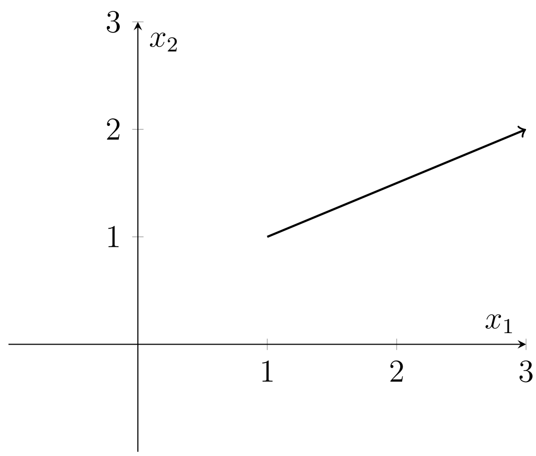
# 1 Vectors

A **vector** is a combination of numbers representing a magnitude and a direction.

They are defined by an origin and an endpoint.

*Example: 2D vector*

> *The origin and the endpoint in the 2D plane can be represented by pairs of numbers in the form ($x_1$,$x_2$)*
>
> *The following shows the vector originating at coordinate ($x_1$,$x_2$) =* $(1, 1)$ *and ending at ($x_1$,$x_2$) =* $(3, 2)$




Each of the numbers is in a coordinate system defined by the $x_1$ and the $x_2$ axis.

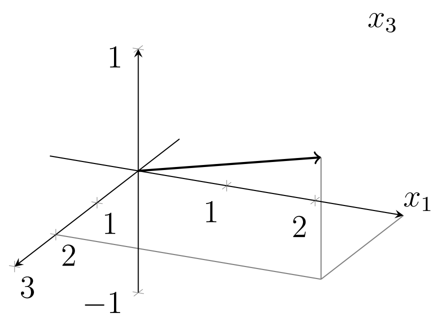
In the 2D vector example above, the coordinate of the origin and endpoint has a **component** $x_1$ and a component $x_2$.

In printed text, vectors are often identified by a bold font, *e.g* $\mathbf{x}$. In written text, vectors are often identified by an arrow above the variable, *e.g.* $\vec{x}$.

A vector can be defined in any number of dimensions (even an *infinite* number of dimensions).

*Example: 3D vector*

> *In a 3-dimensional space, a vector can have origin at (0,0,0) and an endpoint at (*$3{,}2{,}1$*).*




More generally, in an N-dimensional space, a vector can have origin at (0,...,0) and an endpoint at ($x_1$,...,$x_N$). Note that vectors in a space of dimensions larger than 3 cannot be easily presented in a graph.

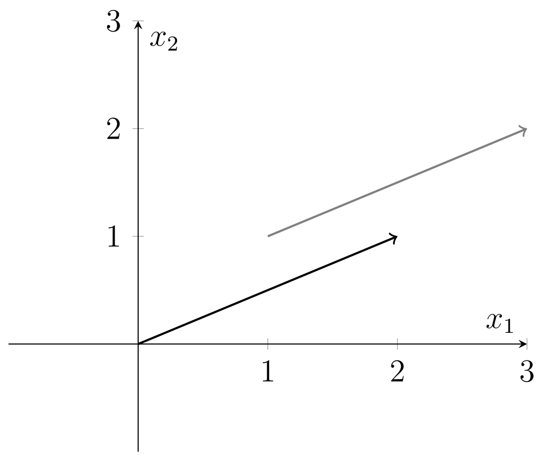
Vectors of similar dimension can be added and subtracted from each other. In $N$-dimensions:

$$\mathbf{y} - \mathbf{x} = (y_1 - x_1, y_2 - x_2, ..., y_N - x_N, )$$

Subtracting a vector pointing from the origin to the start point of a vector **translates** this vector to the origin.

*Example: Translation of a 2D vector*

> *A 2D vector starts at the point (*$x_1$*,*$x_2$*) =* $(1, 1)$ *and ends at the point (*$x_1$*,*$x_2$*) =* $(3, 2)$ *Subtracting the start point from the end point yields:* $(3-1, 2-1) = (2, 1)$.

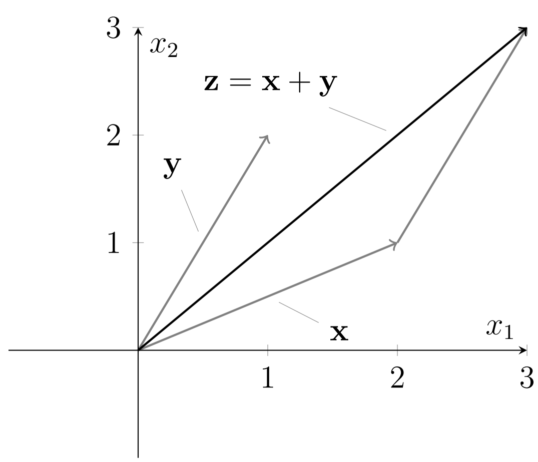



When a vector is defined by $\mathbf{x} = (x_1, x_2, \cdots, x_N)$, it is implied that the origin is as $\mathbf{0} = (\underbrace{0, \cdots, 0}_{N zeros})$.

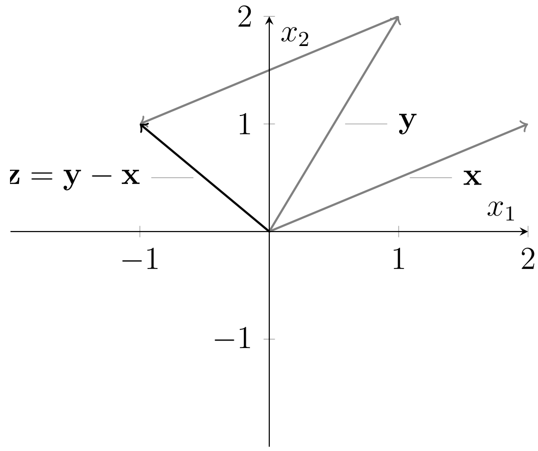
Almost all operations on vectors assume that the origin has been translated to $\mathbf{0}$

*Example: Addition and Subtraction of two 2D vectors*

> *Adding a vector is equivalent to "concatenating" them. So if* $\mathbf{x} = (2, 1)$ *and* $\mathbf{y} = (1, 2)$*, then* $\mathbf{z} = \mathbf{y} + \mathbf{x} = (3, 3)$.




> *Similarly for subtraction: So if* $\mathbf{x} = (2, 1)$ *and* $\mathbf{y} = (1, 2)$*, then* $\mathbf{z} = \mathbf{y} - \mathbf{x} = (-1, 1)$.




A **scalar** multiplication with a vector is:

$$r\mathbf{x} = (rx_1, rx_2)$$

For positive scalar $r$, the direction of the vector is preserved and the vector is stretched or shrunk. For negative $r$ the direction of the vector is reversed.

The length of a vector is called the **norm**. It is the distance between the origin and the end point (for vectors starting at $\mathbf{0}$). The norm is defined as:

$$||\mathbf{x}|| = \sqrt{x_1^2 + ... + x_N^2} = \sqrt{\underbrace{\sum_{i=1}^{N}}_{\text{summation symbol}} x_i^2}$$

*Example: Norm of a 2D vector*

*The norm of* $\mathbf{x} = (3, 4)$ *is* $||\mathbf{x}|| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$

A **unit vector** is a vector of length 1, *i.e.* $||\mathbf{x}|| = 1$

Note that the norm of a vector multiplied by a scalar $r$ is the norm of the vector times the scale:

$$||r\mathbf{x}|| = r||\mathbf{x}||$$

Any non-unit vector can be transformed into a unit vector (or **normalized**) by dividing the vector components by its norm:

$$\mathbf{u} = \frac{\mathbf{x}}{||\mathbf{x}||}$$

$\mathbf{u}$ is in fact a unit vector since:

$$||\mathbf{u}|| = ||\frac{\mathbf{x}}{||\mathbf{x}||}|| = \frac{||\mathbf{x}||}{||\mathbf{x}||} = 1$$

Note that this is an example of scalar multiplication of a vector as seen above.

*Example: Unit vectors*

> $\mathbf{u} = (1, 0, 0)$ *is a unit vector.* $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ *is a unit vector*

The coordinate axes are sometimes written as unit vectors: $\mathbf{u}_1 = (1, \cdots, 0), \mathbf{u}_2 = (0, 1, \cdots, 0), \cdots \mathbf{u}_N = (0, \cdots, 1)$.

Then any vector $\mathbf{x} = (x_1, \cdots, x_N)$ can be written as:

$$\mathbf{x} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_N\mathbf{u}_2$$

The **dot product** of two vectors is defined as:

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{N} x_i y_i$$

Note that the dot product of the vector itself is the square of its norm:

$$\mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^{N} x_i x_i = ||x||^2$$

The dot product quantifies the “projection” of one vector to another, such that:

$$\cos(\theta) = \frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}||||\mathbf{y}||}$$

where $\theta$ is the angle between the two vectors.

For two perpendicular vectors (*i.e.* the angle is $\theta = 90\,\mathrm{deg} = \frac{\pi}{2}$), their dot product is zero).

Two vectors are **orthogonal** if their dot product is zero:

$$\mathbf{x} \cdot \mathbf{y} = 0$$

Two vectors are **orthonormal** if their dot prodict is zero and they have lentgth 1: $||\mathbf{x}|| = ||\mathbf{y}|| = 1$ and they are orthogonal $\mathbf{x} \cdot \mathbf{y} = 0$ .

Vectors and reals can be defined in many differents ways. In general, they are defined as elements of **Vector spaces**, a collection of vectors that can be added together and multiplied by a scalar that verify certain properties and conditions. The **Euclidean space** is a special case where the components of each vectors are real numbers.

The components of the vector do not need to be reals. They can be complex numbers, rational numbers (but not integers), functions, polynomials or in general from any field that satisfies a number of axioms. In this course, we will only use reals.

The following conditions hold for any vector in a vector space:

- Commutative: $\mathbf{x}$+$\mathbf{y}$ = $\mathbf{y}$+$\mathbf{x}$
- Associative ($\mathbf{x}$+$\mathbf{y}$) + $\mathbf{z}$ = $\mathbf{x}$+($\mathbf{y}$+$\mathbf{z}$)
- Zero element $\mathbf{x}$+$\mathbf{0}$ = $\mathbf{x}$
- Inverse $\mathbf{x}$-$\mathbf{x}$ = $\mathbf{0}$
- Scalar Multiplication: $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$, $\alpha(\beta\mathbf{x}) = (\alpha\beta\mathbf{x})$

Note that adding a scalar to a vector: $\mathbf{x} + r$ is not allowed. MATLAB and other vector based programming frameworks will interpret this as $\mathbf{x} + r\mathbf{1}$, where $\mathbf{1}$ is a vector of ones with the same dimension as $\mathbf{x}$. Similarly, adding two vectors of different dimensions is not allowed and will result in an error in MATLAB.

**MATLAB** vectors and functions:

- $i$th component of vector $\mathbf{x}$ is x(i)
- vectors can be defined on rows or columns: x = [ 1 2 1] is a row vector, y = [1; 2; 1] is a column vector. Column vectors can be transformed into row vector using x = y'.
- Additions and multiplications of scalars are written x+y and k*x
- Multiplying two vectors in MATLAB can be of two types: dot product x*y' or element-wise x.*y (component by component multiplication).
- Norm norm(x)
- Absolute value abs(x)
- Generate vectors: x = 1:3, x = 1:.1:3 ,
- Random vectors with rand, randn, randi

A **Linear Combination** of vectors is defined as:

$$\sum_{i=1}^{N} \alpha_i \mathbf{x}_i$$

We've seen that the **axes** are orthonormal vectors themselves. Together they **span** a **vector space**. A vector in this space can be written a linear combination of the unit vectors $\mathbf{u}_1$,$\mathbf{u}_2 \ldots \mathbf{u}_N$:

$$\mathbf{x} = a\mathbf{u}_1 + b\mathbf{u}_2$$

The set of vectors $\mathbf{u}_1$,$\mathbf{u}_2 \ldots \mathbf{u}_N$ are said to form a **basis** of the vector space. If the basis vectors are othonormal, then it is called an **orthonormal basis**.

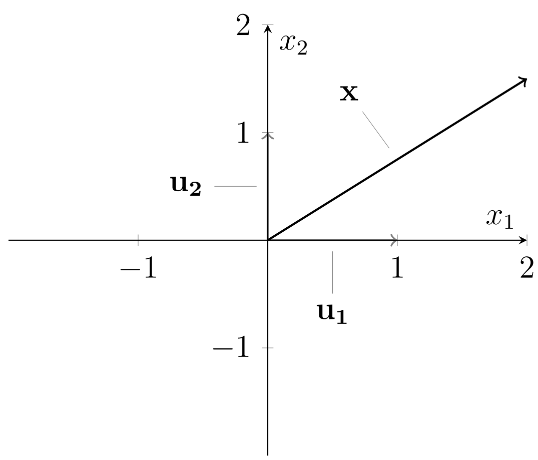
The inner product is a generalization of the dot product in non-euclidean vector spaces. It is a way to multiply vectors together such that the result is a scalar.

The inner product of two vectors is denoted $\langle x, y \rangle$. It is equal to the dot product in Euclidean space.

A vector space combined with a inner product is called an **inner product space**.

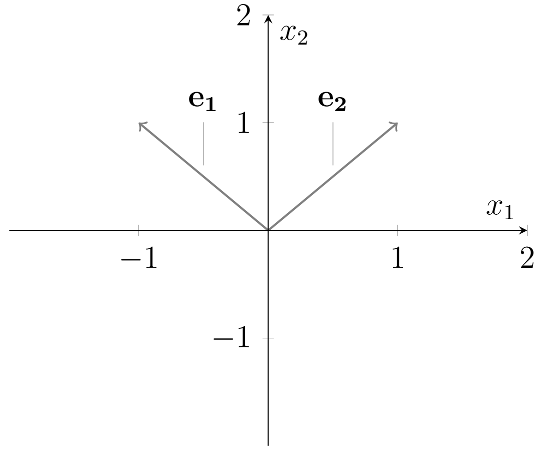
*Example: 2D vector Space*

> *For example, in 2D, the vector* $\mathbf{x} = (2, 1.5)$ *can be written:*
>
> $$\mathbf{x} = 2\mathbf{u}_1 + 1.5\mathbf{u}_2$$
>
> *where* $\mathbf{u}_1 = (1, 0)$*, and* $\mathbf{u}_2 = (0, 1)$.



*Any vector in 2D can be written in this fashion.*

Note that the following vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ also form a basis:




A vector space can be spanned by (infinitely) many basis vectors. But note that any collection of vectors does not necessarily span a space of dimensions equal to the number of vectors. For example, we cannot write $\mathbf{x} = (2, 3)$ as a linear combination of vectors $(1, 0)$ and $(2, 0)$

To form a basis for a $N$ dimensional vector space, the basis must be formed by $N$ vectors **linearly independent** vectors.

Vectors can be tested for linear independence. For example, let's assume that:

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_3\mathbf{x}_3 = 0 \tag{1}$$

has only solution $\alpha_1 = \alpha_2 = \alpha_3 = 0$. In other words $x_1, x_2$ or $x_3$ cannot be expressed as a combination of the others. The test for linear independence can be extended to $N$ dimensions.

*Example: Linear independence in 3D*

$$
\begin{aligned}
\mathbf{x} &= (3, 2, 3) \\
\mathbf{y} &= (1, 1, 0) \\
\mathbf{z} &= (0, 1, -3)
\end{aligned} \tag{2}
$$

*Test for independence:*

$$\alpha_x \mathbf{x} + \alpha_y \mathbf{y} + \alpha_z \mathbf{z} = \mathbf{0}$$

*Solve the equations along each component:*

$$
\begin{aligned}
0 &= 3\alpha_x + \alpha_y \\
0 &= 2\alpha_x + \alpha_y + \alpha_z \\
0 &= 3\alpha_x - 3\alpha_z
\end{aligned} \tag{3}
$$

*We see that any $\alpha_x = \alpha_z$ and $\alpha_y = -3\alpha_x$ is a solution. Therefore the vectors are not linearly independent.*

A vector can be expressed in a different basis. This is called a **change of basis**. A change of basis can be done by projecting the vector along the new basis using the dot product.

*Example: Change of Basis in 2D*

*Basis 1:* $\{\mathbf{u}_1 = (1, 0), \mathbf{u}_2 = (0, 1)\}$
*Basis 2:* $\{\mathbf{u'}_1 = (\sqrt{2}/2, \sqrt{2}/2), \mathbf{u'}_2 = (-\sqrt{2}/2, \sqrt{2}/2)\}$

*Assume that* $\mathbf{z} = (a, b)$ *in Basis 1,* i.e.

$$
\begin{aligned}
\mathbf{z} \cdot \mathbf{u}_1 &= a \\
\mathbf{z} \cdot \mathbf{u}_2 &= b
\end{aligned} \tag{4}
$$

*These are the components of the vector in the original basis (Basis 1). In basis 2,*

$$
\begin{aligned}
\mathbf{z} \cdot \mathbf{u'}_1 &= a\frac{\sqrt{2}}{2} + b\frac{\sqrt{2}}{2} \\
\mathbf{z} \cdot \mathbf{u'}_2 &= -a\frac{\sqrt{2}}{2} + b\frac{\sqrt{2}}{2}
\end{aligned} \tag{5}
$$

# 2 Matrices

A **Matrix** is a collection of vectors of the same vector space. The following defines a $m \times n$ matrix. $m$ is the number of rows, and $n$ is the number of columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

Matrices are often (but not always) denoted by uppercase letters. The element of a matrix is referred to as $a_{ij}$, where $i$ can be any of $1$ to $m$, and $j$ can be any of $1$ to $n$.

A vector is a special case with $m = 1$ (row vector) or $n = 1$ (column vector).

If $A = B$, then $a_{ij} = b_{ij}$ for all $i$ and all $j$.

**Addition/Subtraction** of matrices: The matrix $C = A + B$ has elements $c_{ij} = a_{ij} + b_{ij}$. The sum of any $m \times n$ matrices is a matrix of size $m \times n$.

**Multiplication** of matrices is a generalization of the inner product (dot product in Euclidean space). To multiply matrices, the dimensionalities much match as follows

$$\underbrace{C}_{m\times p} = \underbrace{A}_{m\times n} \cdot \underbrace{B}_{n\times p}$$

*i.e.* the number of columns of $A$ must match the number of rows of $B$.

A matrix can be multiplied by a vector:

$$\underbrace{\mathbf{c}}_{m\times 1} = \underbrace{A}_{m\times n} \cdot \underbrace{\mathbf{b}}_{n\times 1}$$

The result is a vector.

$$\mathbf{c} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ \dots \\ b_{m1} \end{bmatrix}$$

Each component of $\mathbf{c}$ is equal to an inner product between rows of $\mathbf{a}$ and columns of $\mathbf{b}$ (here we have only one column).

$$c_{11} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ \dots \\ b_{n1} \end{bmatrix} = a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1}$$

For all components of $\mathbf{c}$:

$$c_i = a_{i1}b_{11} + a_{i2}b_{21} + \dots + a_{in}b_{n1}$$

The multiplication of two matrices $\underbrace{C}_{m\times p} = \underbrace{A}_{m\times n} \cdot \underbrace{B}_{n\times p}$

is defined in a similar manner:

$$C_{ij} = a_{ij}b_{ij} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}$$

Note that matrices do not commute in general, *i.e.* $AB \neq BA$.

*Example: Matrix Multiplication*

*Multiply the following two matrices*

$$A = \begin{bmatrix} 1 & 3 & 5 \\ -2 & 4 & -6 \end{bmatrix}$$

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

*Calculate* $C = AB$

The transpose of a matrix denoted $A^\top$ is:

$$C = A^\top \leftrightarrow c_{ij} = a_{ji}$$

So if $A$ is $m \times p$, $C$ will be $p \times m$.

*Example: Matrix transpose*

$$A = \begin{bmatrix} 1 & 3 & 5 \\ -2 & 4 & -6 \end{bmatrix}$$

$$A^\top = \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ 5 & -6 \end{bmatrix}$$

**Properties of matrices**

- $A + B = B + A$, $A, B$ are $m \times n$
- $(A + B) + C = A + (B + C)$, $A, B, C$ are $m \times n$
- $A(B + C) = AB + AC$
- $\alpha(A + B) = \alpha A + \alpha B$, $\alpha$ is a scalar
- $A(BC)$ = $(AB)C$
- $(A^\top)^\top = A$
- $(AB)^\top = B^\top A^\top$
- the transpose of a row vector is a column vector
- in MATLAB A' means $A^\top$
- if $\mathbf{x}$ is an eigenvector, $\alpha\mathbf{x}$ is also an eigenvector with the same eigenvalue, where $\alpha$ is a scalar.

## 2.1 Special Matrices

A **square matrix** of size $n \times n$ is a matrix whose number of columns match the number of rows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

The elements $a_{11}, a_{22}, \cdots, a_{nn}$ are the **diagonal** elements. The elements $a_{ij}$, $i \neq j$ are called the **off-diagonal** elements. A **diagonal matrix** is one where off-diagonal elements are zero.

*Example: Diagonal Matrix*

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

*is a diagonal matrix.*

A **symmetric matrix** is one for which $a_{ij} = a_{ji}$. ( Symmetry is with respect to the diagonal ).

*Example: Symmetric Matrix*

$$\begin{bmatrix} 2 & -2 & 1 \\ -2 & 1 & -3 \\ 1 & -3 & -2 \end{bmatrix}$$

*is a symmetric matrix.*

An **identity matrix** is a matrix that has ones on the diagonal and zeros otherwise. It is often denoted $I$ or $I$:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Note that $AI = A = IA$

**Matrices in MATLAB**

- A = [] is an empty matrix of size $0 \times 0$
- diag(a) creates a matrix with elements of vector $\mathbf{a}$ on the diagonal.
- diag(A) Returns the diagonal of matrix $A$.
- eye(n) Creates an identity matrix of size $n \times n$.
- ones(m,n) Creates an $m \times n$ matrix filled with ones ($a_{ij} = 1$ for all $i, j$).
- rand(m,n) Creates an $m \times n$ matrix filled with random numbers.
- zeros(m,n) Creates an $m \times n$ matrix filled with zeros.

## 2.2 Matrix inverse

$A^{-1}$ is the inverse of a matrix $A$ if:

$$AA^{-1} = A^{-1}A = I$$

**Not all matrices can be inverted!** If a matrix has an inverse, it is called **invertible** or **nonsingular**. If a matrix does not have an inverse, it is **singular** or **non invertible**.

Invertibility (or singularity) has to do with the **determinant** of a matrix. The **determinant** can be viewed as the generalization of length applied to matrices. The determinant of matrix $A$ is denoted $\det A$ or $|A|$. If the determinant of a matrix is zero, then it is non-invertible:

$$|A| = 0 \leftrightarrow A \text{ is singular \& noninvertible}$$

This is because the formula for calculating the inverse involves $\frac{1}{|A|}$. Intuitively, if $|A| = 0$ then the inverse cannot be calculated (since $\frac{1}{0}$ is indetermined)

The general formula for calculating the determinant is complex. However of $2 \times 2$ matrices, the formula is straightforward:

$$det(\begin{bmatrix} a & b \\ c & d \end{bmatrix}) = ad - bc$$

For a $3 \times 3$ matrix:

$$det \begin{vmatrix} a\,b\,c \\ d\,e\,f \\ g\,h\,i \end{vmatrix} = a\,det \begin{vmatrix} e\,f \\ h\,i \end{vmatrix} - b\,det \begin{vmatrix} d\,f \\ g\,i \end{vmatrix} + c\,det \begin{vmatrix} d\,e \\ g\,h \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh$$

**Properties of the determinant:**

- If any row or column of matrix $A$ is all zeros, then $|A| = 0$.

- If the off-diagonal terms are all zeros, then the determinant is the product of the diagonal elements.

- $|A^\top| = |A|$
- $|AB| = |A||B|$
- **Only square matrices have determinants and inverses!** Non-square matrices have "pseudo-inverses". Similarly only square matrices have determinants. Pseudo inverses are commonly used for linear regression. Recall that in linear regression, one minimizes least squares error: $argmin(X\beta - \mathbf{y})^2$. The optimal $\beta$ to this equation is the pseudoinverse $\beta = (X^\top X)^{-1}Y$

- The following MATLAB functions calculate the inverse and the determinant respectively: inv(A), det(A).

An **Orthogonal Matrix** $Q$ is a $m \times n$ matrix that is invertible, whose inverse is its transpose:

$$Q^{-1} = Q^\top$$

So $Q^\top Q = QQ^\top = I$.

*Example: Rotation*

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

*is an orthogonal matrix,* i.e. $QQ^\top = I$. *This is a matrix that does a rotation by angle $\theta$ in Euclidean space (shown later).*

An orthogonal matrix has columns (or rows) that are an orthonormal basis.

Let $\mathbf{u}_i, i = 1, \cdots, n$ be a set of orthonormal vectors, *i.e.* $\mathbf{u}_i \cdot \mathbf{u}_i = 1$ and $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $i \neq j$.

$$Q = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ u_{21} & u_{22} & u_{23} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & u_{n3} & \dots & u_{nn} \end{bmatrix}$$

$$Q^\top Q = \begin{bmatrix} u_{11} & u_{21} & u_{31} & \dots & u_{n1} \\ u_{12} & u_{22} & u_{32} & \dots & u_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ u_{1n} & u_{2n} & u_{3n} & \dots & u_{nn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ u_{21} & u_{22} & u_{23} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & u_{n3} & \dots & u_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

The determinant of an orthogonal matrix is 1:

$$\det(Q) = |Q| = I$$

The **rank** of a matrix is the number of linearly independent rows or columns. The rank of a $n \times n$ invertible matrix is $n$ (also called full rank). This is equivalent to saying that the row vectors constituting the matrix are linearly independent.

The rank of a $n \times n$ matrix cannot be greater than $n$. If $rank(A) = n$, the matrix is full rank. For non square matrices, the rank of a $m \times n$ matrix cannot be greater than either $m$ or $n$.

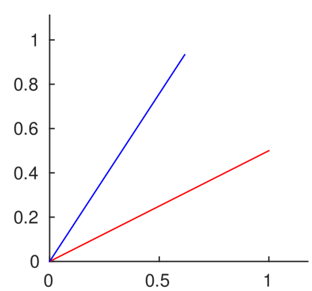
When an orthogonal matrix $Q$ is applied to a vector, it preserves its length (*i.e.* the value of its inner product)

$$\mathbf{y} = Q\mathbf{x}$$

$$\mathbf{y} \cdot \mathbf{y} = \mathbf{y}^\top \mathbf{y} = (Q\mathbf{x})^\top (Q\mathbf{x}) = \mathbf{x}^\top \underbrace{Q \top Q}_{I} \mathbf{x} = \mathbf{x}^\top \mathbf{x}$$

So what does an orthogonal matrix do? it rotates the vectors!

*Example: Rotations in the Plane*




$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

*MATLAB function that does the rotation in the plane.*

./scripts/rotation2d.m

```
theta = pi/6;
v = [1;.5];
Q = [cos(theta) -sin(theta); sin(theta) cos(theta)];
rv = Q*v;
 figure('Position', [100, 100, 400, 400]);
axes
hold on;
line([0;v(1)],[0,v(2)], 'color', 'r');
line([0;rv(1)],[0,rv(2)], 'color', 'b');
xlim([0,2]);
ylim([0,2]);
```

# 3 Eigenvalues and Eigenvectors

We are interested in solving the **Eigenvalue Equation**:

$$A\mathbf{x} = \lambda\mathbf{x},$$

where $\lambda$ is a scalar, $\mathbf{x}$ is a vector different than $\mathbf{0}$ and $A$ is a square matrix. In fact, $\mathbf{0}$ will always be a solution of the eigenvalue equation.

Solutions to the above equation have the special property that $A\mathbf{x}$ leads to the same vector $\mathbf{x}$, scaled by $\lambda$. In this case, $\lambda$ is called the **eigenvalue** and $\mathbf{x}$ is the **eigenvector**.

The eigenvalue equation can be written:

$$\begin{aligned} A\mathbf{x} - \lambda\mathbf{x} &= 0, \\ A\mathbf{x} - \lambda I\mathbf{x} &= 0, \quad && I \text{ is an identity matrix} \\ (A - \lambda I)\mathbf{x} &= 0, \end{aligned}$$

An equation such as the one above has non-trivial solutions (*i.e.* $\mathbf{x} \neq 0$ and $\lambda \neq 0$), only if:

$$det(A - \lambda I) = 0$$

This is called the **characteristic equation** and can be used to find eigenvalues. In general, this equation has $n$ eigenvalues and $n$ eigenvectors

Note that two or more eigenvalues can be the same.

*Example: Eigenvalues and Eigenvectors of $2 \times 2$ Matrices*

*The characteristic equation of matrix $A$:*

$$det(\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix}) = (a-\lambda)(d-\lambda) - cb = 0$$
$$\lambda^2 - (a+d)\lambda + (ad - cb) = 0$$

*has solutions:*

$$\lambda_{\pm} = \frac{1}{2}[(a+d) \pm \sqrt{4bc + (a-d)^2}],$$

*Plugging the above solutions, $\lambda_+$ or $\lambda_-$, back into the characteristic equation give the associated eigenvectors.*

$$A\mathbf{x} = \lambda_+ \mathbf{x}$$

*Thus the eigenvalue associated to $\lambda_+$ will verify the equations:*

$$ax_1 + bx_2 = \lambda_+ x_1$$
$$cx_1 + dx_2 = \lambda_+ x_2$$

*Note that if matrix $A$ only has only one distinct eigenvalue, then the two equations will be the same. In this $2 \times 2$ case, any eigenvector will be an eigenvector.*

If all the eigenvalues are distinct, then the eigenvectors are linearly independent and form a basis. If an $n \times n$ matrix, has $n$ distinct eigenvlues, all the eigenvectors are linearly independent and form a basis for $\mathbb{R}^n$.

This is a very interesting basis: For example vector $\mathbf{c}$ expressed in the basis of eigenvectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ of matrix $A$:

$$A\mathbf{c} = Ac_1\mathbf{x}_1 + Ac_2\mathbf{x}_2 + .. + Ac_N\mathbf{x}_N = \lambda_1 c_1 \mathbf{x}_1 + \lambda_2 c_2 \mathbf{x}_2 + ... + \lambda_N c_N \mathbf{x}_N$$

*i.e.* Each dimension has been stretched by a a scale factor $\lambda_i$

**Properties of eigenvalues**

- $\lambda$ is an eigenvalue $\leftrightarrow det(A - \lambda I) = 0$
- if $\lambda$ is an eigenvalue, $\mathbf{x}$ is an eigenvector $(A - \lambda I)\mathbf{x} = 0$
- Eigenvalues can be positive, negative or complex.
- The determinant of a matrix is the product of its eigenvalues: $det(A) = \lambda_1 \lambda_2 ... \lambda_n$
- A is singular $\leftrightarrow$ one or more eigenvalues are zero
- In MATLAB: [V,D] = eig(A), where V is the matrix of eigenvectors and D is a matrix whose diagonals are eigenvalues.

## 3.1 Similar Matrices

Two $n \times n$ matrices are **similar** if there exists a $n \times n$ matrix $P$ such that:

$$A = P^{-1}BP$$

Two similar matrices have the same eigenvalues.

## 3.2 Diagonalizable Matrices

A $n \times n$ matrix is **diagonalizable** if it is similar to a diagonal matrix, *i.e.* matrix $P$ in the equation

$$A = P^{-1}BP$$

exists and $B$ is diagonal.

A $n \times n$ matrix is diagonalizable if it has $n$ linearly independent eigenvectors (or all eigenvalues are distinct)

If all the eigenvalues of $A$ are distinct, and $P$ is the matrix consisting of respective eigenvectors, then $B$ is the diagonal matrix with eigenvalues are the diagonal.

The following steps demonstrate this:

$$P = \left[\mathbf{x}_1, ..., \mathbf{x}_n\right]$$

where $\mathbf{x}_i$ are eigenvectors.

$$B = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

$$\begin{aligned} AP &= A\left[\mathbf{x}_1, ..., \mathbf{x}_n\right] \\ &= \left[A\mathbf{x}_1, ..., A\mathbf{x}_n\right] \\ &= \left[\lambda_1\mathbf{x}_1, ..., \lambda_n\mathbf{x}_n\right] \\ &= PB \end{aligned}$$

Finally:

$$\begin{aligned} B &= P^{-1}AP \\ PBP^{-1} &= A \end{aligned}$$

*Example: Rotation Matrix*

*What are the eigenvalues of $Q$*

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

## 3.3 Multivariate Data

$n$ observation of $p$ variables can be represented as a $n \times p$ matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix}$$

Here, we will call $X$ the **data matrix**. $X$ can be written as $n$ measurements of $p$ variable/feature (*i.e.* $p$ column vectors):

$$X = \begin{bmatrix} \mathbf{x}_1 & ... & \mathbf{x}_p \end{bmatrix}$$

where $\mathbf{x}_i$ is the $n \times 1$ vector of $n$ observations of the $i^{th}$ feature.

The sample mean of the data matrix denoted $\bar{X}$ is a vector whose coordinates are given by averaging each column of $X$. The components of $\bar{X}$ is written $\bar{x}_i$:

$$\bar{x}_i = u^\top \mathbf{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ji}$$

Geometrically, the sample mean is a projection of the vectors $\mathbf{x}_i$ onto the vector $\mathbf{u} = \frac{1}{n}[1, .., 1]$.

The deviation is the vector with means subtracted from each column.

$$\mathbf{d}_i = \mathbf{x}_i - \bar{x}_i, \quad i = 1, ..., p$$

The **variance** is calculated as the norm of the deviations:

$$\Sigma_{ii} = \frac{1}{n-1} \mathbf{d}_i^\top \mathbf{d}_i = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2$$

Note that in the above equation, the $\mathbf{x}$ vector components have been explicitely written in terms of $x_{ji}$.

Rows are different observations, columns are different variables/features.

The **covariance** matrix is:

$$\Sigma_{ik} = \frac{1}{n-1} \mathbf{d_i}^\top \mathbf{d_k}$$

The **Correlation Coefficient** is the covariance of two variables divided by the product of their standard deviations:

$$\rho_{ik} = \frac{\Sigma_{ik}}{||\mathbf{d}_i|| ||\mathbf{d}_k||} = \frac{\frac{1}{n-1} \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)}{\sqrt{\Sigma_{ii} \Sigma_{kk}}}.$$

*Example: Multivariate Normal Data*

$$X = \begin{bmatrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_n \end{bmatrix}$$

*are $n$ multivariate rvs ($p$ dimensions each) distributed as*

$$X \sim \mathcal{N}(\boldsymbol{\mu},\, \boldsymbol{\Sigma})\,.$$

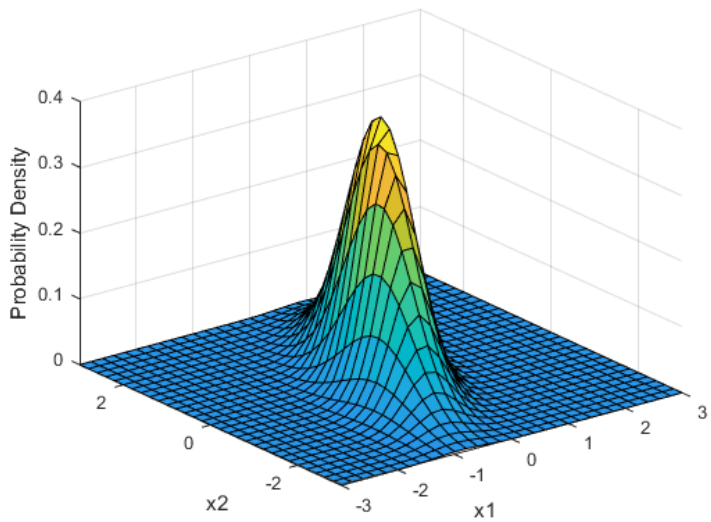
*where the $p \times p$ matrix $\boldsymbol{\Sigma}$ is the covariance matrix and $\boldsymbol{\mu}$ is the location.*

$$p(X) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(X - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(X - \boldsymbol{\mu})\right)$$

*MATLAB code in 2D:*

scripts/gaussian_density_multivariate.m

```
mu = [0 0];
Sigma = [.25 .3; .3 1];
x1 = -3:.2:3; x2 = -3:.2:3;
[X1,X2] = meshgrid(x1,x2);
F = mvnpdf([X1(:) X2(:)],mu,Sigma);
F = reshape(F,length(x2),length(x1));
surf(x1,x2,F);
caxis([min(F(:))-.5*range(F(:)),max(F(:))]);
axis([-3 3 -3 3 0 .4])
xlabel('x1'); ylabel('x2'); zlabel('Probability Density');
```



0.4
0.3
0.2
0.1
0
Probability Density
2
0
-2
x2
-3
-2
-1
0
1
2
3
x1

## Additional Reading


- Strang, Gilbert, et al. Introduction to linear algebra. Vol. 3. Wellesley, MA: Wellesley-Cambridge Press, 1993.
- **For definitions:** `http://mathworld.wolfram.com/`