ETF2700/ETF5970 Mathematics for Business

Lecture 2

Monash Business School, Monash University, Australia

Outline

Last week:

- Real numbers and fraction
- Linear equations, inequalities and intervals
- Percentage and percentage change
- Quadratic equations
- Linear functions and linear equations
- Solving system of linear equations using matrix

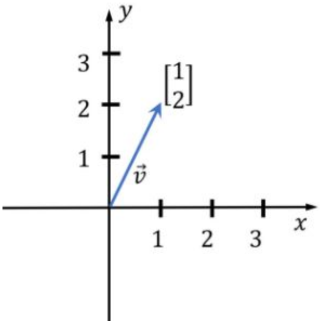
This week:

- Vectors
- Linear dependence and independence
- Orthonormal basis

Linear Algebra

Scalar and vector

- A scalar is a quantity that only has magnitude or size. For example, any real number is a scalar.
- A vector is a list of quantities, and therefore, it has a magnitude and a direction.
- A vector can be written as a column such as $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, or a row such as $(1, 2)$.



Elements of a vector

- Elements of a vector are the entries inside the brackets
- size (also called dimension or length) of a vector is the number of elements it contains.
- A vector of size n can be called an n-vector
- We denote an n-vector using the symbol $\vec{a}$ (or a if it won't cause confusion), the ith element of the vector $\vec{a}$ is denoted as a_i, where the subscript i is an integer index that runs from 1 to n, the size of the vector
- Two vectors $\vec{a}$ and $\vec{b}$ are equal, which we denote $\vec{a} = \vec{b}$, if they have the same size and $a_i = b_i$, for $i = 1, 2, \ldots, n$
- Elements of a vector are *scalars*. If these scalars are real numbers, we call the vector a *real vector*
- The set of all real numbers is written as $\boldsymbol{R}$, and the set of all real n-vectors is denoted as $\boldsymbol{R}^n$
- If we write $\vec{a} \in \boldsymbol{R}^n$, this is another way to say that $\vec{a}$ is an n-vector with real entries

Special vectors

- A **zero vector** is a vector with all elements being zero
- A **unit vector** is a vector with all elements being equal to zero, except for one element being equal to one

Transpose of a vector

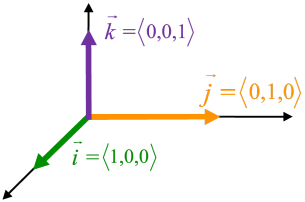
- If $\vec{a}$ denotes an n-vector, then by default, it is a column vector, which has one column with n elements in an order listed from top to bottom
- Transpose of $\vec{a}$, denoted as $\vec{a}^\top$ or $\vec{a}'$, is to express the elements of $\vec{a}$ in a row with the elements in the same order listed from left to the right

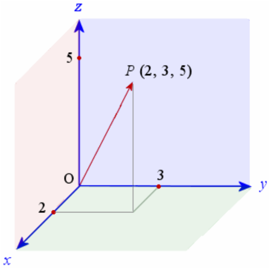
Example: Unit vectors in 3-dimensional space

- Any point can be represented as $\vec{x} = (x_1, x_2, x_3)'$, where the elements are coordinates

- There are three different unit vectors:

$$\vec{\boldsymbol{i}} = (1, 0, 0)^\top, \qquad \vec{\boldsymbol{j}} = (0, 1, 0)^\top, \qquad \vec{\boldsymbol{k}} = (0, 0, 1)^\top.$$





- Any vector $\vec{x} = (x_1, x_2, x_3)^\top$ can be represented as

$$\vec{x} = x_1\vec{\boldsymbol{i}} + x_2\vec{\boldsymbol{j}} + x_3\vec{\boldsymbol{k}}.$$

- Example: A vector given by $\vec{p} = (2, 3, 5)^\top$ is expressed as

$$\vec{p} = 2\vec{\boldsymbol{i}} + 3\vec{\boldsymbol{j}} + 5\vec{\boldsymbol{k}}.$$

Vector addition

- Two vectors of the same size (dimension) can be added together by adding the corresponding elements, to form another vector of the same size, which is called the sum of the two vectors
- Suppose $\vec{a} = (0, 4, 3)'$ and $\vec{b} = (1, 2, 1)'$, and then $\vec{a} + \vec{b} = (1, 6, 4)'$
- Vector subtraction is similar $\vec{a} - \vec{b} = (-1, 2, 2)'$.
- The result of vector subtraction is called the difference of the two vectors

Properties of vector addition

- Vector addition is commutative: $\vec{a} + \vec{b} = \vec{b} + \vec{a}$
- Vector addition is associative: $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$
- $\vec{a} + \vec{0} = \vec{0} + \vec{a} = \vec{a}$, and $\vec{a} - \vec{a} = \vec{0}$, where $\vec{0}$ has the same dimension as $\vec{a}$

Example: Market clearing

- Suppose the k-vector $\vec{q}_i$ represents the quantities of k goods or resources produced (when positive) or consumed (when negative) by Agent i, for $i = 1, 2, \ldots, N$
- For instance, $(\vec{q}_5)_4 = -2.7$ means that Agent 5 consumes 2.7 units of Resource 4
- The sum $\vec{s} = \vec{q}_1 + \ldots + \vec{q}_N$ is the k-vector of total net surplus of the resources (or shortfall when entries are all negative)
- When $\vec{s} = \vec{0}$, we have a closed market, which means that the total quantity of each resource produced by the agents balances the total quantity consumed.
- In this case, the k resources are exchanged among the agents, and we say that the market clears (with the resource vectors $\vec{q}_1, \ldots, \vec{q}_N$)

Multiplication of a vector by a scalar

- Multiplication of a vector by a scalar is conducted by multiplying every element of the vector by the scalar

$$(-2) \times \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} \times 1.5 = \begin{bmatrix} 1.5 \\ 13.5 \\ 9 \end{bmatrix}.$$

- Scalar multiplication on the left has the same meaning as that on the right, which is obtained by multiplying each component by the scalar
- Note that $0 \times \vec{a} = \vec{a} \times 0 = \vec{0}$
- By definition, we have $c \times \vec{a} = \vec{a} \times c$, for any scalar c and any vector $\vec{a}$. It is called the *commutative property* of scalar-vector multiplication
- Scalar-vector multiplication can be written in either order

Scalar multiplication laws

Let $\vec{a}$ and $\vec{b}$ denote 2 vectors and c_1 and c_2 denote 2 scalars

- $(c_1 \times c_2) \times \vec{a} = c_1 \times (c_2 \times \vec{a})$
- Left-distributive property: $(c_1 + c_2) \times \vec{a} = c_1 \times \vec{a} + c_2 \times \vec{a}$
- Right-distributive property: $\vec{a} \times (c_1 + c_2) = \vec{a} \times c_1 + \vec{a} \times c_2$
- Another version of right-distributive property: $c_1 \times (\vec{a} + \vec{b}) = c_1 \times \vec{a} + c_1 \times \vec{b}$

Linear combinations

If $\vec{a}_1, \ldots, \vec{a}_n$ are p-vectors, and $c_1, \ldots, c_n$ are scalars, then the new vector

$$c_1\vec{a}_1 + \ldots + c_n\vec{a}_n.$$

is called a linear combination of the vectors $\vec{a}_1, \ldots, \vec{a}_n$. The scalars $c_1, \ldots, c_n$ are *coefficients* of the linear combination

Linear combination of unit vectors

Let $\vec{a} = (a_1, \ldots, a_k)^\top$ denote a k-vector, and $\vec{e}_1, \ldots, \vec{e}_k$ denote k different unit vectors (of the dimension k). We can express $\vec{a}$ as

$$\vec{a} = a_1\vec{e}_1 + \ldots + a_k\vec{e}_k$$

A simple example is

$$\begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix} = (-1) \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 5 \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Special forms of linear combination $c_1\vec{a}_1 + \ldots + c_n\vec{a}_n$

- Sum of vectors: $c_1 = \ldots = c_n = 1$
- Average of vectors: $c_1 = \ldots = c_n = 1/n$
- Affine combination: $c_1 + \ldots + c_n = 1$
- Convex combination (aka as a mixture or a weighted average): $c_1 + \ldots + c_n = 1$, and $c_i \geq 0$, for $i = 1, \ldots, n$

Inner Product

The (standard) inner product (also called dot product) of two n-vectors $\vec{a} = (a_1, \ldots, a_n)^\top$ and $\vec{b} = (b_1, \ldots, b_n)^\top$ is defined as

$$\vec{a}^\top \vec{b} = a_1 b_1 + \ldots + a_n b_n$$

which is the sum of the products of corresponding entries, also denoted as $\vec{a} \cdot \vec{b}$ or $\langle \vec{a}, \vec{b} \rangle$

- Here is a simple example:

$$[-1, 2, 3] \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = (-1) \times 1 + 2 \times 0 + 3 \times (-2) = -7$$

- It is the sum of element-by-element multiplication of a row vector and a column vector
- Dimensions of the two vectors are compatible: $(1 \times n) \times (n \times 1)$

Properties of Inner Product

- Commutativity: $\vec{a}^\top \vec{b} = \vec{b}^\top \vec{a}$
- Associativity with scalar multiplication: $(c\vec{a})^\top \vec{b} = c(\vec{a}^\top \vec{b})$
- Distributivity with vector addition: $(\vec{a}_1 + \vec{a}_2)^\top \vec{b} = \vec{a}_1^\top \vec{b} + \vec{a}_2^\top \vec{b}$
- For any vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{b}_1$ and $\vec{a}_2$ of the same size (dimension), we have:

$$(\vec{a}_1 + \vec{a}_2)^\top (\vec{b}_1 + \vec{b}_2) = \vec{a}_1^\top \vec{b}_1 + \vec{a}_1^\top \vec{b}_2 + \vec{a}_2^\top \vec{b}_1 + \vec{a}_2^\top \vec{b}_2$$

Can you prove this?

General examples

- Unit vector: $\vec{e}_i^\top \vec{a} = a_i$ (inner product of vector $\vec{a}$ with the ith standard unit vector gives the ith element $\vec{a}$)
- Sum: $\vec{1}'\vec{a} = a_1 + \cdots + a_n$ (inner product of vector $\vec{a}$ with the vector of ones gives the sum of the elements of $\vec{a}$)

General examples

- Average: $\vec{a}^\top \vec{1}/n$ (inner product of an n-vector $\vec{a}$ with the vector $\vec{1}/n$ gives the average or mean of the elements of $\vec{a}$)
- Sum of squares: $\vec{a}^\top \vec{a} = a_1^2 + \cdots + a_n^2$ (inner product of a vector with itself gives the sum of the squares of the elements)
- Selective sum: Let $\vec{b}$ be a vector with all entries being either 0 or 1. Then $\vec{b}^\top \vec{a}$ is the sum of the elements in $\vec{a}$ for which $b_i = 1$

Example: Polynomial evaluation

Suppose the n-vector $\vec{c}$ represents the coefficients of a polynomial $p(x)$ of degree $n - 1$

$$p(x) = c_1 + c_2 x + c_3 x^2 + \cdots + c_n x^{n-1}$$

Let r be a real number, and let $\vec{z} = (1, r, r^2, \ldots, r^{n-1})$ be an n-vector of powers of r. Then $\vec{c}^\top \vec{z}$ is the value of $p(x)$ at $x = r$

Example: Discounted total

Let $\vec{c}$ be an n-vector representing a cash flow with c_i being the cash to be received (when $c_i > 0$) by the end of the ith year. Let $\vec{d}$ be an n-vector defined as

$$\vec{d} = \left(1, \frac{1}{1+r}, \ldots, \frac{1}{(1+r)^{n-1}}\right)^{\top}$$

where r is an interest rate. Then

$$\vec{d}'\vec{c} = c_1 + \frac{c_2}{1+r} + \cdots + \frac{c_n}{(1+r)^{n-1}}$$

is the discounted total of future-year cash flow, that is, its net present value (NPV) at the end of the current year with interest rate r.

Linear functions

- The notation $f : \boldsymbol{R}^n \to \boldsymbol{R}$ means that f is a function that maps real an n-vector to a real number, that is, it is a scalar-valued function of an n-vector
- If $\vec{x}$ is an n-vector, then $f(\vec{x})$, which is a scalar, denotes the value of the function f at $\vec{x}$
- In the notation $f(\vec{x})$, $\vec{x}$ is referred to as the argument of the function
- We can also interpret f as a function of n scalar arguments, the entries of the vector argument, in which case we write $f(\vec{x})$ as

$$f(\vec{x}) = f(x_1, x_2, \cdots, x_n)$$

Inner product function

Suppose $\vec{a} = (a_1, \ldots, a_n)^\top$ is an n-vector. We can define a scalar-valued function f of an n-vector $\vec{x} = (x_1, \ldots, x_n)^\top$:

$$f(\vec{x}) = \vec{a}^\top \vec{x} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

Linear functions: Superposition and linearity

- The inner product function function satisfies

$$\begin{aligned} f(\alpha\vec{x} + \beta\vec{y}) &= \vec{a}^\top(\alpha\vec{x} + \beta\vec{y}) \\ &= \vec{a}^\top(\alpha\vec{x}) + \vec{a}^\top(\beta\vec{y}) \\ &= \alpha(\vec{a}^\top\vec{x}) + \beta(\vec{a}^\top\vec{y}) \\ &= \alpha f(\vec{x}) + \beta f(\vec{y}) \end{aligned}$$

  This property is called *superposition*
- A function $f(\vec{x})$ is called a linear function if it satisfies the superposition property:

$$f(\alpha\vec{x} + \beta\vec{y}) = \alpha f(\vec{x}) + \beta f(\vec{y})$$

- On the left-hand side, the term $\alpha\vec{x} + \beta\vec{y}$ involves scalar-vector multiplication and vector addition
- On the right-hand side, $\alpha f(\vec{x}) + \beta f(\vec{y})$ involves ordinary scalar multiplication and scalar addition

Linear functions: Superposition and linearity

- If a function f is linear, superposition extends to linear combinations of any number of vectors, and we have

$$f(\alpha_1\vec{x}_1 + \ldots + \alpha_k\vec{x}_k) = \alpha_1 f(\vec{x}_1) + \ldots + \alpha_k f(\vec{x}_k)$$

  for any n-vectors $\vec{x}_1, \ldots, \vec{x}_k$ and any scalars $\alpha_1, \ldots, \alpha_k$.
- This more general k-term form of superposition reduces to the 2-term form given above when $k = 2$.

Superposition for $f(\alpha\vec{x} + \beta\vec{y})$

- Homogeneity: For any n-vector $\vec{x}$ and any scalar α, we have $f(\alpha\vec{x}) = \alpha f(\vec{x})$
- Additivity: For any n-vectors $\vec{x}$ and $\vec{y}$, we have $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$
- Homogeneity: scaling $\vec{x}$ is the same as scaling $f(\vec{x})$
- Additivity: adding the vector arguments is the same as adding the function values

Inner product representation of a linear function

Suppose f is a scalar-valued function of an n-vector, and is linear, that is,

$$f(\alpha\vec{x} + \beta\vec{y}) = \alpha f(\vec{x}) + \beta f(\vec{y})$$

holds for any n-vectors $\vec{x}$ and $\vec{y}$, and for any scalars α and β. Then there exists an n-vector $\vec{a}$, such that

$$f(\vec{x}) = \vec{a}^\top \vec{x}, \text{ for all } \vec{x}.$$

We call $\vec{a}^\top \vec{x}$ the inner product representation of f (can you prove this statement? try to do it!)

Affine functions

- A linear function plus a constant is called an affine function
- A function $f : \boldsymbol{R}^n \to \boldsymbol{R}$ is affine if and only if it can be expressed as $f(\vec{x}) = \vec{a}^\top \vec{x} + b$ for some n-vector $\vec{a}$ and scalar b

Affine functions

- Example: the function of $\vec{x} = (x_1, x_2, x_3)^\top$ defined by

$$f(\vec{x}) = 2.3 + 2x_1 + 1.3x_2 - x_3$$

  is affine, with $b = 2.3$ and $\vec{a} = (2, 1.3, -1)^\top$
- Any affine scalar-valued function satisfies the following variation on the superposition property:

$$f(\alpha\vec{x} + \beta\vec{y}) = \alpha f(\vec{x}) + \beta f(\vec{y})$$

  for all n-vectors $\vec{x}$ and $\vec{y}$, and all scalars α and β that satisfy $\alpha + \beta = 1$ (can you prove this statement? try to do it!)

Norm

- The Euclidean norm of an n-vector $\vec{x}$ (named after the Greek mathematician Euclid), denoted as $\|\vec{x}\|$, is the square root of the sum of the squares of all elements:

$$\|\vec{x}\| = \sqrt{x_1^2 + \ldots + x_n^2}$$

- Euclidean norm can also be expressed as the square root of the inner product of the vector with itself $\|\vec{x}\| = \sqrt{\vec{x}^\top \vec{x}}$
- Euclidean norm of a scalar c is $\|c\| = \sqrt{c^2} = |c|$.
- Euclidean norm can be considered a generalisation or extension of the absolute value or magnitude, that applies to vectors

Properties of Norm

- Nonnegative homogeneity: $\|c\vec{x}\| = |c| \times \|\vec{x}\|$.
  The norm of multiplying a vector by a scalar c equals the norm of the vector multiplied by the scalar.

Properties of Norm

- Nonnegative homogeneity: $\|c\vec{x}\| = |c| \times \|\vec{x}\|$.
- Triangle inequality: $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$. Euclidean norm of a sum of two vectors is no more than the sum of their norms.
- Non-negativity: $\|\vec{x}\| \geq 0$.
- Definiteness: $\|\vec{x}\| = 0$ if and only if $\vec{x} = \vec{0}$.

The properties of non-negativity and definiteness is called *positive definiteness*, which states that the norm is always non-negative, and zero only when the vector is zero.

Norm of a sum

Euclidean norm of $\vec{a} + \vec{b}$ is

$$\|\vec{a} + \vec{b}\| = \sqrt{\|\vec{a}\|^2 + 2\vec{a}^\top \vec{b} + \|\vec{b}\|^2}$$

Distance

Euclidean distance between $\vec{a}$ and $\vec{b}$ is

$$\mathrm{dist}(\vec{a}, \vec{b}) = \|\vec{a} - \vec{b}\|$$

Distance: Example

Consider three 4-vectors

$$\vec{u} = \begin{bmatrix} 1.8 \\ 2.0 \\ -3.7 \\ 4.0 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} 0.6 \\ 2.1 \\ 1.9 \\ -1.4 \end{bmatrix}, \quad \vec{w} = \begin{bmatrix} 2.0 \\ 1.9 \\ -4.0 \\ 4.6 \end{bmatrix}$$

$$\begin{aligned} \mathrm{dist}(\vec{u}, \vec{v}) &= \sqrt{(1.8-0.6)^2 + (2-2.1)^2 + (-3.7-1.9)^2 + (4+1.4)^2} \\ &\approx 8.368 \end{aligned}$$

$\mathrm{dist}(\vec{u}, \vec{w}) \approx 0.387,$ and $\mathrm{dist}(\vec{v}, \vec{w}) \approx 8.533$

So, $\vec{u}$ is much closer to $\vec{w}$ than it is to $\vec{v}$.

Distance: Triangle inequality

For any three vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$, which are of the same dimension (size), we have

$$\|\vec{a}-\vec{c}\| \leq \|\vec{a}-\vec{b}\| + \|\vec{b}-\vec{c}\|.$$

To show how it holds, we have

$$\|\vec{a}-\vec{c}\| = \|(\vec{a}-\vec{b}) + (\vec{b}-\vec{c})\|.$$

According to the **triangle inequality** of Euclidean norm, we have

$$\|(\vec{a}-\vec{b}) + (\vec{b}-\vec{c})\| \leq \|(\vec{a}-\vec{b})\| + \|(\vec{b}-\vec{c})\|.$$

Therefore, the triangle inequality holds for the distance measure.

Cauchy-Schwarz Inequality

An important inequality that relates norms and inner products is the *Cauchy-Schwarz Inequality*:

$$|\vec{a}^\top \vec{b}| \le \|\vec{a}\| \times \|\vec{b}\|,$$

for any n-vectors $\vec{a}$ and $\vec{b}$.
Expressing in terms of elements of vectors, this inequality becomes

$$|a_1 b_1 + \ldots + a_n b_n| \le (a_1^2 + \ldots + a_n^2)^{1/2} (b_1^2 + \ldots + b_n^2)^{1/2}.$$

To show that this inequality holds, we have the following 2 scenarios.

1) If $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$, the equality sign clearly holds.
2) Assume $\vec{a} \neq \vec{0}$ and $\vec{b} \neq \vec{0}$ and let $c_1 = \|\vec{a}\|$ and $c_2 = \|\vec{b}\|$. We observe that

Cauchy-Schwarz Inequality

$$\begin{aligned} 0 &\leq \|c_2\vec{a} - c_1\vec{b}\|^2 \\ &= \|c_2\vec{a}\|^2 - 2(c_2\vec{a})^\top(c_1\vec{b}) + \|c_1\vec{b}\|^2 \\ &= c_2^2\|\vec{a}\|^2 - 2c_2c_1(\vec{a}^\top\vec{b}) + c_1^2\|\vec{b}\|^2 \\ &= \|\vec{b}\|^2\|\vec{a}\|^2 - 2\|\vec{b}\|\|\vec{a}\|(\vec{a}^\top\vec{b}) + \|\vec{a}\|^2\|\vec{b}\|^2, \end{aligned}$$

which results in

$$0 \leq 2\|\vec{a}\|^2\|\vec{b}\|^2 - 2\|\vec{b}\|\|\vec{a}\|(\vec{a}^\top\vec{b}).$$

Dividing both sides by $2\|\vec{b}\|\|\vec{a}\|$, we obtain that

$$\vec{a}^\top\vec{b} \leq \|\vec{b}\|\|\vec{a}\|.$$

Applying this inequality to $-\vec{a}$ and $\vec{b}$, we obtain

$$-\vec{a}^\top\vec{b} \leq \|\vec{b}\|\|\vec{a}\|.$$

Cauchy-Schwarz Inequality

Putting these 2 inequalities together, we obtain

$$|\vec{a}^\top \vec{b}| \leq \|\vec{b}\|\|\vec{a}\|.$$

Prove triangle inequality of Euclidean norm

We can use the Cauchy-Schwarz inequality to prove the triangle inequality for vectors. Let $\vec{a}$ and $\vec{b}$ be any vectors, then we have

$$\begin{aligned}
\|\vec{a}+\vec{b}\|^2 &= \|\vec{a}\|^2 + 2\vec{a}^\top \vec{b} + \|\vec{b}\|^2 \\
&\leq \|\vec{a}\|^2 + 2\|\vec{a}\|\|\vec{b}\| + \|\vec{b}\|^2 \\
&= \left(\|\vec{a}\| + \|\vec{b}\|\right)^2
\end{aligned}$$

Linear Dependence

- A collection of n-vectors $\vec{a}_1, \ldots, \vec{a}_k$ (with $k \geq 1$) is called linearly dependent if

$$\beta_1 \vec{a}_1 + \cdots + \beta_k \vec{a}_k = \vec{0}$$

  holds for some $\beta_1, \ldots, \beta_k$ that are not all zero.
- The zero vector can be formed as a linear combination of the vectors, with coefficients that are not all zero
- Linear dependence of a list of vectors does not depend on the ordering of the vectors in the list
- When a collection of vectors is linearly dependent, at least one of the vectors can be expressed as a linear combination of the other vectors. For instance, if $\beta_i \neq 0$, then

$$\vec{a}_i = (-\beta_1/\beta_i)\vec{a}_1 + \ldots + (-\beta_{i-1}/\beta_i)\vec{a}_{i-1} + (-\beta_{i+1}/\beta_i)\vec{a}_{i+1} + \ldots + (-\beta_k/\beta_i)\vec{a}_k$$

Linear dependence

- If any vector in a collection of vectors is a linear combination of the other vectors, then the collection of vectors is linearly dependent
- Linear dependence is an attribute of a collection of vectors, and not individual vectors

Linear independence

- A collection of n-vectors $\vec{a}_1, \ldots, \vec{a}_k$ (with ≥ 1) is called linearly independent if it is not linearly dependent, which means

$$\beta_1 \vec{a}_1 + \cdots + \beta_k \vec{a}_k = \vec{0}$$

  only holds for $\beta_1 = \cdots = \beta_k = 0$
- Linear independence is an attribute of a collection of vectors, and not individual vectors

Linear independence: Examples

- A list consisting of a single vector is linearly dependent only if the vector is zero. It is linearly independent only if the vector is nonzero.
- Any list of vectors containing the zero vector is linearly dependent. For instance,

$$0\vec{a}_1 + \cdots + 0\vec{a}_k + \beta_{k+1}\vec{0} = \vec{0},$$

  where the coefficients attached to $\vec{a}_1, \ldots, \vec{a}_k$ are 0, but the coefficient attached to $\vec{0}$ is non-zero
- A list of two vectors is linearly dependent if and only if one vector is a multiple of the other vector
- A list of vectors is linearly dependent if any one of the vectors is a multiple of another vector

Linear independence: Examples

- The following vectors

$$\vec{a}_1 = \begin{bmatrix} 0.2 \\ -7.0 \\ 8.6 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} -0.1 \\ 2.0 \\ -1.0 \end{bmatrix}, \quad \vec{a}_3 = \begin{bmatrix} 0.0 \\ -1.0 \\ 2.2 \end{bmatrix}$$

  are linearly dependent, because $\vec{a}_1 + 2\vec{a}_2 - 3\vec{a}_3 = \vec{0}$.
- The following vectors

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \vec{a}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

  are linearly independent.
  To show this, let $\beta_1\vec{a}_1 + \beta_2\vec{a}_2 + \beta_3\vec{a}_3 = 0$, which becomes

$$\begin{bmatrix} \beta_1 - \beta_3 \\ -\beta_2 + \beta_3 \\ \beta_2 + \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Linear independence: Examples

1) Take the sum of the last two equations and obtain that $\beta_3 = 0$
2) Plug-in $\beta_3 = 0$ into the 1st-and 2nd-equations and obtain $\beta_2 = \beta_1 = 0$
3) $\beta_1 \vec{a}_1 + \beta_2 \vec{a}_2 + \beta_3 \vec{a}_3 = \vec{0}$ if and only if $\beta_1 = \beta_2 = \beta_3 = 0$
4) Therefore, $\vec{a}_1, \vec{a}_2$ and $\vec{a}_3$ are linearly independent

Unit vectors are linear Independent

The following unit n-vectors

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdots \end{bmatrix}, \quad \cdots, \vec{e}_n = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ 1 \end{bmatrix}$$

are linearly independent. Let $\beta_1 \vec{e}_1 + \beta_2 \vec{e}_2 + \cdots + \beta_n \vec{e}_n = \vec{0}$. Its left-hand-side is actually $(\beta_1, \beta_2, \cdots, \beta_n)^\top$ and equals $\vec{0}$

Linear combinations of linearly independent vectors

- Suppose a vector $\vec{x}$ is a linear combination of $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$:

$$\vec{x} = \beta_1\vec{a}_1 + \beta_2\vec{a}_2 + \ldots + \beta_k\vec{a}_k$$

  When the vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ are linearly independent, the coefficients that form $\vec{x}$ are *unique*

- If $\vec{x}$ can be a linear combination of $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ with another set of coefficients

$$\vec{x} = \gamma_1\vec{a}_1 + \gamma_2\vec{a}_2 + \ldots + \gamma_k\vec{a}_k,$$

  then $\gamma_i = \beta_i$, for $i = 1, 2, \ldots, k$.

- We can find the coefficients that form a vector $\vec{x}$ as a linear combination of linearly independent vectors.
- To prove this, we have

$$\vec{0} = \vec{x} - \vec{x} = (\gamma_1 - \beta_1)\vec{a}_1 + (\gamma_2 - \beta_2)\vec{a}_2 + \ldots + (\gamma_k - \beta_k)\vec{a}_k$$

Independence and dimension inequality

- If the n-vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ are linearly independent, then $k \leq n$.
- A linearly independent collection of n-vectors can have at most n elements (vectors).
- Any collection of $n + 1$ or more n-vectors is linearly dependent

Basis

- A collection of n linearly independent n-vectors is called a *basis*
- If the n-vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$ are a basis, then any n-vector $\vec{x}$ can be written as a linear combination of them.
- Any n-vector $\vec{b}$ can be written in a *unique* way as a linear combination of a basis $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$.

Expansion in a basis

- When we express an n-vector $\vec{x}$ as a linear combination of a basis $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$, we refer to

$$\vec{x} = \beta_1 \vec{a}_1 + \ldots + \beta_n \vec{a}_n$$

  as the expansion of $\vec{x}$ in the basis of $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$.
- The numbers $\beta_1, \beta_2, \ldots, \beta_n$ are called the coefficients of the expansion of $\vec{x}$ in the basis of $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$

Examples

- The n standard unit n-vectors $\vec{e}_1, \ldots, \vec{e}_n$ are a basis. Any n-vector $\vec{b} = (b_1, \ldots, b_n)^\top$ can be expressed as linear combination:

$$\vec{b} = b_1 \vec{e}_1 + \ldots + b_n \vec{e}_n$$

  When we express an n-vector $\vec{x}$ as a linear combination of a basis $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$, we refer to $\vec{x} = \beta_1 \vec{a}_1 + \ldots + \beta_n \vec{a}_n$ as the expansion of $\vec{x}$ in the basis of $\vec{a}_1, \ldots, \vec{a}_n$.

Orthonormal vectors

- Orthogonal vectors: A collection of vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ is orthogonal if $\vec{a}_i^\top \vec{a}_j = 0$ for any i and j with $i \neq j$, and $i, j = 1, \ldots, k$
- Orthonormal vectors: A collection of vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ is orthogonal if it is orthogonal and $\|\vec{a}_i\| = 1$, for $i = 1, \ldots, k$
- These two conditions can be combined into one statement about the inner products of any pair of vectors in the collection: $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$ is orthonormal means that

$$\vec{a}_i^\top \vec{a}_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

- Orthonormal property is an attribute of a collection of vectors, and not an attribute of vectors individually.

Linear independence of orthonormal vectors

Orthonormal vectors are linearly independent.

Linear independence of orthonormal vectors

To prove this, suppose $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ are orthonormal, and let

$$\beta_1 \vec{a}_1 + \ldots + \beta_k \vec{a}_k = \vec{0}.$$

Taking the inner product of this equality with $\vec{a}_i$, we have

$$\begin{aligned} 0 = \vec{a}_i^\top \vec{0} &= \vec{a}_i^\top (\beta_1 \vec{a}_1 + \ldots + \beta_k \vec{a}_k) \\ &= \beta_1 (\vec{a}_i^\top \vec{a}_1) + \ldots + \beta_k (\vec{a}_i^\top \vec{a}_k) = \beta_i \end{aligned}$$

Linear combinations of orthonormal vectors

Suppose a vector $\vec{x}$ is a linear combination of orthonormal vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$ are orthonormal:

$$\vec{x} = \beta_1 \vec{a}_1 + \ldots + \beta_k \vec{a}_k.$$

Taking the inner product of this equality with $\vec{a}_i$, we have

$$\vec{a}_i^\top \vec{x} = \vec{a}_i^\top (\beta_1 \vec{a}_1 + \ldots + \beta_k \vec{a}_k) = \beta_i$$

Linear combinations of orthonormal vectors

For any $\vec{x}$ that is a linear combination of orthonormal vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k$, we have

$$\vec{x} = \beta_1 \vec{a}_1 + \ldots + \beta_k \vec{a}_k.$$

Taking the inner product of the left- and right-hand sides of this equation with $\vec{a}_i$, we have

$$\vec{a}_i^\top \vec{x} = \vec{a}_i^\top (\beta_1 \vec{a}_1 + \ldots + \beta_k \vec{a}_k) = \beta_i,$$

because $\vec{a}_i^\top \vec{a}_j = 0$ for any $j \neq i$.
Replacing β_i with $\vec{a}_i^\top \vec{x}$ in linear combination eqn, we have

$$\vec{x} = (\vec{a}_1^\top \vec{x})\vec{a}_1 + \ldots + (\vec{a}_k^\top \vec{x})\vec{a}_k. \tag{1}$$

This identity is a method to check if an n-vector $\vec{y}$ is a linear combination of orthonormal vectors $\vec{a}_1, \ldots, \vec{a}_k$.

Orthonormal basis

For a given orthonormal basis of $\vec{a}_1, \ldots, \vec{a}_k$, if an n-vector $\vec{y}$ satisfies

$$\vec{y} = (\vec{a}_1^\top \vec{y})\vec{a}_1 + \ldots + (\vec{a}_k^\top \vec{y})\vec{a}_k,$$

then $\vec{y}$ is linear combination of $\vec{a}_1, \ldots, \vec{a}_k$.

Orthonormal basis: Example

Show that the following three 3-vectors are orthonormal

$$\vec{a}_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \vec{a}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{a}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

- Simple calculation shows that $\vec{a}_1^\top \vec{a}_2 = \vec{a}_1^\top \vec{a}_3 = \vec{a}_2^\top \vec{a}_3 = 0$
- $\|\vec{a}_1\| = \sqrt{0^2 + 0^2 + 1^2} = 1$,
  $\|\vec{a}_2\| = \sqrt{(1/\sqrt{2})^2 + (1/\sqrt{2})^2 + 0^2} = 1$, and
  $\|\vec{a}_3\| = \sqrt{(1/\sqrt{2})^2 + (-1/\sqrt{2})^2 + 0^2} = 1$

Linear combination of orthonormal basis: Example

Show that the 3-vector $\vec{y} = (1, 2, 3)^\top$ is a linear combination of the orthonormal basis given above.

Answer: Simple calculation shows that

$$\vec{a}_1^\top \vec{y} = -3, \quad \vec{a}_2^\top \vec{y} = 3/\sqrt{2}, \quad \vec{a}_3^\top \vec{y} = -1/\sqrt{2}$$

$$(\vec{a}_1^\top \vec{y})\vec{a}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \quad (\vec{a}_2^\top \vec{y})\vec{a}_2 = \begin{bmatrix} 3/2 \\ 3/2 \\ 0 \end{bmatrix}, \quad (\vec{a}_3^\top \vec{y})\vec{a}_3 = \begin{bmatrix} -1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

Therefore, the sum of these 3 vectors are

$$(\vec{a}_1^\top \vec{y})\vec{a}_1 + (\vec{a}_2^\top \vec{y})\vec{a}_2 + (\vec{a}_3^\top \vec{y})\vec{a}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

which is exactly $\vec{y}$. This concludes that $\vec{y}$ is a linear combination of the the orthonormal basis $\vec{a}_1$, $\vec{a}_2$ and $\vec{a}_3$.