First we justify that the largest singular value is the operator norm:

Let $x \in V$ and $\|x\| \leq 1$. Suppose

$$ T = \sum_{k=1}^{r} \sigma_k w_k v_k^* $$

is a singular value decomposition with respect to two orthonormal bases $\mathcal{B} = \{v_i|i = 1, 2, ..., n\}$ and $\mathcal{C} = \{w_i|i = 1, 2, ..., m\}$ of $V$ and $W$ respectively. Without loss of generality assume that the singular values are arranged in a non-increasing order.

Let $[x]_{\mathcal{B}} = (c_1 \ c_2 \ \cdots \ c_n)$. (ELFY) $\sum_{i=1}^{n} c_i^2 \leq 1$.

Then $[T]_{\mathcal{C}}[x]_{\mathcal{B}} = (\sigma_1 c_1 \ \sigma_2 c_2 \ \cdots \ \sigma_n c_n)$.

Then

$$ \|Tx\|^2 = ||[T]_{\mathcal{C}}[x]_{\mathcal{B}}||^2 = \sum_{i=1}^{n} \sigma_i^2 c_i^2 \leq \sigma_1^2 \sum_{i=1}^{n} c_i^2 \leq \sigma_1^2 \rightarrow \|Tx\| \leq \sigma_1. $$

Moreover

$$ \|Tv_1\| = ||[T]_{\mathcal{C}}[v_1]_{\mathcal{B}}|| = \sigma_1. $$

Therefore $\|T\| = \sigma_1$.

(ELFY) Show that the operator norm satisfies the four fundamental properties of a norm:

1. Homogeneity: For all $T \in \mathcal{L}(V,W)$ and for all $c \in \mathbb{F}$, $\|cT\| = |c|\|T\|$.
2. Triangle Inequality: For all $T, S \in \mathcal{L}(V,W)$, $\|T + S\| \leq \|T\| + \|S\|$.
3. Positivity (aka non-negativity): For all $T \in \mathcal{L}(V,W)$, $\|T\| \geq 0$.
4. Definiteness (aka non-degeneracy): Let $T \in \mathcal{L}(V,W)$. Then $\|T\| = 0$ iff $T = 0 \in \mathcal{L}(V,W)$. 

An important and immediate consequence of the homogeneity of the operator norm:

Let $x \in V$ be non-zero and $c = \|x\|$. Note $c \in \mathbb{R}$ and $c > 0$. Set $y = \frac{1}{c}x$. Then $\|Ty\| \leq \|T\|$ by the definition of the operator norm.

That implies that $\|Tx\| = \|Tcy\| = c\|Ty\| \leq c\|T\| = \|T\|\|x\|$.

The so-called Frobenius norm on $L(V, W)$ is given by $\|T\|_{\text{Frob}} = \sqrt{\text{trace}(A^*A)}$ where $A = [T]_{CB}$ for some orthonormal bases $\mathcal{B}, \mathcal{C}$ of $V, W$ respectively.

This does not depend on the choice of orthonormal bases (so it’s well-defined) because a change of basis matrix between orthonormal bases is unitary, and for an $m \times n$ matrix $D$, $(UDV)^*(UDV) = V^*D^*DV$ has the same characteristic polynomial (and thus the same trace) as $D^*D$ for all unitary matrices $U \in M_{mm}$ and for all unitary matrices $V \in M_{nn}$.

How does this one compare to the operator norm?

Again, suppose

$$T = \sum_{k=1}^{r} \sigma_k w_k v_k^*$$

is a singular value decomposition with respect to two orthonormal bases $\mathcal{B} = \{v_i|i = 1, 2, ..., n\}$ and $\mathcal{C} = \{w_i|i = 1, 2, ..., m\}$ of $V$ and $W$ respectively. Without loss of generality assume that the singular values are arranged in a non-increasing order.

Then $[T]_{CB}$ has zeros everywhere except that its upper left $r \times r$ block is the diagonal matrix $D = \text{diag}\{\sigma_1, \sigma_2, ..., \sigma_r\}$.

So $\|T\|_{\text{Frob}}^2 = \text{trace}([T]_{CB}^*[T]_{CB}) = \sum_{i=1}^{r} \sigma_i^2 = \|T\|^2 + \sum_{i=1}^{r} \sigma_i^2 \rightarrow \|T\| \leq \|T\|_{\text{Frob}}$.

[We could also just recall that the trace of $A^*A$ is the sum of the eigenvalues.]
Suppose we have an invertible matrix $A \in M_{nn}$ and we want to solve the equation $Ax = b$, but we only have an estimate of $b$.

Let $\Delta b$ denote a small perturbation accounting for the error, so instead of solving $Ax = b$ for an exact solution $x$ (because $b$ is not known exactly) we solve $Ay = b + \Delta b$ for an approximate solution $y$.

Let $\Delta x = y - x$. So $A(x + \Delta x) = b + \Delta b \rightarrow A\Delta x = \Delta b \rightarrow \Delta x = A^{-1}\Delta b$. Then

$$\|\Delta x\| \|x\| = \|A^{-1}\Delta b\| \|b\| = \|A^{-1}\Delta b\| \|Ax\| \leq \|A^{-1}\|\|\Delta b\| \|A\|\|x\| = \|A^{-1}\|\|A\| \|\Delta b\| \|b\|$$

Note that $\|A\|$ and $\|A^{-1}\|$ denote operator norms $[A, A^{-1}$ are matrices of operators in $L(\mathbb{R}^n)$ with respect to the standard (orthonormal) basis of $\mathbb{R}^n]$.

So, $\|\Delta x\| \|x\|$ is the relative error in the solution, is bounded by the product of the (real) number $C = \|A^{-1}\|\|A\|$ and $\|\Delta b\| \|b\|$, which is the relative error in the estimate of $b$.

This motivates the following definition: Let $A \in M_{nn}$ be invertible. The condition number of $A$ is $C = \|A^{-1}\|\|A\|$. The condition number tells you by how much relative error in the estimate of $b$ is amplified into the relative error in the approximate solution.

How does the condition number relate to singular values?

Let $A \in M_{nn}$ be invertible. Let $\sigma_1, \sigma_2, ..., \sigma_n$ be the singular values of $A$ in non-increasing order, so $\sigma_1$ is the largest and $\sigma_n$ is the smallest. We already know that $\|A\| = \sigma_1$.

In particular, $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ are the eigenvalues of $A^*A$. Let $B = \{v_1, v_2, ..., v_n\}$ be an orthonormal basis of $F^n$ consisting of eigenvectors of $A^*A$.

(ELFY) $A^*A$ is invertible, $(A^*)^{-1} = (A^{-1})^*$, and $(A^*A)^{-1} = A^{-1}(A^{-1})^*$.

Since $A^*A$ is invertible, none of the singular values can be 0 (recall: a square matrix is invertible iff 0 is not an eigenvalue).
Now let $C = \{w_i = \frac{1}{\sigma_i} Av_i | i = 1, 2, \ldots, n\}$. Of course these bases $B, C$ are the orthonormal bases in a singular value decomposition of $A$.

Then for $i = 1, 2, \ldots, n$, we have $A^{-1}\sigma_i w_i = v_i$ and thus

$$A^* Av_i = \sigma_i^2 v_i \rightarrow \frac{1}{\sigma_i} Av_i = (A^{-1})^* \sigma_i v_i \rightarrow w_i = (A^{-1})^* A^{-1} \sigma_i^2 w_i.$$ 

Thus for $i = 1, 2, \ldots, n$

$$(A^{-1})^* A^{-1} w_i = \frac{1}{\sigma_i^2} w_i,$$

and thereby the singular values of $A^{-1}$ are $\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_n}$. So $\|A^{-1}\| = \frac{1}{\sigma_n}$ as that would be the largest singular value of $A^{-1}$.

In conclusion, the condition number of an invertible matrix $A \in M_{nn}$ is $C = \|A^{-1}\| \|A\| = s_1 \overline{s_n}$, where $s_1$ is the largest singular value of $A$ and $s_n$ is the smallest singular value of $A$. In other words, its the ratio of the largest to smallest singular value.

(ELFY) If we pick $b = w_1$ and $\Delta b = cw_n$ for $c \in F$ then the inequality $(\ast)$ on the last page becomes an equality.

**Suggested Homework:** 4.1.

We will not cover the last two sections of this chapter.