Matrix Representation

We first discuss how to represent some linear transformations $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{W}$ by matrices. Think of $\mathbf{V}$ and $\mathbf{W}$ as $\mathbb{R}^n$ or $\mathbb{C}^n$. It is a process akin to representing vectors by coordinates. Recall that if $v \in \mathbf{V}$ and $\mathcal{B} = v_1, \ldots, v_n$ is a basis of $\mathbf{V}$, we have a unique expression

$$v = x_1 v_1 + \cdots + x_n v_n.$$ 

We say that the $x_i$ are the coordinates of $v$ with respect to $\mathcal{B}$. We write as

$$[v]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$ 

If $\mathcal{C} = \{w_1, \ldots, w_m\}$ is a basis of $\mathbf{W}$, we would like to find the coordinates of $\mathbf{T}(v)$ in the basis $\mathcal{C}$

$$[\mathbf{T}(v)]_{\mathcal{C}} = \begin{bmatrix} ? \end{bmatrix}.$$
In other words, if $\mathbf{v} = x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n$, 

$$\mathbf{T}(\mathbf{v}) = y_1 \mathbf{w}_1 + \cdots + y_m \mathbf{w}_m,$$

we want to describe the $y_i$ in terms of the $x_j$. The process will be called a matrix representation. It comes about as follows:

$$\sum y_i w_i = \mathbf{T}(\sum x_j \mathbf{v}_j) = \sum x_j \mathbf{T}(\mathbf{v}_j)$$

Thus if we have the coordinates of the $\mathbf{T}(\mathbf{v}_j)$,

$$\mathbf{T}(\mathbf{v}_j) = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

we have

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \sum x_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$
More pictorially

\[
[T(v)]_C = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [T]_B^C \cdot [v]_B
\]

The \( n \times m \) matrix \([T]_B^C\)

is called the matrix representation of \( T \) relative to the bases \( B \) of \( V \) and \( C \) of \( W \).
Quickly: Once bases \( v_1, \ldots, v_n \) and \( w_1, \ldots, w_m \) have been chosen, \( T \) is represented by

\[
\begin{bmatrix}
a_{ij}
\end{bmatrix}
\]

where the entries come from

\[ T(v_j) = \sum_{i=1}^{m} a_{ij}w_i. \]
Example

Recall the transpose operation on a square matrix $A$: if $a_{ij}$ is the $(i, j)$-entry of $A$, the $(i, j)$-entry of $A^t$ is $a_{ji}$. This is a linear transformation $T$ on the space $M_n(F)$:

$$(A + B)^t = A^t + B^t, \quad (cA)^t = cA^t.$$ 

Let us find its matrix representation on $M_2(F)$. This space has the basis

$$v_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad v_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$
Since

\[ T(v_1) = v_1, \quad T(v_2) = v_3, \quad T(v_3) = v_2, \quad T(v_4) = v_4, \]

the matrix representation of transposing is

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
Let $\mathbb{R}_3[x]$ be the space of real polynomials of degree at most 3 and $T$ the differentiation operator.

A basis here are the polynomials $1, x, x^2, x^3$. The corresponding matrix representation is

$$
\begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 3 \\
0 & 0 & 0 & 0
\end{pmatrix}
$$
Consider the following differential equations (or systems of)

\[ y' = ay, \quad a \in \mathbb{R} \]

\[ y'' + ay' + by = 0, \quad a, b \in \mathbb{R} \]

\[
\begin{bmatrix}
  y_1' \\
  y_2'
\end{bmatrix}
= 
\begin{bmatrix}
  10y_1 + 3y_2 \\
  3y_1 + 2y_2
\end{bmatrix}
\]

**Question:** What are their resemblances? Which ones can we solve directly?

They are equations, or systems, of linear differential equations with constant coefficients.
The first equation, \( y' = ay \), is the easiest to deal with: \( y = ce^{at} \) is the general solution.

We will argue that the others, with a formulation using vectors and matrices, have the same kind of solution. Let us do the last one first. Set

\[
Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad Y' = \begin{bmatrix} y'_1 \\ y'_2 \end{bmatrix}, \quad A = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}
\]

Now observe:

\[ Y' = AY. \]

**Question:** This looks like \( y' = ay \), which has \( y = ce^{at} \) for solution. You should be tempted to expect the solution to be

\[ Y = Ce^{tA}. \]

What is \( e^{tA} \), the **exponential** of the matrix \( tA \)? What could it be?
Let us turn to the second order D.E.

\[ y'' + ay' + by = 0 \]

If we set \( z_1 = y \) and \( z_2 = y' = z_1' \),
\[ z_2' = y'' = -ay' - by = -bz_1 - az_2 \]
which can be written in matrix formulation as

\[
\begin{bmatrix}
  z_1 \\
  z_2
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
  z_1' \\
  z_2'
\end{bmatrix}
\]

\[ \mathbf{A} = \begin{bmatrix} 0 & -b \\ 1 & -a \end{bmatrix} \]

We get

\[ \mathbf{Z}' = \mathbf{A}\mathbf{Z} \]

as above \( \mathbf{Z} = \mathbf{C} e^{t\mathbf{A}} \) if we could make sense of then exponential of a matrix.
The function $e^x$ has a power series expansion

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + \cdots$$

If we replace $x$ by the square matrix $A$ (and 1 by $I$), we get

$$e^A = I + A + \frac{A^2}{2} + \cdots + \frac{A^n}{n!} + \cdots,$$

We just must make sure that a theory of series of makes sense. The answer will be sure. Think about the adjustments to be made.
Just for fun let us calculate the exponential of $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

$A^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, $A^3 = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$, $A^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$

$e^A = \begin{bmatrix} e & e \\ 0 & e \end{bmatrix}$
Convergence of $e^A$

That

$$e^A = I + A + \frac{A^2}{2} + \cdots + \frac{A^n}{n!} + \cdots$$

makes sense is due to the power of $n!$:

Suppose $A = [a_{ij}]$ is $m \times m$ and that the absolute value of its entries $|a_{ij}| \leq r$. This implies that the entries of $A^2$

$$\left| \sum_{k=1}^{m} a_{ik}a_{kj} \right| \leq mr^2$$

Similarly one finds that the entries of $A^n$ are bounded by

$$m^{n-1}r^n$$
This implies that the series in any entry of $e^A$ is bounded by the series

$$\sum_{n=0}^{\infty} \frac{m^{n-1} r^n}{n!}$$

that is convergent [e.g. use ratio test].

This proves $e^A$ makes sense since the series in each of its entries is absolutely convergent.
Let us show a long application:

$$\det(e^A) = e^{\text{Trace}(A)}$$

This is obvious if $A$ is a diagonal matrix,

$$A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}, \quad e^A = \begin{bmatrix} e^a & 0 & 0 \\ 0 & e^b & 0 \\ 0 & 0 & e^c \end{bmatrix}, \quad \det(e^A) = e^{a+b+c},$$

but in general...
Let $\mathbf{V}$ be a finite dimensional vector space and 

$$\mathbf{T} : \mathbf{V} \rightarrow \mathbf{V}$$

a linear transformation. 

**Question:** Is there a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of $\mathbf{V}$ so that the matrix representation

$$[\mathbf{T}]_{\mathcal{B}}$$

is as ‘simple’ [e.g. with plenty of 0’s] as possible? 

**Answer:** Well... but for the most ‘interesting’ matrices the answer is YES.
Invariant subspace

Let \( V \) be a finite dimensional vector space and \( T : V \to V \) a linear transformation. If \( W \subset V \) is a subspace, it is of interest to know whether for \( w \in W \) its image \( T(w) \in W \). Clearly this will not happen often.

**Definition**

W is a **T-invariant subspace** if \( T(W) \subset W \). That is, the restriction of (the function) \( T \) to \( W \) is a linear transformation of it. We denote the restriction of \( T \) to \( W \) by \( T_W \).
Let us see what this implies for the matrix representation of $T$. Let $B = \{w_1, \ldots, w_r\}$ be a basis of $W$, and complete it to a basis of $V$

$$A = \{w_1, \ldots, w_r, v_{r+1}, \ldots, v_n\}.$$ 

Since $T(w_i) \in W$, it is a linear combination of the first $r$ vectors, the first $r$ columns of the matrix is

$$[T]_A = \begin{bmatrix} [T_{w_i}]_B & \ast & \cdots & \ast \\ O_{(n-r)\times r} & \ast & \cdots & \ast \\ \end{bmatrix} = \begin{bmatrix} a & b & \ast & \cdots & \ast \\ c & d & \ast & \cdots & \ast \\ 0 & 0 & \ast & \cdots & \ast \\ 0 & 0 & \ast & \cdots & \ast \\ \end{bmatrix}$$
Blocks

Suppose $T$ is a L.T. of vector space $V$ with a basis $\mathcal{A} = v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$. Suppose $T(v_i)$ for $i \leq r$, is a linear combination of the first $r$ basis vectors, and $T(v_i)$ for $i > r$, is a linear combination of the last $n - r$ basis vectors.

Claim: The matrix representation has the block format

$$[T]_{\mathcal{A}} = \begin{bmatrix}
\begin{array}{c}
 r \times r \\
 O
\end{array} &
\begin{array}{c}
 O \\
(n-r) \times (n-r)
\end{array}
\end{bmatrix}$$

This can be refined to more than two blocks. The extreme case is when all blocks are $1 \times 1$. The representation is then said to be diagonal.
The extreme case of an invariant subspace is one of the top 5 notions of L.A.:

**Definition**

An **eigenvector** of the linear transformation $T$ is a **nonzero** vector $v$ such that

$$T(v) = \lambda \cdot v.$$ 

The scalar $\lambda$ is called the (corresponding) **eigenvalue**.

**Means:** The line $Fv$ is an invariant subspace of $T$. Note that $v$ must be **nonzero**, but that $\lambda$ could be zero. Observe who comes first: **eigenvector** $\rightarrow$ **eigenvalue**.
To keep in mind:

\[ \mathbf{v} \neq \mathbf{O}, \quad \mathbf{T}(\mathbf{v}) = \lambda \mathbf{v} \]

Note: Any nonzero multiple of \( \mathbf{v} \) is also an eigenvector [with the same eigenvalue]

\[ a\mathbf{v} \neq 0 \quad \mathbf{T}(a\mathbf{v}) = a\mathbf{T}(\mathbf{v}) = a\lambda \mathbf{v} = \lambda (a\mathbf{v}) \]

The subspace spanned by \( \mathbf{v} \) is \textbf{invariant} under \( \mathbf{T} \)
Examples

- One of the most important L.T. of Mathematics is $T := \frac{d}{dt}$. (On the appropriate V.S.) Its eigenvectors are
  \[ \frac{d}{dt}(f(t)) = \lambda \cdot f(t), \]
  that is $f(t) = e^{\lambda t}$ and its nonzero scalar multiples $ce^{\lambda t}$.
- Let $T$ be the identity L.T. $I$. Then any nonzero vector is a eigenvector. Same property for the [null] $O$ mapping.
• For an angle $0 < \alpha < \pi$, let

$$\mathbf{T}(x, y) = (x \cos \alpha + y \sin \alpha, -x \sin \alpha + y \cos \alpha)$$

This is a rotation in the plane by $\alpha$ degrees. Clearly there is no nonzero vector $v$ in the real plane $\mathbb{R}^2$ that is aligned with $\mathbf{T}(v)$.

• Let $\mathbf{T}$ be the L.T.

$$\begin{bmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 0
\end{bmatrix}$$

Its eigenvectors are (and their nonzero multiples)

$$\mathbf{T}(i) = 1 \cdot i, \quad \mathbf{T}(j) = 2 \cdot j, \quad \mathbf{T}(k) = 0 \cdot k$$
If $T$ is a linear transformation of $F^2$ with a matrix representation

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

we know that

$$A^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Thus, if

$$A(v) = \lambda v, \quad v \neq 0$$

so $\lambda = 0$ since $v \neq O$. 
Let $V$ be the vector space of all $n \times n$ real matrices, and let $T$ be the transformation

$$ T(A) = A^t $$

$T$ is a linear transformation. If $A \neq O$ is one of its eigenvectors,

$$ A^t = \lambda A $$

So, transposing again we get

$$ A = (A^t)^t = \lambda A^t = \lambda^2 A $$

$$ (\lambda^2 - 1)A = O $$

This means that $\lambda = \pm 1$
- If $\lambda = 1$, $A$ is symmetric
- If $\lambda = -1$, $A$ is skew-symmetric
Given a $n$-by-$n$ matrix $A$ [usually representing some linear transformation $T$], how are the eigenvectors to be found? Although the eigenvalues come after the eigenvectors, in some approaches they will appear first. Look at the following analysis: $A\mathbf{v} = \lambda \mathbf{v}$, for $\mathbf{v} \neq \mathbf{0}$ means that

$$(A - \lambda I_n)\mathbf{v} = \mathbf{0},$$

Conclusion: $\mathbf{v}$ is a nonzero vector of the nullspace of $A - \lambda I_n$ and therefore $\text{rank}(A - \lambda I_n) < n$. This in turn means that

$$\det(A - \lambda I_n) = 0.$$
The **characteristic polynomial** of the $n$-by-$n$ matrix $A = [a_{ij}]$ is the polynomial

$$p(x) = \det(A - xI_n) = \det \begin{bmatrix} a_{11} - x & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} - x \end{bmatrix}.$$ 

The equation $p(x) = 0$ is called the **characteristic equation**.

Observe that $\det(A - xI_n)$ is a polynomial of degree $n$,

$$\det(A - xI_n) = (-1)^n x^n + c_{n-1} x^{n-1} + \cdots + c_0.$$
The characteristic polynomial of $\mathbf{A} = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}$ is

$$\det \begin{bmatrix} 10 - x & 3 \\ 3 & 2 - x \end{bmatrix} = (10 - x)(2 - x) - 9 = x^2 - 12x + 11$$

Its roots are

$$\lambda = \frac{12 \pm \sqrt{12^2 - 4 \times 11}}{2} = 6 \pm 5$$
With the eigenvalues in hand we solve for the eigenvectors.

$\lambda = 11$: Will determine the nullspace of $A - 11I_2$

\[
\begin{bmatrix}
10 - 11 & 3 \\
3 & 2 - 11
\end{bmatrix}
\rightarrow
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\]

$v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

$\lambda = 1$: Will determine the nullspace of $A - I_2$

\[
\begin{bmatrix}
10 - 1 & 3 \\
3 & 2 - 1
\end{bmatrix}
\rightarrow
\begin{bmatrix}
3 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

$v_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$
Let us Verify that it will work out for any real symmetric matrix

\[ A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \]

The characteristic polynomial is

\[
\det \begin{bmatrix} a-x & b \\ b & c-x \end{bmatrix} = (a-x)(c-x)-b^2 = x^2-(a+c)x+ac-b^2,
\]

whose roots are

\[
\lambda = \frac{a+c \pm \sqrt{(a+c)^2-4(ac-b^2)}}{2}
\]

Incredibly (?) the quantity under the sign is \((a - c)^2 + 4b^2 \geq 0\), so either there are two distinct real roots or \(a = c, b = 0\). In both cases the matrix is diagonalizable.
A different kind is the rotation $R_\alpha$ by $\alpha$ degrees in the plane $\mathbb{R}^2$:

$$
\begin{bmatrix}
\cos \alpha & -\sin \alpha \\
\sin \alpha & \cos \alpha
\end{bmatrix}
.$$  

Its characteristic polynomial is

$$
\det \begin{bmatrix}
\cos \alpha - x & -\sin \alpha \\
\sin \alpha & \cos \alpha - x
\end{bmatrix} = (\cos \alpha - x)^2 + \sin^2 \alpha = x^2 - (2 \cos \alpha) x + 1.
$$

Its roots are

$$
\lambda = \frac{2 \cos \alpha \pm \sqrt{4 \cos^2 \alpha - 4}}{2},
$$

which is not real unless $\alpha = 0, \pi$. 

We already know that rotations $0 < \alpha < \pi$ have no real eigenvalues. Let us try $\alpha = \pi/2$ anyway: $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. The characteristic polynomial is $x^2 + 1$, so the (complex) eigenvalues are $\lambda = \pm i$.

$\lambda = i$: Will determine the nullspace of $A - iI_2$

$$\begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \rightarrow \begin{bmatrix} -i & 1 \\ 0 & 0 \end{bmatrix}, \quad v_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$$

$\lambda = -i$: Will determine the nullspace of $A + iI_2$

$$\begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \rightarrow \begin{bmatrix} i & 1 \\ 0 & 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$$
**Proposition**

Let $A$ be a $n$-by-$n$ matrix over the field $F$. A scalar $\lambda \in F$ is an eigenvalue for some eigenvector $v \in F^n$ iff $\lambda$ is a root of the polynomial $\det(A - xI_n)$.

**Proof.**

We have already observed that if $Av = \lambda v$, $v \neq 0$, then $\lambda$ is a root of the char polynomial. Conversely, if $\det(A - \lambda I_n) = 0$, then $\text{rank}(A - \lambda I_n) < n$. This implies, by the dimension formula, that the nullspace of $A - \lambda I_n \neq O$. Any nonzero vector in this nullspace will satisfy

$$Av = \lambda v.$$
Corollary

The number of distinct eigenvalues of the n-by-n matrix $A$ is at most $n$. (The set of eigenvalues of a matrix—or of a linear transformation is called its spectrum).
It seems that we have only defined the characteristic polynomial for matrices. Suppose $T$ is a L.T. If we have two bases $A, B$ of the vector space, we have two representations

$$A = [T]_A, \quad B = [T]_B$$

and therefore we have, apparently, two possibly different polynomials

$$\det(A - xI_n), \quad \det(B - xI_n).$$

But we proved that $A$ and $B$ are related: There is an invertible matrix $P$ such that $B = P^{-1}AP$. Now observe
\[
\begin{align*}
\det(B - xI_n) & = \det(P^{-1}AP - xI_n) = \det(P^{-1}AP - P^{-1}xI_nP) \\
& = \det(P^{-1})(A - xI_n)P) \\
& = \det(P^{-1}) \det(A - xI_n) \det(P) \\
& = \det(A - xI_n)
\end{align*}
\]

**Conclusion:** The characteristic polynomial is the same for all representations of \( T \).
Eigenspaces

Definition

If $\lambda$ is an eigenvalue of $A$, the nullspace of $A - \lambda I_n$, denoted by $E_\lambda$, is called the eigenspace associated to $\lambda$.

Observe that $E_\lambda$ is invariant under $A$: If $v \in E_\lambda$ then $Av \in E_\lambda$. 
If $f(x) = a_nx^n + \cdots + a_0$ is a polynomial of degree $n$, with coefficients in the field $\mathbb{F}$ a root is a scalar $r$ such that $f(r) = 0$. It is a hard problem to find $r$.

**Proposition**

If $f(x)$ and $g(x)$ are two polynomials, then there exist polynomials $q(x)$ and $r(x)$ where

$$f(x) = q(x)g(x) + r(x),$$

where $r(x) = 0$ or degree $r(x) < \text{degree } g(x)$.

$q(x)$ is called the **quotient**, and $r(x)$ the **remainder** of the division of $f(x)$ by $g(x)$. They are found by the **long division** algorithm.
Corollary

If \( r \) is a root of the nonzero polynomial \( f(x) \), then
\[
  f(x) = (x - r)q(x),
\]
where \( \deg q(x) = \deg f(x) - 1 \). As a consequence, a polynomial \( f(x) \) of degree \( n \) has at most \( n \) roots.

Proof.

Any other root \( s \) of \( f(x) \) satisfies
\[
  f(s) = q(s)(s - r) = 0,
\]
so \( q(s) = 0 \) since \( s - r \neq 0 \).
Algebraic multiplicity of a root

If \( f(x) = a_n x^n + \cdots + a_0 \) is a nonzero polynomial and \( r \) is one of its roots,

\[
f(x) = (x - r)g(x).
\]

It may occur that \( r \) is a root of \( g(x) \), \( g(x) = (x - r)h(x) \). As the degrees of the quotients decrease, we eventually have

\[
f(x) = (x - r)^s q(x), \quad q(r) \neq 0.
\]

**Definition**

We say that \( r \) is a root of \( f(x) \) of **order** or **multiplicity** \( s \).
Let $\lambda$ be an eigenvalue of the matrix $\mathbf{A}$. There are two notions of multiplicity associated to $\lambda$:

- If $\lambda$ is a root of order $s$ of the characteristic polynomial $\det(\mathbf{A} - x\mathbf{I}_n)$, we say that $\lambda$ has **algebraic multiplicity** $s$.
- If the eigenspace $E_\lambda$ has dimension $t$, we say that $\lambda$ has **geometric multiplicity** $t$. 
Proposition

For any eigenvalue $\lambda$ of a matrix $A$,

\[ \text{algebraic multiplicity} \geq \text{geometric multiplicity}. \]

Proof.

Assume $v_1, \ldots, v_t$ is a basis of $E_\lambda$, and we use it as the beginning of a basis for the whole vector space, the representation of the L.T. has the block format

\[
\begin{bmatrix}
\lambda I_t & B \\
O & C
\end{bmatrix},
\]

\[ \det(A - xI_n) = (\lambda - x)^t \det(C - xI_{n-t}). \]
Properties of eigenvalues

Let \( A \) be a square matrix.

1. If \( \lambda \) is an eigenvalue of \( A \), then \( \lambda^2 \) is an eigenvalue of \( A^2 \):

\[
A^2(v) = A(A(v)) = A(\lambda v) = \lambda A(v) = \lambda \lambda v = \lambda^2 v.
\]

2. More generally, if \( g(x) \) is a polynomial (e.g. \( x^2 - 2x + 1 \)) then

\[
g(A)(v) = g(\lambda)v, \quad (A^2 - 2A + I)(v) = (\lambda^2 - 2\lambda + 1)(v).
\]

3. If \( A \) is invertible, \( A^{-1}(v) = \frac{1}{\lambda}v \).
1. If \( p(x) = \det(A - xI_n) = (-1)^n x^n + \cdots + a_0 \) is the characteristic polynomial of \( A \), then \( a_0 = \det(A) \). Plug in \( x = 0 \) in \( p(x) \).

2. If \( \lambda_1, \ldots, \lambda_n \) are the eigenvalues of \( A \), then

\[
\det(A) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n.
\]

In the decomposition of \( p(x) \),

\[
p(x) = (-1)^n (x - \lambda_1) \cdots (x - \lambda_n),
\]

plug in \( x = 0 \) and use the observation above.
If the field is the complex number filed $\mathbb{C}$, any polynomial $f(x) \in \mathbb{C}[x]$ factors completely

$$f(x) = a_n(x - r_1) \cdots (x - r_n)$$

As a consequence, the eigenvalues of a complex matrix always exist in the field.

If $A$ is a real matrix, its characteristic polynomial $p(x) = \det(A - xI_n)$ is a real polynomial and always have a full set $\lambda_1, \ldots, \lambda_n$ of complex eigenvalues, some of which may be real.
If \( \lambda = a + bi \), is a complex root of \( f(x) \), \( f(\lambda) = 0 \), observe that

\[
f(a + bi) = 0 \Rightarrow f(a - bi) = 0,
\]
because all coefficients of \( f(x) \) are real. Let us explain: Say

\[
7(a + bi)^3 - 2(a + bi)^2 + 117(a + bi) + \pi = 0.
\]

Complex conjugation, \( a + bi \rightarrow \overline{a + bi} = a - bi \) has the property: \( \overline{z_1z_2} = \overline{z_1} \cdot \overline{z_2} \). But if \( z_1 \), say, is real (like the coefficients of the polynomial), \( \overline{z_1} = z_1 \), so they are not affected by changing all \( a + bi \) into \( a - bi \). So if one is a root, so will be the other.

Thus the complex conjugate \( a - bi \) of an eigenvalue \( a + bi \) is also an eigenvalue: So complex eigenvalues of a real matrix occur in pairs.
Groups

Let $G$ be a finite group. There are many injective homomorphisms

$$\varphi : G \rightarrow GL_n(\mathbb{C})$$

Thus we have many ways to view $G$ as a group of linear transformations. It helps a lot to know

**Theorem**

*Every $T \in G$ is diagonalizable.*

You should ask how come, when being diagonalizable is kind of dicey.
Let $\mathbf{T}$ be a L.T. (or matrix). Suppose there is a basis made up of eigenvectors, say $\mathcal{B} = \{v_1, \ldots, v_n\}$, $\mathbf{T}(v_i) = \lambda_i v_i$. The corresponding matrix representation is

$$[\mathbf{T}]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

This is not always possible: Let $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ whose characteristic polynomial is $x^2$. There is just one eigenvalue, $\lambda = 0$. But the corresponding eigenspace $E_0$ has for basis $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. We do not have a basis of eigenvectors, so $\mathbf{A}$ is not diagonalizable.
Let us explore what is needed to have a basis of eigenvectors.

**Proposition**

Let $T$ be a linear transformation and let $v_1, \ldots, v_r$ be a set of eigenvectors of $T$, associated to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then the $v_i$ are linearly independent.
**Proof.** Suppose $c_1 v_1 + \cdots + c_r v_r = O$. Using induction on $r$, we are going to show that all $c_i = 0$. We are going to multiply the equation by $\lambda_1$ and apply $T$ to it to obtain the following two equations:

$$\lambda_1 (c_1 v_1 + \cdots + c_r v_r) = \lambda_1 c_1 v_1 + \cdots + \lambda_1 c_r v_r = 0$$
$$T(c_1 v_1 + \cdots + c_r v_r) = \lambda_1 c_1 v_1 + \cdots + \lambda_r c_r v_r = 0$$

If we subtract one from the other we get the shorter equation,

$$(\lambda_2 - \lambda_1) c_2 v_2 + \cdots + (\lambda_r - \lambda_1) c_r v_r = 0$$

By the induction hypothesis, all $c_i(\lambda_i - \lambda_1) = 0$, for $i > 1$. Since $\lambda_i \neq \lambda_1$, this means $c_i = 0$ for $i > 1$. Finally, since $v_1 \neq 0$ this will imply $c_1 = 0$ as well.
Let $\lambda_1, \ldots, \lambda_r$ be the set of eigenvalues of $T$, and let $E_{\lambda_1}, \ldots, E_{\lambda_r}$ be the corresponding set of eigenspaces. For each of these we pick a basis $B_i$. For simplicity, take 3 eigenvalues and assume the bases chosen for the 3 eigenspaces are

$$\{ u_1, u_2, u_3 \}, \{ v_1, v_2 \}, \{ w_1, w_2 \}$$

Claim: These 7 vectors are linearly independent. Suppose

$$a_1 u_1 + a_2 u_2 + a_3 u_3 + b_1 v_1 + b_2 v_2 + c_1 w_1 + c_2 w_2 = 0,$$

which we write as $1 \cdot u + 1 \cdot v + 1 \cdot w = 0$. Note that if $u \neq 0$ it is an eigenvector (and $v$ and $w$ as well), by the Proposition, $u = v = w = 0$, and then that $a_1 = \cdots = c_2 = 0$, by the linear independence of the respective bases.
Theorem

Let \( \mathbf{A} \) be a \( n \times n \) matrix with \( n \) eigenvalues (maybe repeated). Then \( \mathbf{A} \) is diagonalizable iff for every eigenvalue its geometric multiplicity is equal to its algebraic multiplicity.

Proof. Let \( \lambda_1, \ldots, \lambda_r \) be the set of DISTINCT eigenvalues of \( \mathbf{A} \), and let \( E_{\lambda_1}, \ldots, E_{\lambda_r} \) be the corresponding set of eigenspaces. We have the equalities

\[
\sum_i \text{geom. mult. of } \lambda_i = \sum_i \dim E_{\lambda_i}
\]

\[
\sum_i \text{alg. mult. of } \lambda_i = n.
\]

Since \text{alg. mult. of } \lambda_i \geq \text{geom. mult. of } \lambda_i, \text{ if equality for each } i \text{ holds, the previous discussion shows that we can have a basis of eigenvectors by collecting bases in the } E_{\lambda_i}. \text{ The converse is clear.}
Corollary
Let $A$ be a $n$-by-$n$ matrix with $n$ distinct eigenvalues. Then $A$ is diagonalizable.

Theorem
Let $A$ be a $n$-by-$n$ matrix. $A$ is invertible iff $\lambda = 0$ is not an eigenvalue.

Proof.
$A$ is invertible iff it is one-one: $A(v) \neq 0 \cdot v$ if $v \neq O$.  

\[
\begin{align*}
A(v) &= 0 \cdot v \\
\Rightarrow & \quad \text{(for } v \neq O) \\
A(v) &\neq 0 \\
\end{align*}
\]
Let $A$ be a $n$-by-$n$ matrix and assume $B = \{v_1, \ldots, v_n\}$ is a basis made up of its eigenvectors, $A(v_i) = \lambda_i v_i$. The matrix

$$P = [v_1 | \cdots | v_n]$$

is invertible since the $v_i$ form a basis. **Claim:**

$$P^{-1}AP = D = \begin{bmatrix}
\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_n
\end{bmatrix}$$

To prove we apply $D$ to the standard basis $e_1, \ldots, e_n$. Note that $P(e_1) = v_1$. For instance

$$D(e_1) = P^{-1}(A(P(e_1))) = P^{-1}(A(v_1)) = P^{-1}(\lambda_1 v_1) = \lambda_1 P^{-1}(v_1) = \lambda_1 e_1$$
Note that if $A$ is diagonalizable, that is there is an invertible matrix $P$ such that $P^{-1}AP = D$ (= diagonal), a host of related matrices are also diagonalizable:

1. Any power of $A$ is diagonalizable (let us do square):

$$D^2 = (P^{-1}AP)(P^{-1}AP) = P^{-1}A P P^{-1} = P^{-1}A^2P$$

and certainly $D^2$ is diagonal.

2. If $A$ is invertible [and diagonalizable!] its inverse $A^{-1}$ is also diagonalizable:

$$D^{-1} = (P^{-1}AP)^{-1} = P^{-1}A^{-1}(P^{-1})^{-1} = P^{-1}A^{-1}P$$

3. If $g(x)$ is any polynomial and $A$ is diagonalizable, then $g(A)$ is diagonalizable (check).
Diagonalization Summary

Let $A$ be a $n$-by-$n$ matrix for which we want to find a possible diagonalization.

1. Find the characteristic polynomial $p(x) = \det(A - xI_n)$. Rating: **Routine**, if at times long.

2. Decompose $p(x)$ and collect factors

   $$p(x) = (-1)^n(x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$$

   Rating: **Very Hard**

3. For each $\lambda_i$ find $\dim E_{\lambda_i}$ and check it is $m_i$. Rating: **Gaussian elim**

Comment: This is kind of vague. We need predictions. That is: Guarantees that certain kinds of matrices are diagonalizable.
Example: Let $A$ be the real matrix

$$
\begin{bmatrix}
  2 & 1 & 1 \\
  0 & 1 & 2 \\
  0 & 0 & c
\end{bmatrix},
$$

where $c$ is some number.

(a) What are the eigenvalues of $A$?

(b) If $c \neq 1, 2$, why is $A$ diagonalizable? What happens when $c = 1$ or $c = 2$?

Answer: (a) The characteristic polynomial is

$$
\text{det}(A - xI_3) = (2 - x)(1 - x)(c - x),
$$

whose roots are the eigenvalues: $1, 2, c$.

(b) If $c \neq 1, 2$, there are [automatically] 3 independent eigenvectors and therefore the matrix is diagonalizable.
If \( c = 1 \) or \( c = 2 \), it may go either way [diagonalizable or not] so we must check further to see whether the geometric multiplicities are equal or not to the algebraic multiplicities. For \( c = 1 \): The nullspace of \( A - I_3 \)

\[
\begin{bmatrix}
1 & 1 & 1 \\
0 & 0 & 2 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

is generated by

\[
\begin{bmatrix}
-1 \\
1 \\
0 \\
\end{bmatrix}
\]

and \( A \) is not diagonalizable.

Doing likewise for \( c = 2 \) will again show that \( A \) is not diagonalizable.
Example:

Given the real matrix

\[
A = \begin{bmatrix}
2 & 0 & 3 \\
0 & 2 & 0 \\
3 & 0 & 5
\end{bmatrix}
\]

\[
A - xI_3 = \begin{bmatrix}
2 - x & 0 & 3 \\
0 & 2 - x & 0 \\
3 & 0 & 5 - x
\end{bmatrix}
\]

(a) Find its characteristic polynomial.
(b) Find its eigenvalues.
(c) Explain why \( A \) is diagonalizable. [You do not have to find the eigenvectors to answer.]

Answer: (a) To find \( \det(A - xI_3) \), we expand along the second column

\[
\det(A - xI_3) = (2 - x)((2 - x)(5 - x) - 9) = (2 - x)(x^2 - 7x + 1).
\]
(b) Use the quadratic formula to find the roots of the factor 
\[ x^2 - 7x + 1: \]
\[
= \frac{7 \pm \sqrt{49 - 4}}{2} = \frac{7 \pm 3\sqrt{5}}{2}
\]
Together with 2 these roots are the eigenvalues.

(c) Since the 3 eigenvalues are distinct, we have a basis of eigenvectors for \( \mathbb{R}^3 \) and \( A \) is diagonalizable.
Let $\lambda$ be an eigenvalue of the matrix $A$: $Av = \lambda v$. To find $v \neq 0$ we find the nullspace of $A - \lambda I_n$.

Suppose a mistake was made and instead of $\lambda$ we have $\lambda + \epsilon$. If this value is not an eigenvalue the nullspace of

$$A - (\lambda + \epsilon)I_n$$

is $O$, not a vector ‘close’ to $v$. What to do?
Some stability

**Question:** Assume $A$ admits a basis of eigenvectors. How can we find one, or more eigenvectors, if we cannot solve the characteristic equation? Here is a popular technique. Let $u \in \mathbb{R}^n$ picked at random [?]. We know that

$$u = u_1 + u_2 + \cdots + u_r, \quad Au_i = \lambda_i u_i$$

where the $u_i$ belong to different eigenspaces. Of course, the right hand of this equality is invisible to us. Let us assume $|\lambda_1| > |\lambda_i|, \quad i > 1$. Observe what happens when we apply $A$ repeatedly to $u$:

$$A^n(u) = \lambda_1^n u_1 + \lambda_2^n u_2 + \cdots + \lambda_r^n u_r$$

The growth in the coordinates of $A^n(u)$ is coming from $\lambda_1^n u_1$. 
If we compare the two vectors

\[ A^n(u) = \lambda_1^n u_1 + \lambda_2^n u_2 + \cdots + \lambda_r^n u_r \]

\[ A^{n+1}(u) = \lambda_1^{n+1} u_1 + \lambda_2^{n+1} u_2 + \cdots + \lambda_r^{n+1} u_r \]

It will follow that

\[ \lim_{n \to \infty} \frac{\|A^{n+1}(u)\|}{\|A^n(u)\|} = |\lambda_1|, \]

more precisely: If we set \( v_n = \frac{A^n(u)}{\|A^n(u)\|} \), then

\[ A(v_n) \approx \lambda_1 v_n, \quad n \gg 0. \]
Let us re-visit a problem and solve it in two different ways: It is the system of differential equations

\[
\begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix}, \quad \begin{bmatrix}
  y_1' \\
  y_2'
\end{bmatrix}, \quad \begin{bmatrix}
  10 & 3 \\
  3 & 2
\end{bmatrix}, \quad \begin{bmatrix}
  y_1' \\
  y_2'
\end{bmatrix} = \begin{bmatrix}
  10 & 3 \\
  3 & 2
\end{bmatrix} \begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix}.
\]

Earlier we found the eigenvalues and bases for the eigenspaces:

\[
\lambda = 11: \quad v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \lambda = 1: \quad v_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}
\]

If we change the coordinates

\[
Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad Y = \begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix} P
\]

Now observe:

\[
Z' = P^{-1} Y' = P^{-1} A Y = (P^{-1} A P) Z = \begin{bmatrix} 11 & 0 \\ 0 & 1 \end{bmatrix} Z.
\]
This is a system that is easy to solve

\[ z_1' = 11z_1 \rightarrow z_1 = c_1 e^{11x} \]
\[ z_2' = z_2 \rightarrow z_2 = c_2 e^x \]

From which we get the solution

\[ Y = \begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} c_1 e^{11x} \\ c_2 e^x \end{bmatrix} \]
Another solution

Let \( Y' = AY \) be a system of differential equations in the variable \( t \). If it is just \( y' = ay \), the solution would be \( y = ce^{at} \):

\[
y = ce^{ta} = c(1 + ta + t^2 \frac{a^2}{2} + \cdots + t^n \frac{a^n}{n!} + \cdots)
\]

Let us try the same with a matrix. If we replace \( a \) by the square matrix \( A \) (and 1 by \( I \)), we get

\[
e^{tA} = I + tA + t^2 \frac{A^2}{2} + \cdots + t^n \frac{A^n}{n!} + \cdots
\]

Note that the derivative of the \( n \)th term is

\[
nt^{n-1} \frac{A^n}{n!} = A(t^{n-1} \frac{A^{n-1}}{(n-1)!}),
\]

and thus if \( Y = e^{tA} \) then \( Y' = AY \).

We just must make sure that a theory of series makes sense and taking derivatives of these expressions makes sense. At the end we will also put in a constant: \( Y = e^{tA}Y_0 \).
The expression we wrote above for $e^{tA}$ is actually a set of $2^2$ series, one for each cell $(i, j)$ of the 2-by-2 matrix. That is, when we consider the sum of the terms

$$t^n \frac{A^n}{n!}$$

we observe that convergence, for one, comes from the fact that the $n!$ factor grows much faster than the entries $A^n_{(i,j)}$. Let us give an example. Suppose $A$ is a 2-by-2 diagonal matrix with 1 and 1 on the diagonal. $A^n$ is also diagonal with entries $11^n$ and $1^n$. Adding the series would give the matrix

$$\begin{bmatrix} e^{11t} & 0 \\ 0 & e^t \end{bmatrix} = \begin{bmatrix} 1 + 11t + 1/2(11t)^2 + \cdots & 0 \\ 0 & 1 + t + 1/2t^2 + \cdots \end{bmatrix}$$

Not only this is a nice computation, but tells us the same would work whenever $A$ is a diagonal matrix. Let us show how it would work when $A$ diagonalizable.
Let us show how compute $e^{tA}$ if $A = PDP^{-1}$, with $D$ diagonal.

Noting that 

$$A^n = PD^nP^{-1},$$

we have

$$e^{tA} = \sum \frac{t^n}{n!} A^n = \sum \frac{t^n}{n!} PD^nP^{-1} = P(\sum \frac{t^n}{n!} D^n)P^{-1} = Pe^{tD}P^{-1}$$

**Exercise:** $\det e^A = e^{\text{Trace} (A)}$. (This is beautiful because while we have a great deal of trouble with $e^A$, its determinant is easy!)
Theorem

The solution of the differential equation $Y' = AY$ is

$$Y = e^{tA}C,$$

for some constant vector $C$.

Observe where the constant goes. If you set $t = 0$, $Y_0 = C$, that is the components of $C$ are the initial condition: $y_1(0), y_2(0)$.

Clearly the method will work for matrices of any size.

If $A$ is diagonalizable we know how to compute $e^{tA}$. If not ... also!
Let \( \mathbf{A} \) be a \( 3 \times 3 \) real matrix with entries 0, \( \pm 1 \). Determine how large \( \det \mathbf{A} \) can be. Care to consider the \( 4 \times 4 \) version?

Prove that for any real \( n \times n \) matrix \( \mathbf{A} \), \( \det(e^{\mathbf{A}}) = e^{\text{trace} (\mathbf{A})} \):

First prove for \( \mathbf{A} \) upper triangular, and then use the fact that there are complex matrices \( \mathbf{P} \) and \( \mathbf{B} \) such that \( \mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \mathbf{B} \), where \( \mathbf{B} \) is upper triangular.
Let $V$ be a vector space over the field $F$. We want to develop a geometry for $V$. For that, it is helpful to have a notion of distance, or length. We will transport and then extend numerous constructions of ordinary geometry and their calculus.

We will restrict ourselves to the cases of $F = \mathbb{R}$, or $F = \mathbb{C}$. In the case of $\mathbb{C}$, we use the standard notation for the complex conjugate of the complex number $z = a + bi$

$$\bar{z} = a - bi.$$ 

Some of its properties are:

\[
\begin{align*}
    z\bar{z} &= a^2 + b^2 \\
    \bar{z_1} + \bar{z_2} &= \bar{z_1} + \bar{z_2} \\
    \bar{z_1} \cdot \bar{z_2} &= \bar{z_1} \cdot \bar{z_2} \\
    \frac{1}{\bar{z}} &= \frac{1}{z} \\
    z \neq 0
\end{align*}
\]
For certain operations, like solving polynomial equations, the polar representation of complex numbers

\[ a + bi = r(\cos \theta + i \sin \theta), \quad r = \sqrt{a^2 + b^2}, \quad \tan \theta = \frac{a}{b} \]

is useful. For instance,

\[
\sqrt{i} = \pm (\cos \frac{\pi}{2} + i \sin \frac{\pi}{2})^{1/2} = \pm (\cos \frac{\pi}{4} + i \sin \frac{\pi}{4}) = \pm \frac{\sqrt{2}}{2} (1+i).
\]
An **inner product vector space** \( V \) is a V.S. over \( \mathbb{R} \) or \( \mathbb{C} \) with a mapping

\[
V \times V \rightarrow F, \quad (u, v) \rightarrow \langle u, v \rangle = u \cdot v \in F
\]
satisfying certain conditions. Let us give an example to guide us in what is needed. Let \( V = \mathbb{R}^n \) and define

\[
\begin{bmatrix}
a_1 \\
\vdots \\
a_n
\end{bmatrix} \cdot \begin{bmatrix}
b_1 \\
\vdots \\
b_n
\end{bmatrix} = a_1 b_1 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i
\]

Note the properties: bi-additive; \( v \cdot v \) is a non-negative real number, so we can use \( \sqrt{v \cdot v} \) to define the **magnitude** of \( v \).

**Question:** Could we use the same formula to define an inner product for \( \mathbb{C}^n \)? Well... \((i) \cdot (i)\) would be \(-1\). Of course the formula still defines a nice bilinear mapping but would not meet our need.
Dot product

**Definition**

An inner product vector space is a vector space with a mapping

\[ V \times V \rightarrow F, \quad (u, v) \rightarrow u \cdot v \in F \]

satisfying:

1. \((u_1 + u_2) \cdot v = u_1 \cdot v + u_2 \cdot v\)
2. \((cu) \cdot v = c(u \cdot v)\)
3. \(u \cdot v = v \cdot u\)
4. \(u \cdot u > 0 \text{ if } u \neq 0\)

The better notation for this product is

\[ u \cdot v = \langle u, v \rangle \]
Examples

Of course, the example above of $\mathbb{R}^n$ is the grandmother of all examples. Let us modify it a bit to get an example for $\mathbb{C}^n$:

$$
\begin{bmatrix}
a_1 \\
\vdots \\
a_n
\end{bmatrix}
\cdot
\begin{bmatrix}
b_1 \\
\vdots \\
b_n
\end{bmatrix}
= a_1 \overline{b_1} + \cdots + a_n \overline{b_n} = \sum_{i=1}^{n} a_i \overline{b_i}.
$$

Note the properties: additive; $v \cdot v$ is a non-negative real number

$$
v \cdot v = \sum_{i=1}^{n} a_i \overline{a_i}
$$

so we can use $\sqrt{v \cdot v}$ to define the magnitude of $v$. Note the lack of full symmetry.
Example of Function Space

Let us give an example from left field: Let $V$ be the vector space of all real continuous functions on the interval $[a, b]$, and define for $f(t), g(t) \in V$,

$$\langle f(t), g(t) \rangle = f(t) \cdot g(t) = \int_a^b f(t)g(t)\,dt.$$  

An important case: If $m, n$ are integers,

$$\langle \sin nt, \cos mt \rangle = \int_0^{2\pi} \sin nt \cos mt \,dt = 0$$

$$\langle \sin nt, \sin mt \rangle = \int_0^{2\pi} \sin nt \sin mt \,dt = 0, \ m \neq n$$

$$\langle \cos nt, \cos mt \rangle = \int_0^{2\pi} \cos nt \cos mt \,dt = 0, \ m \neq n$$

$$\langle \sin nt, \sin nt \rangle = \int_0^{2\pi} \sin^2 nt \,dt = \pi, \ n \neq 0$$
Example: $\mathbb{M}_n(\mathbb{F})$

Let $\mathbf{V} = \mathbb{M}_n(\mathbb{F})$ be the V.S. of all $n$-by-$n$ matrices. For any such matrix $\mathbf{A} = [a_{ij}]$ define the adjoint of $\mathbf{A}$ (unfortunately we have already used the word for a very different notion!) to be the matrix

$$\mathbf{A}^* = [\overline{a_{ji}}],$$

that is, we transpose $\mathbf{A}$ and take the complex conjugate of each entry. Define the product (Frobenius product)

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{AB}^*) = \sum_i (\mathbf{AB}^*)_{ii}.$$ 

It is clear that this product has the properties of an inner product. We just check the positivity condition:

$$\langle \mathbf{A}, \mathbf{A} \rangle = \text{trace}(\mathbf{AA}^*) = \sum_i (\mathbf{AA}^*)_{ii}$$

$$\sum \sum a_{ij} \overline{a_{ij}} = \sum |a_{ij}|^2 > 0.$$
Proposition

If \( V \) is an inner product space, the following hold:

1. \( \langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle \)
2. \( \langle u, cv \rangle = \overline{c} \langle u, v \rangle \)
3. \( \langle u, O \rangle = \langle O, v \rangle = 0 \)
4. \( \langle u, u \rangle = 0 \) iff \( u = O \)
5. \( \langle u, v \rangle = \langle u, w \rangle \) for all \( u \in V \) then \( v = w \)

Proof of 1: Note

\[
\langle u, v + w \rangle = \overline{\langle v + w, u \rangle} = \overline{\langle v, u \rangle + \langle w, u \rangle} = \langle v, u \rangle + \langle w, u \rangle = \langle u, v \rangle + \langle u, w \rangle
\]
**Length of a vector**

**Definition**

Let $V, \langle \cdot, \cdot \rangle$ be an inner product space. If $v \in V$, the **length** or **norm** of $v$ is the real number $\|v\| = \sqrt{\langle v, v \rangle}$.

If $V = \mathbb{C}^n$, $v = (a, \ldots, a_n)$,

$$\|v\| = \left[ \sum_{i=1}^{n} |a_i|^2 \right]^{1/2}$$

If $V$ is the space of real continuous functions on $[0, 1]$ and inner product is that we defined previously,

$$\|f(t)\|^2 = \int_0^1 f(t)^2 \, dt.$$
The following assertions permits the construction of ‘recognizable’ objects in any inner product space:

**Theorem**

*If* \( \mathbf{V} \) *is an inner product space, then for all* \( u, v \in \mathbf{V} \)

1. [Cauchy-Schwarz Inequality]

\[
|\langle u, v \rangle| \leq \|u\| \cdot \|v\|
\]

2. [Triangle Inequality]

\[
\|u + v\| \leq \|u\| + \|v\|.
\]
The Cauchy-Schwarz Inequality will allow the introduction of \textbf{angles} and its \textbf{trigonometry} in $V$, while the Triangle Inequality will lead to many constructions extending those we are familiar with in 2- and 3-space.
To prove Cauchy-Schwarz Inequality: Note that for ANY $c \in \mathbb{F}$, $v \neq 0$

\[
0 \leq \|u - cv\|^2 = \langle u - cv, u - cv \rangle = \langle u, u - cu \rangle - c\langle v, u - cv \rangle \\
= \langle u, u \rangle - \bar{c}\langle u, v \rangle - c\langle v, u \rangle + c\bar{c}\langle v, v \rangle
\]

If we set $c = \frac{\langle u, v \rangle}{\langle v, v \rangle}$ the inequality becomes

\[
0 \leq \langle u, u \rangle - \frac{|\langle u, v \rangle|^2}{\|v\|^2},
\]

which proves the assertion.
For the $\Delta$-inequality: Consider

\[
\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle
\]

\[
= \|u\|^2 + 2\langle u, v \rangle + \|v\|^2 = \|u\|^2 + 2\Re\langle u, v \rangle + \|v\|^2
\]

\[
\leq \|u\|^2 + 2\|u\| \cdot \|v\| + \|v\|^2 \quad \text{by C-S inequality}
\]

\[
= (\|u\| + \|v\|)^2.
\]

We used that for any complex number $z = a + bi$, its real part $\Re z = a \leq |z| = \sqrt{a^2 + b^2}$.
Example

To illustrate the power of the axiomatic method, compare the proof above [which holds for ALL examples] with the work needed to check the inequalities just the case of the following example:

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leq \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2} \left[\sum_{i=1}^{n} |b_i|^2\right]^{1/2}$$

$$\left[\sum_{i=1}^{n} |a_i + b_i|^2\right]^{1/2} \leq \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2} + \left[\sum_{i=1}^{n} |b_i|^2\right]^{1/2}$$
Angles and Distances

Equipped with these results, we can define angles and distances, with many of the usual properties, in any inner product space. For example, for a real inner product space, the Cauchy-Schwarz inequality says that for any two [will assume nonzero] vectors \( u, v \),

\[
\langle u, v \rangle \leq \| u \| \cdot \| v \|,
\]

that is

\[
-1 \leq \frac{\langle u, v \rangle}{\| u \| \cdot \| v \|} \leq 1
\]

This means that the ratio can be identified to the cosine, \( \cos \alpha \), of a unique angle \( 0 \leq \alpha \leq \pi \): So we can write

\[
\langle u, v \rangle = \| u \| \cdot \| v \| \cos \alpha
\]

and say that \( \alpha \) is the angle between the vectors \( u \) and \( v \).
An important relationship between two vectors \( u, v \) is when \( \langle u, v \rangle = 0 \): We then say that \( u \) and \( v \) are **orthogonal** or **perpendicular**. One notation for this situation is:

\[
\mathbf{u} \perp \mathbf{v}
\]

The **distance** between the vectors \( u, v \) is defined by

\[
\text{dist}(u, v) = \|u - v\| = \langle u - v, u - v \rangle^{1/2}
\]

One of its properties follow from the triangle inequality: If \( u, v, w \) are three vectors

\[
\text{dist}(u, w) \leq \text{dist}(u, v) + \text{dist}(v, w).
\]
These notions have numerous consequences. Let us begin with:

**Proposition**

Let $v_1, \ldots, v_n$ be nonzero vectors of the inner product space $V$. If $v_i \perp v_j$ for $i \neq j$, then these vectors are linearly independent.

**Proof.**

Suppose we have a linear combination

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0.$$

We claim all $c_i = 0$. To prove, say $c_1 = 0$, take the inner product of the linear combination with $v_1$:

$$c_1 \langle v_1, v_1 \rangle + c_2 \langle v_2, v_1 \rangle + \cdots + c_n \langle v_n, v_1 \rangle = \langle O, v_1 \rangle = 0.$$
A vector $v$ of length $||v|| = 1$ is called a **unit** vector. They are easy to find: given a nonzero vector $u$, $v = \frac{u}{||u||}$ is a unit vector.

A set of vectors $v_1, \ldots, v_n$ is said to be **orthonormal** if $v_i \perp v_j$, for $i \neq j$ and $||v_i|| = 1$ for any $i$. Of course, a good example are the ordinary coordinate vectors of 3-space.
Proposition

Let \( V \) be an inner product space with an orthonormal basis \( v_1, \ldots, v_n \). Then for any \( v \in V \),

\[
v = c_1 v_1 + \cdots + c_n v_n,
\]

where \( c_i = \langle v, v_i \rangle \). The \( c_i \) are called the Fourier coefficients of \( v \) relative to the basis.

Proof.

To get \( c_i \), it suffices to form the inner product of \( v \) with \( v_i \):

\[
\langle v, v_i \rangle = c_i \langle v_i, v_i \rangle = c_i,
\]

since \( \langle v_i, v_i \rangle = 1 \) and all other \( \langle v_j, v_i \rangle = 0 \).
Orthonormal bases are also useful in finding the matrix representation of a L.T. $T : V \rightarrow V$:

Let $\mathcal{A} = \{v_1, \ldots, v_n\}$ be such a basis. Then $[T]_{\mathcal{A}} = [a_{ij}]$ where $a_{ij}$ are the coefficients in the expression

$$T(v_j) = a_{1j}v_1 + \cdots + a_{ij}v_i + \cdots + a_{nj}v_n$$

To select $a_{ij}$ it suffices to ‘dot’ with $v_i$

$$\langle T(v_j), v_i \rangle = a_{1j} \langle v_1, v_i \rangle + \cdots + a_{ij} \langle v_i, v_i \rangle + \cdots + a_{nj} \langle v_n, v_i \rangle$$

$$= 0 \quad = 1 \quad = 0$$

$$[T]_{\mathcal{A}} = [\langle T(v_j), v_i \rangle]$$
**Exercise:** If $u, v$ are vectors of an inner product space $V$, verify the parallelogram law:

$$||u + v||^2 + ||u - v||^2 = 2(||u||^2 + ||v||^2).$$

Draw a picture to illustrate this equality.
We will prove that every finite-dimensional vector space $W$ of an inner product space $V$ has an orthonormal basis.

This will allow us to express the distance from a vector $v \in V$ to the subspace $W$. For instance, if

$$Ax = b$$

is a consistent system of linear equations, that is, if there is some solution $Ax_0 = b$, we know that the solution set is the set

$$x_0 + N(A),$$

where $N(A)$ is the nullspace of $A$. Now we will be able to find the solution of smallest length, if need be.
Let us show how to obtain an orthonormal basis of a vector space from an arbitrary basis $A = \{u_1, \ldots, u_n\}$.

If $n = 1$, $v_1 = \frac{u_1}{\|u_1\|}$ is the answer.

Assume now that we have a basis of two vectors $u_1, u_2$. We need to find two nonzero vectors $v_1, v_2$ in the span of $u_1, u_2$ so that $v_1 \perp v_2$. We use a projection trick: we set $v_1 = u_1$ and look for $c$ so that

$$v_2 = u_2 - cu_1 \perp v_1,$$

that is

$$\langle v_2, v_1 \rangle = \langle u_2, v_1 \rangle - c\langle u_1, v_1 \rangle = 0$$

$$c = \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle}$$

Observe that $v_1, v_2$ have same span as $u_1, u_2$. Now replace $v_i$ by $v_i/\|v_i\|$. 
\[ v - w \perp u \]

\( w = \text{Projection of } v \text{ along } u \)
Projection formula

If $L$ is a line defined by the vector $u \neq O$ and $v$ is another vector,

$$w = \frac{\langle v, u \rangle}{\langle u, u \rangle} u$$

is the **projection** of $v$ along $L$ or $u$.

**Proposition**

$v - w$ is perpendicular to $L$ and the smallest distance from $v$ to any vector of $L$ is $\|v - w\|$.

**Proof.**

We have already seen that $v - w \perp v$. If $cu$ is a vector of $L$, the square distance from $v$ to $cu$ is $(v - w \perp L$, so will use Pythagorean Theorem)

$$\|v - cu\|^2 = \|(v - w) + (w + cu)\|^2 = \|v - w\|^2 + \|w + cu\|^2.$$
Gram-Schmidt Algorithm

The routine to obtain a basis that is orthogonal from another basis [Gram–Schmidt process]:

1. **Input:** \( A = \{u_1, \ldots, u_n\} \) given basis
2. **Set** \( v_1 = u_1 \)
3. **Compute** \( v_2, \ldots, v_n \) successively, one at a time, by

\[
v_i = u_i - \left( \frac{u_i \cdot v_1}{v_1 \cdot v_1} \right) v_1 - \left( \frac{u_i \cdot v_2}{v_2 \cdot v_2} \right) v_2 - \cdots - \left( \frac{u_i \cdot v_{i-1}}{v_{i-1} \cdot v_{i-1}} \right) v_{i-1}
\]

4. **Set** \( w_i = \frac{v_i}{\|v_i\|} \)
5. **Output:** \( B = \{w_1, \ldots, w_n\} \) is an orthonormal basis.
Hadamard’s Inequality

Let $\mathbf{A}$ be a matrix whose columns form a basis $\{u_1, u_2, \ldots, u_n\}$ of $\mathbb{R}^n$ (put $n = 3$ for simplicity)

$$\mathbf{A} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$$

Now consider the matrix

$$\mathbf{B} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 - a_1 u_1 & u_3 - b_1 u_1 - b_2 u_2 \end{bmatrix}$$

where the coefficients are chosen so that the $v_i$'s are perpendicular to one another. Note that $\mathbf{B}$ is obtained from $\mathbf{A}$ by adding scalar multiples of columns to another, so

$$\det(\mathbf{A}) = \det(\mathbf{B}).$$

Furthermore, for each $i$

$$||v_i|| \leq ||u_i||$$

by the projection formula.
Let us calculate $\det(A)^2$:

$$
\det(A)^2 = \det(B)^2 = \det(B) \det(B^t) = \det[v_1 \mid v_2 \mid v_3] \det[v_1 \mid v_2 \mid v_3]^t
$$

$$
= \begin{bmatrix}
\langle v_1, v_1 \rangle & 0 & 0 \\
0 & \langle v_2, v_2 \rangle & 0 \\
0 & 0 & \langle v_3, v_3 \rangle
\end{bmatrix}
$$

$$
= \prod \langle v_i, v_i \rangle
$$
Theorem (Hadamard)

For any square real matrix \( A = [u_1, \ldots, u_n] \),

\[
|\det(A)|^2 \leq \prod_{i=1}^{n} \langle u_i, u_i \rangle.
\]

For instance, if \( A \) is a 4 \( \times \) 4 whose entries are 0, 1, \(-1\), its column vectors have length at most 2, so that \( \det(A) \leq 16 \). According to Joe, there is a such a matrix.
General Projection Formula

Proposition

Let $\mathbf{W}$ be a subspace with an orthonormal basis $A = \{u_1, \ldots, u_n\}$. For any vector $v$, the vector of $\mathbf{W}$

$$ w = \text{proj}_W(v) = \langle v, u_1 \rangle u_1 + \cdots + \langle v, u_n \rangle u_n $$

is the projection of $v$ onto $\mathbf{W}$. It has the following properties

1. $v - w$ is perpendicular to any vector of $\mathbf{W}$. (We say that it is perpendicular to $\mathbf{W}$)
2. $\|v - w\|$ is the shortest distance from $v$ to $\mathbf{W}$.

The proof is like above.
Orthogonal Complement

If $W$ is a subspace of an inner product space $V$, its \textbf{orthogonal complement} $W^\perp$ is the set of all vectors $v$ that are perpendicular to each vector $w$ of $W$. In ordinary 3-space $\mathbb{R}^3$, the $z$-axis is the orthogonal complement of the $xy$-plane.

**Proposition**

$W^\perp$ is a subspace of $V$.

**Proof.**

Clearly $O \in W^\perp$. If $v_1, v_2 \in W^\perp$, for any vector $w \in W$

$$\langle c_1 v_1 + c_2 v_2, w \rangle = c_1 \langle v_1, w \rangle + c_2 \langle v_2, w \rangle = 0,$$

so $W^\perp$ passes the subspace test.
Example

Let $\mathbf{A}$ be an $m \times n$ real matrix. The nullspace of $\mathbf{A}$ is the set of all $n$-tuples $\mathbf{x}$ such that

$$\mathbf{Ax} = 0.$$ 

This means that the nullspace is the orthogonal complement of the row space of $\mathbf{A}$:

$$N(\mathbf{A}) = \text{row space}^\perp.$$ 

Similarly, the left nullspace of $\mathbf{A}$, left $N(\mathbf{A})$, are the $m$-tuples $\mathbf{y}$ such that

$$\mathbf{yA} = \mathbf{O}$$

that is the orthogonal complement of the column space of $\mathbf{A}$. 
These observations suggest several properties of the $\perp$ operation:

1. Let $V$ be a vector space with a basis $e_1, \ldots, e_n$. If $W$ is spanned by $u_1, \ldots, u_m$, $W^\perp$ is the set of all vectors $x_1 e_1 + \cdots + x_n e_n$ such that

$$x_1 \langle e_1, u_i \rangle + \cdots + x_n \langle e_n, u_i \rangle = 0, \quad i = 1, \ldots, m.$$

Thus we find $W$ by solving a system of linear equations.

2. $W \cap W^\perp = (O)$.

3. $\dim W + \dim W^\perp = \dim V$.

4. $(W^\perp)^\perp = W$.
Proposition

\[ \dim W + \dim W^\perp = \dim V. \]

Proof.

Let \( u_1, \ldots, u_m \) be an orthonormal basis of \( W \). We define a mapping \( T : V \to V \) as follows

\[ T(v) = \langle v, u_1 \rangle u_1 + \cdots + \langle v, u_m \rangle u_m. \]

\( T \) is clearly a linear transformation: This is the orthogonal projection of \( V \) onto \( W \). Its range \( R(T) \) is \( W \). Its nullspace \( N(T) \) is the set of vectors \( v \) such that \( \langle v, u_i \rangle = 0 \) for each \( u_i \). This is precisely \( W^\perp \). From the dimension formula

\[ \dim V = \dim R(T) + \dim N(T) = \dim W + \dim W^\perp. \]
Let $G$ be a finite subgroup of $GL_n(\mathbb{C})$. Prove that every $T \in G$ is diagonalizable.
If $V$ is a vector space over the field $F$, a linear functional is a linear transformation

$$f : V \rightarrow F.$$  

For example, if $V = F^n$ and $a = [a_1, \ldots, a_n]$ is a matrix, then for every column vector $v \in F^n$, the function

$$v \mapsto a \cdot v$$

is a linear functional. In fact, every linear functional $f$ has this description.

Inner product spaces, finite/infinite dimensional have a natural method to define linear functionals. Let us exploit it.
Let $V$ be an inner product space. If $u \in V$, the mapping 

$$f : V \to \mathbb{F}, \quad f(v) = \langle v, u \rangle$$

is a linear functional. Observe that if $\langle v, u \rangle = \langle v, w \rangle$, for all $v$, then $\langle v, u - w \rangle = 0$ and therefore $u = w$.

**Proposition**

If $V$ is a finite-dimensional inner product space, for every linear functional $f$ on $V$, there is a unique vector $u$ such that $f(v) = \langle v, u \rangle$ for all $v \in V$.

**Proof.**

Let $v_1, \ldots, v_n$ be an orthonormal basis of $V$, and let 

$$u = \overline{f(v_1)}v_1 + \cdots + \overline{f(v_n)}v_n.$$ 

Note that for each $v_j$, $\langle v_j, u \rangle = \overline{f(v_j)} = f(v_j)$, so the functionals defined by $u$ and $f$ agree on each basis vector, so are equal.
Adjoint of a Linear Transformation

Let $T$ be a L.T. of the inner product space $V$. We are going to build another L.T. associated to $T$, which will be called the adjoint of $T$. It is the parent [or child] of the transpose!

Fix the vector $u \in V$. Consider the mapping $v \to \langle T(v), u \rangle$. This is a linear functional. According to the previous Proposition, there is a unique $w$ such that

$$\langle T(v), u \rangle = \langle v, w \rangle, \quad \forall v \in V.$$  

We set $w = S(u)$. This gives a function $S: V \to V$. It is routine to check that if $w_1 = S(u_1)$ and $w_2 = S(u_2)$, then $S(u_1 + u_2) = w_1 + w_2$, and also $S(cu) = cS(u)$. This L.T. is denoted $T^*$ and termed the adjoint of $T$. 
Proposition

Let $\mathbf{T}$ be a L.T. and let $\mathbf{A} = [a_{ij}]$ be its matrix representation relative to the orthonormal basis $v_1, \ldots, v_n$. Then the matrix representation of the adjoint $\mathbf{T}^*$ is $\overline{\mathbf{A}}^t = [\overline{a_{ji}}]$, the conjugate transpose of $\mathbf{A}$.

Proof.

To find the matrix representation $[b_{ij}]$ of $\mathbf{T}^*$ we write $\mathbf{T}^*(v_j) = \sum_i b_{ij}v_i$, so that

$$\overline{b_{ij}} = \langle v_i, \mathbf{T}^*(v_j) \rangle = \langle \mathbf{T}(v_i), v_j \rangle = a_{ji},$$

as desired.
Problem

Given 3 (or more) points \( P_1 = (x_1, y_1) \), \( P_2 = (x_2, y_2) \), \( P_3 = (x_3, y_3) \) in \( \mathbb{R}^2 \), find the best fit line (what does this mean?):
\[ Y = at + b, \quad Y_i = at_i + b, \quad \text{error} = |Y_i - y_i| \]

| \( t \) | \( y \) | \( Y \) \\
|---|---|---|
| \( t_1 \) | \( y_1 \) | \( Y_1 \) \\
| \vdots | \vdots | \vdots \\
| \( t_n \) | \( y_n \) | \( Y_n \) \\

\[ E = \text{Square Error} = \sum_{i=1}^{n} |Y_i - y_i|^2 = \sum_{i=1}^{n} |at_i + b - y_i|^2 \]

**Problem:** Find \( a \) and \( b \) so that the square error is as small as possible. To answer, we first write the problem in vector notation.
\[
y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad A = \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}, \quad x = \begin{bmatrix} a \\ b \end{bmatrix}
\]

\[
E = ||y - Ax||^2
\]

We are going to do much better: Given a \(m \times n\) matrix \(A\) and a vector \(y \in F^m\), we are going to find a vector \(x_0 \in F^n\) such that

\[
||y - Ax_0||^2 \leq ||y - Ax||^2
\]

for all \(x \in F^n\).
We know that the answer to this will be affirmative: Let $W$ be the range of $A$, that is the set of all vectors $Ax$, for $x \in F^n$. There is a vector $w \in W$, that is $w = Ax_0$ such that

$$||y - Ax_0||^2 \leq ||y - Ax||^2.$$ 

The issue is how to find $x_0$ more explicitly. For this we use the notion of the adjoint of a linear transformation:

$$T : F^n \rightarrow F^m, \quad T^* : F^m \rightarrow F^n$$

$$\langle T(u), v \rangle_m = \langle u, T^*(v) \rangle_n$$

To derive the desired formula (known as the projection formula) we need two properties of $T^*$. 
Proposition

Let $A$ be an $m \times n$ complex matrix and $A^*$ its adjoint (conjugate transpose). Then

1. $\text{rank}(A) = \text{rank}(A^*A)$.
2. If $\text{rank}(A) = n$ then $A^*A$ is invertible.

Proof.

It will suffice to show that $A$ and $A^*A$ have the same nullspace. Why?

If $A^*A(x) = 0$, then for all $z \in F^n$

$$0 = \langle A^*A(x), z \rangle_n = \langle Ax, (A^*)^*z \rangle_m = \langle Ax, Az \rangle_m$$

so $Ax = O$ by choosing $z = x$.

The second assertion now follows: Since $A^*A$ is an $n \times n$ matrix of rank $n$, it is invertible.
Projection Formula

**Theorem**

Let $A$ be an $m \times n$ complex matrix and let $y \in F^m$. Then there exists $x_0 \in F^n$ such that $A^*A(x_0) = A^*y$ and $\|Ax_0 - y\| \leq \|Ax - y\|$ for all $x \in F^n$. If $A$ has rank $n$ then

$$x_0 = (A^*A)^{-1}A^*y.$$ 

**Proof.**

Since $Ax_0 - y$ is perpendicular to the range of $A$,

$$0 = \langle Ax, Ax_0 - y \rangle_m = \langle x, A^*(Ax_0 - y) \rangle = \langle x, ((A^*A)x_0 - A^*y) \rangle$$

for all $x \in F^n$. Thus $(A^*A)x_0 - A^*y = 0$ and therefore

$$x_0 = (A^*A)^{-1}A^*y,$$
Illustration

\[ A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}, \quad \text{rank}(A) = 2, \quad y = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix} \]

\[ A^*A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix} \]

\[ (A^*A)^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix} \]
\[
x_0 = \begin{bmatrix}
a \\
b \\
\end{bmatrix} = \frac{1}{20} \begin{bmatrix}
4 & -10 \\
-10 & 30 \\
\end{bmatrix} \begin{bmatrix}
1 & 2 & 3 & 4 \\
1 & 1 & 1 & 1 \\
\end{bmatrix} \begin{bmatrix}
2 \\
3 \\
5 \\
7 \\
\end{bmatrix} = \begin{bmatrix}
1.7 \\
0 \\
\end{bmatrix}
\]

**Answer:** The least squares line is

\[
y = 1.7t
\]

The error is

\[
E = \|Ax_0 - y\|^2 = 0.3
\]
The method is very general: Suppose we are given a number of points and we want to fit a quadratic polynomial

\[ Y = at^2 + bt + c \]

to the data.

\[
A = \begin{bmatrix}
  t_1^2 & t_1 & 1 \\
  \vdots & \vdots & \vdots \\
  t_n^2 & t_n & 1
\end{bmatrix}, \quad 
\mathbf{x}_0 = \begin{bmatrix}
  a \\
  b \\
  c
\end{bmatrix}, \quad 
\mathbf{y} = \begin{bmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{bmatrix}
\]

Now \( \text{rank}(A) = 3 \) if there are 3 distinct values of \( t \).
We are going to find the **shortest** solution of a consistent system of equations \((m \times n)\)

\[
Ax = b.
\]

This will be a solution \(u\) such that \(\|u\|\) is minimal. The argument will also show that \(u\) is unique.

Let \(x_0\) be a special solution and denote by \(N(A)\) the nullspace of \(A\). The solution set is

\[
x_0 + N(A) = \{x_0 + v, \quad v \in N(A)\}.
\]

To pick out of this set the vector \(x_0 + v\) of smallest length, note that \(\|x_0 + v\|\) is the distance from \(x_0\) to \(-v\). So we have our answer: Pick for \(-v\) the projection \(w\) of \(x_0\) into \(N(A)\). Then \(s = x_0 - w\) is the desired solution:
\[ \mathbf{x}_0 - \mathbf{w} \perp N(A) \]

\( \mathbf{w} = \text{Projection of } \mathbf{x}_0 \text{ along } N(A) \)
One algorithm for the shortest solution

1. Find an orthonormal basis $u_1, \ldots, u_r$ for $N(A)$
2. Determine the projection $w$ of $x_0$ onto $N(A)$:

$$w = \sum_{i=1}^{r} \langle x_0, u_i \rangle u_i$$

3. $x_0 - w$ is the shortest solution of $Ax = b$
This solution requires the calculation of the projection of $x_0$ into $N(A)$. Let us discuss another, more direct, approach. If $v \in N(A)$, $A(v) = O$,

$$0 = \langle x, A(v) \rangle = \langle A^*(x), u \rangle$$

which means $v \perp A^*(x) = 0$ for all $x$. This means that the range of $A^*$, $R(A^*)$, is contained in the orthogonal complement $N(A) \perp$ of $N(A)$. By the dimension formula we have $N(A) \perp = R(A^*)$.

**Summary:** The minimal vector $s$ satisfies

$$As = b, \quad s \in R(A^*)$$

That is, pick any solution of

$$AA^*y = b,$$

and set

$$s = A^*y.$$
Homework #9

Section 6.3: 3a, 6, 10, 13, 18, 22a, 23
1. Normal Operators ($TT^* = T^*T$): real symmetric/skew symmetric
2. Hermitian Operator
4. Spectral Theorem
5. Goodies: Applications
Interesting diagonalizable operators

We are going to show a class of linear transformations that are diagonalizable. It will include the class represented by real symmetric matrices.

Let $T : V \rightarrow V$ be a L.T. of a complex inner product space. We have defined the adjoint $T^*$ of $T$ as the L.T. with the property

$$\langle T(u), v \rangle = \langle u, T^*(v) \rangle, \quad \forall u, v \in V.$$ 

Let us compare the eigenvalues and eigenvectors of $T$ and $T^*$.
Proposition

If $\lambda$ is an eigenvalue of $T$ then $\overline{\lambda}$ is an eigenvalue of $T^*$.

Proof: Suppose $T(u) = \lambda u$, $u \neq O$. Then for any $v \in V$,

$$0 = \langle O, v \rangle = \langle (T - \lambda I)(u), v \rangle = \langle u, (T - \lambda I)^*(v) \rangle = \langle u, (T^* - \overline{\lambda} I)(v) \rangle$$

This says that $O \neq u \perp \text{range}(T^* - \overline{\lambda} I)$, so the range of $T^* - \overline{\lambda} I$ is not the whole of $V$, which implies nullspace of $T^* - \overline{\lambda} I \neq O$. This means that $\overline{\lambda}$ is an eigenvalue of $T^*$. 
Let us use this result to decide when a L.T. $T$ of an inner product space $V$ admits a basis $A$ such that

$$[T]_A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix},$$

that is, $T$ admits a matrix representation that is upper triangular.

Note that the characteristic polynomial has all of its roots in the field

$$\det(T - xI) = (a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x),$$

that is the characteristic polynomial splits. Recall that this is always the case when the field is $\mathbb{C}$. 
Theorem (Schur)

Let \( T \) be a L.T. of the inner product space \( V \). If the characteristic polynomial of \( T \) splits, then \( V \) admits an orthonormal basis \( A \) such that \( [T]_A \) is upper triangular.

**Proof:** We will argue by induction on \( \dim V = n \). If \( n = 1 \), the assertion is obvious. Let us assume that the assertion holds for dimension \( n - 1 \). By the Proposition above, we know that \( T^* \) has one eigenvalue \( \lambda \). Let \( u \) be a unit vector so that \( T^*(u) = \lambda u \), and set \( W \) for the subspace spanned by \( u \). We claim that \( W^\perp \) is \( T \)-invariant: If \( v \in W^\perp \)

\[
\langle T(v), u \rangle = \langle v, T^*(u) \rangle = \langle v, \lambda u \rangle = \bar{\lambda} \langle v, u \rangle = 0
\]

So \( T(v) \in W^\perp \).
We also have \( \dim W + \dim W^\perp = \dim V = n \), so \( \dim W^\perp = n - 1 \). Now we apply the induction hypothesis to the restriction of \( T \) to \( W^\perp \): Let \( v_1, \ldots, v_{n-1} \) be an orthonormal basis of \( W^\perp \) for which the restriction of \( T \) is upper triangular. If we add to the \( v_i \) the vector \( u \), we get the orthonormal basis \( A = v_1, \ldots, v_{n-1}, u \). The matrix representation

\[
[T]_A = \begin{bmatrix}
[T]_{W^\perp} & \begin{bmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{bmatrix} \\
0 & 0 & \cdots & 0
\end{bmatrix},
\]

which has the desired form.
Normal operator

Observe that if there is an orthonormal basis \( \mathcal{A} \) of eigenvectors of \( T \), \( [T]_{\mathcal{A}} \) is a diagonal matrix, and since \( [T^*]_{\mathcal{A}} = [T]_{\mathcal{A}}^* \), this matrix is also diagonal. Since diagonal matrices commute, we have \( TT^* = T^*T \).

**Definition**

A linear transformation \( T \) of an inner product space is **normal** if \( TT^* = T^*T \).

**Example:** If \( A \) is a symmetric real matrix, \( A^* = A^t = A \), so \( A \) commutes with itself! Skew-symmetric real matrices, \( A^* = -A \), are also normal.
Theorem

If $T$ is a normal operator ($TT^* = T^*T$) of a complex inner vector space $V$, then there is an orthonormal basis of eigenvectors of $T$. (The converse was proved already so this is a characterization of normal operators.)

This is an important result, it has many useful consequences. To prove it we shall need some properties of normal operators.
**Proposition**

Let $T$ be a normal operator ($TT^* = T^*T$) of the inner vector space $V$. Then:

1. $\|T(u)\|^2 = \|T^*(u)\|^2$ for every $u \in V$.
2. $T - cI$ is normal for every $c \in F$.
3. If $T(u) = \lambda u$ then $T^*(u) = \bar{\lambda}u$.
4. If $\lambda_1$ and $\lambda_2$ are distinct eigenvalues of $T$ with corresponding eigenvectors $u_1$ and $u_2$, then $u_1 \perp u_2$.

**Proof:**

1. For any vector $u \in V$,

$$\|T(u)\|^2 = \langle T(u), T(u) \rangle = \langle T^*T(u), u \rangle = \langle TT^*(u), u \rangle = \langle T^*(u), T^*(u) \rangle = \|T^*(u)\|^2$$

2. $(T - cI)(T^* - \bar{c}I) = (T^* - \bar{c}I)(T - cI)$: check
3. Suppose $T(u) = \lambda u$. Let $U = T - \lambda I$. Then $U(u) = 0$ so by 2. $U$ is normal and by 1. $U^*(u) = 0$. That is $T^*(u) = \overline{\lambda}u$.

4. Let $\lambda_1$ and $\lambda_2$ be distinct eigenvalues of $T$ with corresponding eigenvectors $u_1$ and $u_2$. Then by 3.

$$\lambda_1 \langle u_1, u_2 \rangle = \langle \lambda_1 u_1, u_2 \rangle = \langle T(u_1), u_2 \rangle = \langle u_1, T^*(u_2) \rangle$$

$$= \langle u_1, \overline{\lambda_2} u_2 \rangle = \lambda_2 \langle u_1, u_2 \rangle.$$

Since $\lambda_1 \neq \lambda_2$, $\langle u_1, u_2 \rangle = 0$. 
We are now in position to prove that a normal operator $T$ admits an orthonormal basis $v_1, v_2, \ldots, v_n$ of eigenvectors. We already know, by Schur theorem, that there is an orthonormal basis for which the matrix representation is upper triangular

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  0 & a_{22} & a_{23} \\
  0 & 0 & a_{33}
\end{bmatrix}
\]

We want to show that the off-diagonal elements are 0, that is, all the $v_i$ are eigenvectors. [For simplicity we take $n = 3$] Note that $T(v_1) = a_{11} v_1$, so $v_1$ is an eigenvector. To show $v_2$ is an eigenvector notice that

\[ T(v_2) = a_{12} v_1 + a_{22} v_2 \]

We must show $a_{12} = 0$. 
\[ \mathbf{T}(v_2) = a_{12}v_1 + a_{22}v_2 \]

We must show \( a_{12} = 0 \):

\[ a_{12} = \langle \mathbf{T}(v_2), v_1 \rangle = \langle v_2, \mathbf{T}^*(v_1) \rangle = \langle v_2, a_{11}v_1 \rangle = a_{11}\langle v_2, v_1 \rangle = 0 \]

as desired. Now with \( v_1, v_2 \) eigenvectors, we show that \( a_{13} = a_{23} = 0 \). We consider

\[ \mathbf{T}(v_3) = a_{13}v_1 + a_{23}v_2 + a_{33}v_3 \]

The proof is similar: For instance

\[ a_{23} = \langle \mathbf{T}(v_3), v_2 \rangle = \langle v_3, \mathbf{T}^*(v_2) \rangle = \langle v_3, a_{22}v_2 \rangle = a_{22}\langle v_3, v_2 \rangle = 0 \]
We have already remarked that real symmetric matrices, $A = A^t$, are normal. It turns out that complex symmetric matrices are not always normal. Truly the complex cousins of real symmetric matrices are called:

**Definition**

Let $T$ be a linear operator of the inner product space $V$. $T$ is called **self-adjoint** (Hermitian) if $T = T^*$. 

$$A = \begin{bmatrix} 2 & 3 + 5i \\ 3 - 5i & 6 \end{bmatrix}$$
Lemma

Let \( T \) be a self-adjoint linear operator of the inner product space \( V \). Then

1. Every eigenvalue is real.
2. If \( V \) is a real vector space then the characteristic polynomial splits.

Proof: 1. Suppose \( T(u) = \lambda u, u \neq O \). By a previous result, \( T^*(u) = \overline{\lambda} u \). Since \( T = T^* \), \( \lambda \) is real.

2. Let \( n = \dim V \), \( \mathcal{B} \) an orthonormal basis of \( V \) and \( A = \begin{bmatrix} T \end{bmatrix}_\mathcal{B} \). Then \( A \) is self-adjoint. Let \( T_A \) be the linear operator of \( \mathbb{C}^n \) defined by \( T_A(u) = Au \) for all \( u \in \mathbb{C}^n \).
Note that $T_A$ is self-adjoint because $[T_A]_C = A$, where $C$ is the standard (orthonormal) basis of $\mathbb{C}^n$. So the eigenvalues of $T_A$ are real. Since the characteristic polynomial of $T_A$ is equal to the characteristic polynomial of $A$, which is equal to the characteristic of $T$, the characteristic polynomial of $T$ splits.

What we are saying is the following: Let $A$ be a $n \times n$ symmetric real matrix and employ it to define a L.T. of the complex vector space $\mathbb{C}^n$

$$ T = T_A : \mathbb{C}^n \rightarrow \mathbb{C}^n, \quad T(u) = A(u). $$

Note $\det(T - xI) = \det(A - xI)$. 
First Main Theorem of the Course

Theorem

Let $T$ be a linear operator on the finite-dimensional inner product space $V$. Then $T$ is self-adjoint if and only if there exists an orthonormal basis of $V$ consisting of eigenvectors of $T$. 
Unitary Operators

**Definition**

A linear operator $T$ of the inner product space $V$ is called **unitary** if $TT^* = T^*T = I$. If $V$ is a real inner product space, $T$ is called **orthogonal**.

The rotation operator

$$T(x, y) = (x \cos \alpha + y \sin \alpha, -x \sin \alpha + y \cos \alpha)$$

is a major example.

If $A$ is a complex $n$-by-$n$ matrix and $AA^* = A^*A = I$, the column vectors of $A$ form an orthonormal basis of $\mathbb{C}^n$.

We now develop quickly some basic properties of these operators.
Let $T$ be a linear operator of the finite-dimensional inner product space $V$. TFAE:

1. $T$ is an unitary operator: $TT^* = T^*T = I$.
2. $\langle T(u), T(v) \rangle = \langle u, v \rangle$ for all $u, v \in V$.
3. For every orthonormal basis $\mathcal{B} = v_1, \ldots, v_n$ of $V$, $T(v_1), \ldots, T(v_n)$ is also an orthonormal basis of $V$.
4. For some orthonormal basis $\mathcal{B} = v_1, \ldots, v_n$ of $V$, $T(v_1), \ldots, T(v_n)$ is also an orthonormal basis of $V$.
5. $\|T(u)\| = \|u\|$ for every $u \in V$.

Proof. $1 \Rightarrow 2, 3, 4, 5$: (Other $\Rightarrow$ LTR)

$$\langle u, v \rangle = \langle T^*T(u), v \rangle = \langle T(u), (T^*)^*(v) \rangle = \langle T(u), T(v) \rangle.$$ 

$$\delta_{ij} = \langle v_i, v_j \rangle = \langle T(v_i), T(v_j) \rangle.$$
Properties of unitary operators

Let $T$ be an unitary operator of the inner product space $V$.

1. The eigenvalues of $T$ have length 1: If $T(u) = \lambda u$,

\[ \langle u, u \rangle = \langle T(u), T(u) \rangle = \langle \lambda u, \lambda u \rangle = \lambda \lambda \langle u, u \rangle \]

and thus $\lambda \lambda = 1$.

2. If $A$ is a matrix representation of $T$, $|\det(A)| = 1: \det(A) \det(A^*) = 1$

3. If $T$ is orthogonal, $\det(A) = \pm 1$.

4. If $T$ and $U$ are unitary operators, then $T^*$ and $T \circ U$ are also unitary operators.
Orthogonal operators of $\mathbb{R}^2$

We have already mentioned rotations, $R_\alpha$. Let us analyze the possibilities. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = [v_1 | v_2], \quad ||v_1|| = ||v_2|| = 1, \quad v_1 \perp v_2$$

be an orthogonal matrix. This means

$$a^2 + c^2 = 1, \quad b^2 + d^2 = 1, \quad ab + cd = 0$$

We can set $a = \cos \alpha, \ c = \sin \alpha$ and $b = \cos \beta, \ d = \sin \beta$ so that

$$ab + cd = \cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\alpha - \beta) = 0.$$ 

This means that $\alpha - \beta = \pm \pi/2$. The two possibilities lead to

$$R_\alpha = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}, \quad T = \begin{bmatrix} \cos \beta & \sin \beta \\ \sin \beta & -\cos \beta \end{bmatrix}$$
To analyze
\[ T = \begin{bmatrix} \cos \beta & \sin \beta \\ \sin \beta & -\cos \beta \end{bmatrix} \]
we look at its eigenvalues:

\[
\text{det}(T - xI) = \begin{bmatrix} \cos \beta - x & \sin \beta \\ \sin \beta & -\cos \beta - x \end{bmatrix} = x^2 - 1
\]

So \( \lambda = \pm 1 \). This means we have an orthonormal basis \( v_1, v_2 \), and \( T(v_1) = v_1, T(v_2) = v_2 \).

Thus the line \( \mathbb{R}v_1 \) is fixed under \( T \), and the perpendicular line \( \mathbb{R}v_2 \) is flipped about \( \mathbb{R}v_1 \). These transformations are called reflections.

**Summary:** If \( A \) is an orthogonal 2-by-2 matrix, then if \( \text{det} A = 1 \), it is a rotation, and if \( \text{det} A = -1 \), it is a reflection.
Matrix product and dot product

Let \( u \) and \( v \) be two vectors of \( \mathbb{R}^n \). Their dot product

\[
\begin{bmatrix}
a_1 \\
\vdots \\
a_n
\end{bmatrix} \cdot \begin{bmatrix}
b_1 \\
\vdots \\
b_n
\end{bmatrix}
\]

can be expressed as a matrix product

\[
\begin{bmatrix}
a_1 & \cdots & a_n
\end{bmatrix}
\begin{bmatrix}
b_1 \\
\vdots \\
b_n
\end{bmatrix}
\]

Keep in mind

\[
u^t v = u \cdot v
\]
Spectral Decomposition

Let \( A \) be a \( n \)-by-\( n \) symmetric real matrix, \( P = [v_1 | \cdots | v_n] \) a matrix whose columns form an orthonormal basis of eigenvectors of \( A \):

\[
A = PDP^t = [v_1 | \cdots | v_n] \cdot \begin{bmatrix}
\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_n
\end{bmatrix} \cdot \begin{bmatrix}
v_1^t \\
\vdots \\
v_n^t
\end{bmatrix}
\]

Instead of this representation of \( A \) as a product of 3 matrices, we are going to express \( A \) as a sum of simple matrices of rank 1.
Expanding we get

\[
A = PDP^t = [v_1 | \cdots | v_n] \cdot \begin{bmatrix}
\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_n
\end{bmatrix} \cdot \begin{bmatrix}
v_1^t \\
\vdots \\
v_n^t
\end{bmatrix}
\]

\[
= [\lambda_1 v_1 | \cdots | \lambda_n v_n] \cdot \begin{bmatrix}
v_1^t \\
\vdots \\
v_n^t
\end{bmatrix} 
\]

\[
= \lambda_1 v_1 v_1^t + \cdots + \lambda_n v_n v_n^t 
\]

\[
= \sum \lambda_i P_i, \quad P_i = v_i v_i^t.
\]

Let us examine the matrices \( P_i \).
1. $P_i$ has rank 1 and is symmetric

$$P_i = v_i v_i^t, \quad P_i^t = (v_i v_i^t)^t = (v_i^t)^t v_i^t = P_i$$

2. $P_i$ is a projection

$$P_i P_i = (v_i v_i^t)(v_i v_i^t) = v_i (v_i^t v_i) v_i^t = v_i v_i^t = P_i$$

since $v_i^t v_i = \langle v_i, v_i \rangle = 1$

3. $P_i P_j = O$ for $i \neq j$

$$P_i P_j = (v_i v_i^t)(v_j v_j^t) = v_i (v_i^t v_j) v_j^t = O$$

since $v_i^t v_j = \langle v_i, v_j \rangle = 0$
The equality

\[ A = \sum \lambda_i P_i, \ P_i = v_i v_i^t \]

is called the **spectral decomposition** of \( A \).

**Example:** Let \( A = \begin{bmatrix} 3 & -4 \\ -4 & -3 \end{bmatrix} \)

The eigenvalues are 5 and \(-5\), with corresponding [normalized] eigenvectors

\[ v_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \]

\[ P_1 = v_1 v_1^t = \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix}, \quad P_2 = v_2 v_2^t = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix} \]
Exercise:

Let \( A \) be a real symmetric matrix. Prove that there is a symmetric matrix \( B \) such that \( B^3 = A \).

We know that there is an orthonormal basis \( v_1, \ldots, v_n \) of eigenvectors of \( A \). The matrix \( P = [v_1 | \cdots | v_n] \) is orthogonal [i.e. \( P^{-1} = P^t \)] and

\[
P^{-1}AP = D
\]

is a real diagonal matrix. Let \( E \) be a real ‘cubic root’ of \( D \) (if a diagonal entry of \( D \) is \( d_{ii} \), the corresponding entry of \( E \) is the real root \( d_{ii}^{1/3} \)).

Set \( B = P^{-1}EP \). Note

\[
B^t = (P^{-1}EP)^t = P^tE^t(P^{-1})^t = P^{-1}EP = B,
\]

\[
B^3 = P^{-1}E^3P = A.
\]
Exercise: Let $A$ be skew-symmetric matrix. Prove that $\det A \geq 0$. *Hint:* Recall that $A$ is normal, then pair up the complex eigenvalues of $A$. Moreover, show that if $A$ has integer entries, then $\det A$ is the square of an integer.
Real quadratic forms

A real **quadratic form** in $n$ variables is a polynomial

$$q(x) = \sum_{i,j} a_{ij} x_i x_j.$$ 

They occur in the elementary theory of conic sections--e.g. what is $10x^2 + 6xy + 2y^2 = 5$, an ellipse, a parabola, or a hyperbola?-- but also in the theory of max and min of functions $f(x_1, \ldots, x_n)$ of several variables. In both endeavors, a solution arises after an appropriate change of variables, $x = P(y)$,

$$q(x) = q(P(y)) = \sum_i d_i y_i^2.$$ 

Let us see how this comes about:
Let us begin with $Ax^2 + Bxy + Cy^2$, which we write as $ax^2 + 2bxy + cy^2$. (For general fields this would require $2 \neq 0$.) Now look:

$$ax^2 + 2bxy + cy^2 = x(ax + by) + y(bx + cy)$$

$$= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= x^tQx$$

where $x = \begin{bmatrix} x \\ y \end{bmatrix}$ and $Q$ is a symmetric matrix.

It is routine to verify that every quadratic form $q(x)$ has such a representation,

$$q(x) = x^tQx, \quad Q = Q^t$$

Now we can apply to $Q$ the spectral theorem we have developed.
Since $\mathbf{Q}$ is (orthogonally) diagonalizable, there is an orthogonal matrix $\mathbf{P}$ (formed by an orthonormal basis of eigenvectors of $\mathbf{Q}$) such that

$$\mathbf{P}^{-1} \mathbf{Q} \mathbf{P} = \mathbf{D} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

This means that in $q(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x}$, if we change the variables by the rule $\mathbf{x} = \mathbf{P} \mathbf{y}$,

$$q(\mathbf{x}) = \mathbf{x}^t \mathbf{Q} \mathbf{x} = \mathbf{y}^t \mathbf{P}^{-1} \mathbf{Q} \mathbf{P} \mathbf{y} = \mathbf{y}^t \mathbf{D} \mathbf{y} = \sum_i \lambda_i y_i^2.$$
Among the potential applications, we mentioned the identification of conics. For example, \(10x_1^2 + 6x_1x_2 + 2x_2^2 = 5\):

The matrix

\[
Q = \begin{bmatrix}
10 & 3 \\
3 & 2
\end{bmatrix}
\]

has for eigenvalues 11, 1 with

\[
P = \frac{1}{\sqrt{10}} \begin{bmatrix}
1 & -3 \\
3 & 1
\end{bmatrix}
\]

The change of variables \(x = Py\) gives

\[11y_1^2 + y_2^2 = 5,
\]

the equation of an ellipse.
Another application, to the theory of max and min appears as follows: If \( \mathbf{a} \) is a critical point of the function \( f(\mathbf{x}) \) – that is all the partial derivatives vanish at \( \mathbf{x} = \mathbf{a} \), \( \frac{\partial f}{\partial x_i}(\mathbf{a}) = 0 \), Taylor’s expansion of \( f \) in a neighborhood of \( \mathbf{a} \) gives

\[
f(\mathbf{x}) = f(\mathbf{a}) + q(h) + \text{error}
\]

where \( q \) is a quadratic polynomial on the vector \( \mathbf{h} = \mathbf{x} - \mathbf{a} \). The corresponding symmetric matrix is

\[
Q = \begin{bmatrix}
\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}(\mathbf{a}) \\
\end{bmatrix}
\]

If all the eigenvalues of \( Q \) are positive [negative], \( q(h) \geq 0 \) Then \( f(\mathbf{x}) \geq f(\mathbf{a}) \) in a neighborhood of \( \mathbf{a} \): local max [local min]. The other cases are saddle points [the higher dimensional analogues of inflection points]
Rigid Motion

A **rigid motion** on the inner product space $\mathbf{V}$ is a mapping

$$ \mathbf{T} : \mathbf{V} \rightarrow \mathbf{V} $$

with the property

$$ ||\mathbf{T}(u) - \mathbf{T}(v)|| = ||u - v||, \quad \forall u, v \in \mathbf{V}. $$

That is, $\mathbf{T}$ preserves distance of the images. A simple example is a translation: If $\mathbf{a}$ is a fixed vector, the function

$$ \mathbf{T}(v) := \mathbf{a} + v $$

is obviously a rigid motion. What else? We have seen that orthogonal transformations $\mathbf{S}$, $\mathbf{S}^\dagger = \mathbf{I}$, preserve distances. Another such motion is obtained by composition: following a translation with an orthogonal mapping. What else? That is it!
Theorem

Any rigid motion $T$ of $V$ decomposes into $T = S \circ U$, where $S$ is an orthogonal transformation and $U$ is a translation.

Proof: Set $a = T(O)$. Then the function $F(u) = T(u) - a$ is a rigid motion and $F(O) = O$. It is enough to prove that $F$ is orthogonal. Note that

$$||F(u) - F(O)|| = ||u - O||,$$

so $F$ preserves lengths, which is the key property of orthogonal transformations. BUT we are NOT assuming that $F$ is linear, we must prove it.

We first prove that $F$ preserves dot products: $\langle F(u), F(v) \rangle = \langle u, v \rangle$: We start from the equality and expand both sides
Thus proving
\[
\langle \mathbf{F}(u), \mathbf{F}(v) \rangle = \langle u, v \rangle.
\]

Now we are going to prove that \( F \) is a linear function by first showing that it is additive:
\[ \| F(u + v) - F(u) - F(v) \|^2 \geq 0 \]

\[ \| F(u + v) \|^2 + \| F(u) \|^2 + \| F(v) \|^2 - 2 \langle F(u + v), F(u) \rangle - 2 \langle F(u + v), F(v) \rangle + 2 \langle F(u), F(v) \rangle = \| (u + v) - u - v \|^2 = 0. \]

Scaling, that \( F(cu) = cF(u) \) for any \( c \in \mathbb{R} \), has a similar proof: Expand

\[ \| F(cu) - cF(u) \|^2 \]
Homework #10

Section 6.4: 2f, 4, 6, 12, 13, 15

Section 6.5: 6, 10, 11, 17, 27a
1. Section 6.5, Problem 27d
Let \( A \) be a \( 3 \times 3 \) orthogonal matrix. Prove that \( A \) is similar to a matrix of the form

\[
\begin{bmatrix}
R & O \\
O & \pm 1
\end{bmatrix}
\]

where \( R \) is a \( 2 \times 2 \) orthogonal matrix.

2. Section 6.3, Problem 22c
Let \( A \) be a skew-symmetric real matrix. If \( A \) diagonalizable, prove that \( A = 0 \).