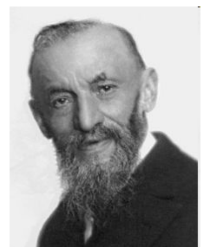
# Chapters 5
## Vector Spaces

David Hilbert (1862–1943, Germany)



Giuseppe Peano (1858–1932, Italy)

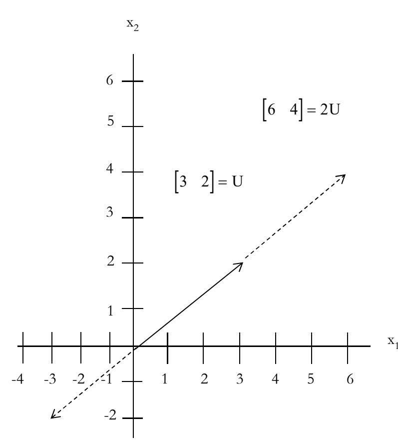




## 5.1 Vector multiplication: Geometric interpretation

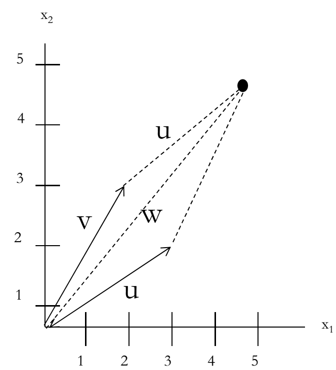
- Think of a vector (an *Euclidian vector*) as a *directed line segment in N-dimensions*! (has "length" and "direction")
- Scalar multiplication ("*scales*" the vector –i.e., changes length)
- Source of linear dependence

$-1 \cdot U = [-3 \quad -2]$



## 5.1 Vector Addition: Geometric interpretation

- **v**' = [2 3]
- **u**' = [3 2]
- **w**' = **v**' + **u**' = [5 5]




- <u>Note</u>: Two vectors plus the concepts of addition and multiplication can create a two-dimensional *space*. (Space = A set of points, say $R^3$, a "universe".)

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

3

## 5.2 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 (of scalars) and a set $V$ of objects (vectors), on which "vector addition" ($V$x$V$→$V$), denoted by "+", and "scalar multiplication" (R$\times V$ →$V$), 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.** $\varnothing \in V$, such that $u + \varnothing = u$ ($\varnothing$ = null element).

**4.** $u + (v + w) = (v + u) + w$ (distributive law of vector addition)

4

## 5.2 Vector Space

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

**6.** $c \, . u \in V$ (closed under scalar multiplication).

**7.** c. $(k \, . \, u) = (c \, . k) \, u$ (scalar multiplication is associative).

**8.** c. $(v + u) = (c. \, v) + (c. \, u)$

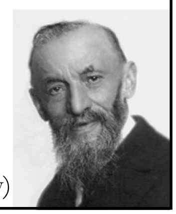
**9.** $(c + k) \, . \, u = (c. \, u) + (k. \, u)$

**10.** $1.u=u$ (unit element).

**11.** $0.u = \emptyset$ (zero element).

We can write S = $\{V, R, +, .\}$ 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)

## 5.2 Vector Space: Examples

**1.** n-dimensional vectors:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in R^n, C^n$$

Note: 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 $R^n$.

**2.** An infinite sequence of real numbers. We will be interested in bounded sequences such that $\{x_k\} < M$, for $M < \infty$.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ \vdots \end{bmatrix} \in R^\infty, C^\infty$$

6

## 5.2 Vector Space: Examples

**3.** The collection of all continuous real valued functions $f(t)$ on the interval C[$a$,$b$] on the real line is a linear vector space. The zero vector is the function identically zero on [$a$,$b$].

**4.** The collection of polynomial functions on the interval [$a$,$b$] is a linear vector space.



## 5.2 Vector Space: Subspace

- Definition: Subspace

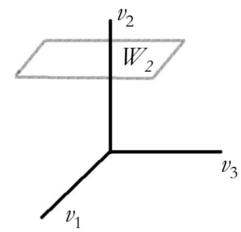
Given the vector space $V$ and $W$ a set of vectors, such that $W \in V$. Then, $W$ is a *subspace* if it also a vector space. That is,

- $u, v \in W \quad \Rightarrow u + v \in V$, and
- $u \in W$ & for every $c \in R \quad \Rightarrow c.u \in V$.
- $W$ contains the **0**-vector of $V$.

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!

## 5.2 Vector Space: Subspace




- **Example**: Subspace

$W_2$ is not a subspace of $V$, it does not include the **0**-vector of $V$.

$W_2$ is just a "*hyperplane*."

- Definition: Linear Combination

Given vectors $u_1, ..., u_k$, the vector $w = c_1 u_1 + .... + c_k u_k$ is called a *linear combination* of the vectors $u_1, ..., u_k$.

Notation: $<u_1, ..., u_k>$ is the set of *all* linear combinations of $u_j$'s.

- Recall that a set of vectors is *linearly dependent* if any one of them can be expressed as a linear combination of the remaining vectors; otherwise, the set is linearly independent (LI).



## 5.2 Vector Space: Rank

- Definition: Maximally linear independent (max-LI) subset

Given a set $U=\{u_1, ..., u_k\}$ of vectors in a vector space $V$. If $W = \{u_{i,1}, ..., u_{i,q}\} \in U$ containing $q$ vectors is LI and every subset with more than $q$ vectors is LD, then $W$ is called *maximally linear independent* subset of U. Moreover, we will call $q$ the *rank* of the set U, written $q = \text{rank}(U)$. (The max-LI subset is not unique.)

Definition: Full rank

Given $\mathbf{A}$ ($m$x$n$). We say $\mathbf{A}$ has *full rank* if $\text{rank}(\mathbf{A}) = \min(m,n)$.

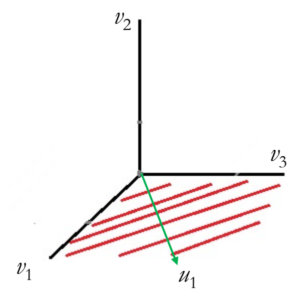
## 5.2 Vector Space: Spanning Set

- Definition: Spanning set

Given the set Z in $V$ and U = $\{u_1, ..., u_k\}$ in Z, we say U spans Z, or U is a *spanning set* for Z, if Z $\in <u_1, ..., u_k>$ or Z $\in$<U>.

That is, a set of vectors *spans* Z if all the vectors in Z can be expressed in terms of this set of vectors.

**Example**: Vectors $v_1$ & $v_3$ span the $v_3 - v_1$ plane. Also, $v_3$ & $u_1$ also span the same $v_3 - v_1$ plane.





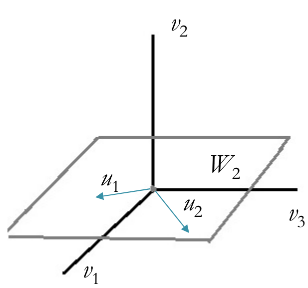

## 5.2 Vector Space: Basis

- Definition: Basis set ("basis")

Given U = $\{u_1, ..., u_k\}$ and a subspace $W \in V$. Then, U is a *basis set* for $W$ if

1) span the subspace $W$,
2) U is linearly independent (LI).

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



## 5.2 Vector Space: Basis and Space Dimension

- **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**:

If $\{u_1, \ldots, u_k\}$ are linearly independent vectors for $k \leq N$, where $N$ is the dimension of the vector space, we can always construct a basis by adding additional independent vectors:

$$\{u_1, \ldots, u_k, u_{k+1}, \ldots, u_N\}.$$



## 5.2 Vector Space: Basis and Space Dimension

- A vector space $V$ can be decomposed into independent subspaces instead of vectors, say $V_1, V_2, \ldots, V_m$. Then,

$$V = V_1 + V_2 + V_3 + \ldots + V_m$$

We call this *direct sum decomposition*. It is similar to a basis decomposition when the $V_i$ all have dimension 1.

**Example**: The 3-dimensional space $V$ can be decomposed as:

$V = v_1 + v_2 + v_3$, where the $v_i$'s are LI.

Alternatively, $V$ can be decomposed as

$V = V_1 + v_3$, where $V_1 = v_1 + v_2$.

- If $V=U+W$, then $dim(V)= dim(U)+dim(W)$, where $U \cap W = \{\mathbf{0}\}$.

## 5.2 Vector Space: Measuring Length

- 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* (≤), 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 ($V$, ‖.‖) where $V$ is a vector space and ‖.‖ is a norm on $V$.

- A normed vector space has a defined mapping from $V \rightarrow R^1$.

15

## 5.3 Notes on Vector Operations

- An ($m$x1) column vector $u$ and a (1x$n$) row vector $v$, yield a product matrix $uv$ of dimension ($m$x$n$).

$$\underset{3\,x\,1}{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \qquad \underset{1x3}{v'} = \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}$$

$$\underset{3\,x\,3}{u\,v'} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 12 & 15 \\ 2 & 8 & 10 \\ 1 & 4 & 5 \end{bmatrix}$$ A matrix

$$\underset{1\,x\,1}{u\,'\,v} = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} = 16$$ A scalar

16

## 5.3 Vector Multiplication: Dot (inner) Product

- The *dot or Inner product (IP)*, "•", is a function that takes pairs of vectors and produces a number. For vectors **c** & **z**, it is defined as:

$$\boldsymbol{c} \bullet \boldsymbol{z} = c_1 * z_1 + c_2 * z_2 + \dots + c_n * z_n = \sum_{i=1}^{n} c_i z_i$$

$$y = \boldsymbol{c} \bullet \boldsymbol{z} = \begin{bmatrix} c_1 & c_2 \dots & c_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_3 \end{bmatrix} = \mathbf{c}'\mathbf{z}$$

- The dot product produces a scalar! y = **c'z** =(1x1=1x*n* *n*x1)= **z'c**. Note that from the definition, the dot product is commutative.

- When **c** is a vector of 1's, usually noted as $\boldsymbol{\iota}$, then:

$$\boldsymbol{\iota} \bullet \boldsymbol{z} = 1 * z_1 + 1 * z_2 + \dots + 1 * z_n = \sum_{i=1}^{n} z_i$$

17

## 5.3 Vectors: Dot Product

- Inner products (IP) in econometrics are common. For example, the Residual Sum of Squares (RSS), where **e** is a vector of residuals:

$$\boldsymbol{e} \bullet \boldsymbol{e} = \boldsymbol{e}'\boldsymbol{e} = e_1 * e_1 + e_2 * e_2 + \dots + e_n * e_n = \sum_{i=1}^{n} e_i^2$$

- 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. For example, for functions *f(x)* & *g(x):*

$$f \bullet g = \int f(x)\, g(x)\, dx$$

## 5.3 Vectors: Dot Product - Intuition

- Some intuition.

- IP is used as a tool to define length or size for vectors:

$$\| \alpha \| = \text{sqrt}[\alpha' \alpha] = \sqrt{\alpha_1{}^2 + \cdots + \alpha_n{}^2}$$

The square root makes the inner product to be expressed in the same units as the original vector.

- Now, it is possible to compare vectors and measure "distances" between vectors and, eventually, convergence!

- IP also can be used to define a notion like angle between vectors, since any two vectors, say $\alpha$ and $\beta$, determine a plane. The IP connects the length and the angle between the vectors $\alpha$ and $\beta$:

$$\alpha' \bullet \beta = \| \alpha \| \, \| \beta \| \cos(\theta)$$

## 5.3 Vectors: Dot Product – Geometry

- There is a geometric interpretation to the IP. The IP connects the length and the angle between the vectors $\alpha$ and $\beta$:

$$\alpha' \bullet \beta = \| \alpha \| \, \| \beta \| \cos(\theta)$$

The IP is related to the angle between the two vectors – but it does not tell us the angle.

- Now, we define orthogonality using the IP. Since $\cos(\theta=90)=0$, the IP of two *orthogonal* (perpendicular *or* "$\perp$") vectors is zero.

**Example**: In the CLM, we have

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} = \text{Projection} + \text{“error”}$$

Then:

$$\mathbf{X'e} = \mathbf{X'}(\mathbf{y} - \mathbf{Xb}) = \mathbf{X'y} - \mathbf{X'X}\,(\mathbf{X'X})^{-1}\mathbf{X'y} = \mathbf{0} \quad \Rightarrow (\mathbf{X} \perp \mathbf{e}).$$

## 5.3 Vectors: Dot Product - Properties

- The dot product fulfils the following properties if α, β, and γ are real vectors and $k$ is a scalar.

1. Commutative: $\alpha \bullet \beta = \langle\alpha, \beta\rangle = \beta \bullet \alpha$

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

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

3. Bilinear: $\alpha \bullet (k\beta + \gamma) = k\,(\alpha\bullet\beta) + \alpha\bullet\gamma$

4. Scalar multiplication: $(k_1\,\alpha) \bullet (k_2\,\beta) = k_1\,k_2\,(\alpha \bullet \beta)$

Note: Nice, intuitive properties.

Notation: $\alpha \bullet \beta = \langle\alpha, \beta\rangle$ (<.,.> is the physics notation).

## 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):

$$\|\,\boldsymbol{\alpha}\,\| = \text{sqrt}[\boldsymbol{\alpha}'\boldsymbol{\alpha}\,] = \sqrt{\alpha_1{}^2 + \alpha_2{}^2 + \cdots + \alpha_n{}^2}$$

- 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 $\mathbf{x}_n$ converge to a point $\mathbf{c}$ if $\|\mathbf{x}_n - \mathbf{c}\|$ 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.

$$\|\,k\boldsymbol{\alpha}\,\| = \text{sqrt}[k\boldsymbol{\alpha}'\,k\boldsymbol{\alpha}\,] = |\,k\,|\;\|\,\boldsymbol{\alpha}\,\|$$

Note: If we set $k = 1/\|\alpha\|$ $\Rightarrow \|(1/\|\alpha\|)\,\alpha\| = 1$.

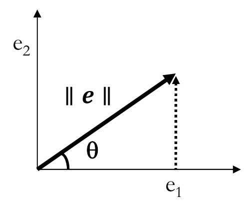
$\Rightarrow$ Nice result, used to normalize vectors.

## 5.3 Vectors: Magnitude and Direction

• We mentioned that vectors have magnitude and direction.

• To talk about magnitude, we need to define the notion of size, length, or distance (from origin) of a vector. In Euclidean spaces, the *magnitude* of the n-dimensional vector **e** can be calculated as:

$$\| \boldsymbol{e} \| = \text{sqrt}[\mathbf{e'e}] = \sqrt{e_1{}^2 + \cdots + e_n{}^2}$$




• This measure is called the *Euclidian norm*.

• The angle ("phase") **θ** gives the direction of the vector. Thus, a vector can also be defined in terms of polar coordinates: ($\| \boldsymbol{e} \|$, **θ**).

• If $\| \boldsymbol{e} \|$=1, $\boldsymbol{e}$ is a *unit vector* (a *pure direction* vector).

## 5.4 Vectors: Norm

• Given a vector space $V$, the function g:$V \rightarrow R$ is called a *norm* iff:

| | |
|---|---|
| 1) $g(x) \geq 0$, | for all $x \in V$ |
| 2) $g(x) = 0$ | iff $x = \emptyset$ (empty set) |
| 3) $g(\alpha x) = \|\alpha\| g(x)$ | for all $\alpha \in R$, $x \in V$ |
| 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 $\mathbf{R}^n$, the *Euclidian norm* of $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is given by

$$\| \boldsymbol{x} \| = \text{sqrt}[\mathbf{x'x}] = \sqrt{x_1{}^2 + x_2{}^2 + \cdots + x_n{}^2}$$

while the *Manhattan (Taxicab) norm* is defined as:

$$\| \boldsymbol{x} \|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

## 5.4 Vectors: Norm

**Example (continuation)**: On C[$a$, $b$]

$$\| \boldsymbol{f}(t) \| = \text{sqrt}[\int_a^b |f(t)|^2 dt]$$

Note: Euclidian norm = $L_2$ *norm* (2-*norm*).
Manhattan norm = $L_1$ *norm* (1-*norm*).

- We can generalize the concept of norm on $\mathbf{R}^n$.

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

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

## 5.4 Vectors: Norm

- An infinite number of functions can be shown to qualify as norms. For vectors in $R^n$, we have the following examples:

$g(x) = \max_i (x_i)$;

$g(x) = \sum_i |x_i|$;

$g(x) = [\sum_i (x_i)^4]^{1/4}$

- 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*.

## 5.4 Vectors: Metric

• Given a vector space $V$, the function d: $V$x$V \rightarrow R$ is called a *metric* or a *distance function* if and only if:

1) d($x$, $y$) ≥ 0, ("positive") for all $x,y \in V$
2) d($x$, $y$) = 0 ("non-degenerate") iff $x = y$
3) d($x$, $y$) = d($y$, $x$) ("symmetry") for all $x,y \in V$
4) d($x$+$y$) ≤ d($x$, $z$) + d($z$, $y$) ("*triangle inequality*") for all $x,y,z \in V$

Given a norm g(.), we can define a metric by the equation:

d($x$, $y$) = g($x$ – $y$).

Check:

1) and 2) follow immediately from properties of g(.)
3) d($x$, $y$) = g($x$-$y$) = g((-1)($y$-$x$)) = |-1| g($y$ – $x$) = g($y$ – $x$) = d($y$, $x$)
4) ($x$ – $y$) = ($x$ – $z$) + ($z$ – $y$) ⇒ g($x$ – $y$) ≤ g($x$ – $z$) + g($z$ – $y$)
⇒ d($x$, $y$) ≤ d($x$, $z$) + d($z$, $y$)

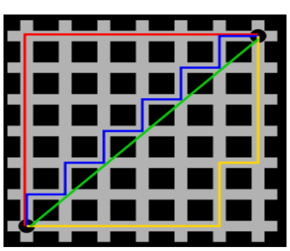
## 5.4 Vectors: Metric

**Example**: On $R^n$, 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.

## 5.4 Vectors: Metric

- **Theorem**: Cauchy-Schwarz Inequality

If **u** and **v** are vectors in a real inner product space, then

$$| \boldsymbol{u} \bullet \boldsymbol{v} | \leq \|\mathbf{u}\| \; \|\boldsymbol{v}\|$$

Note: This result can be written as

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

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

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

General Proof: Trivial proof when $v = 0$. We assume that $v \bullet v \neq 0$.

## 5.4 Vectors: Metric

- **Theorem**: $| \boldsymbol{u} \bullet \boldsymbol{v} | \leq \|\mathbf{u}\| \; \|\boldsymbol{v}\|$

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

$0 \leq \| \mathbf{u} - \delta \mathbf{v} \|^2 = (\mathbf{u} - \delta\, \mathbf{v}) \bullet (\mathbf{u} - \delta\, \mathbf{v}) = \mathbf{u} \bullet \mathbf{u} - 2\, \delta\, \mathbf{u} \bullet \mathbf{v} + |\delta|^2\, \mathbf{v} \bullet \mathbf{v}$

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

$$\delta = \boldsymbol{u} \bullet \boldsymbol{v} / \boldsymbol{v} \bullet \boldsymbol{v}$$

(Use the trick of thinking of $F$ as R. and get f.o.c. and solve:
$d(\|\mathbf{u} - \delta \boldsymbol{v}\|^2)/dk = -2\, \mathbf{u} \bullet \mathbf{v} + 2\, k\, \mathbf{v} \bullet \mathbf{v} = 0 \qquad \Rightarrow \delta = \boldsymbol{u} \bullet \boldsymbol{v} / \boldsymbol{v} \bullet \boldsymbol{v}$.)

Then, we get $\quad 0 \leq \mathbf{u} \bullet \mathbf{u} - 2\, (\boldsymbol{u} \bullet \boldsymbol{v} / \boldsymbol{v} \bullet \boldsymbol{v})\, \mathbf{u} \bullet \mathbf{v} + | \boldsymbol{u} \bullet \boldsymbol{v} / \boldsymbol{v} \bullet \boldsymbol{v} |^2\, \mathbf{v} \bullet \mathbf{v}$

$$\Rightarrow 0 \leq \mathbf{u} \bullet \mathbf{u} - | \boldsymbol{u} \bullet \boldsymbol{v} |^2\, (\mathbf{v} \bullet \mathbf{v})^{-1}$$

which is true if and only if $| \boldsymbol{u} \bullet \boldsymbol{v} |^2 \leq \mathbf{u} \bullet \mathbf{u}\, (\mathbf{v} \bullet \mathbf{v})$

or equivalently: $| \boldsymbol{u} \bullet \boldsymbol{v} | \leq \|\mathbf{u}\| \; \|\boldsymbol{v}\|$

## 5.4 Vectors: Metric Space & Cauchy Sequence

<u>Definition</u>: Metric Space

A metric on a space $M$ is a mapping d(., .): $M \times M \rightarrow [0, \infty)$ 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*.

<u>Definition</u>: Cauchy sequence

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

In other words a sequence is Cauchy if, eventually, all the terms are all very close to each other. Moreover, every convergent sequence is Cauchy (this is easy to prove. Do it!).

Q: Does every Cauchy sequence converge to a limit? Consider a sequence approximating π.

## 5.4 Vectors: Complete Metric Space

- Consider the sequence 3, 3.14, 3.141, 3.1415, . . . This sequence is clearly Cauchy. When considered as a sequence in $R$, it does converge to π. But, as a sequence in $Q$ (rational numbers) it does not converge, since $\pi \notin Q$.

- A metric space is *complete* if every Cauchy sequence in the space converges to some point $v$ in the vector space $V$.

**Examples – Complete**: Real numbers (rational + irrationals) are complete on the real line ($R^1$). In $R^p$ with finite dimension $p$ every Cauchy sequence converges to a limit in $R^p$.

**Examples – Non-complete**: The rational numbers, $Q$. The open interval (0,1) with |.| as a metric is not complete. The sequence defined by $\{x_n = 1/n\}$ is Cauchy, but does not have a limit in the given space. On the other hand, [0,1] is complete.

## 5.4 Vectors: Complete Vector Space

• Intuitively, a space is complete if there are no "tiny holes" on it (inside or at the boundary).

• We want the vector 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 vectors to a point $p$, then, $p$ is in the space. Now, we can approximate functions. Now, the techniques of calculus can be used.

Note: In many problems, we find a solution by approximating the answer. We need completeness to ensure the approximations actually converge to something in the space. For example, continuous functions on [0,1] can be approximated by polynomials.

## 5.5 Vectors: Banach and Hilbert Spaces

Definition: Banach space
A *Banach space* $B$ is a complete normed vector space.

• Completeness makes a Banach space closed under convergence.

**Examples**: ($V = R^1$ (real line), $d = \|x - y\|_p$) is a Banach space.
($V = R^N$, $d = \|x - y\|_p$) is also a Banach space.

• Since any finite dimensional vector space can be mapped in a one to one fashion to $R^N$, we have the following result:

**Theorem**: In a normed linear vector space, any finite-dimensional subspace is complete and thus it forms a Banach space.

## 5.5 Vectors: Hilbert Space

- 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 specify a norm, the inner product (IP), $(x \bullet y)$.

Definition: Hilbert space
A *Hilbert space* $H$ is a vector space endowed with an IP, $(x \bullet y)$, associated norm $\|x\|$ = sqrt$(y \bullet x)$, and metric $\|x - y\|$ such that every Cauchy sequence in $H$ has a limit in $H$.

- 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 $R^n$ or $C^n$. These are also Hilbert spaces.

Note: The space we live on, $R^3$, is a Hilbert space!

## 5.5 Vectors: Hilbert Space

- It can be shown that a Banach space is a Hilbert space if and only if its norm satisfies the Parallelogram Law:

$$\|x+y\|^2 + \|x-y\|^2 = 2 * (\|x\|^2 + \|y\|^2)$$

**Example**: A Banach space, but not a Hilbert space
The space C[0,1] of continuous functions $f$:[0,1] $\rightarrow R$, with the supremum norm, $\|.\|_\infty$, is a Banach space, but not a Hilbert space.

Let $f(x) = x$ for $x \in [0,1]$ and $g(x) = 1$ for $x \in [0,1]$. We check if the parallelogram law is not satisfied. Using the supremum norm:

$$\|f\|_\infty = 1; \ \|g\|_\infty = 1; \ \|f+g\|_\infty = 2; \ \& \ \|f-g\|_\infty = 1.$$

Then,

$$5 = \|f+g\|_\infty{}^2 + \|f-g\|_\infty{}^2 \neq 2*(\|f\|_\infty{}^2 + \|g\|_\infty{}^2) = 4.$$

## 5.5 Vectors: Hilbert Space - Examples

• Hilbert spaces appear frequently in mathematics, statistics, and physics (SS, state space for quantum mechanics), typically as infinite-dimensional function spaces. Hilbert space methods helped in the development of functional analysis.

**Example I**: The space $V$ of random variables defined on a common probability space $\{\Omega, F, P\}$ with finite second moments, endowed with IP, $X \bullet Y = E[XY]$, associated norm $\|X\| =$ sqrt($X \bullet X$) and metric $\|X - Y\|$.

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

$$f \bullet g = \int f(x)\, g(x)\, dx$$

## 5.5 Vectors: Hilbert Space - Summary

• Banach spaces, and their special case, Hilbert spaces, are usually studied in the context of infinite dimensional vector spaces, as they are tailor made to study spaces of functions and sequences.

• Typical examples of Hilbert spaces: Euclidean spaces, spaces of square-integrable functions, spaces of sequences.

• Hilbert Space - Summary:
- Generalization of Euclidian 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)

## 5.6 Vectors: Orthogonality

- **Theorem:** (Generalized Law of Pythagoras)

If **u** and **v** are orthogonal vectors in an IP space, then

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$

**Proof**: It follows from $\boldsymbol{u} \bullet \boldsymbol{v} = 0$.

- Definition: Orthogonal Complement

Let $W$ be a subspace of an inner product space $V$. A vector $\boldsymbol{u}$ 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".]

## 5.6 Vectors: Orthogonality

- **Theorem:** Properties of Orthogonal Complements

If $W$ is a subspace of a finite-dimensional IP 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 $\mathrm{R}^N$, 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 $R^N$.

## 5.6 Vectors: Projections

- **Theorem**: Projection

  Given $\boldsymbol{v}$ a vector in $V$ and a subspace $V_1$, there is a unique vector $\boldsymbol{v_1}$ such that $|\mathbf{e}| = \|\boldsymbol{v} - \boldsymbol{v_1}\|$ is minimized.

  Proof:

  Decompose $V = V_1 + V_2 = V_1 + V_1^{\perp}$ $(V_2 = V_1^{\perp})$

  Let $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2$, where $\boldsymbol{v}_1 \in V_1$ and $\boldsymbol{v}_2 \in V_1^{\perp}$.

  Pick an arbitrary "error" vector, $\mathbf{e}$, where $\boldsymbol{p}_1 \in V_1$, as:

  $$\mathbf{e} = \boldsymbol{v} - \boldsymbol{p_1} = (\boldsymbol{v}_1 + \boldsymbol{v}_2) - \boldsymbol{p_1} = (\boldsymbol{v}_1 - \boldsymbol{p_1}) + \boldsymbol{v}_2$$

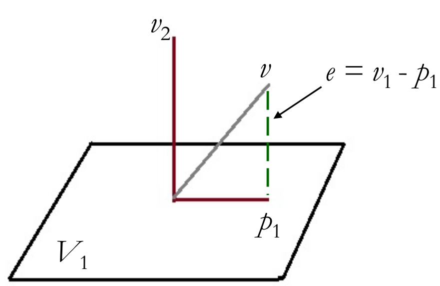
  Then,

  $$\|\mathbf{e}\|^2 = \mathbf{e} \bullet \mathbf{e} = \langle(\boldsymbol{v}_1 - \boldsymbol{p_1}) + \boldsymbol{v}_2, (\boldsymbol{v}_1 - \boldsymbol{p_1}) + \boldsymbol{v}_2\rangle$$
  $$= \|(\boldsymbol{v}_1 - \boldsymbol{p_1})\|^2 + \|\boldsymbol{v}_2\|^2$$

  which is minimized when $\boldsymbol{v}_1 = \boldsymbol{p_1}$ $\Rightarrow \mathbf{e} \in V_1^{\perp}$ (or $\mathbf{e} \perp \boldsymbol{v_i} \in V_1$).

## 5.6 Vectors: Projections

- **Theorem**: Projection




Note: Many optimization problems in Hilbert spaces use this idea.

## 5.6 Vectors: Projections

- Definition: Projection

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two non-zero vectors in an inner product space $V$. Then, the *scalar projection* of $\boldsymbol{u}$ onto $\boldsymbol{v}$ is defined as

$$k = \boldsymbol{u}\bullet\boldsymbol{v}/\boldsymbol{v}\bullet\boldsymbol{v}$$

The *vector projection* of $\boldsymbol{u}$ onto $\boldsymbol{v}$ is $\boldsymbol{p} = k\,\boldsymbol{v} = (\boldsymbol{u}\bullet\boldsymbol{v}/\boldsymbol{v}\bullet\boldsymbol{v})\,\boldsymbol{v}$

Derivation: Given two vectors: $\boldsymbol{v}$ in S & $\boldsymbol{u}$ in $R^n$. We want to find $\mathbf{p}$, the vector in S closest to $\boldsymbol{u}$. Let $\mathbf{p} = k\mathbf{v}$.

To minimize $\|\boldsymbol{u}-\boldsymbol{p}\|$ with respect to $k$.

$\|\boldsymbol{u}-\boldsymbol{p}\|^2 = \|\mathbf{u}-k\mathbf{v}\|^2 = (\mathbf{u}-k\mathbf{v})\bullet(\mathbf{u}-k\mathbf{v}) = \mathbf{u}\bullet\mathbf{u} - 2k\,\mathbf{u}\bullet\mathbf{v} + k^2\,\mathbf{v}\bullet\mathbf{v}$

$d(\|\mathbf{u}-\mathbf{p}\|^2)/dk = -2\mathbf{u}\bullet\mathbf{v} + 2k\,\mathbf{v}\bullet\mathbf{v} = 0 \quad \Rightarrow k = \boldsymbol{u}\bullet\boldsymbol{v}/\boldsymbol{v}\bullet\boldsymbol{v}$

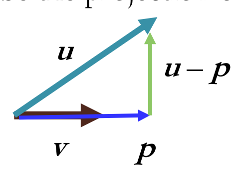
$\Rightarrow \boldsymbol{p} = (\boldsymbol{u}\bullet\boldsymbol{v}/\boldsymbol{v}\bullet\boldsymbol{v})\,\boldsymbol{v}$

## 5.6 Vectors: Projections

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

(i) $(\boldsymbol{u}-\boldsymbol{p}) \perp \boldsymbol{p}$

(ii) $\boldsymbol{u} = \boldsymbol{p} \Leftrightarrow \mathbf{u} = k\,\mathbf{v}$ for some $k$




**Proof**: Recall $\boldsymbol{p} = k\mathbf{v} = (\boldsymbol{u}\bullet\boldsymbol{v}/\boldsymbol{v}\bullet\boldsymbol{v})\,\boldsymbol{v}$

(i) $\boldsymbol{p}\bullet(\boldsymbol{u}-\boldsymbol{p}) = \boldsymbol{p}\bullet\boldsymbol{u} - \boldsymbol{p}\bullet\boldsymbol{p} =$

$= |\boldsymbol{u}\bullet\boldsymbol{v}|^2/\|\boldsymbol{v}\| - |\boldsymbol{u}\bullet\boldsymbol{v}|^2/\|\boldsymbol{v}\| = 0$

$\Rightarrow (\boldsymbol{u}-\boldsymbol{p}) \perp \boldsymbol{p}$

(ii) straightforward.

## 5.6 Vectors: Projections - Examples

**Example**: Let $V = R^2$ be spanned by

$$\boldsymbol{v_1} = \begin{bmatrix}1\\0\end{bmatrix} \quad \text{and} \quad \boldsymbol{v_2} = \begin{bmatrix}2\\1\end{bmatrix}$$

The projection of $\boldsymbol{v_2}$ onto $\boldsymbol{v_1}$ is:

$$\boldsymbol{p} = (\boldsymbol{v_2} \bullet \boldsymbol{v_1})/(\boldsymbol{v_1} \bullet \boldsymbol{v_1})\, \boldsymbol{v_1} = \frac{2}{1}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$$

Also, we calculate

$$(\boldsymbol{v_2} - \boldsymbol{p}) = \begin{bmatrix}2\\1\end{bmatrix} - \begin{bmatrix}2\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}$$

Note that $\boldsymbol{v_1}$ and $(\boldsymbol{v_2} - \boldsymbol{p})$ form a standard basis for $R^2$.

Check property (i): $(\boldsymbol{v_2} - \boldsymbol{p}) \perp \boldsymbol{p}$ (IP = 0)

$$\boldsymbol{p} \bullet (\boldsymbol{v_2} - \boldsymbol{p}) = 2*0 + 0*1 = 0.$$

## 5.6 Vectors: Projections - Examples

**Example**: Find the best approximation, $\boldsymbol{p}$, in subspace S be spanned by the columns of $\mathbf{X}$ ($N$x$k$).

Project a vector $\boldsymbol{y}$ onto S (spanned by columns $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$):

$$\boldsymbol{y}^p = \mathbf{x}_1 \mathrm{b}_1 + \mathbf{x}_2 \mathrm{b}_2 + .. + \mathbf{x}_k \mathrm{b}_k = \mathbf{X}\,\mathbf{b}$$

The error vector $\mathbf{e}$ will be perpendicular to all vectors in S. Then, for $\mathbf{e} = \boldsymbol{y} - \mathbf{X}\,\mathbf{b}$ (Note: $\{\mathbf{X}\}$: independent columns.)

$$\begin{cases} x_1' e = 0 \\ x_2' e = 0 \\ \quad \ldots \\ x_N' e = 0 \end{cases} \Rightarrow \mathbf{X}'\,(\boldsymbol{y} - \mathbf{X}\,\mathbf{b}) = \mathbf{0} \quad \Rightarrow \mathbf{X}'\,\boldsymbol{y} - \mathbf{X}'\,\mathbf{X}\,\mathbf{b} = \mathbf{0}$$

## 5.6 Vectors: Projections - Examples

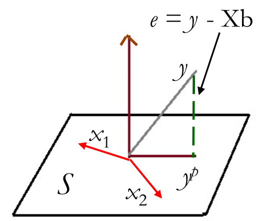
$$\begin{cases} x_1'e = 0 \\ x_2'e = 0 \\ \quad \dots \\ x_N'e = 0 \end{cases} \quad \Rightarrow \mathbf{X}'\,(y - \mathbf{X}\,\mathbf{b}) = \mathbf{0} \quad \Rightarrow \mathbf{X}'\,y - \mathbf{X}'\,\mathbf{X}\,\mathbf{b} = 0$$

$$\Rightarrow \mathbf{X}'\,y = \mathbf{X}'\,\mathbf{X}\,\mathbf{b}$$

$$\Rightarrow \mathbf{b} = (\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'y$$

Then,

$$\Rightarrow \; y^p = \mathbf{X}\,\mathbf{b} = \mathbf{X}(\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'y$$




This is the general result of projecting a vector $y$ onto a subspace S with a basis specified by the columns of a matrix **X**. The matrix $\mathbf{X}(\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'$ is called *projection matrix*, **P** ($N$x$N$ matrix).

## 5.6 Vectors: Projections - CLM

- In the CLM, we have a "Projection matrix", **P**:

  $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (**X** is $N$x$k$ $\Rightarrow$ **P** is $N$x$N$)

- Features

  $\mathbf{Py} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{Xb} = \hat{\mathbf{y}}$ (fitted values)

  **Py**: The projection of **y** into the column space of **X**.

  $\mathbf{PM} = \mathbf{P}[\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,] = \mathbf{MP} = \mathbf{0}$ (**M**: residual maker)

  $\mathbf{PX} = \mathbf{X}$

- Properties
  - **P** is symmetric – $\mathbf{P} = \mathbf{P}'$
  - **P** is idempotent – $\mathbf{P}^*\mathbf{P} = \mathbf{P}$
  - **P** is singular – $\mathbf{P}^{-1}$ does not exist. $\Rightarrow$ rank(**P**)=$k$

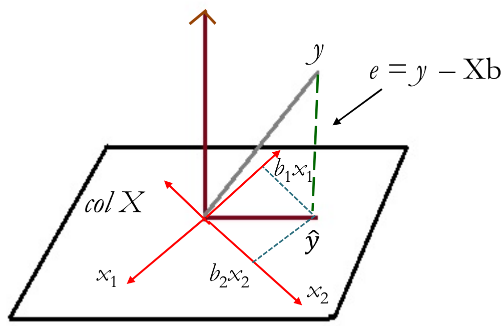
## 5.6 Vectors: Projections - CLM

- We have two ways to look at **y**:

  **y** = **Xβ** + **ε** = Conditional mean + disturbance

  **y** = **Xb** + **e** = Projection + residual

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




## 5.6 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, do not have to be; basis are not unique!)

$$x = [1 \quad 0 \quad 0]^T \qquad x \cdot y = 0$$
$$y = [0 \quad 1 \quad 0]^T \qquad x \cdot z = 0$$
$$z = [0 \quad 0 \quad 1]^T \qquad y \cdot z = 0$$

- The *Gram-Schmidt process* is a popular method to orthonormalize a set of vectors in Hilbert spaces (actually, IP spaces). A method based on the Cholesky decomposition can also be used.

## 5.6 Vectors: Orthonormal Basis

- If **X**, **Y**, **Z** form an orthonormal basis in $R^3$, we can describe any 3D point as a linear combination of these vectors.

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

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

(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?"