In general terms, our aim in this first part of the course is to use vector space theory to study the geometry of Euclidean space. A good knowledge of the subject matter of the Matrix Applications course is assumed.

The dot product

Let $\mathbb{R}^n$ denote the vector space of n-tuples of real numbers:

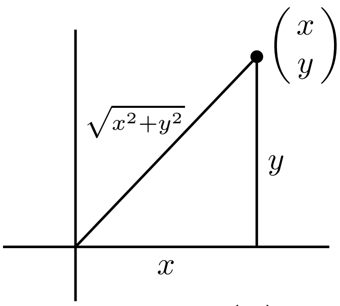
$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \;\middle|\; x_1,\, x_2,\, \ldots,\, x_n \in \mathbb{R} \right\}.$$

Note that we write elements of $\mathbb{R}^n$ as column vectors. We shall also on occasion deal with row vectors, and we denote the space of n-component row vectors by $(\mathbb{R}^n)'$:

$$(\mathbb{R}^n)' = \{\, (x_1 \; x_2 \; \ldots \; x_n) \mid x_1,\, x_2,\, \ldots,\, x_n \in \mathbb{R} \,\}.$$

Everyone should be familiar with the idea of specifying a point in the Euclidean plane by a pair of numbers, giving the coordinates of the point relative to a fixed Cartesian coordinate system. Thus points in the plane correspond in a natural way to vectors in $\mathbb{R}^2$. Similarly, points in three dimensional Euclidean space correspond to vectors in $\mathbb{R}^3$. Although $\mathbb{R}^n$ does not correspond to physical space when $n > 3$, it is very similar mathematically to $\mathbb{R}^2$ and $\mathbb{R}^3$, and we can use geometrical intuition based on our familiarity with physical space as a guide to help us understand $\mathbb{R}^n$. In particular, we sometimes refer to vectors in $\mathbb{R}^n$ as "points", and we shall deal with geometrically-motivated concepts such as the distance between two such points.

By Pythagoras' Theorem, the distance from the origin to the point $\begin{pmatrix} x \\ y \end{pmatrix}$ is $\sqrt{x^2 + y^2}$.



Similarly, in $\mathbb{R}^3$ the distance from the origin to $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is $\sqrt{x^2 + y^2 + z^2}$. So if

$$\underset{\sim}{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

then it is natural to say that the distance from the origin to $\underset{\sim}{v}$ is $\sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$.

Definition. Let $\underset{\sim}{v} \in \mathbb{R}^n$ as above. The *length* of the vector $\underset{\sim}{v}$ is the scalar $\|\underset{\sim}{v}\|$ given by the formula $\|\underset{\sim}{v}\| = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$.

Obviously, $\|\underset{\sim}{v}\| \geq 0$, and if $\|\underset{\sim}{v}\| = 0$ then $\underset{\sim}{v} = \underset{\sim}{0}$. (Recall that the zero vector, $\underset{\sim}{0}$, is the vector all of whose components are zero.)

Now suppose that

$$\underset{\sim}{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \underset{\sim}{w} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

are two vectors in $\mathbb{R}^n$.

Definition. The *dot product* of $\underset{\sim}{v}$ and $\underset{\sim}{w}$ is the scalar quantity $\underset{\sim}{v} \cdot \underset{\sim}{w}$ given by

$$\underset{\sim}{v} \cdot \underset{\sim}{w} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n.$$

For example,

$$\begin{pmatrix} -2 \\ 3 \\ -4 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix} = ((-2) \times 1) + (3 \times 1) + ((-4) \times 2) + (1 \times 2) = -5.$$

Note that the dot product of a vector with itself gives the square of the length of the vector: $\underset{\sim}{v} \cdot \underset{\sim}{v} = \|\underset{\sim}{v}\|^2$.

Properties of the dot product. Suppose that $\underset{\sim}{u}$, $\underset{\sim}{v}$ and $\underset{\sim}{w}$ are vectors in $\mathbb{R}^n$, and that $k \in \mathbb{R}$ (a scalar). Then the following hold true.

1) $(\underset{\sim}{u} + \underset{\sim}{v}) \cdot \underset{\sim}{w} = \underset{\sim}{u} \cdot \underset{\sim}{w} + \underset{\sim}{v} \cdot \underset{\sim}{w}$.
2) $\underset{\sim}{u} \cdot (\underset{\sim}{v} + \underset{\sim}{w}) = \underset{\sim}{u} \cdot \underset{\sim}{v} + \underset{\sim}{u} \cdot \underset{\sim}{w}$.
3) $\underset{\sim}{u} \cdot \underset{\sim}{v} = \underset{\sim}{v} \cdot \underset{\sim}{u}$.
4) $k(\underset{\sim}{u} \cdot \underset{\sim}{v}) = (k\underset{\sim}{u}) \cdot \underset{\sim}{v} = \underset{\sim}{u} \cdot (k\underset{\sim}{v})$.
5) $\underset{\sim}{u} \cdot \underset{\sim}{u} \geq 0$, and if $\underset{\sim}{u} \cdot \underset{\sim}{u} = 0$, then $\underset{\sim}{u} = \underset{\sim}{0}$.

In the computer tutorials we shall be using the computer algebra program known as MAGMA. Here are some magma commands that you can use to get MAGMA to calculate the dot product of two vectors.

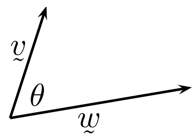
```
> R := RealField();
> V := VectorSpace(R,4);
> a := V![1,2,3,4];
> b := V![1,-1,1,-1];
> R;
Real Field
> V;
Full Vector space of degree 4 over Real Field
> a, b;
(1 2 3 4)
( 1 -1 1 -1)
> InnerProduct(a,b);
-2
```

Although we prefer to use column vectors, MAGMA unfortunately assumes that vectors are row vectors. This will not be a serious problem for us, although occasionally we have to modify formulas a little when using the computer.

If $\utilde{v}$, $\utilde{w} \in \mathbb{R}^n$ and θ is the angle between $\utilde{v}$ and $\utilde{w}$, then

$$\utilde{v} \cdot \utilde{w} = \|\utilde{v}\|\,\|\utilde{w}\| \cos\theta \tag{1}$$

(a result that should be familiar for $\mathbb{R}^2$ and $\mathbb{R}^3$). The diagram below illustrates what is meant by the angle "between" $\utilde{v}$ and $\utilde{w}$. Note that $0 \le \theta \le \pi$.



If $\utilde{v} = \begin{pmatrix} -2 \\ 3 \\ -4 \\ 1 \end{pmatrix}$ and $\utilde{w} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix}$ then $\utilde{v} \cdot \utilde{w} = -5$ (see above), and we find that

$$\cos\theta = \frac{\utilde{v} \cdot \utilde{w}}{\|\utilde{v}\|\,\|\utilde{w}\|} = \frac{-5}{\sqrt{2^2+3^2+4^2+1^2}\sqrt{1^2+1^2+2^2+2^2}} = \frac{-5}{10\sqrt{3}} \approx -.2887,$$

and so $\theta \approx 1.864$ (radians). (This is roughly $106^\circ\,47'$.)

In $\mathbb{R}^2$ and $\mathbb{R}^3$ the formula in Eq. (1) above can be proved by an application of the cosine rule in the triangle whose vertices are the origin and the points $\utilde{v}$ and $\utilde{w}$. For larger values of n, Eq. (1) should be regarded as the definition of the angle between $\utilde{v}$ and $\utilde{w}$. But since $|\cos\theta|$ is at most 1, to define $\cos\theta = (\utilde{v} \cdot \utilde{w})/(\|\utilde{v}\|\,\|\utilde{w}\|)$ is only legitimate if $\big|(\utilde{v} \cdot \utilde{w})/(\|\utilde{v}\|\,\|\utilde{w}\|)\big| \le 1$. So we need to prove this.

Theorem [The Cauchy-Schwarz Inequality]. *If $\utilde{v}$, $\utilde{w} \in \mathbb{R}^n$ then $|\utilde{v} \cdot \utilde{w}| \le \|\utilde{v}\|\,\|\utilde{w}\|$.*

Proof. Let $\utilde{v}$, $\utilde{w} \in \mathbb{R}^n$ be arbitrary. If $\utilde{v} = \utilde{0}$ then $|\utilde{v} \cdot \utilde{w}| = 0 = \|\utilde{v}\|\,\|\utilde{w}\|$; so the inequality is satisfied. Hence we may assume that $\utilde{v} \ne \utilde{0}$.

Let x be any scalar. Then

$$\begin{aligned} 0 &\le (x\utilde{v} + \utilde{w}) \cdot (x\utilde{v} + \utilde{w}) \\ &= x^2(\utilde{v} \cdot \utilde{v}) + 2x(\utilde{v} \cdot \utilde{w}) + (\utilde{w} \cdot \utilde{w}) \\ &= Ax^2 + 2Bx + C \end{aligned} \tag{2}$$

where $A = \utilde{v} \cdot \utilde{v}$, $B = \utilde{v} \cdot \utilde{w}$ and $C = \utilde{w} \cdot \utilde{w}$. Note that $A = \|\utilde{v}\|^2 > 0$ (since $\utilde{v} \ne \utilde{0}$). The inequality in (2) above holds for all x, and in particular it holds when $x = -B/A$. (We choose this value for x since it is the value that minimizes the quadratic expression $Ax^2 + 2Bx + Cx$, as can easily be shown by use of elementary calculus.) So we deduce that

$$0 \le A\Big(\frac{-B}{A}\Big)^2 + 2B\Big(\frac{-B}{A}\Big) + C = \Big(\frac{B^2}{A}\Big) - 2\Big(\frac{B^2}{A}\Big) + C = C - \frac{B^2}{A},$$

and hence that $B^2/A \le C$. Multiplying through by the positive number A, we deduce that $B^2 \le AC$. That is,

$$(\utilde{v} \cdot \utilde{w})^2 \le (\utilde{v} \cdot \utilde{v})(\utilde{w} \cdot \utilde{w}) = \|\utilde{v}\|^2\|\utilde{w}\|^2.$$

Taking square roots gives $|\utilde{v} \cdot \utilde{w}| \le \|\utilde{v}\|\,\|\utilde{w}\|$, as required. □

Transposes of matrices

If A is an $n \times m$ matrix then its *transpose* A^T is the $m \times n$ matrix obtained by changing rows into columns. For example,

$$\begin{pmatrix} 1 & 1 & 3 & 7 \\ -1 & -2 & -3 & -4 \end{pmatrix}^T = \begin{pmatrix} 1 & -1 \\ 1 & -2 \\ 3 & -3 \\ 7 & -4 \end{pmatrix}.$$

More formally, the entry in row i and column j in A becomes the entry in row j and column i in A^T. Thus in the above example, the $(2,3)$-entry of A is -3, and this is also the $(3,2)$-entry of A^T.

Properties of the transpose operation. Let A and B be matrices.

1) If the product AB is defined (which is the case if the number of columns of A equals the number of rows of B) then $B^T A^T$ is defined, and $(AB)^T = B^T A^T$.
2) If A and B have the same shape then $(A+B)^T = A^T + B^T$.
3) If k is any scalar then $(kA)^T = k(A^T)$.
4) In all cases, $(A^T)^T = A$.
5) If A is square then $\det A^T = \det A$.
6) If A has an inverse (which of course can only happen when A is square) then (A^T) has an inverse, and $(A^T)^{-1} = (A^{-1})^T$.

The first of these is the one that causes students most trouble. So make sure you remember it: transposing reverses multiplication.

Definition A matrix A is said to be *symmetric* if $A^T = A$; it is said to be *skew-symmetric* if $A^T = -A$.

Notice that symmetric and skew-symmetric matrices are necessarily square (since this must hold whenever the matrix and its transpose have the same shape).

Here is an example of a symmetric matrix and an example of a skew-symmetric matrix.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 8 & 7 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 3 & -2 \\ -3 & 0 & -4 \\ 2 & 4 & 0 \end{pmatrix}$$

The entries on the main diagonal in a skew-symmetric matrix are always zero. Can you see why?

Vectors can be thought of as matrices, and in particular the transpose of a column vector is a row vector, and vice versa. Thus if $\underset{\sim}{v} \in \mathbb{R}^n$ then we can think of $\underset{\sim}{v}$ as an $n \times 1$ matrix, and $\underset{\sim}{v}^T$ is therefore a $1 \times n$ matrix; in other words, $\underset{\sim}{v}^T \in (\mathbb{R}^n)'$.

Observe that a $1 \times n$ matrix and an $n \times 1$ matrix can always be multiplied, giving a 1×1 matrix—that is, a scalar—as the answer. So if $\underset{\sim}{v}$, $\underset{\sim}{w} \in \mathbb{R}^n$ then $\underset{\sim}{v}^T \underset{\sim}{w}$ is defined. It is easily seen that in fact $\underset{\sim}{v}^T \underset{\sim}{w} = \underset{\sim}{v} \cdot \underset{\sim}{w}$.

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = (x_1 \ x_2 \ \ldots \ x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Let us see how to use MAGMA for some matrix calculations.

```
> Q := RationalField();
> M := KMatrixSpace(Q,2,4);

> A := M![1,1,1,1,2,4,6,8];
> B := Transpose(A);

> Q;
Rational Field
> M;
Full KMatrixSpace of 2 by 4 matrices over Rational Field

> A;
[1 1 1 1]
[2 4 6 8]
> B;
[1 2]
[1 4]
[1 6]
[1 8]

> M := KMatrixSpace(Q,1,4);
> a := M![1,2,3,4];
> b := M![1,-1,1,-1];
> a * Transpose(b);
> [-2]
```

The orthogonal projection

Definition. If $\underset{\sim}{v}$, $\underset{\sim}{w} \in \mathbb{R}^n$ then $d(\underset{\sim}{v}, \underset{\sim}{w})$, the *distance* from the point $\underset{\sim}{v}$ to the point $\underset{\sim}{w}$, is defined by $d(\underset{\sim}{v}, \underset{\sim}{w}) = \|\underset{\sim}{v} - \underset{\sim}{w}\|$.

Suppose that W is a subspace of $\mathbb{R}^n$. We shall investigate the following problem: given a vector $\underset{\sim}{v} \in \mathbb{R}^n$ that is not in W, find the element of W that is the closest possible approximation to $\underset{\sim}{v}$ by an element of W. In other words, find the element $\underset{\sim}{p} \in W$ such that the distance from $\underset{\sim}{v}$ to $\underset{\sim}{p}$ is as small as possible. Before we can solve this, however, we need some preliminary results.

The *triangle inequality* says that the sum of the lengths of two sides of a triangle always exceeds the length of the third side. (This is intuitively reasonable, since the shortest distance between two points is given by a straight line.) So it should be true that $d(\underset{\sim}{x}, \underset{\sim}{y}) + d(\underset{\sim}{y}, \underset{\sim}{z}) \geq d(\underset{\sim}{x}, \underset{\sim}{z})$.

Recall from 1st year that addition of vectors can be performed by the so-called Triangle Law: given $\underset{\sim}{v}$ and $\underset{\sim}{w}$, let A be any point, choose B so that $\overrightarrow{AB} = \underset{\sim}{v}$, and choose C so that $\overrightarrow{BC} = \underset{\sim}{w}$. Then $\overrightarrow{AC} = \underset{\sim}{v} + \underset{\sim}{w}$. So, in vector terminology, the statement that the length of AB plus the length of BC is greater than or equal to the length of AC becomes $\|\underset{\sim}{v}\| + \|\underset{\sim}{w}\| \geq \|\underset{\sim}{v} + \underset{\sim}{w}\|$. It is quite straightforward to use the properties of the dot product to prove this form of the triangle inequality.

Proposition. *For all* $\underset{\sim}{v}$, $\underset{\sim}{w} \in \mathbb{R}^n$, *we have* $\|\underset{\sim}{v} + \underset{\sim}{w}\| \leq \|\underset{\sim}{v}\| + \|\underset{\sim}{w}\|$.

Proof. Let $\underset{\sim}{v}$, $\underset{\sim}{w} \in \mathbb{R}^n$. Then

$$\begin{aligned}
\|\underset{\sim}{v} + \underset{\sim}{w}\|^2 &= (\underset{\sim}{v} + \underset{\sim}{w}) \cdot (\underset{\sim}{v} + \underset{\sim}{w}) \\
&= \underset{\sim}{v} \cdot \underset{\sim}{v} + 2\underset{\sim}{v} \cdot \underset{\sim}{w} + \underset{\sim}{w} \cdot \underset{\sim}{w}, \\
&= \|\underset{\sim}{v}\|^2 + 2\|\underset{\sim}{v}\|\,\|\underset{\sim}{w}\| \cos\theta + \|\underset{\sim}{w}\|^2 && . \\
&\le \|\underset{\sim}{v}\|^2 + 2\|\underset{\sim}{v}\|\,\|\underset{\sim}{w}\| + \|\underset{\sim}{w}\|^2 && \text{(since } \cos\theta \le 1) \\
&= (\|\underset{\sim}{v}\| + \|\underset{\sim}{w}\|)^2
\end{aligned}$$

Taking square roots gives the required result. □

If $\underset{\sim}{x}$, $\underset{\sim}{y}$, $\underset{\sim}{z} \in \mathbb{R}^n$ are arbitrary, and if we define $\underset{\sim}{v} = \underset{\sim}{x} - \underset{\sim}{y}$ and $\underset{\sim}{w} = \underset{\sim}{y} - \underset{\sim}{z}$, then we see that $\underset{\sim}{v} + \underset{\sim}{w} = \underset{\sim}{x} - \underset{\sim}{z}$, and so the inequality $\|\underset{\sim}{v} + \underset{\sim}{w}\| \le \|\underset{\sim}{v}\| + \|\underset{\sim}{w}\|$ becomes $\|\underset{\sim}{x} - \underset{\sim}{z}\| \le \|\underset{\sim}{x} - \underset{\sim}{y}\| + \|\underset{\sim}{y} - \underset{\sim}{z}\|$. This can be rephrased as $d(\underset{\sim}{x}, \underset{\sim}{z}) \le d(\underset{\sim}{x}, \underset{\sim}{y}) + d(\underset{\sim}{y}, \underset{\sim}{z})$, which is the other version of the triangle inequality given above.

Here are some more MAGMA calculations.

```
> Length := func< u | Sqrt(InnerProduct(u,u)) >;
> Angle := func< u,v | Arccos( InnerProduct(u,v) /
> (Length(u) * Length(v)));
> Distance := func< u,v | Length(u-v) >;

> R := RealField();
> V := VectorSpace(R,4);

> a := V![1,2,3,4];
> b := V![1,-1,1,-1];

> Length(a);
5.477225575051661134569697828006
> Length(b);
2.000000000000000000000000000000
> Angle(a,b);
1.754400337073815190484058148 45
> Angle(b,a);
1.754400337073815190484058148 45
> Distance(a,b);
6.164414002968976450250192381425
> Distance(b,a);
6.164414002968976450250192381425
```

Let $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$, ..., $\underset{\sim}{a}_k$ be a basis for the subspace W of $\mathbb{R}^n$, and let A be the matrix that has $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$, ..., $\underset{\sim}{a}_k$ as its columns:

$$A = (\underset{\sim}{a}_1 \ \underset{\sim}{a}_2 \ \cdots \ \underset{\sim}{a}_k).$$

This is an $n \times k$ matrix whose columns are linearly independent. Note that $k \le n$, since $k = \dim W$ and $W \subseteq \mathbb{R}^n$.

Recall that the set of all linear combinations of the columns of a matrix is called the *column space* of the matrix; thus W is the column space of A. If $\underset{\sim}{v} \in W$ then there exist scalars $x_1, x_2, \ldots, x_k \in \mathbb{R}$ such that

$$\underset{\sim}{v} = x_1\underset{\sim}{a}_1 + x_1\underset{\sim}{a}_1 + \ldots + x_k\underset{\sim}{a}_k = (\underset{\sim}{a}_1 \ \underset{\sim}{a}_2 \ \cdots \ \underset{\sim}{a}_k)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} = A\underset{\sim}{x},$$

where $\underset{\sim}{x} = (x_1, x_2, \ldots, x_k)^T \in \mathbb{R}^k$.

It is a consequence of the fact that the columns of A are linearly independent that the $k \times k$ matrix A^TA is invertible. To prove this, we make use of the following result from the first year linear algebra course:
a system of k linear equations in k unknowns has a unique solution if and only if the coefficient matrix is invertible.
Thus, to prove that A^TA is invertible it is sufficient to show that $\underset{\sim}{x} = \underset{\sim}{0}$ is the only solution of $A^TA\underset{\sim}{x} = \underset{\sim}{0}$.

So, suppose that $A^TA\underset{\sim}{x} = \underset{\sim}{0}$. Then

$$(\underset{\sim}{x}^TA^T)(A\underset{\sim}{x}) = \underset{\sim}{x}^T(A^TA\underset{\sim}{x}) = \underset{\sim}{x}^T\underset{\sim}{0} = 0,$$

and since $\underset{\sim}{x}^TA^T = (A\underset{\sim}{x})^T$, this gives $(A\underset{\sim}{x})^T(A\underset{\sim}{x}) = 0$. But $(A\underset{\sim}{x})^T(A\underset{\sim}{x}) = (A\underset{\sim}{x}) \cdot (A\underset{\sim}{x})$, and by the properties of the dot product we know that this can only be zero when $A\underset{\sim}{x} = \underset{\sim}{0}$. So

$$x_1\underset{\sim}{a}_1 + x_2\underset{\sim}{a}_2 + \cdots + x_k\underset{\sim}{a}_k = A\underset{\sim}{x} = \underset{\sim}{0},$$

and since $\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_k$ are linearly independent it follows that $x_1 = x_2 = \cdots = x_k = 0$, as required.

Definition. We say that the vectors $\underset{\sim}{v}, \underset{\sim}{w} \in \mathbb{R}^n$ are *orthogonal* if $\underset{\sim}{v} \cdot \underset{\sim}{w} = 0$.

Assuming that $\underset{\sim}{v}$ and $\underset{\sim}{w}$ are both nonzero, they are orthogonal if and only if the angle between them is $\pi/2$.

Pythagoras' Theorem. *If $\underset{\sim}{v}$ and $\underset{\sim}{w}$ are orthogonal then $\|\underset{\sim}{v} + \underset{\sim}{w}\|^2 = \|\underset{\sim}{v}\|^2 + \|\underset{\sim}{w}\|^2$.*

This follows by a short calculation with dot products:

$$\begin{aligned} \|\underset{\sim}{v} + \underset{\sim}{w}\|^2 &= (\underset{\sim}{v} + \underset{\sim}{w}) \cdot (\underset{\sim}{v} + \underset{\sim}{w}) \\ &= \underset{\sim}{v} \cdot \underset{\sim}{v} + 2\underset{\sim}{v} \cdot \underset{\sim}{w} + \underset{\sim}{w} \cdot \underset{\sim}{w} \\ &= \|\underset{\sim}{v}\|^2 + \|\underset{\sim}{w}\|^2 \qquad (\text{since } \underset{\sim}{v} \cdot \underset{\sim}{w} = 0). \end{aligned}$$

We are now ready to tackle the problem we mentioned above: finding the element of the subspace W that is the closest to a given $\underset{\sim}{v} \in \mathbb{R}^n$. Our next result is the key to this.

Theorem. *Suppose that $\underset{\sim}{p} \in W$ is such that $\underset{\sim}{v} - \underset{\sim}{p}$ is orthogonal to all elements of W. Then $d(\underset{\sim}{v}, \underset{\sim}{p}) \le d(\underset{\sim}{v}, \underset{\sim}{w})$ for all $\underset{\sim}{w} \in W$, with equality holding if and only if $\underset{\sim}{w} = \underset{\sim}{p}$.*

Proof. Let $\underset{\sim}{w} \in W$. Then $\underset{\sim}{p} - \underset{\sim}{w} \in W$ also, since $\underset{\sim}{p} \in W$ and W is closed under addition and scalar multiplication. So $\underset{\sim}{v} - \underset{\sim}{p}$ is orthogonal to $\underset{\sim}{p} - \underset{\sim}{w}$. Thus

$$\begin{aligned} \|\underset{\sim}{v} - \underset{\sim}{w}\|^2 &= \|(\underset{\sim}{v} - \underset{\sim}{p}) + (\underset{\sim}{p} - \underset{\sim}{w})\|^2 \\ &= \|(\underset{\sim}{v} - \underset{\sim}{p})\|^2 + \|(\underset{\sim}{p} - \underset{\sim}{w})\|^2 \qquad (\text{by Pythagoras' Theorem}) \\ &\ge \|(\underset{\sim}{p} - \underset{\sim}{w})\|^2. \end{aligned}$$

Moreover, if $\underset{\sim}{w} \neq \underset{\sim}{p}$ then $\|\underset{\sim}{p} - \underset{\sim}{w}\| > 0$, and so the inequality is strict, as required. $\square$

The above proof shows that if there exists an element $\underset{\sim}{p} \in W$ satisfying the condition that $\underset{\sim}{v} - \underset{\sim}{p}$ is orthogonal to all elements of W, then it is unique. However, we still have the problem of showing that such an element does actually exist. We do this now by determining a formula for $\underset{\sim}{p}$.

Since $\underset{\sim}{p}$ is to be in W it must have the form $A\underset{\sim}{x}$ for some $\underset{\sim}{x} \in \mathbb{R}^k$, and so the condition to be satisfied is that $(\underset{\sim}{v} - A\underset{\sim}{x})$ and $A\underset{\sim}{y}$ should be orthogonal for all $\underset{\sim}{y} \in \mathbb{R}^k$. In other words, we require that $(A\underset{\sim}{y}) \cdot (\underset{\sim}{v} - A\underset{\sim}{x}) = 0$, or, equivalently, $(A\underset{\sim}{y})^T(\underset{\sim}{v} - A\underset{\sim}{x}) = 0$, for all $\underset{\sim}{y} \in \mathbb{R}^k$. Now since

$$(A\underset{\sim}{y})^T(\underset{\sim}{v} - A\underset{\sim}{x}) = y^T A^T(\underset{\sim}{v} - A\underset{\sim}{x}),$$

the desired condition will certainly be satisfied if $A^T(\underset{\sim}{v} - A\underset{\sim}{x}) = \underset{\sim}{0}$. So we want the vector $\underset{\sim}{x}$ to satisfy $A^T\underset{\sim}{v} = A^T A\underset{\sim}{x}$. But we have shown above that the matrix $A^T A$ is invertible; so $\underset{\sim}{x} = (A^T A)^{-1}(A^T\underset{\sim}{v})$ is a solution of this this system of equations. So the vector $\underset{\sim}{p}$ is given by $\underset{\sim}{p} = A\underset{\sim}{x} = A(A^T A)^{-1}(A^T\underset{\sim}{v})$.

The following theorem summarizes what we have proved.

Theorem. *Let W be a subspace of $\mathbb{R}^n$, and let $\underset{\sim}{v}$ be an arbitrary element of $\mathbb{R}^n$. If A is a matrix whose columns form a basis for W then the vector $\underset{\sim}{p}$ given by*

$$\underset{\sim}{p} = A(A^T A)^{-1}A^T\underset{\sim}{v}$$

is the unique element of W such that $d(\underset{\sim}{v}, \underset{\sim}{p}) \leq d(\underset{\sim}{v}, \underset{\sim}{w})$ for all $\underset{\sim}{w} \in W$.

The vector $\underset{\sim}{p}$ constructed above is called the *orthogonal projection* of $\underset{\sim}{v}$ onto the subspace W. Note that it is uniquely determined by $\underset{\sim}{v}$ and W; so if we replace A by any other matrix whose columns form a basis of W then we will get the same answer.

Here is a simple example with $n = 4$ and $k = 2$. Let W be the subspace of $\mathbb{R}^4$ spanned by $(1, 0, 0, 1)^T$ and $(1, 1, 0, 2)^T$. These vectors are obviously linearly independent, and hence form a basis for W. We shall calculate the projection onto W of the vector $\underset{\sim}{v} = (0, 1, 0, 0)^T$.

As a first step we must find $A^T A$ and its inverse, where A is a matrix whose columns form a basis of W. Choosing the given basis, we obtain

$$A^T A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 3 & 6 \end{pmatrix},$$

and this gives

$$(A^T A)^{-1} = \frac{1}{3}\begin{pmatrix} 6 & -3 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2/3 \end{pmatrix}.$$

So $\underset{\sim}{p}$ is given by

$$\begin{aligned}\underset{\sim}{p} &= \begin{pmatrix}1 & 1\\ 0 & 1\\ 0 & 0\\ 1 & 2\end{pmatrix}\begin{pmatrix}2 & -1\\ -1 & 2/3\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & 1\\ 1 & 1 & 0 & 2\end{pmatrix}\begin{pmatrix}0\\ 1\\ 0\\ 0\end{pmatrix}\\ &= \begin{pmatrix}1 & 1\\ 0 & 1\\ 0 & 0\\ 1 & 2\end{pmatrix}\begin{pmatrix}2 & -1\\ -1 & 2/3\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix}\\ &= \begin{pmatrix}1 & 1\\ 0 & 1\\ 0 & 0\\ 1 & 2\end{pmatrix}\begin{pmatrix}-1\\ 2/3\end{pmatrix} = \begin{pmatrix}-1/3\\ 2/3\\ 0\\ 1/3\end{pmatrix}.\end{aligned}$$

We leave it to the reader to check that choosing another basis of W, such as $\underset{\sim}{b}_1 = \underset{\sim}{a}_1 + \underset{\sim}{a}_2$ and $\underset{\sim}{b}_2 = \underset{\sim}{a}_1 - \underset{\sim}{a}_2$, gives the same answer for $\underset{\sim}{p}$.

Some more MAGMA

Using MAGMA we need to transpose the whole formula for $\underset{\sim}{p}$, remembering that transposing reverses multiplication. To be specific, $\underset{\sim}{p}^T = \underset{\sim}{x}^T A^T$, where $\underset{\sim}{x}^T = (\underset{\sim}{v}^T A)(A^T A)^{-1}$. Here is how it can be done in MAGMA.

```
> Projection := func< W,v | x*B
>   where x is Solution(B*A,v*A)
>   where A is Transpose(B)
>   where B is BasisMatrix(W)>;

> R := RealField();
> V := VectorSpace(R,4);

> a := V![1,2,3,4];
> b := V![1,-1,1,-1];
> c := V![6,7,8,9];

> W := sub< V | a,b >;
> Projection(W,a);
(1 2 3 4)
> Projection(W,c);
(104/29 133/29 262/29 291/29)
> Projection(sub< V | a,b >,a);
(1 2 3 4)
> Projection(sub< V | a,b >,c);
(104/29 133/29 262/29 291/29)
```

(Note that the command
    `W := sub< V | a,b >;`
defines `W` to be the subspace of `V` spanned by `a` and `b`.)

Projection: summary

To find the projection of the vector $\underset{\sim}{v}$ onto a subspace W of $\mathbb{R}^n$, proceed as follows:

1) Find a basis of W, and write down the matrix A that has these vectors as its columns. This will be an $n \times k$ matrix, where, k is the number of vectors in the basis of W. That is, $k = \dim W$.

2) Compute $A^T A$. This will be a $k \times k$ matrix.

3) Compute $A^T \underset{\sim}{v}$. This will be a k-component column.

4) Solve the equations $A^T A\underset{\sim}{x} = A^T\underset{\sim}{v}$. Note that the coefficient matrix (on the left-hand side) is the $k \times k$ matrix you found in Step 2 and the right-hand side is the k-component column you found in Step 3. So you have to solve a system of k equations in k unknowns.

5) The answer is $\underset{\sim}{p} = A\underset{\sim}{x}$, where $\underset{\sim}{x}$ is the solution you found in Step 4. Note that $\underset{\sim}{p}$, the projection of $\underset{\sim}{v}$ onto W, is an n-component column.

Example

Let $\underset{\sim}{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, and W the subspace of $\mathbb{R}^4$ with basis $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}$.

1) $$A = \begin{pmatrix} 1 & 2 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

2) $$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 6 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

3) $$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix}.$$

4)
$$\left(\begin{array}{ccc|c} 4 & 0 & 0 & 2 \\ 0 & 6 & -1 & -2 \\ 0 & -1 & 2 & -1 \end{array}\right) \xrightarrow[R_2:=(1/6)R_2]{R_1:=(1/4)R_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1/2 \\ 0 & 1 & -1/6 & -1/3 \\ 0 & -1 & 2 & -1 \end{array}\right)$$
$$\xrightarrow{R_3:=R_3+R_2} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1/2 \\ 0 & 1 & -1/6 & -1/3 \\ 0 & 0 & 5/6 & -4/3 \end{array}\right).$$

Now by back substitution we find that

$$x_3 = -\tfrac{4}{3}/\tfrac{5}{6} = -\tfrac{8}{5}, \qquad x_2 = -\tfrac{1}{3} + \tfrac{1}{6}x_3 = -\tfrac{1}{3} - \tfrac{4}{15} = -\tfrac{3}{5}, \qquad x_1 = \tfrac{1}{2}.$$

where x_1, x_2 and x_3 are the components of $\underset{\sim}{x}$.

5) $$\begin{pmatrix} 1 & 2 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1/2 \\ -3/5 \\ -8/5 \end{pmatrix} = \begin{pmatrix} -7/10 \\ -1/2 \\ 11/10 \\ 21/10 \end{pmatrix}$$ is the projection.