5.1 - Linear Models & Matrix Algebra:
Summary

- Matrix algebra can be used:
  a. to express the system of equations in a compact notation;
  b. to find out whether solution to a system of equations exist; and
  c. to obtain the solution if it exists.

  d. If \( n \) is small, we can find \( A^{-1} \), but in general, we will avoid this step. We will resort to more efficient methods to solve for \( x^* \), likely using a Cholesky decomposition with Gaussian elimination.

\[
Ax = d \\
x^* = A^{-1}d \\
A^{-1} = \frac{adjA}{\det A} \\
x^* = \frac{adjA}{|A|}d
\]
5.1 - Notation and Definitions: Summary

- **A** (Upper case letters) = matrix
- **b** (Lower case letters) = vector
- \(n \times m\) = \(n\) rows, \(m\) columns
- \(\text{rank}(A)\) = number of linearly independent vectors of \(A\)
- \(\text{trace}(A) = \text{tr}(A)\) = sum of diagonal elements of \(A\)
- Null matrix = all elements equal to zero.
- Diagonal matrix = all off-diagonal elements are zero.
- \(I\) = identity matrix (diagonal elements: 1, off-diagonal: 0)
- \(|A| = \det(A)\) = determinant of \(A\)
- \(A^{-1}\) = inverse of \(A\)
- \(A' = A^T\) = Transpose of \(A\)
- \(|M_{ij}|\) = Minor of \(A\)
- \(A = A^T\) \(\Rightarrow\) Symmetric matrix
- \(A^T A = A A^T\) \(\Rightarrow\) Normal matrix
- \(A^T = A^{-1}\) \(\Rightarrow\) Orthogonal matrix
- \(A = A^2\) \(\Rightarrow\) Idempotent matrix

5.2 Eigenvalues and Diagonal Systems

- A set of linear simultaneous equations:
  \[A b = d\] \(\Rightarrow\) \(A\) is a non-singular and \(b\) and \(d\) are comformable vectors.

- Under certain circumstances we can diagonalize this system:
  \[\Lambda \nu = \nu\] where \(\Lambda\) is a diagonal matrix.

- Eigenvalues (Characteristic Roots)
  \[A x = \lambda x\] This is the eigenvalue problem
  \(\lambda\) is the eigenvalue (characteristic root)
  \(x\) is the eigenvector (characteristic vector)

- Cauchy discovered them studying how to find new coordinate axes for the graph of the quadratic equation \(ax^2 + 2bxy + cy^2 = d\) so that the equation with the new axes would be of the form \(Ax^2 + Cy^2 = D.\)
5.2 Eigenvalues and Diagonal Systems

• \( Ax = \lambda \ x \)  (Basic equation of eigenvalue problem)

• For the square matrix \( A \), there is a vector \( x \) such that the product of \( Ax \) such that the result is a scalar, \( \lambda \), that, when multiplied by \( x \), results in the same product.

• The multiplication of vector \( x \) by a scalar is the same as stretching or shrinking the coordinates by a constant value. (The matrix \( A \) just scales the vector \( x \)!

• \( Ax = \lambda \ Ix \)  \( \Rightarrow [ A - \lambda \ I] x = 0 \)
• \( K = [ A - \lambda \ I] \) Characteristic matrix of matrix \( A \)
• \( K x = 0 \) Homogeneous equations.

5.2 Eigenvalues and Diagonal Systems

• Homogeneous equations: \( K x = 0 \)
  - Trivial solution \( x = 0 \) (If \( |K| \neq 0 \), from Cramer's rule)
  - Nontrivial solution \( (x \neq 0) \) can occur if \( |K| = 0 \).

• That is, do all matrices have eigenvalues? No. They must be square and \( |K| = |A - \lambda I| = 0 \).

• Eigenvectors are not unique. If \( x \) is an eigenvector, then \( \beta x \) is also an eigenvector: \( A(\beta x) = \lambda (\beta x) \)

• To calculate eigenvectors and eigenvalues, expand the equation \( |A - \lambda I| = 0 \)
• The resulting equation is called characteristic equation.
5.2 Eigenvalues and Diagonal Systems

• Characteristic equation: \(|\mathbf{A} - \lambda \mathbf{I}| = 0\)

Example: For a 2x2 matrix:

\[
[\mathbf{A} - \lambda \mathbf{I}] = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix}
a_{11} - \lambda & a_{12} \\
a_{21} & a_{22} - \lambda
\end{bmatrix}
\]

\[
|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix}
a_{11} - \lambda & a_{12} \\
a_{21} & a_{22} - \lambda
\end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0
\]

\[
a_{11}a_{22} - a_{12}a_{21} - \lambda(a_{11} + a_{22}) + \lambda^2 = 0
\]

• For a 2-dimensional problem, we have a simple quadratic equation with two solutions for \(\lambda\).

5.2 Eigenvalues and Diagonal Systems

• For \(n=2\), we have a simple quadratic equation with two solutions for \(\lambda\). In fact, there is generally one eigenvalue for each dimension, but some may be zero, and some complex.

\[
0 = a_{11}a_{22} - a_{12}a_{21} - (a_{11} + a_{22})\lambda + \lambda^2
\]

\[
\lambda = \frac{(a_{11} + a_{22}) \pm \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}}{2}
\]

• Note: The solution for \(\lambda\) can be written as:

\[
\lambda = \frac{1}{2}\text{trace}(\mathbf{A}) \pm \frac{1}{2} \sqrt{(\text{trace}(\mathbf{A})^2 - 4|\mathbf{A}|)^{1/2}}
\]

Three cases:

1) Real different roots: \(\text{trace}(\mathbf{A})^2 > 4|\mathbf{A}|\)
2) One real root: \(\text{trace}(\mathbf{A})^2 = 4|\mathbf{A}|\)
3) Complex roots: \(\text{trace}(\mathbf{A})^2 < 4|\mathbf{A}|\)
5.2 Eigenvalues and Diagonal Systems

- **Note:** If $A$ is symmetric, the eigenvalues are real. That is, we need to have $\text{trace}(A)^2 > 4|A|$. For $n=2$, we check this condition:

$$
\begin{align*}
(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21}) &> 0 \\
a_{11}^2 + a_{22}^2 + 2a_{11}a_{22} - 4a_{11}a_{22} + 4a_{12}^2 &> 0 \\
a_{11}^2 + a_{22}^2 - 2a_{11}a_{22} + 4a_{12}^2 &> 0 \\
(a_{11} - a_{22})^2 + 4a_{12}^2 &> 0
\end{align*}
$$
5.2 Eigenvalues and Diagonal Systems

• Geometric Interpretation:

The $x,y$ values of $A$ can be seen as representing points on an ellipse centered at (0,0). The eigenvectors are in the directions of the major and minor axes of the ellipse, and the eigenvalues are the lengths of these axes to the ellipse from (0,0).

Example: A correlation matrix

$$ A = \begin{bmatrix} 1 & 0.75 \\ 0.75 & 1 \end{bmatrix} $$

$$ \lambda = \frac{1}{2} \text{trace}(A) \pm \frac{1}{2} \sqrt{[\text{trace}(A)^2 - 4 |A|]}^{1/2} $$

$$ = \frac{1}{2} 2 \pm \frac{1}{2}[2^2 - 4 \times 0.4735]^{1/2} = 1 \pm \frac{1}{2}[2.25]^{1/2} $$

$$ = 1 \pm \frac{1}{2}[1.5] = 0.25; 1.75 $$

$$ x = [-0.7071, 0.7071]; [0.7071, 0.7071] $$

Note: $x$ is not unique. It is usually imposed that $\|x\|=1$. 
5.2 Eigenvalues and Diagonal Systems

- Graphical interpretation: Correlation as an ellipse, whose major axis is one eigenvalue and the minor axis length is the other:

No correlation yields a circle, and perfect correlation yields a line.

5.2 Eigenvalues and Diagonal Systems: Reig

- Command “eigen,” recover values with $

```r
> A <- cbind( rbind(1,.75), rbind(.75,1) )
> A
   [,1] [,2]
[1,] 1.00 0.75
[2,] 0.75 1.00
> eigen(A)
eigen() decomposition
$values
[1] 1.75 0.25
$vectors
[,1]       [,2]
[1,] 0.7071068 -0.7071068
[2,] 0.7071068  0.7071068.
> lamb <- eigen(A)
> lambda <- lamb$values
> lambda
[1] 1.75 0.25
```
5.2 Eigenvalues: Example

• 2nd order multivariable equations: \(ax^2 + 2kxy + by^2 = c\)

• Represented in a quadratic form with symmetric matrix \(A\):

\[
x^T A x = c, \quad \text{where} \quad x = \begin{pmatrix} x \\ y \end{pmatrix}, \quad A = \begin{pmatrix} a & k \\ k & b \end{pmatrix}
\]

\[
5x^2 + 4xy + 3y^2 = 10 \quad A = \begin{pmatrix} 5 & 2 \\ 2 & 3 \end{pmatrix}
\]

• Eigenvector decomposition:

\[
\lambda_1 = 1.764, \quad x_1 = [0.5257, -0.8507] \\
\lambda_2 = 6.236, \quad x_2 = [-0.8507, -0.5257]
\]

• Symmetric \(A\) \(\Rightarrow\) orthogonal e-vectors!

• Geometrical interpretation: Principal Axes of Ellipse

• Positive Definite \(A\) \(\Rightarrow\) positive real eigenvalues!

5.2 Eigenvalues and Diagonal Systems

• General \(n \times n\) case:

The characteristic determinant \(D(\lambda) = \det(A - \lambda I)\)

is clearly a polynomial in \(\lambda\):

\[
D(\lambda) = \alpha_n \lambda^n + \alpha_{n-1} \lambda^{n-1} + \alpha_{n-2} \lambda^{n-2} + \ldots + \alpha_1 \lambda + \alpha_0
\]

• Characteristic equation:

\[
D(\lambda) = \alpha_n \lambda^n + \alpha_{n-1} \lambda^{n-1} + \alpha_{n-2} \lambda^{n-2} + \ldots + \alpha_1 \lambda + \alpha_0 = 0
\]

There are \(n\) solutions to this polynomial. The set of eigenvalues

is called the spectrum of \(A\). The largest of the absolute values

of the eigenvalues of \(A\) is called the spectral radius of \(A\).

• Eigenvalues are computed using the QR algorithm (1950s) or

the divide-and-conquer eigenvalue algorithm (1990s). They are

computationally intensive. They take \(4n^3/3\) flops.
5.2 Diagonal (Eigen) decomposition

- Let \( A \) be a square \( nxn \) matrix with \( n \) linearly independent eigenvectors, \( x_i \) \((i=1,2,...,n)\). Then \( A \) can be factorized as \( A=X \Lambda X^{-1} \)

where \( X \) is the square \((nxn)\) matrix whose \( i \)th column is the eigenvector \( x_i \) of \( A \) and \( \Lambda \) is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, \( \text{i.e., } \Lambda_{ii} = \lambda_i \).

- The eigenvectors are usually normalized, but they need not be. A non-normalized set of eigenvectors can also be used as the columns of \( X \).

**Proof:**

\[ Ax = \lambda x \quad \Rightarrow AX = X \Lambda \quad \Rightarrow A = X \Lambda X^{-1} \quad (X^{-1} \text{ exists}) \]

- Conversely: \( X^{-1} AX = A \)
- If \( X^T X = I \), \( A \) is orthogonally diagonalizable.

5.2 Diagonal decomposition: Example

- Let \( A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \); \( \lambda_1 = 1, \lambda_2 = 3 \).

The eigenvectors: \( x_1 = [1, -1], x_2 = [1, 1] \).

- Let \( X \) be the matrix of eigenvectors: \( X = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \)

- Inverting, we have \( X^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \)

- Then, \( A = X \Lambda X^{-1} = \begin{bmatrix} 1 & 1 & 1/2 & -1/2 \\ -1 & 1/2 & 1/2 \end{bmatrix} \)
5.2 Diagonal decomposition: Example

- Diagonalizing a system of equations:
  \[ Ax = y \]
  Pre-multiply both sides by \( X^{-1} \):
  \[ X^{-1}Ax = X^{-1}y = v \]
  \[ X^{-1}A(XX^{-1})x = v \] (Let \( v = X^{-1}x \))
  \[ \Rightarrow \Lambda v = v \]
- Using the (2x2) previous example:
  \[
  \begin{bmatrix}
  1 & 0 \\
  0 & 3 
  \end{bmatrix}
  \begin{bmatrix}
  1/2 \\
  1/2 
  \end{bmatrix}
  =
  \begin{bmatrix}
  1/2 \\
  1/2 
  \end{bmatrix}
  \begin{bmatrix}
  y_1 \\
  y_2 
  \end{bmatrix}
  \]

  \[
  u_1 = \frac{1}{2} x_1 - \frac{1}{2} x_2 \\
  u_2 = \frac{1}{2} x_1 + \frac{1}{2} x_2 
  \]

  \[
  \frac{1}{2} y_1 - \frac{1}{2} y_2 = v_1 \\
  \frac{1}{2} y_1 + \frac{1}{2} y_2 = v_2 
  \]

  \[
  u_1 = v_1 \\
  3u_2 = v_2 
  \]

5.2 Diagonal decomposition: Application

- Let \( M \) be the square \( n \times n \) matrix defined by: \( M = I_n - Z(Z'Z^{-1})Z' \), where \( Z \) is an \( n \times k \) matrix, with rank(\( Z \)) = \( k \).

  Let’s calculate the trace(\( M \)):
  \[ \text{trace}(M) = \text{trace}(I_n - Z(Z'Z^{-1})Z') = \text{trace}(I_n) - \text{trace}(Z(Z'Z^{-1})Z') = n - \text{trace}(Z'Z^{-1}Z') = n - \text{trace}(I_k) = n - k. \]

  It is easy to check that \( M \) is idempotent (\( \lambda_i \)'s are all 0 or 1) and symmetric (\( \lambda_i \)'s are all real and \( x \) are orthogonal).

  Write an orthogonal diagonalization: \( M = XX^{-1} \) \( (XX^{-1} = I) \).

  Again, let’s calculate the trace(\( M = XX^{-1} \)):
  \[ \text{trace}(M) = \text{trace}(XX^{-1}X) = \text{trace}(XX^{-1}X) = \text{trace}(\Lambda) = \sum \lambda_i \]
  That is, \( M \) has \( n - k \) non-zero eigenvalues.
5.2 Why do Eigenvalues/vectors matter?

- Eigenvectors are invariants of $A$
  - Don't change direction when operated $A$

- Use to determine the definiteness of a matrix.

- A singular matrix has at least one zero eigenvalue.

- Solutions of multivariable differential equations (the bread-and-butter in linear systems) correspond to solutions of linear algebraic eigenvalue equations.

- Eigenvalues are used to study the stability of autoregressive time series models.

- The orthogonal basis of eigenvectors forms the core of principal components analysis (PCA).
5.2 Sign of a quadratic form: Eigenvalue tests

• Suppose we are interesting in an optimization problem for \( z = f(x,y) \). We set the first order conditions (f.o.c.), solve for \( x^* \) and \( y^* \), and, then, check the second order conditions (s.o.c.).

• Let’s re-write the s.o.c. of \( z = f(x,y) \):

\[
d^2 z = q = f_{xx} dx^2 + 2 f_{xy} dx dy + f_{yy} dy^2
\]

\[
q = \begin{bmatrix} dx & dy \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = u' Hu
\]

• The s.o.c. of \( z = f(x,y) \) is a quadratic form, with a symmetric matrix, \( H \).

• To determine what type of extreme points we have, we need to check the sign of the quadratic form.

5.2 Sign of a quadratic form: Eigenvalue tests

• Quadratic form:

\[
q = u' H u
\]

(note: the Hessian (\( H \)) is a symmetric matrix)

• Let \( u = Ty \), where \( T \) is the matrix of eigenvectors of \( H \), such that \( T' T = I \)

• Then,

\[
q = y' T' H T y = y' \Lambda y \quad \text{(note: } T' H T = \Lambda)\]

\[
q = \lambda_1 y_1^2 + \lambda_2 y_2^2 + ... + \lambda_i y_i^2 + ... + \lambda_n y_n^2
\]

\[\Rightarrow\] The sign(q) depends on the \( \lambda_i \)'s only.

• We say:

- \( q \) is positive definite iff \( \lambda_i > 0 \) for all i.
- \( q \) is positive semi-definite iff \( \lambda_i \geq 0 \) for all i.
- \( q \) is negative semi-definite iff \( \lambda_i \leq 0 \) for all i.
- \( q \) is negative definite iff \( \lambda_i < 0 \) for all i.
- \( q \) is indefinite if some \( \lambda_i > 0 \) and some \( \lambda_i < 0 \).
5.2 Sign of a quadratic form: Eigenvalue tests

- Example: Find extreme values for \( z = f(x,y) \), and determine if they are a max or min.

\[
z = x^2 + xy + 2y^2 + 3
\]

F.o.c.

\[
\begin{align*}
\frac{\partial f}{\partial x} &= 2x + y = 0 \\
\frac{\partial f}{\partial y} &= x + 4y = 0
\end{align*}
\]

\( y^* = 0, \quad x^* = 0, \quad z^* = 3 \)

Calculate matrix of second derivatives

\[
|H| = \begin{bmatrix}
f_{xx} & f_{xy} \\
f_{yx} & f_{yy}
\end{bmatrix} = \begin{bmatrix}
2 & 1 \\
1 & 4
\end{bmatrix}, \quad \lambda_1 = 1.58582; \quad \lambda_2 = 4.4142
\]

\( \Rightarrow \lambda_1 \) and \( \lambda_2 \) are positive, \( q \) is positive definite

\( \Rightarrow \) \( z^* \) is minimum

5.3 Vector multiplication: Geometric interpretation

- Think of a vector (an *Euclidian vector*) as a directed line segment in \( N \)-dimensions! (has “length” and “direction”)

- Scalar multiplication (“scales” the vector –i.e., changes length)

- Source of linear dependence

\[-1.u = [-1 \quad -2] \]

\[
\begin{bmatrix}
6 \\ 4
\end{bmatrix} = 2u
\]

\[
\begin{bmatrix}
3 \\ 2
\end{bmatrix} = u
\]
5.3 Vector Addition: Geometric interpretation

- \( v' = [2 \ 3] \)
- \( u' = [3 \ 2] \)
- \( w' = v' + u' = [5 \ 5] \)

**Note:** Two vectors plus the concepts of addition and multiplication can create a two-dimensional space.

A vector space is a mathematical structure formed by a collection of vectors, which may be added together and multiplied by scalars. (It's closed under multiplication and addition.) Giuseppe Peano in 1888 gave a precise definition to this concept.

5.3 Vector (Linear) Space

- We introduce an algebraic structure called vector space over a field. We use it to provide an abstract notion of a vector: an element of such algebraic structure.

- Given a field \( R \) and a set \( V \) of objects, on which “vector addition” \( (V \times V \rightarrow V) \), denoted by “\( + \)”, and “scalar multiplication” \( (R \times V \rightarrow V) \), denoted by “\( . \)”, are defined.

  If the following axioms are true for all objects \( u, v, \) and \( w \in V \) and all scalars \( c \) and \( k \) in \( R \), then \( V \) is called a vector space and the objects in \( V \) are called vectors.

  1. \( u + v \in V \) (closed under addition).

  2. \( u + v = v + u \) (vector addition is commutative).

  3. \( \emptyset \in V \), such that \( u + \emptyset = u \) (\( \emptyset \) = null element).

  4. \( u + (v + w) = (v + u) + w \) (distributive law of vector addition).
5.3 Vector Space

5. For each \( v \), there is a \(-v\) such that \( v + (-v) = \emptyset \)

6. \( c \cdot u \in V \) (closed under scalar multiplication).

7. \( c \cdot (k \cdot u) = (c \cdot k) \cdot u \) (scalar multiplication is associative).

8. \( c \cdot (v + u) = (c \cdot v) + (c \cdot u) \)

9. \((c + k) \cdot u = (c \cdot u) + (k \cdot u) \)

10. \( 1 \cdot u = u \) (unit element).

11. \( 0 \cdot u = \emptyset \) (zero element).

We can write \( S = \{ V, R, +, \cdot \} \) to denote an abstract vector space.

This is a general definition. If the field \( R \) represents the real numbers, then we define a real vector space.

Giuseppe Peano (1858 – 1932, Italy)

5.3 Vector Space

- **Definition:** Linear Combination

Given vectors \( u_1, \ldots, u_k \), the vector \( w = \epsilon_1 u_1 + \ldots + \epsilon_k u_k \) is called a linear combination of the vectors \( u_1, \ldots, u_k \).

Notation: \( \langle u_1, \ldots, u_k \rangle \) is the set of all linear combinations of \( u_i \)’s.

- **Definition:** Subspace

Given the vector space \( V \) and \( W \) a set of vectors, such that \( W \subseteq V \). Then, \( W \) is a subspace iff:

- \( u, v \in W \Rightarrow u + v \in W \), and

- \( c \cdot u \in W \) for every \( c \in R \).

Thus, a nonempty subset \( W \) of a vector space \( V \) that is closed under addition and scalar multiplication (and contains the 0-vector of \( V \)) is a subspace of \( V \). That is, a subspace is a subset of \( V \) that can be considered a vector space!
5.3 Vector Space

- **Definition:** Spanning set
  Given the set $Z$ in $V$ and $U = \{u_1, \ldots, u_k\}$ in $Z$, we say $U$ spans $Z$, or $U$ is a *spanning set* for $Z$, if $Z \in \langle u_1, \ldots, u_k \rangle$ or $Z \in \langle U \rangle$.

- **Definition:** Basis set (“basis”)
  Given $U = \{u_1, \ldots, u_k\}$ and a subspace $W \subseteq V$. We say that if $U$ is LI and it is a spanning set for $W$, then $U$ is called a *basis* set for $W$.

  **Example:** The $N$-dimensional subspace $W_N$ of the $V$ space ($N=2$).

\[
W_N = \text{span}\{u_1, \ldots, u_k\}
\]

5.3 Vector Space

- **Theorem:**
  If a vector space has a basis with a finite number $N$ of elements, then every other basis also has $N$ elements.

- **Definition:** Space dimension
  If a vector space $V$ has a basis with $N < \infty$ elements, we say that $V$ is a finite dimensional vector space and that $V$ has *dimension* $N$, or $N = \dim(V)$.

- **Theorem:**
  The vector space consisting of $n$-column vectors, with vector addition and multiplication corresponding to matrix operations is an $n$-dimensional vector space (*Euclidean n-space*), which we will denote $\mathbb{R}^n$. 

32
Definition: Maximally linear independent (max-LI) subset
Given a set \( U = \{u_1, \ldots, u_k\} \) of vectors in a vector space \( V \). If \( T = \{u_i, \ldots, u_q\} \in U \) containing \( q \) vectors is LI and every subset with more than \( q \) vectors is LD, then \( T \) is called maximally linear independent subset of \( U \). Moreover, we will call \( q \) the rank of the set \( U \), written \( q = \text{rank}(U) \).

It is common to say the rank of a matrix \( A \) equals “the number of LI columns (rows)” in \( A \). This is OK, but keep in mind that in general there is not a unique subset of \( q \) LI columns or rows that are max-LI.

Definition: Full rank
Given \( A \) \((m \times n)\). We say \( A \) has full rank if \( \text{rank}(A) = \min(m,n) \).

5.3 Vector Space

- As we defined them, vector spaces do not provide enough structure to study issues in real analysis, for example convergence of sequences. More structure is needed.

- For example, we can introduce as an additional structure the concept of order \((\leq)\), to compare vectors. This additional structure creates ordered vector spaces.

- We can introduce a norm, which we will use to measure the length or magnitude of vectors. This creates a normed vector space, denoted as a pair \((V, \|\cdot\|)\) where \( V \) is a vector space and \( \|\cdot\| \) is a norm on \( V \).
5.3 Notes on Vector Operations

• An \((m\times1)\) column vector \(u\) and a \((1\times n)\) row vector \(v\), yield a product matrix \(uv\) of dimension \((m\times n)\).

\[
\begin{bmatrix}
3 \\
2 \\
1
\end{bmatrix}_{3 \times 1}
\begin{bmatrix}
1 & 4 & 5
\end{bmatrix}_{1 \times 3}
\]

\[
u'v = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}_{3 \times 1} \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}_{1 \times 3} = \begin{bmatrix} 3 & 12 & 15 \\ 2 & 8 & 10 \\ 1 & 4 & 5 \end{bmatrix}
\]

A matrix

\[
u'v = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}_{1 \times 3} = 16
\]

A scalar

5.3 Vector Multiplication: Dot (inner) Product

• The dot product, “\(\cdot\)”, is a function that takes pairs of vectors and produces a number. For two vectors, \(c\) and \(z\), it is defined as:

\[
c \cdot z = c_1z_1 + c_2z_2 + c_3z_3 + \ldots + c_nz_n = \sum_{i=1}^{n} c_i z_i
\]

\[
y = c \cdot z = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = c'z
\]

• The dot product produces a scalar! \(y = c'z = 1\times1(=1\times n\times 1) = z'c\). Note that from the definition, the dot product is commutative.

• When \(c\) is a vector of 1’s, usually noted as \(t\), then:

\[
t \cdot z = 1 \cdot z_1 + 1 \cdot z_2 + 1 \cdot z_3 + \ldots + 1 \cdot z_n = \sum_{i=1}^{n} z_i
\]
5.3 Vectors: Dot Product

• Dot products in econometrics are common. For example, the Residual Sum of Squares (RSS), where \( e \) is a vector of residuals:

\[
\mathbf{e} \cdot \mathbf{e} = e' e = e_1^2 + e_2^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2
\]

• It is possible to define an inner product for functions. Instead of a sum over the corresponding elements of a vector, the inner product on functions is defined as an integral over some interval.

• Some intuition.
- The dot product is used as a tool to define size for vectors:

\[
\|\alpha\| = (\alpha^T \alpha)^{1/2} = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2}
\]
- Now, we can compare vectors and measure “distances” between vectors and, eventually, convergence!

5.3 Vectors: Dot Product - Properties

• The dot product fulfills the following properties if \( \alpha, \beta, \) and \( \gamma \) are real vectors and \( k \) is a scalar.

1. Commutative: \( \alpha \cdot \beta = \langle \alpha, \beta \rangle = \beta \cdot \alpha \)

   **Note:** We say \( \alpha \) and \( \beta \) (non-zero vectors) are **orthogonal** iff \( \alpha \cdot \beta = 0 \)

2. Distributive over vector addition: \( \alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma \)

3. Bilinear: \( \alpha \cdot (k\beta + \gamma) = k \cdot (\alpha \cdot \beta) + \alpha \cdot \gamma \)

4. Scalar multiplication: \( (k_1 \alpha) \cdot (k_2 \beta) = k_1 k_2 \cdot (\alpha \cdot \beta) \)

   **Note:** Nice, intuitive properties.

**Notation:** \( \alpha \cdot \beta = \langle \alpha, \beta \rangle \)
5.3 Vectors: Dot Product & Size

• The magnitude (length or size) is the square root of the dot product of a vector with itself (just like the Pythagorean theorem):

\[ \|\mathbf{a}\| = [a_a^T a_a]^{1/2} = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2} \]

• Now, we can talk about the size of vectors. We can apply this definition to define other concepts, for example, convergence, in a similar fashion as in calculus: A sequence of vectors \( x_n \) converge to a point \( c \) if \( \|x_n - c\| \) decreases to 0 as \( n \) increases.

• Useful property: If \( k \) is a scalar, then the size of a vector times \( k \) is \(|k|\) times the size of the vector.

\[ \|k\mathbf{a}\| = [k a_a^T a_a]^{1/2} = |k| \|\mathbf{a}\| \]

Note: If we set \( k = 1/\|\mathbf{a}\| \) \( \Rightarrow \) \( (1/\|\mathbf{a}\|) \mathbf{a} \) \( = 1 \).

\[ \Rightarrow \text{Nice result, used to normalize vectors.} \]

5.3 Vectors: Dot Product – Geometry

• There is a geometric interpretation to the dot product.

• Any two vectors, say \( \mathbf{a} \) and \( \mathbf{b} \), determine a plane. We can measure the angle between the two vectors. The inner product connects the length and the angle between the vectors:

\[ \mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos(\theta) \]

The dot product is related to the angle between the two vectors – but it doesn’t tell us the angle.

Notes: As the \( \cos(\theta = 90) = 0 \), the dot product of two perpendicular (orthogonal, or “\( \perp \)”) vectors is zero.

• In the CLM, we have \( \mathbf{y} = \mathbf{Xb} + \mathbf{e} = \text{Projection + residual} \)

Then: \( \mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{y} - \mathbf{Xb}) = \mathbf{X}'\mathbf{y} - \mathbf{XX}(\mathbf{XX})^{-1}\mathbf{X}'\mathbf{y} = 0 \quad \Rightarrow (\mathbf{X} \perp \mathbf{e}). \)
5.3 Vectors: Magnitude and Phase (direction)

- **Magnitude:**
  \[ \|v\| = \sqrt{\sum_{i=1}^{n} x_i^2} \]  
  (Magnitude or “2-norm”)
  
  If \( \|v\| = 1 \), \( v \) is a unit vector.
  
  If \( u = \frac{1}{\|v\|} v \), then \( u \) is a unit vector.
  
  (unit vector \( \Rightarrow \) pure direction)

- **Phase:**
  \[ \cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} \]

5.3 Vectors: Norm

- **Given a vector space** \( V \), the function \( g: V \rightarrow \mathbb{R} \) is called a norm iff:
  1) \( g(x) \geq 0 \), for all \( x \in V \)
  2) \( g(x) = 0 \) iff \( x = \theta \) (empty set)
  3) \( g(\alpha x) = |\alpha| g(x) \) for all \( \alpha \in \mathbb{R}, x \in V \)
  4) \( g(x+y) \leq g(x) + g(y) \) (“triangle inequality”) for all \( x, y \in V \)

  The norm is a generalization of the notion of size or length of a vector.

**Example:** On \( \mathbb{R}^n \), the **Euclidian norm** of \( x = (x_1, x_2, ..., x_n) \) is given by
\[ \|x\| = \|x\|^{1/2} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} \]

while the **Manhattan (Taxicab) norm** is defined as:
\[ \|x\|_1 = \sum_{i=1}^{n} |x_i| \]

**Note:**
- Euclidian norm = \( L_2 \) norm (2-norm).
- Manhattan norm = \( L_1 \) norm (1-norm).
### 5.3 Vectors: Norm

- We can generalize the concept of norm.

**Definition:** $L^p$ norm

For a real number $p \geq 1$, the $L^p$-norm (or $p$-norm) of $\mathbf{x}$ is defined by

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + ... + |x_n|^p)^{1/p}$$

- An infinite number of functions can be shown to qualify as norms. For vectors in $\mathbb{R}^n$, we have the following examples:
  
  $g(\mathbf{x}) = \max_i (x_i)$; $g(\mathbf{x}) = \sum_i |x_i|$; $g(\mathbf{x}) = [\sum_i (x_i)^4]^{1/4}$

- Given a norm on a vector space, we can define a measure of “how far apart” two vectors are, using the concept of a **metric**.

### 5.3 Vectors: Metric

- Given a vector space $V$, the function $d: V \times V \rightarrow \mathbb{R}$ is called a **metric** or a **distance function** if and only if:
  1) $d(\mathbf{x}, \mathbf{y}) \geq 0$, (“positive”) for all $\mathbf{x}, \mathbf{y} \in V$
  2) $d(\mathbf{x}, \mathbf{y}) = 0$ (“non-degenerate”) iff $\mathbf{x} = \mathbf{y}$
  3) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ (“symmetric) for all $\mathbf{x}, \mathbf{y} \in V$
  4) $d(\mathbf{x} + \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ (“triangle inequality”) for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$

Given a norm $g(\cdot)$, we can define a metric by the equation:

$$d(\mathbf{x}, \mathbf{y}) = g(\mathbf{x} - \mathbf{y}).$$

Check:

1) and 2) follow immediately from properties of $g(\cdot)$

3) $d(\mathbf{x}, \mathbf{y}) = g(\mathbf{x} - \mathbf{y}) = g((-1)(\mathbf{y} - \mathbf{x})) = |-1| g(\mathbf{y} - \mathbf{x}) = g(\mathbf{y} - \mathbf{x}) = d(\mathbf{y}, \mathbf{x})$.

4) $(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{z}) + (\mathbf{z}, \mathbf{y})$ 

$$\Rightarrow g(\mathbf{x}, \mathbf{y}) \leq g(\mathbf{x}, \mathbf{z}) + g(\mathbf{z}, \mathbf{y})$$

$$\Rightarrow d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$$
5.3 Vectors: Metric

Example: On $\mathbb{R}^n$, the distance between two points is usually given by the 2-norm distance. But, other distances are possible.

1-norm distance: $d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$

2-norm distance: $d(x, y) = (\sum_{i=1}^{n} |x_i - y_i|^2)^{1/2}$

$p$-norm distance: $d(x, y) = (\sum_{i=1}^{n} |x_i - y_i|^p)^{1/p}$

The red, yellow, and blue lines have the same length (12) —i.e., same $L_1$ distance.

The green lines is the $L_2$ distance, the shortest distance, which is unique.

5.3 Vectors: Metric

• **Theorem**: Cauchy-Schwarz Inequality

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a real inner product space, then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| \cdot ||\mathbf{v}||$$

Note: This result can be written as

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \cdot \langle \mathbf{v}, \mathbf{v} \rangle$$

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq ||\mathbf{u}||^2 \cdot ||\mathbf{v}||^2$$

General Proof: Trivial proof when $\nu = 0$, so we assume that $<\nu, \nu> \neq 0$. 

23
5.3 Vectors: Metric

• Theorem: \( |\langle u, v \rangle| \leq \| u \| \| v \| \)

Proof: Let \( \delta \) be any number in the field \( F \). Then,

\[
0 \leq \| u - \delta v \|^2 = \langle u - \delta v, u - \delta v \rangle = \langle u, u \rangle - \delta \langle v, u \rangle - \delta \langle u, v \rangle + \delta^2 \langle v, v \rangle
\]

Choose the value of \( \delta \) that minimizes this quadratic form:

\[
\delta = \frac{\langle u, v \rangle}{\langle v, v \rangle}
\]

(A quick way to remember this value of \( \delta \) is to imagine \( F \) to be the real numbers, and set the derivative equal to zero to pick \( \delta \).)

We get

\[
0 \leq \langle u, u \rangle - |\langle u, v \rangle|^2 \langle v, v \rangle^{-1}
\]

which is true if and only if

\[
|\langle u, v \rangle|^2 \leq \langle u, u \rangle < \langle v, v \rangle
\]

or equivalently:

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

5.3 Vectors: Metric Space

• Definition: Metric Space

A metric on a space \( M \) is a mapping \( d(\cdot, \cdot): M \times M \rightarrow [0, \infty) \) satisfying the metric properties (1) through (4) for all \( x, y \) and \( z \) in \( M \). A space endowed with a metric is called a metric space.

• \( M \) is not necessarily a vector space. By definition, any space endowed with a metric is a metric space. For example, the space of density functions on \([0,1]\) endowed with the metric:

\[
d(f,g) = \int_0^1 \left( \sqrt{f(u)} - \sqrt{g(u)} \right)^2 du.
\]

• This space is not a vector space (& not possible to define \( <.>! \))
5.3 Vectors: Hilbert Space

- **Definition:** pre-Hilbert space
  A vector space \( V \) endowed with an inner product \(<x,y>\) and associated norm \( \|x\| = \sqrt{<y,x>} \) and metric \( \|x-y\| \) is called a pre-Hilbert space.

- We say “pre” because a fundamental property is still missing, namely that every Cauchy sequence has a limit in \( V \).

- **Definition:** Cauchy sequence
  A sequence of elements \( x_n \) of a metric space with metric \( d(.,.) \) is called a Cauchy sequence if for every \( \varepsilon > 0 \) there exists an \( n_0(\varepsilon) \) such that for all \( k,m \geq n_0(\varepsilon) \), \( d(x_k, x_m) < \varepsilon \).

  **Example:** In \( \mathbb{R}^p \) with finite dimension \( p \) every Cauchy sequence converges to a limit in \( \mathbb{R}^p \).

- **We want the Hilbert space to be complete —i.e., every Cauchy sequence has a limit in the space. A very useful property.**

- If we have a convergent sequence of points, then it actually converges to a point in the space. Contrast that with, say, the rational numbers, \( \mathbb{Q} \). We can have a convergent sequence —say, approximations to \( \pi \) — that do not converge to a point in \( \mathbb{Q} \), because, obviously \( \pi \) is irrational.

- This situation does not occur in a Hilbert space: If we have a convergent sequence of vectors to a point \( p \), then \( p \) is in the space. Now, the techniques of calculus can be used.

- Completeness also makes a Hilbert space closed under convergence, which generates useful properties.
5.3 Vectors: Hilbert Space

- **Definition**: Hilbert space
  
  A *Hilbert space* $H$ is a vector space endowed with an inner product and associated norm and metric such that every Cauchy sequence in $H$ has a limit in $H$.

- A Hilbert space is a special case of a Banach space. A Banach space is a complete normed vector space. In a Hilbert space we specified a norm, the inner product.

- If the space is not complete, $H$ is known as an *inner product space*.

- Usually, in linear algebra, we are familiar with some vector spaces. They are $\mathbb{R}^n$ or $\mathbb{C}^n$. These are also Hilbert spaces.

---

**Example I**: The space $V$ of random variables defined on a common probability space $\{\Omega, F, P\}$ with finite second moments, endowed with the inner product $\langle X, Y \rangle = \mathbb{E}[XY]$ and associated norm $\|X\| = \sqrt{\langle X, X \rangle}$ and metric $\|X-Y\|$. The space $V$ is a Hilbert space.

**Example II**: $L^2$, the set of all functions $f: \mathbb{R} \to \mathbb{R}$ such that the integral of $f^2$ over the whole real line is finite. In this case, the inner product is

$$\langle f, g \rangle = \int f(x) g(x) \, dx$$
5.3 Vectors: Hilbert Space


- Hilbert Space - Summary:
  - Generalization of Euclidean space (R^2, R^3).
  - Abstract vector (linear) space with an inner product, complete.
  - Nice properties: linear space, inner product, sums that should converge do converge, calculus can be used.

David Hilbert (1862–1943, Germany)

5.3 Vectors: Orthogonality

- **Theorem:** (Generalized Law of Pythagoras)
  If \( \mathbf{u} \) and \( \mathbf{v} \) are orthogonal vectors in an inner product space, then
  \[
  \| \mathbf{u} + \mathbf{v} \|^2 = \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2
  \]
  **Proof:** It follows from \( \langle \mathbf{u}, \mathbf{v} \rangle = 0 \).

- **Definition:** Orthogonal Complement
  Let \( W \) be a subspace of an inner product space \( V \). A vector \( \mathbf{u} \) in \( V \) is said to be orthogonal to \( W \) if it is orthogonal to every vector in \( W \). The set of all vectors in \( V \) that are orthogonal to \( W \) is called the orthogonal complement of \( W \).

  **Notation:** We denote the orthogonal complement of a subspace \( W \) by \( W^\perp \). [read “\( W \) perp”]
5.3 Vectors: Orthogonality

- **Theorem:** Properties of Orthogonal Complements
  If \( W \) is a subspace of a finite-dimensional inner product space \( V \), then
  - \( W \perp \) is a subspace of \( V \).
  - The only vector common to both \( W \) and \( W \perp \) is \( 0 \).
  - The orthogonal complement of \( W \perp \) is \( W \); that is \( (W \perp) \perp = W \).

- **Theorem:**
  If \( W \) is a subspace of \( \mathbb{R}^n \), then \( \dim(W) + \dim(W \perp) = n \).
  Furthermore, if \( \{u_1, ..., u_k\} \) is a basis for \( W \) and \( \{u_{k+1}, ..., u_n\} \) is a basis for \( W \perp \), then \( \{u_1, ..., u_k, u_{k+1}, ..., u_n\} \) is a basis for \( \mathbb{R}^n \).

5.3 Vectors: Projections

- **Definition:** Projection
  Let \( u \) and \( v \) be two non-zero vectors in an inner product space \( V \). Then, the *scalar projection* of \( u \) onto \( v \) is defined as
  \[
  k = \frac{\langle u, v \rangle}{\| v \|} = \frac{\langle u, v \rangle}{\| v \|^2}
  \]
  The *vector projection* of \( u \) onto \( v \) is
  \[
  p = k \frac{v}{\| v \|} = \frac{\langle u, v \rangle}{\| v \|^2} v
  \]

Derivation: Given two vectors: \( v \) in \( S \) & \( u \) in \( \mathbb{R}^n \). We want to find \( p \), the vector in \( S \) closest to \( u \). Let \( p = kv \).

To minimize \( \| u - p \| \) with respect to \( k \).
\[
\| u - p \|^2 = \| u - kv \|^2 = (u - kv) \cdot (u - kv) = u \cdot u - 2ku \cdot v + k^2 v \cdot v
\]
\[
d(\| u - p \|^2) / dk = -2u \cdot v + 2k v \cdot v = 0 \quad \Rightarrow \quad k = u \cdot v / v \cdot v
\]
\[
\Rightarrow p = (u \cdot v / v \cdot v) v
\]
5.3 Vectors: Projections

Lemma: Let $v$ be a non-zero vector and $p$ be the projection of $u$ onto $v$. Then,

(i) $(u - p) \perp p$
(ii) $u = p \Leftrightarrow u = k \cdot v$ for some $k$

Proof: Recall $p = k \cdot v = (u \cdot v / v \cdot v) \cdot v$

(i) $< p, u - p > = < p, u > - < p, p > =
\quad = | < u, v > |^2 / ||v|| - | < u, v > |^2 / ||v|| = 0$
\quad \Rightarrow (u - p) \perp p$

(ii) straightforward.

5.3 Vectors: Projections - CLM

- In the CLM, we have a “Projection matrix”, $P$: $P = X(X^T)X^T$ \hspace{1cm} (X is $N \times k$ \hspace{0.5cm} $\Rightarrow$ \hspace{0.5cm} $P$ is $N \times N$)

- Features
  $Py = X(X^T)X^T = Xb = \hat{y}$ \hspace{0.5cm} (fitted values)
  $Py$ is the projection of $y$ into the column space of $X$.
  $PM = P[ I - X(X^T)^{-1}X ] = MP = 0$ \hspace{0.5cm} ($M$: residual maker)
  $PX = X$

- Properties
  - $P$ is symmetric \hspace{0.5cm} $P = P'$
  - $P$ is idempotent \hspace{0.5cm} $P^2 = P$
  - $P$ is singular \hspace{0.5cm} $P^{-1}$ does not exist. \hspace{0.5cm} $\Rightarrow$ \hspace{0.5cm} rank($P$) = $k$
5.3 Vectors: Projections - CLM

- We have two ways to look at \( y \):
  \[
y = X\beta + \varepsilon = \text{Conditional mean + disturbance}
  \]
  \[
y = Xb + e = \text{Projection + residual}
  \]

- Note: \( X'e = X'(y-Xb) = X'y - XX(X'X)^{-1}X'y = 0 \).

5.3 Vectors: Orthonormal Basis

- **Basis**: A space is totally defined by a set of vectors – any point is a linear combination of the basis
- **Ortho-Normal**: orthogonal + normal
- **Orthogonal**: dot product is zero – i.e., vectors are perpendicular
- **Normal**: magnitude is one

- Example: \( X, Y, Z \) (but don’t have to be!)
  \[
  x = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \quad x \cdot y = 0
  \]
  \[
  y = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \quad x \cdot z = 0
  \]
  \[
  z = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \quad y \cdot z = 0
  \]

- \( X, Y, Z \) is an orthonormal basis. We can describe any 3D point as a linear combination of these vectors.
### 5.3 Vectors: Orthonormal Basis

- How do we express any point as a combination of a new basis \( U, V, N \), given \( X, Y, Z \)?

\[
\begin{bmatrix}
    a & 0 & 0 \\
    0 & b & 0 \\
    0 & 0 & c \\
\end{bmatrix}
\begin{bmatrix}
    u_1 & v_1 & n_1 \\
    u_2 & v_2 & n_2 \\
    u_3 & v_3 & n_3 \\
\end{bmatrix}
= \begin{bmatrix}
    a \cdot u + b \cdot v + c \cdot u \\
    a \cdot v + b \cdot v + c \cdot v \\
    a \cdot n + b \cdot n + c \cdot n \\
\end{bmatrix}
\]

(not an actual formula – just a way of thinking about it)

- To change a point from one coordinate system to another, compute the dot product of each coordinate row with each of the basis vectors.

---

**You know too much linear algebra when...**

You look at the long row of milk cartons at Whole Foods --soy, skim, .5% low-fat, 1% low-fat, 2% low-fat, and whole-- and think: "Why so many? Aren't soy, skim, and whole a basis?"