I : Topics from linear algebra

> Linear algebra [is] an indispensable tool for any modern treatment of geometry.
>
> Ryan, p. 6

Many introductions to vector algebra show how one can use vectors to prove results in elementary geometry. In fact, vector algebra plays an important role in most if not all current scientific work involving geometry, from the abstract study of the subject for its own sake to theoretical and applied work on computer graphics and the uses of geometry in the sciences and engineering. Vectors are often a valuable tool when working problems using coordinates, for they provide various means to simplify all sorts of computations and formulas. The following heavily edited version of a remark by J. Dieudonné (1906 – 1992) summarizes the substance of the modern perspective fairly well:

> There are **…** people who do linear geometry **…** by taking coordinates, and they call this analytical geometry**.**
>
> [***Note:*** The quotation is available in its entirety at the online site http://www-history.mcs.st-andrews.ac.uk/Quotations/Dieudonne.html and parts of it can be disputed on grounds of historical accuracy as well as its controversial across – the – board disapproval of working with numerical coordinates.]

Because of the fundamental role that linear algebra plays in the modern study of geometry, we shall begin by covering a few topics from elementary linear algebra that are particularly useful in geometric work.

The official prerequisites for this course include linear algebra at least through the theory of determinants, and most students have probably also seen dot products before, either in a linear algebra course or a calculus course. However, since dot products may not have been treated in prerequisite courses, we shall cover them briefly and include some points that are often omitted in calculus and linear algebra courses. Many basic facts from linear algebra are summarized in Appendix **D** of Ryan. Another review of some topics from linear algebra is given in Sections **I.A** and **I.B** of the following online document:

http://math.ucr.edu/~res/math132/linalgnotes.pdf

A few additional facts from linear algebra are discussed in Section **I.0** of the exercises:

http://math.ucr.edu/~res/math133/exercises.pdf

Needless to say, we shall use basic facts from the linear algebra prerequisites as necessary.

Numbering conventions. In mathematics it is often necessary to use results that were previously established. Throughout these notes we shall refer to results from earlier sections by notation like Proposition **I.5.9**, which will denote Proposition 9 from Section **I.5** (this particular example does not actually exist, but it should illustrate the key points adequately).

Ends of proofs. In classical writings mathematicians used the initials **Q. E. D.** (for the Latin phrase, *that which was to be demonstrated*) or **Q. E. F.** (for the Latin phrase, *that which was to be constructed*) to indicate the end of a proof or construction. Some writers still use this notation, but more often the end of a proof or line of reasoning is now indicated by a large black square, which is sometimes known as a "tombstone" or "Halmos (big) dot." We shall also use the symbol "■" to mark the end of an argument or to indicate that nothing more will be said about the proof of a statement.

I.1 : Dot products

Much if not all of this material has probably been seen in previous courses, but since it is not necessarily covered in one of the prerequisite alternatives we shall present it here for the sake of completeness. Some parts are adapted and expanded from the following online sources:

http://en.wikipedia.org/wiki/Dot_product

http://en.wikipedia.org/wiki/Cauchy-Schwarz_inequality

http://en.wikipedia.org/wiki/Triangle_inequality

http://en.wikipedia.org/wiki/Gram-Schmidt_process

Definition. The **dot product**, also known as the **scalar product** or **inner product**, is a binary operation which assigns a real – valued scalar quantity to an ordered pair of vectors in $\mathbf{R}^n$. Specifically, the dot product of two vectors

$$\mathbf{a} = (a_1, a_2, \ldots, a_n) \text{ and } \mathbf{b} = (b_1, b_2, \ldots, b_n) \text{ in } \mathbf{R}^n$$

is given by

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

where as usual Σ denotes summation notation. Frequently a dot or inner product is also written $\langle \mathbf{a}, \mathbf{b} \rangle$, particularly in more advanced textbooks. As an ***example*** to illustrate the definition, the dot product of the three-dimensional vectors $(1, 3, -2)$ and $(4, -2, -1)$ is equal to $(1, 3, -2) \cdot (4, -2, -1) = 1(4) + 3(-2) + (-2)(-1) = 4 - 6 + 2 = 0$.

Note that if $\mathbf{a} = \mathbf{b}$, then $\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{a}$ is the square of its ***length***; namely, $\|\mathbf{a}\|^2$.

Properties of dot products

The following two properties hold if **a, b**, and **c** are vectors and r is a scalar.

The dot product is **commutative:** $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$

The dot product is **bilinear:** $\mathbf{a}\cdot(k\mathbf{b} + \mathbf{c}) = k(\mathbf{a}\cdot\mathbf{b}) + (\mathbf{a}\cdot\mathbf{c})$

The dot product is **distributive:** $\mathbf{a}\cdot(\mathbf{b} + \mathbf{c}) = (\mathbf{a}\cdot\mathbf{b}) + (\mathbf{a}\cdot\mathbf{c})$

When multiplied by scalars, it is **homogeneous:** $(c_1\mathbf{a})\cdot(c_2\mathbf{b}) = (c_1 c_2)(\mathbf{a}\cdot\mathbf{b})$

<u>Note.</u> The last two follow from commutativity and the previous bilinearity identity**.**

Last but certainly not least, the dot product is **positive definite:**

$$\mathrm{a}\cdot\mathrm{a}\ \textit{ is nonnegative, and it is equal to zero if and only if }\ \mathrm{a} = 0.$$

This follows because the dot product of **a** with itself is just the sum of the terms a_i^2 and since the latter are all nonnegative it follows that there sum is nonnegative. If **a** is the zero vector then clearly this sum is equal to zero, but if **a** is not the zero vector then at least one of these terms is positive and therefore the entire sum must be positive.

The positive definiteness property allows us to introduce a notion of ***distance between two vectors*** for $\mathbf{R}^n$**,** or more generally for any vector space which has a suitable notion of inner product satisfying the properties given above. Specifically, if **a** and **b** are vectors in such a vector space, then the *<u>distance</u>* between the vectors **a** and **b** is defined by

$$d(\mathbf{a}, \mathbf{b}) = ||\mathbf{b} - \mathbf{a}||\,.$$

This definition satisfies the standard abstract properties of distance, the first two or which are given below:

$$d(\mathrm{a}, \mathrm{b})\ \textit{ is nonnegative, and it equals zero if and only if }\ \mathrm{a} = \mathrm{b}.$$

$$d(\mathrm{a}, \mathrm{b}) = d(\mathrm{b}, \mathrm{a})\ \textit{for all}\ \mathrm{a}\ \textit{ and }\ \mathrm{b}.$$

The third basic property of an abstract distance, known as the ***Triangle Inequality,*** will be verified in the next part of this section.

The Cauchy – Schwarz Inequality

This basic fact is also known as the ***Cauchy Inequality***, the ***Schwartz Inequality***, or the ***Cauchy – Bunyakovski – Schwarz*** Inequality (frequently "Schwarz" is misspelled "Schwartz" in books and papers). It states that if x and y are elements of a real (or complex) vector space with a suitable notion of inner product then

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle.$$

Furthermore, the two sides are equal if and only if x and y are linearly dependent (hence either at least one of them is zero or else each is a nonzero scalar multiple of the other). Another form of this inequality, involving the lengths of vectors, is given by taking the square roots of both sides of the previous inequality:

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

For ordinary 2 – and 3 – dimensional vectors, there is a simple geometric interpretation of the second form if both x and y are nonzero; in this case the quotient of the left hand side divided by the right hand side corresponds to the cosine of the angle x **0** y, and we expect this number to have an absolute value less than or equal to 1, with equality if and only if x and y are nonzero scalar multiples of each other. More generally, the Cauchy – Schwarz inequality allows one to define a notion of "the angle between the two vectors" for an arbitrary inner product, where the extendibility of such concepts from Euclidean geometry may not be intuitively clear. In particular, since the angle x **0** y is a right angle if and only if the cosine is zero, we have the following:

Definition. Two vectors **a** and **b** are ***perpendicular*** or ***orthogonal*** if and only if they satisfy the equation $\langle \mathbf{a}, \mathbf{b} \rangle = 0$.

Proof of the Cauchy – Schwarz Inequality. Since the inequality is trivially true when $y = \mathbf{0}$, we may as well assume $\langle y, y \rangle$ is nonzero. Let λ be a scalar. Then we have

$$0 \leq \|x - \lambda y\|^2 = \langle x - \lambda y, x - \lambda y \rangle$$
$$= \langle x, x \rangle - \lambda \langle x, y \rangle - \lambda \langle y, x \rangle + |\lambda|^2 \langle y, y \rangle.$$

Choosing

$$\lambda = \langle x, y \rangle \cdot \langle y, y \rangle^{-1}$$

we obtain

$$0 \leq \langle x, x \rangle - |\langle x, y \rangle|^2 \cdot \langle y, y \rangle^{-1}$$

which is true if and only if

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \cdot \langle y, y \rangle$$

or equivalently:

$$|\langle x, y \rangle| \leq \|x\| \, \|y\| .$$

This completes the proof.■

One immediate consequence of the Cauchy – Schwarz Inequality is the ***Triangle Inequality*** for vector lengths:

$$\| \boldsymbol{x} + \boldsymbol{y} \| \leq \| \boldsymbol{x} \| + \| \boldsymbol{y} \| \quad \textit{for all vectors} \quad \boldsymbol{x}, \boldsymbol{y}$$

The latter immediately yields a corresponding Triangle Inequality for distances:

$$d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b}) \quad \textit{for all vectors} \quad \mathbf{a}, \mathbf{b}, \mathbf{c}$$

Derivation of the second inequality from the first. This follows by making the substitutions $\boldsymbol{x} = \mathbf{c} - \mathbf{a}$ and $\boldsymbol{y} = \mathbf{b} - \mathbf{c}$ in the first inequality.■

Derivation of the first inequality. Given vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we have

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y \rangle \\ &= \|x\|^2 + \langle x, y \rangle + \langle y, x \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2|\langle x, y \rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2\end{aligned}$$

Taking the square roots yields the Triangle Inequality for vector lengths.■

The following consequences of the triangle inequalities are often useful; they give lower bounds instead of upper bounds:

Proposition 1. *For all vectors* **a, b, c,** $\boldsymbol{x}, \boldsymbol{y}$ *we have the following:*

$$\big|\, \| \boldsymbol{x} \| - \| \boldsymbol{y} \| \,\big| \leq \| \boldsymbol{x} - \boldsymbol{y} \|$$

$$\big|\, d(\mathbf{b}, \mathbf{c}) - d(\mathbf{c}, \mathbf{a}) \,\big| \leq d(\mathbf{a}, \mathbf{b})$$

Proof. As above, the main point is to verify the first inequality and to derive the second from it. Two applications of the Triangle Inequality for vector lengths show that

$$\| \boldsymbol{x} \| \leq \| \boldsymbol{x} - \boldsymbol{y} \| + \| \boldsymbol{y} \|, \quad \| \boldsymbol{y} \| \leq \| \boldsymbol{x} \| + \| \boldsymbol{y} - \boldsymbol{x} \| = \| \boldsymbol{x} \| + \| \boldsymbol{x} - \boldsymbol{y} \|$$

and these may be rewritten as follows:

$$\| \boldsymbol{x} \| - \| \boldsymbol{y} \| \leq \| \boldsymbol{x} - \boldsymbol{y} \|, \quad \| \boldsymbol{y} \| - \| \boldsymbol{x} \| \leq \| \boldsymbol{x} - \boldsymbol{y} \|$$

These are equivalent to the first inequality in the proposition, and the second follows by making the substitutions $\boldsymbol{y} = \mathbf{c} - \mathbf{a}$ and $\boldsymbol{x} = \mathbf{b} - \mathbf{c}$ in the first one. Of course, by the symmetry properties of distance and the expression $\big|\, \| \boldsymbol{x} \| - \| \boldsymbol{y} \| \,\big|$ this inequality can also be rewritten numerous other ways, including $\big|\, d(\mathbf{a}, \mathbf{c}) - d(\mathbf{b}, \mathbf{c}) \,\big| \leq d(\mathbf{a}, \mathbf{b})$.■

Equality in the Triangle Inequalities

The preceding material is covered in many linear algebra courses, but we shall now go one step beyond such courses. As already noted, one has an equality

$$|\langle x,\ y\rangle| \;=\; ||x||\cdot||y||$$

corresponding to the Cauchy – Schwarz Inequality if and only if the vectors x and y are linearly dependent. For our purposes it will be important to know the analogous conditions under which one has the equations

$$||x+y|| \;=\; ||x||+||y|| \qquad d(\mathbf{a},\mathbf{b}) \;=\; d(\mathbf{a},\mathbf{c}) + d(\mathbf{c},\mathbf{b})$$

associated to the Triangle Inequalities.

Proposition 2. *Two nonzero vectors x and y satisfy $||x+y|| = ||x||+||y||$ if and only if each is a nonnegative multiple of the other.*

Proposition 3. *Three vectors **a**, **b** and **c** satisfy $d(\mathbf{a},\mathbf{b}) = d(\mathbf{a},\mathbf{c}) + d(\mathbf{c},\mathbf{b})$ if and only if $\mathbf{c}-\mathbf{a} = s(\mathbf{b}-\mathbf{a})$ where $0 \le s \le 1$.*

Proof of Proposition 2. If one of the vectors is nonzero, then it is clear that equality holds and that one of the vectors is a nonnegative multiple of the other, and this is why we assume that both x and y are nonzero.

If we look back at the derivation of the Triangle Inequality, we see that the crucial step in deriving an inequality comes from applying the Cauchy – Schwarz Inequality to x and y. In particular, it follows that $||x+y|| = ||x||+||y||$ will hold if and only if we have

$$\langle x,\ y\rangle \;=\; ||x||\,||y||$$

(note that this condition is stronger than equality in the Cauchy – Schwarz Inequality, for we the latter involves the absolute value of the inner product and not the inner product itself). It will suffice to show that this stronger equation holds if and only if the nonzero vectors x and y are positive multiples of each other. If $y = tx$ where t is positive, then we have

$$\langle x,\ y\rangle \;=\; \langle x,\ tx\rangle \;=\; t\,||x||^{2} \;=\; ||x||\,||tx|| \;=\; ||x||\,||y||$$

which shows the "if" direction. Conversely, if equality holds then the conclusion of the Cauchy – Schwarz Inequality shows that x and y are nonzero multiples of each other, so let $y = tx$ where t is nonzero; we need to show that t is positive. But now we have

$$t\,||x||^{2} \;=\; \langle x, tx\rangle \;=\; \langle x,\ y\rangle \;=\; ||x||\,||y|| \;=\; ||x||\,||tx|| \;=\; |t|\,||x||^{2}$$

which implies that $t = |t|$ and hence that t is positive.■

Proof of Proposition 3. Suppose first that $\mathbf{a} = \mathbf{b}$. Then $0 = d(\mathbf{a},\mathbf{b}) = d(\mathbf{a},\mathbf{c}) + d(\mathbf{c},\mathbf{b})$ if and only if both summands on the right hand side are equal to zero, which is equivalent to $\mathbf{a} = \mathbf{b} = \mathbf{c} = \mathbf{0}$, so that the conclusion is true for trivial reasons.

Suppose now that **a** and **b** are unequal. By Proposition A and the definition of distance, we have an equation $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b})$ if and only if either (1) one of $\mathbf{a} = \mathbf{c}$ or $\mathbf{b} = \mathbf{c}$ is true, (2) the vectors $\mathbf{c} - \mathbf{a}$ and $\mathbf{b} - \mathbf{c}$ are positive multiples of each other. In the first cases we have $\mathbf{c} - \mathbf{a} = 0 \cdot (\mathbf{b} - \mathbf{a})$ and $\mathbf{c} - \mathbf{a} = 1 \cdot (\mathbf{b} - \mathbf{a})$, so the conclusion is also true in these cases. Thus we are left with the case where $\mathbf{b} - \mathbf{c} = t(\mathbf{c} - \mathbf{a})$ for some positive scalar t. We then have $\mathbf{b} - \mathbf{a} = (\mathbf{c} - \mathbf{a}) + (\mathbf{b} - \mathbf{c}) = (1+t)(\mathbf{c} - \mathbf{a})$, so that $\mathbf{c} - \mathbf{a} = (1+t)^{-1}(\mathbf{b} - \mathbf{a})$ for some positive scalar t. To conclude the argument, note that the latter is equivalent to $\mathbf{c} - \mathbf{a} = s(\mathbf{b} - \mathbf{a})$ for some scalar s satisfying the conditions $0 < s < 1$.■

Orthogonality and the Gram – Schmidt process

One geometric way of thinking about n – dimensionality is that it implies the existence of n distinct and mutually perpendicular directions, ***no more and no less.*** At least this is clear for $n = 2$ or 3, and it leads more generally to the following question: Suppose that **W** is an n – dimensional subspace of $\mathbf{R}^m$. Is it true that **W** has n mutually perpendicular nonzero vectors but does not have $n + 1$ such vectors?

There are actually two parts to this question. We shall first show that an n – dimensional subspace cannot contain $n + 1$ mutually perpendicular nonzero vectors.

Proposition 4. *Suppose that* $\{ \mathbf{v}_1, \mathbf{v}_2, \dots , \mathbf{v}_n \}$ *is a set of nonzero mutually perpendicular vectors in* $\mathbf{R}^m$. *Then* $\{ \mathbf{v}_1, \mathbf{v}_2, \dots , \mathbf{v}_n \}$ *is linearly independent.*

Corollary 5. *If* **W** is a k *– dimensional subspace of* $\mathbf{R}^m$ *and* $\{ \mathbf{v}_1, \mathbf{v}_2, \dots , \mathbf{v}_n \}$ *is a set of nonzero mutually perpendicular vectors in* **W**, *then* $n \leq k$.

The corollary follows because a linearly independent subset of **W** contains at most $k = \dim \mathbf{W}$ vectors.■

Proof of Proposition 4. Suppose we have an equation of the form

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}.$$

We need to prove that all the coefficients a_i must be equal to zero. To do this, take the dot products of both sides of the equation above with some vector $\mathbf{v}_i$. The right hand side yields a value of zero, and thus we have

$$0 = (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n) \cdot \mathbf{v}_i =$$
$$a_1 (\mathbf{v}_1 \cdot \mathbf{v}_i) + a_2 (\mathbf{v}_2 \cdot \mathbf{v}_i) + \dots + a_n (\mathbf{v}_n \cdot \mathbf{v}_i).$$

Now the terms $(\mathbf{v}_j \cdot \mathbf{v}_i)$ are zero unless $j = i$, in which case the term is positive. Thus the right hand side of the displayed equation is equal to $a_i (\mathbf{v}_i \cdot \mathbf{v}_i)$, which we now know must be equal to zero. Since the second factor is positive this means that a_i must be zero. Finally, since we chose i arbitrarily it follows that all the coefficients a_i must be equal to zero, and hence the given set of nonzero mutually orthogonal vectors must be linearly independent.■

We must now prove the other half. In fact, given an arbitrary ordered set of linearly independent vectors in $\mathbf{R}^n$ there is a recursive method called the **Gram – Schmidt (orthogonalization) process,** which takes a finite, linearly independent set of vectors $\mathbf{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and yields an orthogonal set $\mathbf{T} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ which spans the same vector subspace as $\mathbf{S}$. We can also modify the set of vectors $\mathbf{T}$ to obtain a set of vectors that is also ***orthonormal;*** *i.e.*, the length of each vector is equal to $\mathbf{1}$.

In fact, one can say more: For each $k \leq n$ the first k vectors of $\mathbf{T}$ span the same subspace as the first k vectors of $\mathbf{S}$, and if the first m vectors of $\mathbf{S}$ are already orthonormal, then they agree with the first m vectors of $\mathbf{T}$.

Description of the Gram–Schmidt process. Let $\mathbf{u}$ be a nonzero vector. We define the *perpendicular projection of* $\mathbf{v}$ *onto the subspace spanned by* $\mathbf{u}$ by the formula

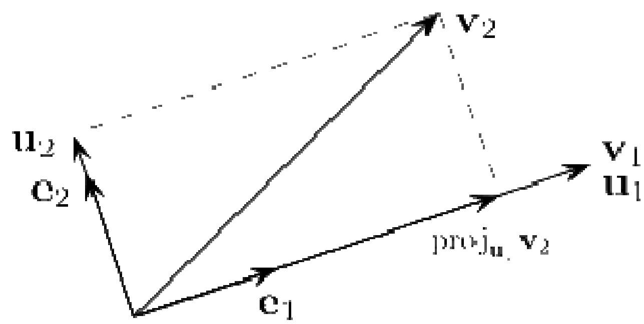
$$\mathrm{proj}_{\mathbf{u}}\,\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u}.$$

This projects the vector $\mathbf{v}$ orthogonally onto the line joining $\mathbf{0}$ and $\mathbf{u}$. In other words, we have a "*resolution of* $\mathbf{v}$ *into perpendicular components*" of the form $\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_1$ where the first summand is the multiple of $\mathbf{u}$ on the right hand side of the displayed equation and the second vector is perpendicular to $\mathbf{u}$ (see the picture below).

The Gram–Schmidt process then goes recursively as follows:

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{||\mathbf{u}_1||} \\
\mathbf{u}_2 &= \mathbf{v}_2 - \mathrm{proj}_{\mathbf{u}_1}\,\mathbf{v}_2, & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{||\mathbf{u}_2||} \\
\mathbf{u}_3 &= \mathbf{v}_3 - \mathrm{proj}_{\mathbf{u}_1}\,\mathbf{v}_3 - \mathrm{proj}_{\mathbf{u}_2}\,\mathbf{v}_3, & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{||\mathbf{u}_3||} \\
&\ \ \vdots & &\ \ \vdots \\
\mathbf{u}_k &= \mathbf{v}_k - \sum_{j=1}^{k-1} \mathrm{proj}_{\mathbf{u}_j}\,\mathbf{v}_k, & \mathbf{e}_k &= \frac{\mathbf{u}_k}{||\mathbf{u}_k||}
\end{aligned}$$

Here is a picture illustrating the first two steps of the Gram–Schmidt process.



We claim that the sequence $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is the required system of orthogonal vectors, and the normalized vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$ form an ***orthonormal*** system.

To check that the formulas above will yield an orthogonal sequence, we first compute $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ by substituting into the above formula for $\mathbf{u}_2$: this computation shows that the given inner product is equal zero. Then we can use this to compute $\langle \mathbf{u}_1, \mathbf{u}_3 \rangle$ again by substituting into the formula for $\mathbf{u}_3$; once again, the value turns out to be zero. The general proof proceeds by (finite) mathematical induction: If we have the orthogonality property for the first k vectors in the set, we can use the formulas for the next vector to verify the property for the first $k + 1$ vectors in the set, and we can continue in this fashion until we have proven it for the entire set.

Geometrically, this method proceeds as follows: to compute $\mathbf{u}_i$, it projects $\mathbf{v}_i$ orthogonally onto the subspace $\mathbf{U}$ generated by $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}$, which is the same as the subspace generated by $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. The vector $\mathbf{u}_i$ is then defined to be the difference between $\mathbf{v}_i$ and this projection, and hence it will be orthogonal to all of the vectors in $\mathbf{U}$.■

Example. Consider the following set of vectors in $\mathbf{R}^2$ (with the usual inner product):

$$S = \left\{ \mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}.$$

Now use the Gram–Schmidt process to obtain an orthogonal set of vectors::

$$\mathbf{u}_1 = \mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$$\mathbf{u}_2 = \mathbf{v}_2 - \mathrm{proj}_{\mathbf{u}_1} \mathbf{v}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \mathrm{proj}_{\left(\begin{smallmatrix} 3 \\ 1 \end{smallmatrix}\right)} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix}.$$

We check that the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are indeed orthogonal:

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \left\langle \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix} \right\rangle = -\frac{6}{5} + \frac{6}{5} = 0.$$

We can then normalize the vectors if we divide by their lengths as shown above:

$$\mathbf{e}_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$$\mathbf{e}_2 = \frac{1}{\sqrt{\frac{40}{25}}} \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

I.2 : Cross products

The ***cross product***, also known as the ***vector product*** or ***outer product,*** is a binary operation on vectors in $\mathbf{R}^3$. It differs from the dot product (or inner product) in that its value is a vector rather than a scalar.

Once again, much of this material may be review, but we shall start from the beginning for the sake of completeness. Some parts of this treatment below are adapted from the following online source:

http://en.wikipedia.org/wiki/Cross_product

Definition. Let **a** and **b** be two vectors in $\mathbf{R}^3$ and write these vectors using coordinates as **a** = (a_1, a_2, a_3) and **b** = (b_1, b_2, b_3). Then the *cross product* **a** × **b** is defined as follows:

$$\mathbf{a} \times \mathbf{b} = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)$$

Probably the best way to remember this definition is to reformulate it using determinants. If we denote the standard unit vectors in $\mathbf{R}^3$ by $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, and $\mathbf{k} = (0, 0,1)$, then the cross product can be written formally as the determinant of the following matrix:

$$\mathbf{a} \times \mathbf{b} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}.$$

Examples. **1.** If $\mathbf{a} = (1, 2, 3)$ and $\mathbf{b} = (4, 5, 6)$ then the cross product $\mathbf{a} \times \mathbf{b}$ is equal to

$$\mathbf{a} \times \mathbf{b} = (1, 2, 3) \times (4, 5, 6) =$$
$$(2\cdot6 - 3\cdot5,\ 3\cdot4 - 1\cdot6,\ 1\cdot4 - 2\cdot3) = (1, 6, -2).$$

2. It follows immediately from the definitions that $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ for all vectors **a (**in such cases the last two rows of the determinant are equal**).** Direct computation yields the following values for the other cross products of the standard unit vectors:

$$\mathbf{i} \times \mathbf{j} = -\,\mathbf{j} \times \mathbf{i} = \mathbf{k} \qquad \mathbf{j} \times \mathbf{k} = -\,\mathbf{k} \times \mathbf{j} = \mathbf{i} \qquad \mathbf{k} \times \mathbf{i} = -\,\mathbf{i} \times \mathbf{k} = \mathbf{j}.$$

Properties of the cross product

We shall begin by listing some basic algebraic properties of the cross product which follow immediately from the definitions.

The cross product is ***anticommutative*****:** $\mathbf{a} \times \mathbf{b} = -\,\mathbf{b} \times \mathbf{a}$

The cross product is ***distributive*****:** $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$

It is ***compatible with scalar multiplication*****:** $(c\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (c\mathbf{b}) = c\,(\mathbf{a} \times \mathbf{b})$**.**

There are also several important identities involving both the dot product and the cross product. Given three vectors **a**, **b**, **c** in $\mathbf{R}^3$**,** we shall define (**a**, **b**, **c**) to be the 3×3 matrix whose rows are **a**, **b**, **c** in that order.

<u>Proposition 1.</u> *We have the following relationship***:**

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = det\,(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

This scalar is often called the (scalar) ***triple product*** or the ***box product*** **[a, b, c].**

The second equation follows from the definition of the cross product and the formula for expanding a 3×3 determinant by minors along the first row.■

The properties of determinants now yield the following simple chain of identities**:**

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = [\mathbf{a}, \mathbf{b}, \mathbf{c}] = [\mathbf{c}, \mathbf{a}, \mathbf{b}] = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$$

The preceding observations lead directly to the following consequence**:**

<u>Proposition 2.</u> *If* **a** *and* **b** *are arbitrary vectors in* $\mathbf{R}^3$**,** *then* $\mathbf{a} \times \mathbf{b}$ *is perpendicular to both* **a** *and* **b.**

Proof. By the preceding identities we have $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{a} = [\mathbf{a}, \mathbf{b}, \mathbf{a}] = \mathbf{0}$, where the latter holds because the determinant of a matrix vanishes if two rows are identical. In a similar fashion we also have $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{b} = [\mathbf{a}, \mathbf{b}, \mathbf{b}] = \mathbf{0}$.■

Similarly, we have the following basic result.

Theorem 3. *If* **a** *and* **b** *are arbitrary vectors in* $\mathbf{R}^3$, *then* $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ *if and only if* **a** *and* **b** *are linearly dependent. Furthermore, if* **a** *and* **b** *are linearly independent, then the vectors* **a, b** *and* $\mathbf{a} \times \mathbf{b}$ *form a basis for* $\mathbf{R}^3$.

In the course of proving this we shall also establish several other important facts regarding the cross product.

Lemma 4. *If* **a** *and* **b** *are arbitrary vectors in* $\mathbf{R}^3$, *then we have the following identity*:

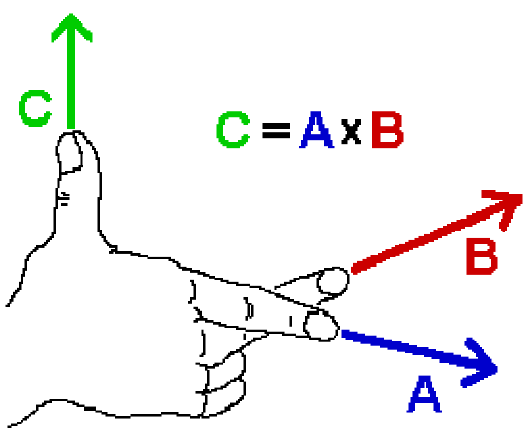
$$|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2$$

The lemma leads directly to the standard geometric interpretation of the cross product as

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

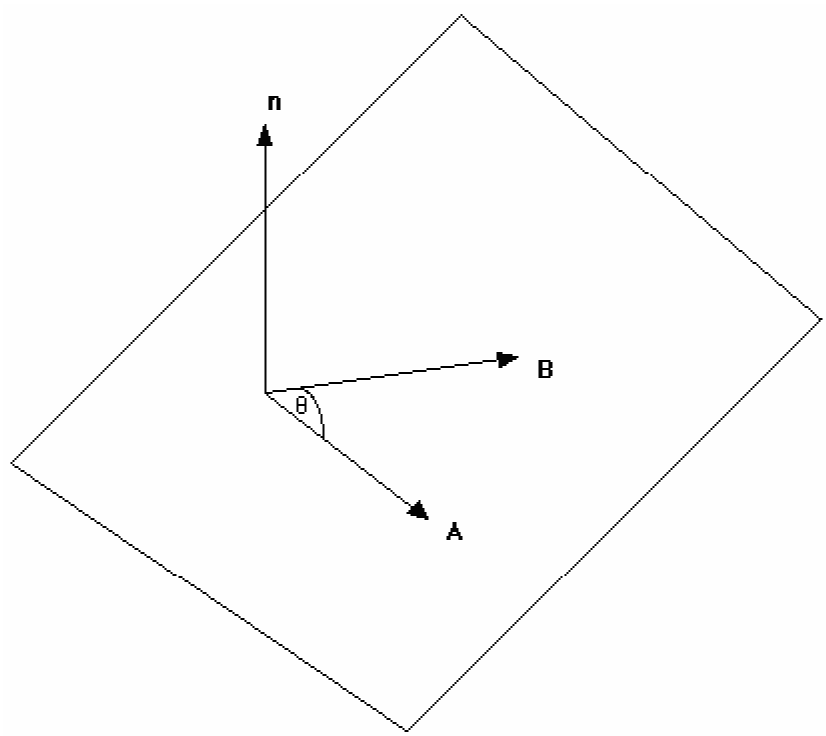
where θ is the measure of the angle between **a** and **b** $(0° \leq \theta \leq 180°)$ on the plane defined by the span of these vectors and the zero vector, and **n** is a unit vector which is perpendicular to both **a** and **b.** Generally there are two choices for **n,** each of which is the negative of the other, and to make the description complete we need to use the ***right hand rule*** to determine the perpendicular (or normal) vector **n.**

One easy way to find this direction is as follows: If one simply points the forefinger of the right hand in the direction of the first factor and the middle finger in the direction of the second, then the thumb points in the direction of **n.**

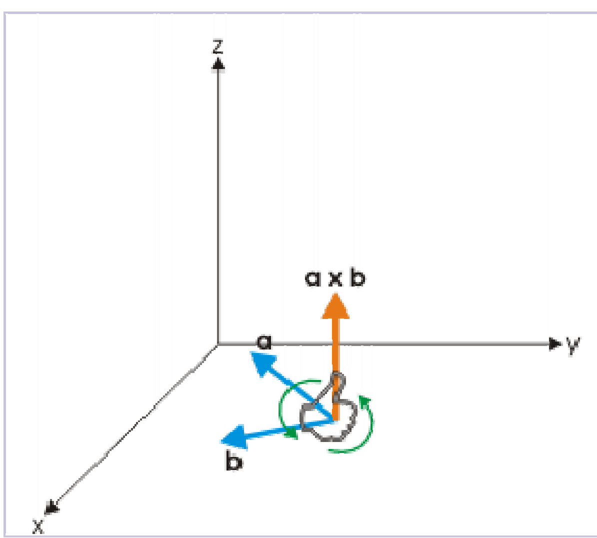


(***Source:*** http://www.physics.udel.edu/~watson/phys345/class/1-right-hand-rule.html)

Here is a second approach: Take the picture at the top of the next page. Point your thumb in the direction of **A,** and point your fingers in the direction of **B**. Your palm will face in the direction of **n**, out of the screen.



Yet another approach is to position your right hand so that your (non – thumb) fingers curl around the perpendicular axis in the direction going from **A** to **B;** if you stick your thumb out, it will point in the direction of **n.** An illustration of this method appears below.



(***Source:*** http://cnx.org/content/m13603/latest/)

***Proof of the Lemma* 4.** We begin by writing out $|\mathbf{a} \times \mathbf{b}|^2$ explicitly**:**

$$|\mathbf{a} \times \mathbf{b}|^2 = (a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - a_1 b_3)^2 + (a_1 b_2 - a_2 b_1)^2$$

Direct computation then shows that the latter is equal to

$$(a_1^{\,2} + a_2^{\,2} + a_3^{\,2})(b_1^{\,2} + b_2^{\,2} + b_3^{\,2}) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2.$$

***Proof of Theorem* 3.** First of all, suppose that **a** and **b** are linearly dependent. Then there is some vector **c** such that **a** and **b** are both scalar multiples of **c,** say $\mathbf{a} = s\,\mathbf{c}$ and $\mathbf{b} = t\,\mathbf{c}$. By anticommutativity we have $\mathbf{c} \times \mathbf{c} = -\mathbf{c} \times \mathbf{c}$, so that $2(\mathbf{c} \times \mathbf{c}) = \mathbf{0}$, which means that $\mathbf{c} \times \mathbf{c} = \mathbf{0}$. Thus we also have

$$\mathbf{a} \times \mathbf{b} = s\mathbf{c} \times t\mathbf{c} = st\,(\mathbf{c} \times \mathbf{c}) = st\,\mathbf{0} = \mathbf{0}$$

and hence the cross product vanishes if the vectors are linearly dependent.

Assume now that **a** and **b** are linearly independent. By the preceding lemma and the Cauchy – Schwarz Inequality, the right hand side is zero if and only if **a** and **b** are linearly dependent, and hence if **a** and **b** are linearly independent, then $\mathbf{a} \times \mathbf{b}$ must be nonzero.

Suppose now that we have an equation of the form $x\,\mathbf{a} + y\,\mathbf{b} + z\,(\mathbf{a} \times \mathbf{b}) = \mathbf{0}$ for suitable scalars x, y, z**.** Taking dot products with $\mathbf{a} \times \mathbf{b}$ yields $z\,|\mathbf{a} \times \mathbf{b}|^2 = 0$, which by the previous paragraph implies that $z = 0$**.** One can now use the linear independence of **a** and **b** to conclude that x and y must also be zero. Therefore the three vectors **a**, **b** and $\mathbf{a} \times \mathbf{b}$ are linearly independent, and consequently they must form a basis for $\mathbf{R}^3$.■

Although the dot and cross products do not satisfy analogs of the standard *cancellation property* for real numbers (if $a \neq 0$ and $a\,b = a\,c$, then $b = c$), there is a ***mixed cancellation property*:**

Proposition 5. *Suppose that* **a, b, c** *are vectors in* $\mathbf{R}^3$ *such that* $\mathbf{a} \neq \mathbf{0}$ *and we have both* $\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{c}$ *and* $\mathbf{a} \times \mathbf{b} = \mathbf{a} \times \mathbf{c}$. *Then* $\mathbf{b} = \mathbf{c}$.

Derivation. The hypotheses imply $\mathbf{a} \cdot (\mathbf{b} - \mathbf{c}) = 0$ and $\mathbf{a} \times (\mathbf{b} - \mathbf{c}) = \mathbf{0}$. Since **a** is nonzero, the second of these and a previous result imply that $\mathbf{b} - \mathbf{c}$ is a scalar multiple of **a,** so write $\mathbf{b} - \mathbf{c} = s\,\mathbf{a}$, where s is a real number. If we substitute this into the first equation we obtain $0 = \mathbf{a} \cdot (\mathbf{b} - \mathbf{c}) = \mathbf{a} \cdot s\mathbf{a} = s(\mathbf{a} \cdot \mathbf{a})$.

Since **a** is nonzero it follows that the second factor on the right hand side is positive, and thus we must have $s = 0$, which means that $\mathbf{b} - \mathbf{c} = \mathbf{0}$ and hence $\mathbf{b} = \mathbf{c}$.■

Cross products of three vectors

The cross product does not satisfy an ***associative law*** for multiplication; in other words, given three vectors **a, b, c** it is possible that $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$. For example, suppose that **a, b** and **c** are the unit vectors **i, i** and **j** respectively. Then we have the following:

$$(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} \;=\; \mathbf{0} \times \mathbf{j} \;=\; \mathbf{0}$$

$$\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) \;=\; \mathbf{i} \times \mathbf{k} \;=\; -\mathbf{j}$$

Fortunately, there is a simple, useful formula for the cross product of three vectors in $\mathbf{R}^3$**:**

Theorem 6. (The "**BAC – CAB**" rule) *If* **a, b, c** *are vectors in* $\mathbf{R}^3$, *then*

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \;=\; \mathbf{b}\,(\mathbf{a} \cdot \mathbf{c}) \;-\; \mathbf{c}\,(\mathbf{a} \cdot \mathbf{b})$$

or in a more standard format the left hand side is equal to

$$(\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}.$$

Derivation. Suppose first that **b** and **c** are linearly dependent. Then their cross product is zero, and one is a scalar multiple of the other. If $\mathbf{b} = x\,\mathbf{c}$, then it is an elementary exercise to verify that the right hand side of the desired identity is zero, and we already know the same is true of the left hand side. If on the other hand $\mathbf{c} = y\,\mathbf{b}$, then once again one finds that both sides of the desired identity are zero.

Now suppose that **b** and **c** are linearly independent, so that $\mathbf{b} \times \mathbf{c} \neq \mathbf{0}$. Note that a vector is perpendicular to $\mathbf{b} \times \mathbf{c}$ if and only if it is a linear combination of **b** and **c.** The forward implication follows from the perpendicularity of **b** and **c** to their cross product and the distributivity of the dot product, while the reverse implication follows because every vector is a linear combination $x\,\mathbf{b} + y\,\mathbf{c} + z\,(\mathbf{b} \times \mathbf{c})$ and this linear combination is perpendicular to the cross product if and only if $z = 0$; *i.e.*, if and only if the given vector is a linear combination of **b** and **c.**

Before proceeding to the general case, we first consider the special cases $\mathbf{b} \times (\mathbf{b} \times \mathbf{c})$ and $\mathbf{c} \times (\mathbf{b} \times \mathbf{c})$. Since $\mathbf{b} \times (\mathbf{b} \times \mathbf{c})$ is perpendicular to $\mathbf{b} \times \mathbf{c}$ we may write it in the form

$$\mathbf{b} \times (\mathbf{b} \times \mathbf{c}) \;=\; u\,\mathbf{b} + v\,\mathbf{c}$$

for suitable scalars u and v. If we take dot products with **b** and **c** we obtain the following equations:

$$0 \;=\; [\mathbf{b}, \mathbf{b}, \mathbf{b} \times \mathbf{c}] \;=\; \big(\mathbf{b} \cdot (\mathbf{b} \times (\mathbf{b} \times \mathbf{c}))\big) \;=\; \mathbf{b} \cdot (u\,\mathbf{b} + v\,\mathbf{c}) \;=\; u\,(\mathbf{b} \cdot \mathbf{b}) + v\,(\mathbf{b} \cdot \mathbf{c})$$

$$|\,\mathbf{b} \times \mathbf{c}\,|^2 \;=\; -[(\mathbf{b} \times \mathbf{c}), \mathbf{b}, \mathbf{c}] \;=\; [\mathbf{b}, (\mathbf{b} \times \mathbf{c}), \mathbf{c}] \;=\; [\mathbf{c}, \mathbf{b}, (\mathbf{b} \times \mathbf{c})] \;=\; \big(\mathbf{c} \cdot (\mathbf{b} \times (\mathbf{b} \times \mathbf{c}))\big) \;=\; \mathbf{c} \cdot (u\,\mathbf{b} + v\,\mathbf{c}) \;=\; u\,(\mathbf{b} \cdot \mathbf{c}) + v\,(\mathbf{c} \cdot \mathbf{c})$$

If we solve these equations for u and v we find that $u = \mathbf{b} \cdot \mathbf{c}$ and $v = -\mathbf{b} \cdot \mathbf{b}$. Therefore we have $\mathbf{b} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{b} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{b} \cdot \mathbf{b})\,\mathbf{c}$. Similarly, we have a second identity of the form $\mathbf{c} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{c} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\,\mathbf{c}$.

If we now write $\mathbf{a} = p\,\mathbf{b} + q\,\mathbf{c} + r\,(\mathbf{b} \times \mathbf{c})$ we have

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \;=\; p\,\mathbf{b} \times (\mathbf{b} \times \mathbf{c}) \;+\; q\,\mathbf{c} \times (\mathbf{b} \times \mathbf{c}) \;=$$

$$(p\,(\mathbf{b} \cdot \mathbf{c}) + q\,(\mathbf{c} \cdot \mathbf{c})\,)\,\mathbf{b} \;-\; (p\,(\mathbf{b} \cdot \mathbf{b}) + q\,(\mathbf{b} \cdot \mathbf{c})\,)\,\mathbf{c}\,.$$

Since **b** and **c** are perpendicular to their cross product, we must have

$$(\mathbf{a} \cdot \mathbf{c}) \;=\; p\,(\mathbf{b} \cdot \mathbf{c}) + q\,(\mathbf{c} \cdot \mathbf{c}) \quad \text{and} \quad (\mathbf{a} \cdot \mathbf{b}) \;=\; p\,(\mathbf{b} \cdot \mathbf{b}) + q\,(\mathbf{b} \cdot \mathbf{c})$$

so that the previous expression for $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ is equal to $(\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} \;-\; (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}$.■

Although the cross product is not associative, it satisfies a condition on threefold products called the ***Jacobi identity:***

Theorem 7. *If* **a, b, c** *are vectors in* $\mathbf{R}^3$, *then*

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \;+\; \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) \;+\; \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) \;=\; \mathbf{0}\,.\,.$$

Derivation. Three applications of the **"BAC – CAB"** rule yield the following equations**:**

$$\begin{aligned}
\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) &= (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c} \\
\mathbf{b} \times (\mathbf{c} \times \mathbf{a}) &= (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c} - (\mathbf{b} \cdot \mathbf{c})\,\mathbf{a} \\
\mathbf{c} \times (\mathbf{a} \times \mathbf{b}) &= (\mathbf{b} \cdot \mathbf{c})\,\mathbf{a} - (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b}
\end{aligned}$$

Since the sum of the three expressions on the right hand side is equal to zero, the same is true for the sum of the three expressions on the left hand side.■

The formula for $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ also yields numerous other identities. Here is one that is sometimes useful.

Proposition 7. *If* **a**, **b**, **c** *and* **d** *are arbitrary vectors in* $\mathbf{R}^3$ *then we have the identity*

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) \;=\; (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) \;-\; (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})\,.$$

Proof. By definition, the expression on the left hand side of the display is equal to the triple product $[\,(\mathbf{a} \times \mathbf{b}),\, \mathbf{c},\, \mathbf{d}]$. As noted above, the properties of determinants imply that the latter is equal to $[\mathbf{d},\, (\mathbf{a} \times \mathbf{b}),\, \mathbf{c}]$, which in turn is equal to

$$\mathbf{d} \cdot (\mathbf{a} \times (\mathbf{b} \times \mathbf{c})\,) \;=\; \mathbf{d} \cdot (\,(\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} \;-\; (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}\,)$$

and if we expand the final term we obtain $(\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) \;-\; (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$.■

Cross products and higher dimensions

Given the relative ease in defining generalizations of the inner (or dot) product on $\mathbf{R}^n$ and the usefulness of the cross product on $\mathbf{R}^3$, it is natural to ask whether there are also generalizations of the cross product. However, it is rarely possible to define good generalizations of the cross product that satisfy most of the latter's good properties (specifically, it can only be done when $n = 7$). Partial but significantly more complicated generalizations can be constructed using relatively sophisticated techniques (for example, from tensor algebra or Lie algebras), but such material goes far beyond the scope of this course. Here are two online references containing further information:

http://www.math.niu.edu/~rusin/known-math/95/prods

http://www.math.niu.edu/~rusin/known-math/96/octonionic http://arxiv.org/PS_cache/math/pdf/0204/0204357.pdf

We shall not use the material in these references subsequently. However, we note that there is an n – ary operation in $\mathbf{R}^{n+1}$ (a special case of the ***wedge product***) which is analogous to the cross product on $\mathbf{R}^3$, and it is given by the following formula:

$$\bigwedge(\mathbf{v}_1, \cdots, \mathbf{v}_n) = \begin{vmatrix} v_1{}^1 & \cdots & v_1{}^{n+1} \\ \vdots & \ddots & \vdots \\ v_n{}^1 & \cdots & v_n{}^{n+1} \\ \mathbf{e}_1 & \cdots & \mathbf{e}_{n+1} \end{vmatrix}.$$

This formula is identical in structure to the determinant formula for the usual cross product in $\mathbf{R}^3$ except that the row of basis vectors is the last row in the determinant rather than the first. A discussion of the reasons for this is beyond the scope of these notes. The vector obtained by this operation is perpendicular to all the vectors $\mathbf{v}_i$, and it turns out to be nonzero if and only if these vectors are linearly independent,

I.3 : Linear varieties

Solutions to systems of linear equations are studied in all basic linear algebra courses. The purpose of this section is to describe several equivalent ways of characterizing those subsets of $\mathbf{R}^n$ that arise as solutions to such systems of equations. Portions of the discussion below are adapted from the following online sites:

http://en.wikipedia.org/wiki/Linear_algebra

http://en.wikipedia.org/wiki/System_of_linear_equations

In general, a system with m linear equations and n unknowns can be written as

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{gathered}$$

where $x_1, x_2, \ldots, x_n$ are the unknowns and the numbers $a_{11}, a_{12}, \ldots, a_{ij}$ are the coefficients of the system. We can display the coefficients in a matrix as follows:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

If we represent each matrix by a single letter, this becomes

$$\mathbf{Ax} = \mathbf{b}$$

where **A** is an $m \times n$ matrix above, **x** is a column vector with n entries and **b** is a column vector with m entries. We may also rewrite this in terms of vectors as a system of m dot product equations

$$\mathbf{a}_i \cdot \mathbf{x} = b_i \qquad i = 1, 2, \dots, m$$

where $\mathbf{a}_i$ denotes the i^{th} row of the matrix **A.** A system of the form $\mathbf{Ax} = \mathbf{0}$ is called a ***homogeneous*** system of linear equations, and given a system of equations $\mathbf{Ax} = \mathbf{b}$, the related homogeneous system $\mathbf{Ax} = \mathbf{0}$ is called the corresponding ***reduced system of linear equations.***

Definition. We shall say that a subset **X** of $\mathbf{R}^{\mathbf{n}}$ is a ***linear variety*** if it is the set of solutions for a system of linear equations as above.

If we are given a system of linear equations, then the methods of elementary linear algebra provide an effective method for computing the corresponding set of solutions. Our emphasis here will be on general descriptions of such sets of solutions. Nnn

Examples. 1. If we have a homogeneous system of equations, the set of solutions is a vector subspace, and its dimension is equal to $n - r$, where r is the dimension of the subspace spanned by the rows $\mathbf{a}_i$ of the matrix **A.**

2. If $m = n$ and **A** is an invertible matrix, then the set of solutions consists of the single vector $\mathbf{A}^{-1}\mathbf{b}$.

3. In some cases the solution set is empty. For example, this happens if we have the ***inconsistent*** or ***overdetermined*** system of equations $\mathbf{a} \cdot \mathbf{x} = b$, $\mathbf{a} \cdot \mathbf{x} = b + c$, where $c \neq 0$. Special cases like the system $x + y = 1$, $x + y = 2$ are usually discussed in elementary algebra textbooks.

4. Other elementary examples with nonempty sets of solutions include single equations like $x + y + z = 1$; geometrically, the set of solutions for this equation corresponds to the plane passing through the three unit vectors $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$.

Not surprisingly, our interest centers on examples for which the set of solutions is nonempty, and for our purposes the following result is the starting point.

Theorem 1. *Suppose that* $\mathbf{x}_0$ *is a particular solution of the system of equations* $\mathbf{Ax} = \mathbf{b}$. *Then the set of solutions for the system consists of all vectors which can be expressed in the form* $\mathbf{x}_0 + \mathbf{y}$, *where* $\mathbf{y}$ *is a solution to the associated reduced system* $\mathbf{Ax} = \mathbf{0}$.

Proof. First of all every vector of the form $\mathbf{x}_0 + \mathbf{y}$ as above is a solution because we have $\mathbf{A}(\mathbf{x}_0 + \mathbf{y}) = \mathbf{Ax}_0 + \mathbf{Ay} = \mathbf{b} + \mathbf{0} = \mathbf{b}$. Conversely, if $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{y} = \mathbf{x} - \mathbf{x}_0$

then $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$ and $\mathbf{Ay} = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{Ax}_0 - \mathbf{Ax} = \mathbf{b} - \mathbf{b} = \mathbf{0}$, so every solution can be expressed in the given form.■

Notation. Given a subspace **W**, the ***translate of* W *by the vector* v** is the set of all vectors expressible in the form $\mathbf{v} + \mathbf{w}$ for some $\mathbf{w} \in \mathbf{W}$, and this set is written $\mathbf{v} + \mathbf{W}$. Using this terminology we can restate the previous result as follows.

Corollary 2. *If the system of equations* $\mathbf{Ax} = \mathbf{b}$ *has a nonempty set of solutions, then this set of solutions has the form* $\mathbf{x}_0 + \mathbf{W}$, *where* $\mathbf{x}_0$ *is an arbitrary vector in the solution set and* **W** *is the subspace of solutions to the associated reduced equation.*■

Our next order of business is to prove a converse to this corollary.

Proposition 3. *If* **W** *is a subspace of* $\mathbf{R}^n$ *and* $\mathbf{x}_0$ *is an arbitrary vector in* $\mathbf{R}^n$, *then there is a system of linear equations* $\mathbf{Ax} = \mathbf{b}$ *whose solution set is equal to* $\mathbf{x}_0 + \mathbf{W}$.

The following characterization of linear varieties follows immediately.

Corollary 4. *A subset of* $\mathbf{R}^n$ *is a linear variety if and only if it is either empty or it is a translate of some vector subspace.*■

Proof of Proposition 3. The most important step is to prove the statement when $\mathbf{x}_0 = \mathbf{0}$ (the homogeneous case). Let d be the dimension of **W**, and let **C** be a $d \times n$ matrix whose rows correspond to a basis for **W**. It follows that the solution subspace $\mathbf{W}^{\perp}$ for the homogeneous linear system $\mathbf{Cy} = \mathbf{0}$ has dimension equal to $n - d$. Now let **A** be the $(n - d) \times n$ matrix whose rows correspond to a basis for $\mathbf{W}^{\perp}$ and consider the subspace **U** of solutions for the system $\mathbf{Ax} = \mathbf{0}$. It follows that **W** is contained in **U**, and we also have $\dim \mathbf{U} = n - (n - d) = d = \dim \mathbf{W}$. Since **W** is contained in **U** and the dimensions are equal, we must have $\mathbf{W} = \mathbf{U}$, and thus we have shown that **W** is the set of solutions to some homogeneous system $\mathbf{Ax} = \mathbf{0}$. Finally, to prove the general case it suffices to set **b** equal to $\mathbf{Ax}_0$; by the preceding proposition and the argument above, it follows that $\mathbf{x}_0 + \mathbf{W}$ is the solution set for the system $\mathbf{Ax} = \mathbf{b}$.■

Hyperplanes

One special case of the preceding discussion is particularly important.

Definition. A subset **H** of $\mathbf{R}^n$ is called a ***hyperplane*** in the latter if it has the form $\mathbf{v} + \mathbf{W}$ where **W** is an $(n - 1)$ – dimensional vector subspace of $\mathbf{R}^n$ and **v** is a vector in $\mathbf{R}^n$.

Note that this definition involves the vector space containing **H** as well as **H** itself. Of course, if $n = 3$ then we often simply call **H** a ***plane***, and if $n = 2$ then **H** corresponds to the usual notion of ***line*** in the coordinate plane (compare the discussion under the heading *Flat subsets* below).

Proposition 5. *A subset* **H** of $\mathbf{R}^n$ *is a hyperplane in the latter if and only if it is the solution set for a nontrivial linear equation* $\mathbf{a} \cdot \mathbf{x} = b$ *where* **a** *is a nonzero vector in* $\mathbf{R}^n$ *and* b *is some scalar.*

Proof. We shall first verify that if **H** is a hyperplane then **H** is the solution set for some nontrivial linear equation, and we shall do so by looking more closely at the proof of the preceding proposition. Using the definition write $\mathbf{H} = \mathbf{v} + \mathbf{W}$ as above.

In the present situation we have $d = 1$, and thus the matrix **C** is just an n – dimensional row vector **a.** Since the rows of **C** are linearly independent by construction, it follows that **a** is nonzero, and by the argument in the previous result we know that **H** is the set of solutions for the nontrivial linear equation $\mathbf{a} \cdot \mathbf{x} = b$, where $b = \mathbf{a} \cdot \mathbf{v}$.

Conversely, if we have a nontrivial linear equation of the form $\mathbf{a} \cdot \mathbf{x} = b$, then the solution set for the associated reduced equation $\mathbf{a} \cdot \mathbf{x} = 0$ is an $(n-1)$ – dimensional vector subspace of $\mathbf{R}^n$, and hence the preceding results show that the set of solutions is either empty or a hyperplane. Thus we only need to show that *the set of solutions for the original equation is nonempty*. Since **a** is nonzero, there is some coordinate a_j which is nonzero. Thus if $\mathbf{e}_j$ denotes the unit vector whose j^{th} coordinate is equal to 1 and whose other coordinates are zero, then it follows that $(b/a_j)\,\mathbf{e}_j$ is a solution to the original equation $\mathbf{a} \cdot \mathbf{x} = b$.■

The following result generalizes the existence portion of the previous result, and proofs of it (or some equivalent statement) appear in virtually all linear algebra texts.

Theorem 6. *Suppose that* $m \leq n$ *and* **A** *is an* $m \times n$ *matrix with linearly independent rows. Then the system of linear equations* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *has at least one solution.*■

Flat subsets

There is another characterization of linear varieties that is geometrically motivated and useful in several contexts. In order to define this we must first give a definition of lines which works for an arbitrary vector space.

Definition. Let **V** be a vector space over the real numbers. A ***line*** in **V** is a subset **L** of the form **v + W,** where **W** is a 1 – dimensional vector subspace of **V** and $\mathbf{v} \in \mathbf{V}$.

In order to justify this terminology we shall prove that such lines share a fundamental geometric property with ordinary geometrical lines.

Proposition 7. If **x** and **y** are distinct vectors in the vector space **V,** then there is a unique line in **V** containing them.

Proof. We first verify that at least one line exists: Since $\mathbf{x} \neq \mathbf{y}$ we know that $\mathbf{y} - \mathbf{x} \neq \mathbf{0}$ and hence the subspace **W** of all scalar multiples of **y – x** is 1 – dimensional. It follows immediately that the line **x + W** contains $\mathbf{x} = \mathbf{x} + \mathbf{0}$ and $\mathbf{y} = \mathbf{x} + (\mathbf{y} - \mathbf{x})$, We now need to show that *there is only one line containing* **x** *and* **y.** Suppose that **v + W** (with **v** and **W** as above) is a line containing **x** and **y,** and write **x = v + a** and **y = v + b** for suitable vectors **a, b** $\in$ **W.** Then **y – x = b – a** and therefore **y – x** lies in **W.** Since **W** is 1 – dimensional there is a nonzero vector **z** in **W** such that every vector in **W** is a scalar multiple of **z.** This means that **a** = s **z** and **b** = t **z** for appropriate scalars s and t, and hence also that **y – x** = $(t-s)$**z.** Since **x** and **y** are distinct, it follows that the quantity $t-s$ is nonzero. Dividing by this quantity, we conclude that **z** is a scalar multiple of **y – x** and hence that **W** is contained in the 1 – dimensional span $\mathbf{W_1}$ of **y – x.** Since **W** and $\mathbf{W_1}$ have the same dimension, they must be equal. To complete the proof, it will suffice to verify that **v + W = x + W,** and we do so as follows: Given a vector **v + w** in **v + W** (with **w** $\in$ **W**), we have

$$\mathbf{v + w} \;=\; \mathbf{v + a - a + w} \;=\; \mathbf{x + (w - a)}$$

and the vector on the right hand side lies in **x + W** because **W** is closed under subtraction. Conversely, given a vector **x + u** in **v + W** (with **u** $\in$ **W**)**,** we have

$$\mathbf{x + u} \;=\; \mathbf{v + a + u}$$

and the vector on the right hand side lies in **v + W** because **W** is closed under addition. Thus the line **v + W** must be equal to the line described in the first paragraph of the proof.■

The preceding result allows us to speak of the ***line joining two distinct vectors*** in a vector space.

Definition. Let **V** be a vector space, and let **F** be a nonempty subspace of **V.** We shall say that **F** is a ***flat subset*** of **V** if for each pair of distinct vectors **x, y** $\in$ **F** the line joining them also lies in **F,**

Example. If **W** is a vector subspace of **V** and **v** $\in$ **V,** then **v + W** is a flat subset. Here is the ***proof*:** Suppose that **x** and **y** are distinct points of **v + W,** and write **x = v + a** and **y = v + b** for suitable vectors **a, b** $\in$ **W.** By the algebraic equations in the proof of the preceding result, a typical vector on the line joining the given two vectors will have the form

$$\mathbf{x} + t\,\mathbf{(y - x)} \;=\; \mathbf{x} + t\,\mathbf{(b - a)} \;=\; \mathbf{v + a} + t\,\mathbf{(b - a)}$$

and clearly the sum of the second and third terms in the right hand expression lies in the vector subspace **W.**■

The final result of this section is a converse to the preceding example.

Theorem 8. *Let* **V** *be a vector space over the real numbers, and let* **F** *be a flat subset of* **V.** *Then there is a vector subspace* **W** *of* **V** *such that* **F = v + W** *for some* **v** $\in$ **F.**

It follows that <u>*a nonempty subset of* $\mathbf{R}^n$ *is a **linear variety** if and only if it is **flat.***</u>

<u>Proof.</u> Since **F** is nonempty it contains some element **v.** Let $\mathbf{F_0}$ be the set of all vectors expressible as **x – v** where **x** $\in$ **F.** We then have **F = v +** $\mathbf{F_0}$, and thus it suffices to show that $\mathbf{F_0}$ is a vector subspace. If **F** consists of only the vector **v,** then $\mathbf{F_0}$ consists of only the vector **0** and hence is trivially a subspace, so we shall assume for the remainder of the proof that **F** contains more than one vector.

We first show that $\mathbf{F_0}$ is closed under scalar multiplication. Suppose that **w** $\in$ $\mathbf{F_0}$ and c is a scalar. If **w = 0** then the conclusion is trivial, so assume that **w** $\neq$ **0.** Then by flatness we know that **F** contains every point on the line joining **x** to **x + w,** and in particular that it contains the point

$$\mathbf{x} + c\,([\mathbf{x} + \mathbf{w}] - \mathbf{x}) \;=\; \mathbf{x} + c\mathbf{w}$$

which means that $c\mathbf{w} \in \mathbf{F_0}$ as desired**.**

Suppose now that **u, v** $\in$ $\mathbf{F_0}$**;** we need to show that their sum also lies in $\mathbf{F_0}$**.** If **u = v** this follows from the previous paragraph because **u + v** = 2**u** = 2**v,** so assume from now on that **u** $\neq$ **v.** By flatness we know that the point

$$(\mathbf{x} + \mathbf{u}) + \tfrac{1}{2}([\mathbf{x}+\mathbf{v}] - [\mathbf{x}+\mathbf{u}]) \;=\; (\mathbf{x}+\mathbf{u}) + \tfrac{1}{2}(\mathbf{v} - \mathbf{u}) \;=\; \mathbf{x} + \tfrac{1}{2}(\mathbf{u}+\mathbf{v})$$

also lies in **F,** so that ½ **(u+v)** $\in$ $\mathbf{F_0}$**.** We can now apply the results of the previous paragraph on scalar multiples to conclude that **u + v = 2 · ½ (u + v)** also belongs to $\mathbf{F_0}$**,** which completes the proof that $\mathbf{F_0}$ is a vector subspace of **V.**■

A result on translates of subspaces

The following result turns out to be an extremely useful tool for many proofs in these notes. We shall call this the ***equivalence class property*** or the ***Coset Property..***

<u>Lemma 9.</u> *Let* **W** *be a subspace of the vector space* **V,** *let* **x** $\in$ **V,** *and let* **z** $\in$ **x + W.** *Then* **z + W = x + W.**

<u>Proof.</u> Suppose that **z + w** $\in$ **z + W,** where **w** $\in$ **W.** By the assumption on **z** we have **z = x + u** where **u** $\in$ **W,** so that **z + w = x + u + w.** Since **W** is a vector subspace it follows that **u + w** $\in$ **W,** which means that **z + w = x + u + w** $\in$ **x + W.**

Conversely, suppose that **x + w** $\in$ **x + W,** where **w** $\in$ **W.** Then **z = x + u** implies that **x = z – u** and hence **x + w = z – u + w = z + w – u.** Since **W** is a vector subspace it follows that **w – u** $\in$ **W,** so that **x + w = z + w – u** $\in$ **z + W.**

Thus every vector in **z + W** is in **x + W** and <u>*vice versa*</u>, so that the two subsets must be equal.■

Remark. The motivation for the name is that the sets **x + W** are *equivalence classes* with respect to the *equivalence relation* on the vector space **V** defined by **u ~ v** if and only if **u – v** $\in$ **W.** A general result on the equivalence classes of an equivalence relation states that they are either disjoint or identical, and the lemma essentially gives a special case of that result.

Multidimensional geometry

The concepts from linear algebra in this section provide a conceptual basis for at least a crude mathematical theory of n – dimensional geometry, where n is an arbitrary positive integer. In dimensions **2** and **3,** we have seen that lines and planes are translates of vector subspaces and that distances and angles can defined in terms of numerical formulas on $\mathbf{R}^n$ which make mathematical sense for all finite values of n. We have already defined ***lines*** in $\mathbf{R}^n$ to be translates of **1** – dimensional vector subspaces, and in a similar manner we can define ***planes***, or more correctly **2** – ***planes***, to be translates of **2** – dimensional subspaces. If n is greater than **3,** one might expect that there is some further "flat subset" structure on $\mathbf{R}^n$ corresponding to translates of k – dimensional vector spaces for all $k = 3, \dots, n - 1$; such objects are frequently called k – ***planes.*** If we compare this definition with previous ones, we see that the previously defined notion of a hyperplane in $\mathbf{R}^n$ is the same as $(n - 1)$ – planes.

During the 19th century, both mathematical and physical considerations motivated interest in the geometry of n – dimensional space for values of n not necessarily equal to **2** or **3,** and a fairly extensive theory of such objects was established fairly quickly by the end of the century. The importance of **4** – dimensional geometry for relativity theory is fairly well known, and multidimensional objects arise in an extremely wide range of other mathematical contexts which are not only theoretical but also involve applications in many different directions. Not surprisingly, simple multidimensional concepts like the "flat subset" structure are often the most useful because it is relatively quick and easy to develop enough intuition about them to work with them effectively.

In graduate level courses and even many undergraduate courses, many (in fact, most) basic concepts are developed in terms of n – dimensional Euclidean space, where n can be any positive integer. Besides the obvious advantages of increased generality and broader applicability, this also provides a unified framework for numerous results in the **2** – dimensional and **3** – dimensional cases that are not necessarily clear if one treats these dimensions separately and ignores all higher dimensions. Linear algebra and its applications are filled with examples of this sort. However, the usefulness of the multivariable approach is also extremely apparent multivariable calculus; in particular, the multidimensional approach provides a systematic and unified approach to results such as the second partial derivative tests for local maxima and minima, the standard results on interchanging the order of integration, the change of variables formula for multiple integrals, Green's Theorem, Stokes' Theorem the Divergence Theorem, and their analogs in higher dimensions.

The role of n – dimensional spaces in present day mathematics is well reflected by nearly every standard text in undergraduate linear algebra as well as higher level texts on the multivariable calculus, and also in nearly every standard graduate text on measure theory, topological spaces, or differentiable manifolds. Some of these would be appropriate recommendations for further information on multidimensional spaces and their role in mathematics. Various online and textbook references for $\mathbf{4}$ – dimensional geometry, or more generally for n – dimensional geometry, provide still further options. However, some of these can be very difficult reading because it is not always easy to develop a good intuition or visual model for many of the objects that are studied.

I.4 : Barycentric coordinates

The material in this section below is not covered in the greatest possible generality, but the treatment will suffice for our purposes. Here are some further references for more information on this and closely related topics such as ***affine geometry.*** In particular, the treatments in these references develop the whole theory for $\mathbf{R}^n$, where n is an arbitrary positive integer, and not just for $\mathbf{R}^2$ as in these notes.

http://www.cut-the-knot.org/triangle/barycenter.shtml

http://mathworld.wolfram.com/BarycentricCoordinates.html

http://mathworld.wolfram.com/MenelausTheorem.html

http://graphics.idav.ucdavis.edu/education/GraphicsNotes/Barycentric-Coordinates/Barycentric-Coordinates.html

G. Birkhoff and S. MacLane, ***A Survey of Modern Algebra.*** (Reprint of the Third 1968 Edition). Chelsea Publishing, New York, NY, 1988. ISBN: 0 – 023 – 74310 – 7.

I. Kaplansky, ***Linear Algebra and Geometry: A Second Course.*** (Reprint of the 2nd Edition). Dover Publications, New York, NY, 2003. ISBN: 0 – 486 – 43233 – 5.

Several of the exercises also discuss barycentric coordinates in more general situations.

The algebraic setting

We shall work entirely within $\mathbf{R}^2$. As noted above, there are generalizations to arbitrary vector spaces, but they require additional formal machinery. Given a subset $\mathbf{Y}$ of a vector space $\mathbf{V}$, we shall say that the subset is ***collinear*** if there is a line containing it and it is ***noncollinear*** if no such line exists.

Theorem 1. *Let* **a, b, c** *be (distinct) noncollinear points in* $\mathbf{R}^2$. *Then every vector* $\mathbf{v}$ *in* $\mathbf{R}^2$ has a unique expression as a linear combination $x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ *such that the sum of the coefficients* $x + y + z$ is equal to $\mathbf{1}$.

Clearly the condition on the sum of the coefficients is crucial for uniqueness. Every set of three vectors in $\mathbf{R}^2$ is linearly dependent, and there are infinitely many ways of writing a vector in the latter as an unrestricted linear combination of, say, the noncollinear vectors $\mathbf{e_1}$, $\mathbf{e_2}$ and $\mathbf{e_1} + \mathbf{e_2}$ (checking that three vectors are not collinear is left to the reader as an exercise).

Definition. The uniquely determined coefficients x, y, z in the theorem are called the ***barycentric coordinates*** of **v** with respect to the points **a, b, c.**

Proof. The first step is to show that **b – a** and **c – a** are linearly independent. If not, then since these vectors are nonzero we can write each as a nonzero scalar multiple of the other; for example, we have $\mathbf{b} - \mathbf{a} = r\,(\mathbf{c} - \mathbf{a})$. But then we also have

$$\mathbf{b} = \mathbf{a} + (\mathbf{b} - \mathbf{a}) = \mathbf{a} + r\,(\mathbf{c} - \mathbf{a})$$

which implies that **b** lies on the line determined by **a** and **c.** Since this is impossible, the two vectors in the first line of this paragraph must be linearly independent. It follows that **b – a** and **c – a** form a basis for the 2 – dimensional vector space $\mathbf{R}^2$.■

Suppose now that $\mathbf{x} \in \mathbf{R}^2$. By the previous paragraph we can find scalars s and t such that $\mathbf{x} - \mathbf{a} = s(\mathbf{b} - \mathbf{a}) + t(\mathbf{c} - \mathbf{a})$. We may rewrite this in the form

$$\mathbf{x} = (1 - s - t)\,\mathbf{a} + s\,\mathbf{b} + t\,\mathbf{c}$$

and thus it follows that **x** has at least one expression as a linear combination of the prescribed type. Suppose now that we have an arbitrary such expansion

$$\mathbf{x} = p\,\mathbf{a} + q\,\mathbf{b} + r\,\mathbf{c}$$

where the sum of the coefficients is equal to 1. Then we may rewrite this as

$$\mathbf{x} = (1 - q - r)\,\mathbf{a} + q\,\mathbf{b} + r\,\mathbf{c}$$

and we may rewrite it still further in the following form:

$$\mathbf{x} - \mathbf{a} = q\,(\mathbf{b} - \mathbf{a}) + r(\mathbf{c} - \mathbf{a})$$

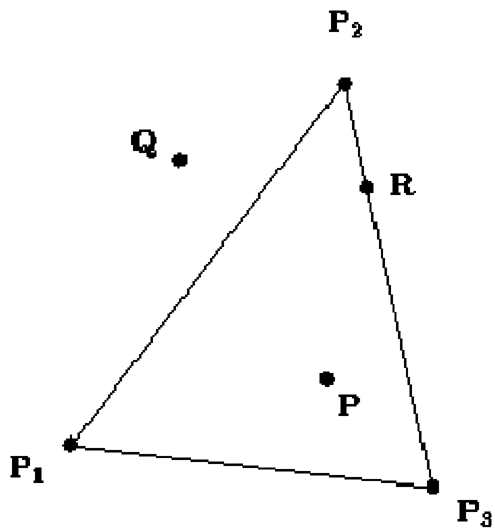
Since **b – a** and **c – a** form a basis for $\mathbf{R}^2$, it follows that the coefficients for the two expressions for **x – a** must be equal; *i.e.*, we have $q = s$ and $r = t$. Finally, since we also have $p + q + r = 1$, it follows that $p = 1 - q - r = 1 - s - t$, and hence the corresponding coefficients in both expressions for **x** must be equal.■

Physical interpretation

The name "***barycentric coordinates***" was given because these numbers have a physical interpretation involving the center of gravity (or ***barycenter***) of a system of weights placed at the three noncollinear points. Specifically, in the illustration below suppose that we place weights of w_i units at each of the points $\mathbf{P}_i$, and suppose also that we normalize our measurement system so that the sum of the weights is equal to $\mathbf{1}$. Then the center of gravity for the system is expressible in vector form as

$$\text{Center of Gravity} = w_1\mathbf{P}_1 + w_2\mathbf{P}_2 + w_3\mathbf{P}_3$$

Ordinarily we think of weights as being positive or at least nonnegative, but if we have some way of thinking about negative weights (for example, attaching a helium balloon at one or two vertices rather than a weight made of iron or lead), then the formula still works provided we still assume that the sum of the three weights is equal to $\mathbf{1}$ unit. In the picture below, the point **P** is the barycenter for a system such that all weights w_i are positive, while **Q** is the barycenter for a system where $w_1 = 0$ but the other two weights are positive, and **R** is the barycenter for a system where w_1 is negative but the other two weights are positive.



(***Source:*** http://graphics.idav.ucdavis.edu/education/GraphicsNotes/Barycentric-Coordinates/Barycentric-Coordinates.html **)**

Examples of points for which w_2 is positive but the remaining weights are negative can also be constructed using this picture; for example, if one draws the perpendicular from $\mathbf{P}_2$ to the line $\mathbf{P}_1\mathbf{P}_3$, then each point **S** on this perpendicular which lie above $\mathbf{P}_2$ in the picture (alternatively, $\mathbf{P}_2$ is between **S** and the foot of the perpendicular on $\mathbf{P}_1\mathbf{P}_3$) will have this property.

Note. The condition $w_1 + w_2 + w_3 = \mathbf{1}$ is mainly a matter of convenience to simplify the mathematical discussion; if k is a nonzero constant then the barycenter will be the same for any weight system with weights of $k\,w_i$ units at the vertices).

Applications to geometric proofs

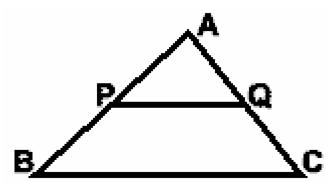
Barycentric coordinates are often very useful for proving geometrical results using vectors. We shall give one simple example here, and there are others in the exercises. For the sake of completeness we start with a standard observation:

Fact 2. *If* **P** *and* **Q** *are points in some* $\mathbf{R}^n$**,** *then the midpoint* **M** *of the points* **P** *and* **Q** *is equal to* **½(P + Q).**

Verification. The point **M** lies on the line joining **P** and **Q** because **M** is algebraically equal to **P + ½(Q – P)**, and since **M – P = ½(Q – P) = Q – M** it follows that **M** satisfies all the conditions to be a midpoint.■

We shall apply this to give a purely algebraic proof of a standard theorem from plane geometry. It would be instructive (but not necessary) to compare the argument given below with the usual proof(s) in classical Euclidean geometry.

Theorem 3. *The line joining the midpoints of two sides of a triangle is parallel to the third side.*



(Source: http://www.ilovemaths.com/1area.htm)

Proof. Let **A, B** and **C** be the vertices of the triangle, and let **P** and **Q** be the midpoints of the sides **[AB]** and **[AC]** respectively. By the preceding observation on midpoints, we know that **P = ½(A + B)** and **Q = ½(A + C).** The latter imply that **Q – P = ½(C – B).** To prove that the lines **PQ** and **BC** are parallel we need to show that they have no points in common, so suppose that we do have a point on both. This means that there are scalars s and t such that **P +** s **(Q – P) = B +** t **(C – B)** and if we rewrite **P** and **Q** using the midpoint formulas this equation becomes

$$\tfrac{1}{2}\mathbf{A} + \tfrac{1}{2}(1-s)\,\mathbf{B} + \tfrac{1}{2}s\,\mathbf{C} \;=\; (1-t)\,\mathbf{B} + t\,\mathbf{C} \;=\; 0\cdot\mathbf{A} + (1-t)\,\mathbf{B} + t\,\mathbf{C}\,.$$

Since **A, B, C** are noncollinear, the coefficients of these vectors on the left and right hand sides of this equation must be equal. However, this is not the case, for the coefficients of **A** on the two sides of the equation are ***not*** equal, and thus we have a contradiction. The source of the contradiction is our assumption that there was a common point on the lines **BC** and **PQ,** and thus no such common point exists; this means that the two lines are parallel.■

The exercises for this section include linear algebraic proofs of some results from Euclidean geometry that are considerably more difficult to prove by classical Greek methods.

IMPORTANT. There will be many other examples of geometric proofs using vectors and barycentric coordinates in these notes, particularly in Units **II** and **III.** Therefore it is important to understand the material in this section very thoroughly and likewise for the exercises. Some typical examples of exercises involving vectors and barycentric coordinates are worked out in the next section, and these provide further illustrations of how vectors and barycentric coordinates are applied to geometric questions.

I.5 : Some examples

In this section we shall work a few problems which illustrate the concepts discussed in Sections **I.3 – I.4** and are similar to homework exercises for the latter sections.

PROBLEM 1. *Let* **L** *be the line in* $\mathbf{R}^3$ *joining the points* **a** = **(1, 0, 0)** *and* **b** = **(0, 1, 0)**, *and let* **M** *be the line joining points* **c** = **(1, 2, 0)** *and* **d** = **(0, 0, 1)**. *Determine whether the lines intersect, and if so find their point of intersection***.**

SOLUTION. The lines intersect if and only if we can find scalars u and v such that

$$\mathbf{a} + u(\mathbf{b} - \mathbf{a}) = \mathbf{c} + v(\mathbf{d} - \mathbf{c})$$

or equivalently if we can write

$$\mathbf{c} - \mathbf{a} = u(\mathbf{b} - \mathbf{a}) - v(\mathbf{d} - \mathbf{c})$$

for suitable scalars u and v. Now we have

$$\mathbf{b} - \mathbf{a} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{d} - \mathbf{c} = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}, \qquad \mathbf{c} - \mathbf{a} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

so the answers reduce to finding solutions for the system of three equations $u - v = 0$, $u + 2v = 2$, and $v = 0$. This system is ***inconsistent***, and therefore ***the two lines do not have any points in common.***■

PROBLEM 2. *Let* **L** *be the line in* $\mathbf{R}^3$ *joining the points* **a** = **(0, 0, 1)** *and* **b** = **(1, 1, 3)**, *and let* **M** *be the line joining points* **c** = **(2, 1, 4)** *and* **d** = **(1, 2, 4)**. *Determine whether the lines intersect, and if so find their point of intersection***.**

SOLUTION. The intersection and solution criteria are the same as before**.** We now have the vector identities

$$\mathbf{b} - \mathbf{a} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \qquad \mathbf{d} - \mathbf{c} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{c} - \mathbf{a} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

so in this case the answer reduces to finding solutions for the system of linear equations $u + v = 2$, $u - v = 1$, and $2u = 3$. In this case one has a unique solution; namely, $u = 3/2$ and $v = ½$. Direct substitution of these values for u and v then yields the desired common point, which is **(3/2, 3/2, 4)**.■

PROBLEM 3. *Describe the line determined by the intersections of the planes with equations* $x + y + z = 1$ *and* $x + 2y + 3z = 6$.

SOLUTION. We shall analyze the set of solutions using linear algebra and the augmented matrix of coefficients:

$$\begin{pmatrix} 1 & 1 & 1 & : & 1 \\ 1 & 2 & 3 & : & 6 \end{pmatrix}$$

If we put this into row reduced echelon form, we obtain a new system of equations which is equivalent to the old one in which two variables are given explicitly as first degree polynomials in the third. Subtracting the first row from the second yields the matrix

$$\begin{pmatrix} 1 & 1 & 1 & : & 1 \\ 0 & 1 & 2 & : & 5 \end{pmatrix}$$

and if we now subtract the second row from the first we obtain

$$\begin{pmatrix} 1 & 0 & -1 & : & -4 \\ 0 & 1 & 2 & : & 5 \end{pmatrix}$$

which leads to the equations $x = z - 4$ and $y = 5 - 2z$. If we add the tautological equation $z = z$ to this list we obtain the parametric equations for the line, which can be rewritten in vector form as

$$\mathbf{s}(z) = (-4, 5, 0) + z\,(1, -2, 1).$$

Taking $z = 1$, we see that this is the line joining $(-4, 5, 0)$ to $(-3, 3, 1)$. — As always, it is reassuring to check the correctness of the calculations by verifying that each of these two points does satisfy the equations for both planes.■

PROBLEM 4 (Special case of Ceva's Theorem). *Suppose we are given noncollinear points* **A, B, C** *in* $\mathbf{R}^2$, *suppose they are the vertices of an isosceles triangle such that* d(**A, B**) = d(**A, C**), *and let* **D** *be the midpoint of* **[BC]**. *Suppose also that we have a pair of points* **E** *and* **F** *on the segments* **[AC]** *and* **[AB]** *respectively such that* d(**C, E**) = $(1/3)$ d(**C, A**) *and that* d(**B, F**) = $(1/3)$ d(**B, A**). *Assume further that the lines* **BE** *and* **CF** *meet at some point* **G** *on the line segment* **[AD]**. *Find the ratio* d(**D, G**) / d(**D, A**).

SOLUTION. In these problems it is always good to start by writing out everything that is given. First, we know that the midpoint **D** satisfies **D** = ½ **B** + ½ **C**. Next, we know that **E** = p**A** + $(1 - p)$**C** and **F** = q**A** + $(1 - q)$**B**; we can conclude that $p = q = 1/3$ from the assumption on distance ratios and the vector equations **E** – **C** = p(**A** – **C**) and **F** – **C** = q(**B** – **C**). Also, since the point **G** lies on **BE** and **CF** it follows that we have **G** = u**B** + $(1 - u)$**E** = v**C** + $(1 - v)$**F**. We may now use the formulas for **E** and **F** to obtain the following expressions for **G**:

$$\mathbf{G} = u\mathbf{B} + [(1-u)/3]\,\mathbf{A} + [2(1-u)/3]\,\mathbf{C} = v\mathbf{C} + [(1-v)/3]\,\mathbf{A} + [2(1-v)/3]\,\mathbf{B}$$

In both of these expressions the coefficients of **A, B,** and **C** add up to 1, and since **A, B,** and **C** are noncollinear the results on barycentric coordinates imply the corresponding coefficients are equal. Thus we have the following equations:

$$(1 - u)/3 \;=\; (1 - v)/3$$
$$u \;=\; 2(1 - v)/3$$
$$v \;=\; 2(1 - u)/3$$

The first of these equations implies $u = v$ and the others then yield $u = 2/5$. If we substitute this into the formula for **G** we see find that

$$\mathbf{G} \;=\; (1/5)\,\mathbf{A} + (2/5)\mathbf{B} + (2/5)\mathbf{C} \;=\; (1/5)\,\mathbf{A} + (4/5)\mathbf{D}$$

so that $\mathbf{G} - \mathbf{D} = (1/5)\,(\mathbf{A} - \mathbf{D})$ and hence the ratio $d(\mathbf{D}, \mathbf{G})\,/\,d(\mathbf{D}, \mathbf{A})$ is equal to $1/5$.■