Linear Algebra lecture note

Let us start with defining basic notation that we will use in this course:

- Small letters are for scalars
- Bold small letters **v**, **w**, or small letters with an arrow on top $\vec{v}, \vec{w}$ are for vectors; bold small letters with a hat $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ represent unit vectors (length = 1)
- Capital italic letter, such as *V*, stands for vector spaces, or *B*, for basis.
- Lowercase indexed letters x_1, x_2, … x_n are components of vectors
- Bold capital letters **A**, **M** are for matrices
- Indexed bold letters $\mathbf{a}_1$, $\mathbf{a}_2$, … are column or row vectors of a matrix
- Bold **0** is a vector of certain dimensionality with 0 in each component

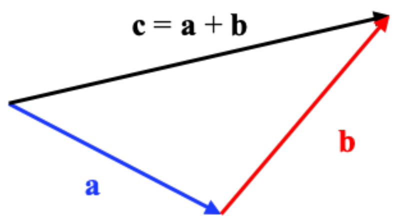
Vector

Vectors have both magnitude (size) and direction, usually denoted as an arrow in space, or a list of numbers.
For example, a 2D vector: an arrow on a flat plane, or a pair of real numbers.

Vectors are ordered lists of N numbers, each number as a component and N is the dimensionality of the vector. We usually arrange the numbers vertically to have a column vector, such as: $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{w} = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$ or horizontally as a row vector: $\mathbf{a} = [a_1, a_2, \ldots, a_n]$

Algebraic properties of vector operations:

- Addition
  Distributive: **u** + (**v** + **w**) = (**u** + **v**) + **w**
  Commutative: **v** + **w** = **w** + **v**
- Scalar multiplication
  a(b**v**) = (ab)**v**



Geometric example of **a** + **b** = **c** (Triangular inequality: ||**a**||+ ||**b**|| <= ||**c**||)
Vector addition: 'Tip-to-tail method' place the initial point (tail) of each successive vector at the terminal point (head) of the previous vector, draw the resultant from tail of the first vector to the head of the last vector.

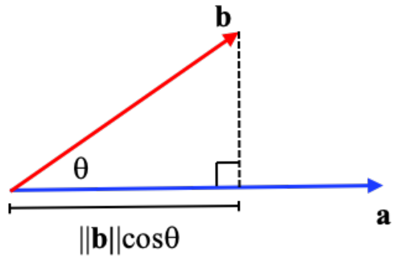
The norm ||**v**|| (or |**v**|) of a vector is its length, we can scale a vector to a unit vector by dividing its norm $\hat{\mathbf{v}} = \frac{\mathbf{v}}{||\mathbf{v}||}$

Vectors can be added and scaled (combination of vectors) to form a vector space.

How do we multiple vectors? There is more than one way!

Dot product (also termed 'inner product') of two vectors **a** = $[a_1,a_2,\ldots,a_n]$, **b** = $[b_1,b_2,\ldots,b_n]$ is a scalar, it is the sum of the pairwise product of corresponding elements:

$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + \ldots + a_nb_n$



Geometrically it is the product of the length of one vector and the length of projection of the 2nd vector on the 1st one:

$\mathbf{a}\cdot\mathbf{b} = ||\mathbf{a}||\,||\mathbf{b}||\cos(\theta)$

This is also a way of defining the angle θ between two vectors!

The dot product of two perpendicular/orthogonal vectors is 0

The dot product of two parallel vectors is the product of their norms

The dot product of a vector with itself is the square of its norm: $\mathbf{a}\cdot\mathbf{a} = ||a||^2$

Euclidean length of a vector is the square root of the dot product with itself, i.e. $||a|| = \sqrt{\mathbf{a}\cdot\mathbf{a}}$

Orthogonal vectors are linearly independent, their dot product is zero, e.g. [1 0] and [0 3]

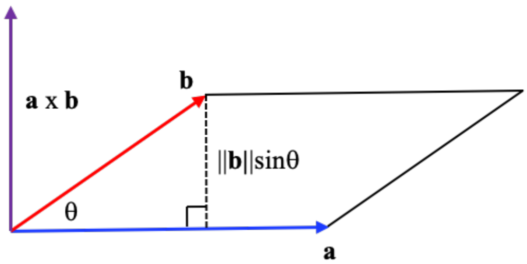
Algebraic properties of dot product:

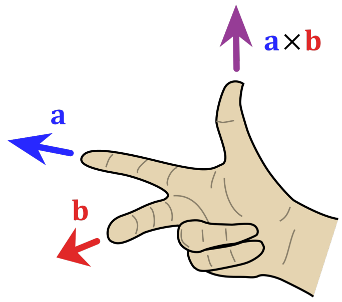
- Commutative: $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$
- Distributive over vector addition: $\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}$
- Scalar multiplication: $(c_1\mathbf{a})\cdot(c_2\mathbf{b}) = c_1c_2(\mathbf{a}\cdot\mathbf{b})$

Cross product (also called vector product) of two vectors **a** and **b** is another vector **c** that is right angles to both. The magnitude of the cross product equals to the area of a parallelogram with vector **a** and **b** for sides.

$\mathbf{a} \times \mathbf{b} = ||\mathbf{a}||\,||\mathbf{b}||\sin(\theta)\mathbf{n}$

n is the unit vector perpendicular to both **a** and **b**, its direction follows right hand rule.





The cross product of two vectors is zero when they point in the same, or opposite direction

The cross product of two vectors reaches maximum length when two vectors are at right angle

Linear combination, Span, Basis vector

https://www.youtube.com/watch?v=k7RM-ot2NWY

Vectors live in vector space.

If vector $\mathbf{v}$ can be written in the form $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$, then $\mathbf{v}$ is said to be a linear combination of the vector $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, where $c_1, c_2, \ldots, c_n$ are scalars.

Span$\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is defined as the set of all linear combinations of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in R^m$
i.e. $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n$

It is all the set of vectors we can reach by just two basic operations: addition and scalar multiplication. You can think it as the space the basis vectors fill out. A set of basis vectors form a basis, *B*.

Example: Span$\{\binom{1}{0}, \binom{0}{1}\} = R^2$
Standard Basis vectors in $\hat{\mathbf{i}}, \hat{\mathbf{j}}$: $\binom{1}{0}$ $\binom{0}{1}$, they are orthogonal.
Every vector is a linear combination of standard basis vectors!
$\mathbf{x}$: $\binom{x1}{x2} = x1\binom{1}{0} + x2\binom{0}{1}$

The coefficients x1, x2 of this linear combination are referred as coordinates on *B* of the vector.

When removing one of the vectors does not affect the span, we say this vector is linearly dependent on other vectors (i.e. it is a linear combination of the others). 'Dependent' means that this vector really does not give new information.

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent if and only if the solution to the equation:
$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n = 0$ is $x_n = 0$ for all n

Recap:

- Any vector in the vector space *V* can be represented as a linear combination of basis vectors: $\mathbf{v} = a_1\mathbf{v_1} + a_2\mathbf{v_2} + \ldots + a_n\mathbf{v_n}$, i.e. using addition and scaling operation
- Example: $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ orthogonal vectors (Q: does orthogonality imply linear independence?)
- Orthonormal basis: basis composed of orthogonal and unit (length equals to 1) vectors, e.g. the standard basis that lie along the axes of the space
- Vector space R^N requires a basis of size N

Exercise: how to determine if one vector is in the span of other vectors

Is $\begin{pmatrix}-1\\4\\11\end{pmatrix} \in$ Span$\left\{\begin{pmatrix}1\\2\\-4\end{pmatrix}, \begin{pmatrix}-3\\-5\\13\end{pmatrix}, \begin{pmatrix}2\\-1\\-12\end{pmatrix}\right\}$?

Linear Transformation (L.T.) and Matrix Multiplication

https://www.youtube.com/watch?v=kYB8IZa5AuE
https://www.youtube.com/watch?v=XkY2DOUCWMU
https://www.youtube.com/watch?v=Ip3X9LOh2dk

A transformation of vector space is linear if:

1. 'Grid lines' remain parallel and evenly spaced
2. Origin remains fixed

*Rotation, scaling, reflection, shearing are common linear transformations

Where do the new vectors corresponding to all of the original vectors land after a transformation? Just keep note of where the basis vectors land! As long as it is a linear transformation, all of the new vectors are still the same linear combination of the basis vectors.

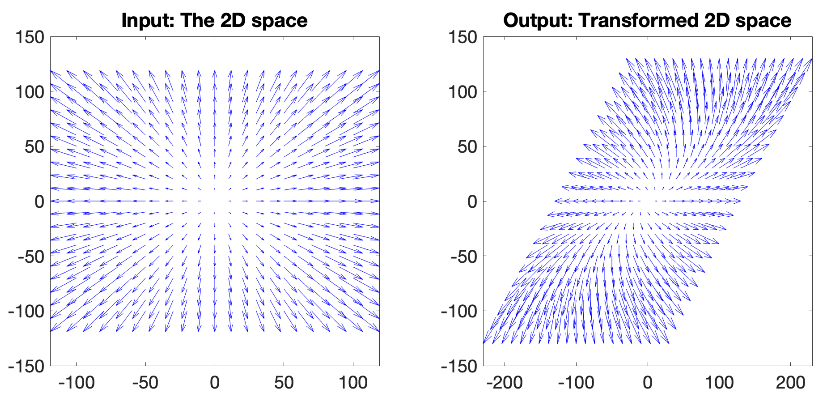
For example, a 2D transformation can be described in terms of just 4 numbers—the new coordinates of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$.

It is common to package them up into a 2 x 2 matrix where the columns can be interpreted as landing points of basis vectors in space.

Matrix-vector multiplication is simply a way of telling what the transformation does to a vector.

Shear transformation example: $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

Visualise in 2D:



Matrix multiplication

<u>Perspective 1</u>: Multiplying matrix **A** by a vector **x** is taking the dot product of each row in **A** and column of **x** (same apply for matrix-matrix multiplication)

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + 2y + 3z \\ 4x + 5y + 6z \\ 7x + 8y + 9z \end{bmatrix}$$

Perspective 2: **Ax** is a linear combination of column vectors of **A**:

$\mathbf{Ax} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3}, \ldots, x_n\mathbf{v_n}$

$\mathbf{A} = [\mathbf{v_1}\ \mathbf{v_2}\ \mathbf{v_3}\ \ldots\ \mathbf{v_n}]$

$\mathbf{x} = [x_1, x_2, x_3, \ldots, x_n]^T$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} + z \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

Basic matrix operations:

- The transpose of an m x n matrix **A** is to swap rows and columns to get an n x m matrix $\mathbf{A}^T$, the i, j element of $\mathbf{A}^T$ is the j, i element of **A**.

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Algebraic properties of taking transpose:

$(\mathbf{W}^T)^T = \mathbf{W}$

$(c\mathbf{W})^T = c(\mathbf{W})^T$

$(\mathbf{M} + \mathbf{N})^T = \mathbf{M}^T + \mathbf{N}^T$

$(\mathbf{MN})^T = \mathbf{N}^T\mathbf{M}^T$

Descriptions like “symmetric”, “diagonal” are the shape and property of the matrix, and influence their transformations:

- If the matrix is square, and the diagonal elements are all one, the matrix does nothing, it is the identify matrix, denoted **I**, e.g. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- A symmetric matrix is a square matrix that is equal to its transpose, i.e. $\mathbf{A}^T = \mathbf{A}$ $\begin{bmatrix} 2 & 5 \\ 5 & 6 \end{bmatrix}$

- A diagonal matrix is one for which only elements along the diagonal can be non-zero

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

<u>A diagonal matrix operates on vector spaces by stretching or compressing axes</u>

- An orthogonal matrix **B** is a square matrix whose columns are pairwise orthogonal unit vectors (i.e. orthonormal vectors):

$\mathbf{B} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n]$

$\mathbf{v}_i \cdot \mathbf{v}_i = 1$; $\mathbf{v}_i \cdot \mathbf{v}_j = 0$, i not equal to j

$\mathbf{BB}^T = \mathbf{B}^T\mathbf{B} = \mathbf{I}$

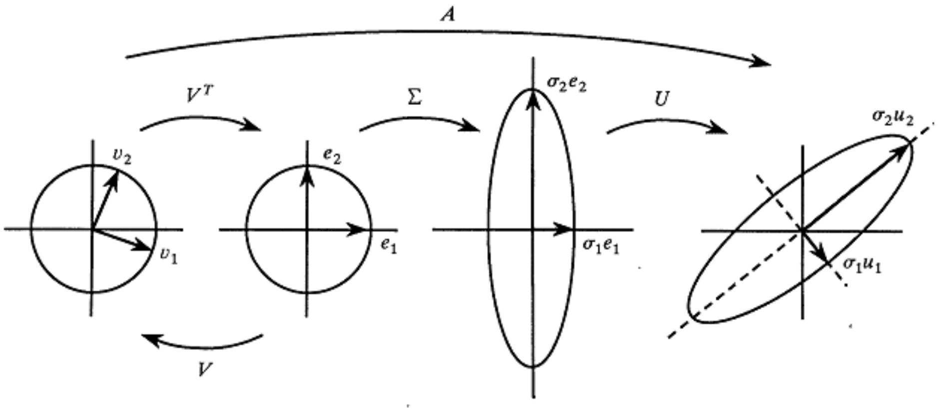
<u>The operation of an orthogonal matrix is a rotation.</u>

Exercise: Can you give some examples of orthogonal matrix?

$\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$ $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

Later we will learn singular value decomposition (SVD). We will be able to decompose any matrix **A** into a product of 3 matrices: **A** = **UΣV**^{T}

U and **V** are orthogonal matrices, and **Σ** is a diagonal matrix, we can therefore easily describe the operation of **A** in terms of a rotation, followed by scaling of the axes, and then another rotation! Here we can see, how matrix **A** as a linear transformation morphs a sphere in its domain to a tilted ellipse:



The sequential application of multiple linear systems is still a linear system!
The matrix associated with the full system is a matrix product of the subsystem matrices:
A(**BC**) = (**AB**)**C** *It is associative but not commutative, i.e. order is important.

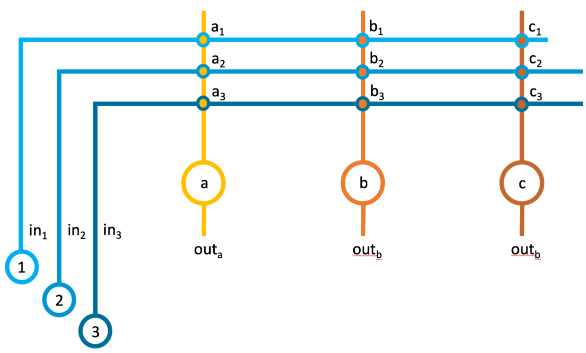
Determinant of a Matrix
det(**A**) tells how much a transformation scales areas, it would be zero if L.F. squished all vectors onto a smaller dimension
if det < 0, L.F. inverts the orientation of space (result in a flipped space)

For a 2 x 2 matrix, det($\begin{bmatrix} a & b \\ c & d \end{bmatrix}$) = ad – bc

det($\mathbf{M}_1\mathbf{M}_2$) = det($\mathbf{M}_1$)det($\mathbf{M}_2$)

A simple neural circuit example (e.g. cerebellar granule cells synapse on Purkinje cells):

Let us measure the behaviour of multiple neurons simultaneously with multiple inputs.



The synaptic weight between output neuron 'a' and input neuron '1' is denoted as a_1
Assume that the response of a neuron to the combinations of multiple presynaptic neurons is just the weighted sum of all individual inputs, e.g. for neuron a: $out_a = a_1 \cdot in_1 + a_2 \cdot in_2 + a_3 \cdot in_3$
Same for neuron b: $out_b = b_1 \cdot in_1 + b_2 \cdot in_2 + b_3 \cdot in_3$; and neuron c: $out_c = c_1 \cdot in_1 + c_2 \cdot in_2 + c_3 \cdot in_3$

Input vector: $\overrightarrow{in} = \begin{pmatrix} in\ 1 \\ in\ 2 \\ in\ 3 \end{pmatrix}$ Output vector: $\overrightarrow{out} = \begin{pmatrix} out\ a \\ out\ b \\ out\ c \end{pmatrix}$

we can use the dot product notation to emphasize that the weights and inputs are vectors:

$\overrightarrow{out}_a = \vec{a} \cdot \overrightarrow{in}$
$\overrightarrow{out}_b = \vec{b} \cdot \overrightarrow{in}$
$\overrightarrow{out}_c = \vec{c} \cdot \overrightarrow{in}$,

$$\begin{bmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \\ c1 & c2 & c3 \end{bmatrix} \begin{pmatrix} in\ 1 \\ in\ 2 \\ in\ 3 \end{pmatrix} = \begin{pmatrix} out\ a \\ out\ b \\ out\ c \end{pmatrix}$$

Matrix multiplication: $\mathbf{W}\overrightarrow{in} = \overrightarrow{out}$
This looks like a function y = wx, right? The entire behaviour of the circuit depends on just the synaptic weights in this example!

Matrix equation: $\mathbf{Ax} = \mathbf{v}$
We are looking for a vector **x**, after applying transformation **A**, lands on vector **v**

Solving systems of linear equations

The classic motivation for the study of linear algebra is the solution of a set of linear equations such as:

$a_{11}\mathbf{v}_1 + a_{12}\mathbf{v}_2 + \ldots + a_{1N}\mathbf{v}_N = \mathbf{b}_1$
$a_{21}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \ldots + a_{2N}\mathbf{v}_N = \mathbf{b}_2$
...
$a_{M1}\mathbf{v}_1 + a_{M2}\mathbf{v}_2 + \ldots + a_{MN}\mathbf{v}_N = \mathbf{b}_M$

"**Av** = **b**"
We can use tools of linear algebra to determine if there is a solution for **v**

Algebraic properties of matrix multiplication/linear transformation:

- Additivity: $\mathbf{L}(\mathbf{v} + \mathbf{w}) = \mathbf{Lv} + \mathbf{Lw}$
- Scaling: $\mathbf{L}(c\mathbf{v}) = c\mathbf{Lv}$

Augmented matrix, Gaussian elimination, and reduced row echelon form (rref)

Exercise1: Solve the following linear equations

$x + 2y + 3z = 1$

$3x + 2y + z = 7$

$2x + y + 2z = 1$

How do we solve linear system equations as such?
Coefficient matrix + constant vector = augmented matrix

Gaussian elimination:
Elementary row operations that allow us to solve complicated linear equations with relatively little hassle, we can:

1. Switch any two rows
2. Multiply a row by a nonzero constant
3. Add one row to another

Reduced row echelon form (rref):

1. All zero rows are at the bottom
2. Each leading 1 is to the right of those above it
3. All entries are zero below a leading 1
4. Zeros directly above leading 1s

Exercise2: Find all solutions of vector equation: $x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3} = 0$

$$\mathbf{v}1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{v}2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{v}3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

What conclusion can be made about linear dependence of the system vectors $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$?