Complements to the Teaching of Linear Algebra

David Mond

May 15, 2013

Contents

1 Introduction

These notes cover some topics in Linear Algebra that I feel are frequently misunderstood by students. They are listed and briefly introduced below. I also include some views on teaching which I think may be helpful, even if it is only by inviting disagreement and discussion. This version of the notes is slightly improved and corrected from the printed version distributed at the start of the lectures. I would be grateful for comments, criticisms and corrections. In particular, if the explanations are not clear at any point, please let me know, by e-mail to `d.m.q.mond@warwick.ac.uk`

The topics covered are

1. How to distinguish between "real vector space" and "$\mathbb{R}^n$". Examples of finite dimensional vector spaces which have no natural choice of basis:
   (a) The kernel of a linear map $\mathbb{R}^n \to \mathbb{R}^p$
   (b) The solution space of a linear ODE. Point out that fundamental set of solutions is basis.
   (c) The space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^p$.

2. Distinguish betwen a vector and its expression with respect to a basis. The key point here is notation. If a vector is labelled v and the basis is labelled E, then expression of v with respect to E is denoted $[v]_E$.

3. Derive the procedure for changing basis. Introduce matrix of base-change.

4. Distinguish between matrix and linear transformation. Go over construction of matrix of linear transformation T with respect to basis E in the source and F in the target, $[T]_F^E$. Explain procedure to change basis:

$$[T]_{F_2}^{E_2} = [I]_{F_2}^{F_1}[T]_{F_1}^{E_1}[I]_{E_1}^{E_2}.$$

5. Applications of 4: find the matrix, with respect to the standard basis of $\mathbb{R}^3$, of a reflection in given hyperplane, or of a rotation about a given axis. This involves first choosing an adapted basis (e.g. with one of the basis vectors along axis of rotation), writing down the matrix with respect to this basis, then conjugating by the change of basis matrix. This gives an opportunity for some interesting geometry: show that the composite of two rotations in $\mathbb{R}^3$ is a rotation. Find the axis and angle of rotation of the composite (this requires use of eigenvectors, so perhaps first a revision of this topic). Similar questions about reflection composed with rotation, etc.

6. Inner products. The L_2 inner product on the function spaces is a limit of "dot products" on finite dimensional approximations.

7. Determinant and volume. Invariance under row and column operations. Geometric proof that

   (a) $|detA| =$ volume of parallelipiped spanned by columns of A.

   (b) $det(AB) = det(A)det(B)$, so det is invariant of a linear operator.

   (c) Heuristic proof of formula for change of variable in multiple integration.

8. The last lecture of Linear Algebra could look at the (Frechet) derivative, at a point, of a smooth map. This provides a good example of the pedagogical point that the fewer the objects, the clearer the picture. The chain rule for maps with many variables and many components is very complicated if expressed in terms of partial derivatives, but very simple expressed in terms of the Frechet derivative. Seeing that the matrix of the Frechet derivative is the matrix of partial derivatives, would revise the discussion planned for the first session, on how to write down the matrix of a linear map.

   Also, there are lots of examples where computing the Frechet derivative is fun. For example, for the map

   $$det : \{n \times n \text{ matrices }\} \to \mathbb{R},$$

   or

   $$\text{inverse} : \{n \times n \text{ matrices }\} \to \{n \times n \text{ matrices }\}.$$

   These provide examples of naturally occuring linear maps - of which there are not enough in most courses on linear algebra.

   If there was time, we could go on to the tangent space of a submanifold of $\mathbb{R}^n$, and, perhaps, explain the technique of Lagrange multipliers.

2 Vector spaces over $\mathbb{R}$ or over "a field"?

In a first course, should all vector spaces be real (i.e. with $\mathbb{R}$ as the field of scalars), or is it preferable to speak from the beginning of vector spaces "over a field k"?

Arguments in favour of teaching vector spaces over a general field

1. All of the vector space axioms, and practically all of the algebraic propeties of vector spaces, are independent of the choice of field. The theorems are just as easily proved for vector spaces over an (unspecified) field k, as for vector spaces over $\mathbb{R}$. Now that we mathematicians have understood this fundamental virtue of the axiomatic method, it would be perverse not to benefit from it.

2. Students of modern mathematics should get used from the very beginning to the axiomatic method, in which a mathematical object is defined by means of its axioms, and all its properties follow from these axioms. Teaching mathematics at an abstract level trains students in this approach, and frees them from reliance on their (usually unreliable) intuition.

3. It's actually useful to be able to talk about vector spaces over $\mathbb{Q}$ and $\mathbb{C}$, as well as vector spaces over $\mathbb{R}$. They give good examples to think about: for instance,

   (a) one can think of $\mathbb{R}$ as a vector space over $\mathbb{Q}$, or $\mathbb{C}$ as a vector space over $\mathbb{R}$, and ask about their dimensions (∞ and 2, respectively).

   (b) One can ask: are the real numbers $1, \sqrt{2}, \sqrt{3}$ linearly independent over $\mathbb{Q}$?

   (c) One can ask: can you show me a basis for $\mathbb{R}$ as vector space over $\mathbb{Q}$?

   (d) Since $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$, every vector space over $\mathbb{C}$ is also a vector space over $\mathbb{R}$ and over $\mathbb{Q}$. One can ask: if $\dim_{\mathbb{C}} V = n$, what is $\dim_{\mathbb{R}} V$? Given a basis $v_1, \ldots, v_n$ for V as vector space over $\mathbb{C}$, can you use it to cook up a basis for V over $\mathbb{R}$?

   And so on.

Arguments in favour of teaching vector spaces over $\mathbb{R}$

1. From the perspective of the teacher the first argument above is correct: it is no harder to state and prove the properties of vector spaces over a general field, than to do this only for vector spaces over $\mathbb{R}$. But it is only the teacher who knows this to be so, not the student. The student has no way of knowing that the theorems and proofs are the same for a general field as for $\mathbb{R}$. When the teacher tells the student that this is the case, the student will not necesarily be reassured. He is quite likely to believe that his teacher is exceptionally intelligent and that he, the student, is not. He may feel that the teacher's assurance is based on the teacher's assumption that he, the student, is just as clever as his teacher.

   The student is much more likely to be convinced that the theory is the same over a general field, *after he has seen some of the theory, and is in a position to make his own judgement.* Therefore it is better to postpone the introduction of "a general field" until the student has become familiar with the beginning of the subject.

   Furthermore: we pride ourselves that in mathematics we can show the students why things are true, by proving them. This emphasis on proving things, and not requiring the students to take things on trust, is characteristic of mathematics courses. This transparency is compromised if we tell the students “you will not understand this now, but the theory is the same over a general field as over $\mathbb{R}$” and expect them to believe it without giving them any evidence.

2. By restricting to vector spaces over $\mathbb{R}$, one can use geometry to motivate and enrich linear algebra. One can draw pictures of lines, planes and their intersections, and use these to illustrate the ideas of the course. One can apply linear algebra methods to geometry, to do interesting calculations and prove interesting statements.

3. Once a student has a confident grasp of the basics of linear algebra, they will easily see that the theory works just the same over a general field. Especially the bright ones will grasp this very quickly. The less bright ones who find this generalisation difficult would have found it even more difficult to learn the whole theory from the start over a general field.

4. Argument (c) in favour of a general field, is not really in favour of a general field, but in favour of a number of specific examples. The interesting exercises sketched there may be just what is needed to help the student to move from “vector spaces over $\mathbb{R}$” to “vector spaces over other fields”. When the student is ready for them, they will be useful and exciting. Too early, and they will give the student the feeling that the whole subject is too complicated and that he will never understand it. The student is likely to believe that in order to understand vector spaces over $\mathbb{C}, \mathbb{R}$ and $\mathbb{Q}$ he has to do three times as much work as to understand only vector spaces over $\mathbb{R}$. This is especially likely if he is at the start of his mathematical career.

In the rest of these notes, all vector spaces will be over $\mathbb{R}$ unless specifically stated otherwise.

3 Examples of vector spaces

Every n- dimensional vector space V over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$. Some people deduce from this that we only need to study $\mathbb{R}^n$. I believe that this is a mistake. The very fact that an n-dimensional space is isomorphic to $\mathbb{R}^n$ is itself non-trivial. An isomorphism between V and $\mathbb{R}^n$ requires a choice of basis for V, and different bases give rise to different isomorphisms. Usually there is no natural choice of basis. Excessive concentration on $\mathbb{R}^n$ restricts the imagination [1] of the student, and leaves him or her unable to work with other examples when they occur.

Therefore it is good to introduce a wide range of examples to broaden the imagination. Here are some which I believe are useful. Many can be set as exercises.

Example 3.1. 1. The set $V := \{(x, y, z) \in \mathbb{R}^3 : 2x - 4y + z = 0\}$ is a vector subspace of $\mathbb{R}^3$. Note that V is the set of vectors orthogonal to $(2, -4, 1)$.

(a) If v_1 and v_2 belong to V and are linearly independent, then the three vectors v_1, v_2 and $(2, -4, 1)$ are linearly independent, and so a basis for $\mathbb{R}^3$.

(b) By the previous step, every vector $v \in \mathbb{R}^3$ can be written as a linear combination

$$v = av_1 + bv_2 + c(2, -4, 1)$$

[1] See the discussion of the notion of volume in Section 5 for an example of the usefulness of giving a wide range of examples

If $v \in V$, the coefficient c must be 0, for

$$v \in V \implies v \cdot (2,-4,1) = 0 \implies \big(av_1 + bv_2 + c(2,-4,1)\big) \cdot (2,-4,1) = 0.$$

Since $v_1 \cdot v = v_2 \cdot v = 0$ this means $c(2,-4,1) \cdot (2,-4,1) = 0$, and thus $c = 0$. We conclude that every vector in V can be written as a linear combination just of v_1 and v_2 – that is, v_1 and v_2 span V. This, together with the fact of their linear independence, shows that they form a basis for V, so that V must have dimension 2.

(c) Suggested exercise: Find an orthonormal basis for V (i.e. a basis consisting of vectors of length 1 which are orthogonal to one another.

2. The space of all polynomial functions on $\mathbb{R}$ with real coefficients, $\mathbb{R}[X]$, is a vector space, under the natural operations of addition, and multiplication by scalars. This space provides a good testing ground for some of the basic ideas of linear algebra. In discussing this space, it is important to distinguish between the *poynomial* **0** and the *number* 0. The polynomial **0** is the polynomial which takes the value 0 at every point.

   (a) There are plenty of natural linear maps with domain $\mathbb{R}[X]$:

      i. The map $\mathrm{ev}_a : \mathbb{R}[X] \to \mathbb{R}$ defined by evaluating each polynomial at the point $a \in \mathbb{R}$, i.e.
      $$\mathrm{ev}_a(F) = F(a).$$

      ii. The map $d : \mathbb{R}[X] \to \mathbb{R}[X]$ defined by differentiation, $F \mapsto F'$.

      iii. The map defined by translation, mapping the polynomial $F(X)$ to $F(X-a)$ for some fixed $a \in \mathbb{R}$.

      iv. More generally, the map defined by substitution: if $Q(X) \in \mathbb{R}[X]$ then the map $Q^* : \mathbb{R}[X] \to \mathbb{R}[X]$ sending the polynomial $F(X)$ to $F(Q(X))$, is linear.

   (b) The polynomials $1, X, X^2, \ldots, X^n$ are linearly independent. Here is a simple proof: suppose
   $$a_0 + a_1X + \cdots + a_nX^n = \mathbf{0} \tag{3.1}$$
   in $\mathbb{R}[X]$ (i.e. $a_0 + a_1X + \cdots + a_nX^n$ is the zero polynomial **0**. Once again, we must be very careful to distinguish between the zero polynomial and the real number 0). Evaluating the left hand side at $0 \in \mathbb{R}$, we get
   $$a_0 = 0$$
   straight away. Now differentiate the equality (3.1). We get
   $$a_1 + 2a_2X + \cdots + na_nX^{n-1} = \mathbf{0};$$
   now evaluate at 0 again. We conclude that $a_1 = 0$. Continuing in this way, we conclude that $0 = a_0 = a_1 = \cdots = a_n$. That is, from the supposition that $a_0 + a_1X + \cdots + a_nX^n$ is the zero polynomial, we are able to deduce that all the coefficients a_i are zero. This is what it means for the vectors $1, X, \ldots, X^n$ to be linearly independent.

   Notice that it is the linear independence of $1, X, \ldots, X^n$ that justifies the "method of equating coefficients". This is the rule that if two polynomials $a_0 + a_1X + \cdots + a_nX^n$ and $b_0 + b_1X + \cdots + b_nX^n$ take the same values at every point, then their coefficients are equal, i.e. $a_0 = b_0, a_1 = b_1, \ldots, a_n = b_n$. For if the two polynomials take the same value at every point, then their difference, the polynomial $(a_0 - b_0) + (a_1 - b_1)X + \cdots + (a_n - b_n)X^n$, takes the value 0 at every point. In other words,

$$(a_0 - b_0) + (a_1 - b_1)X + \cdots + (a_n - b_n)X^n = \mathbf{0}.$$

From this we can conclude, *by the linear independence of* $1, X, \ldots, X^n$, that $a_0 - b_0 = 0, \ldots, a_n - b_n = 0$, i.e. $a_0 = b_0, \ldots, a_n = b_n$.

(c) The infinite collection of polynomials, $1, X, \ldots, X^n, \ldots$, are linearly independent. Students may need to be reminded that this means that every *finite* relation

$$\sum_{k \leq N} a_k X^k = \mathbf{0}$$

is necessarily trivial. Linear algebra is never concerned with infinite sums. With this understood, linear independence of the infinite set $1, X, \ldots, X^n, \ldots$ is equivalent to linear independence of each finite subset, which we have shown above.

(d) The polynomials $1, X, \ldots, X^n, \ldots$ span $\mathbb{R}[X]$. This is obvious from the meaning of the term "polynomial". From this and their linear independence, we deduce that $1, X, \ldots, X^n, \ldots$ form a basis for $\mathbb{R}[X]$. In particular, $\mathbb{R}[X]$ has infinite dimension.

3. Slightly more sophisticated: for any interval $I \subset \mathbb{R}$, the space $C^0(I)$ of all continuous functions $I \to \mathbb{R}$, the space $C^1(I)$ of all differentiable functions $I \to \mathbb{R}$ with continuous derivative, $\ldots$,the space $C^k(I)$ of all functions $I \to \mathbb{R}$ which are k times differentiable and have continuous k'th derivative,$\ldots$, the space $C^\infty(I)$ of all functions $I \to \mathbb{R}$ which are infinitely differentiable — all of these are vector spaces over $\mathbb{R}$. It is an interesting and perplexing fact that (provided the interval I has positive length!) it is not possible to exhibit a basis for any one of these spaces, even though, by a well-known application of Zorn's Lemma, every vector space has a basis. The fact that it is impossible to find a basis is a consequence of a sophisticated argument involving the independence of the Axiom of Choice of the other axioms of Set Theory. It is not something one can expect to teach undergraduates! But they might enjoy being told this, since it seems so mysterious.

   By playing around with these spaces one can, once again, illustrate many basic ideas of linear algebra. For example:

   (a) The bilinear pairing $C^0([a, b]) \to \mathbb{R}$ defined by

   $$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

   is an inner product. It is known as the L_2 inner product and denoted $\langle f, g \rangle_{L_2}$. Proving this requires some knowledge of Analysis. But that is good - all of mathematics is interconnected. In fact, although the L_2 inner product looks very different from the standard Euclidean inner product in $\mathbb{R}^n$ defined by

   $$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle_E = x_1 y_1 + \cdots + x_n y_n,$$

   it is not really so far from it. We can see this by approximating the functions f and g by step functions. Subdivide the interval $[a, b]$ into n equal subintervals and let $f_i^{(n)}$

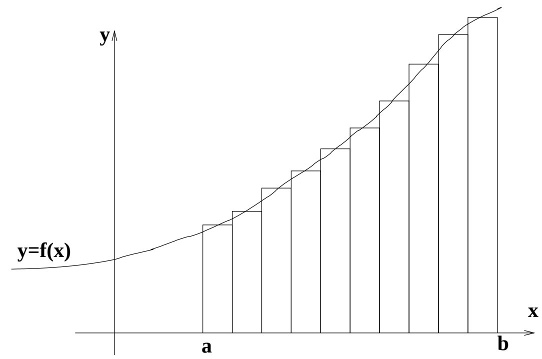


Figure 1: Approximating f by a step function

and $g_i^{(n)}$ be the values of f and g at the midppoint of the i'th of these n sub-intervals. Approximate f by the step function $f^{(n)}$ which takes the constant value $f_i^{(n)}$ on the whole of the i'th subinterval, and approximate g by the step function $g^{(n)}$ defined similarly. Then

$$\begin{aligned}
\langle f^{(n)}, g^{(n)} \rangle_{L_2} &= \int_a^b f^{(n)}(x) g^{(n)}(x) dx \\
&= \sum_{i=1}^{n} f_i^{(n)} \times g_i^{(n)} \times \frac{b-a}{n} \\
&= \frac{b-a}{n} \langle (f_1, \ldots, f_n), (g_1, \ldots, g_n) \rangle_E \qquad (3.2)
\end{aligned}$$

If f and g are both continuous then as $n \to \infty$, the approximations $f^{(n)}$ and $g^{(n)}$ converge to f and g, and $\langle f^{(n)}, g^{(n)} \rangle_{L^2}$, which by (3.2) is really a version of a Euclidean inner product, tends to $\langle f, g \rangle_{L_2}$.

(b) The infinite collection of functions $\sin x, \sin 2x, \ldots, \sin nx, \ldots, n \in \mathbb{N}$, are linearly independent —

(c) — but they are not a basis for $C^\infty(\mathbb{R})$. This may be shown by finding a linear map $\ell : C^\infty(\mathbb{R}) \to \mathbb{R}$ which kills $\sin nx$ for every $n \in \mathbb{N}$, together with some other function $f \in C^\infty(\mathbb{R})$ such that $\ell(f) \neq 0$. Any suggestions?

(d) If we add, to this collection of functions, the functions $\cos nx$ for $n \in \mathbb{N}$, then we still have a linearly independent set in $C^\infty(\mathbb{R})$. And once again it is not a basis for $C^\infty(\mathbb{R})$. Both may be shown by similar methods to the previous case.

4 Linear Maps, Bases and Matrices

One topic in Linear Algebra that causes a lot of confusion is the link between linear maps and matrices. Given vector spaces V and W, a linear map $T : V \to W$ and bases for V and W, we can write down the matrix of T with respect to these bases. But how does this matrix change if we change our bases? How can we tell if two different matrices represent the same linear transformation? In order to answer such questions, it turns out that what we need above all else is a good notation! We will develop a good notation here, and then put it through its paces.

Let V be an n-dimensional vector space and let $E := e_1, \ldots, e_n$ be a basis. Then by assigning, to $v \in V$, its n coefficients in the basis E, we define a map $V \to \mathbb{R}^n$. It is useful to have a clear and concise way of referring to the column vector formed by these coefficients. We denote it by $[v]_E$. For example, if V is the example at the start of Section 3,

$$V = \{(x, y, z) \in \mathbb{R}^3 : 2x - 4y + z = 0\}$$

and E is the basis $e_1 := (2, 1, 0), e_2 := (0, 1, 4)$, then for $v = (4, -1, -12)$ we have

$$v = 2e_1 - 3e_2$$

and therefore

$$[v]_E = \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

One of the questions we will answer is the following: if $E' = e'_1, e'_2$ is another basis for V, what is the relation between $[v]_E$ and $[v]_{E'}$?

Two important properties of the representation of a vector in a basis:

1. If $u, v \in V$ then
$$[u + v]_E = [u]_E + [v]_E. \tag{4.1}$$

2. If $v \in V$ and $\lambda \in \mathbb{R}$, then
$$[\lambda v]_E = \lambda[v]_E. \tag{4.2}$$

These two properties can be summarised by saying that the map

$$V \to \mathbb{R}^n \quad \text{defined by} \quad v \mapsto [v]_E$$

is a *linear* map of vector spaces. Of course, it is a specially nice linear map – it is an isomorphism. That is, besides being linear, it is a bijection. It is *injective* because for no two different vectors v_1 and v_2 do we have $[v_1]_E = [v_2]_E$; for if $[v_1]_E = (\lambda_1, \ldots, \lambda_n)^t = [v_2]_E$ then $v_1 = \lambda_1 e_1 + \cdots + \lambda_n e_n = v_2$. It is *surjective* because for every n-tuple $(\lambda_1, \ldots, \lambda_n)^t$ in $\mathbb{R}^n$, there is a vector $v \in V$ such that $[v]_E = (\lambda_1, \ldots, \lambda_n)^t$, namely $v = \lambda_1 e_1 + \cdots + \lambda_n e_n$.

Let V and W be vector spaces of dimension m and n, let $E = e_1, \ldots, e_m$ and $F = f_1, \ldots, f_n$ be bases for V and W, and let $T : V \to W$ be a linear map. If $v \in V$ then, in the notation introduced in the preceding paragraph, the column vector consisting of the coefficients of $T(v)$ with respect to the basis F of W is denoted $[T(v)]_F$. What is the relation between $[v]_E$ and $[T(v)]_F$? The answer, not surprisingly, is that one obtains $[T(v)]_F$ from $[v]_E$ by multiplying it by *a matrix*. Let us recall that by definition of matrix multiplication,

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & \cdots & a_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} v_1 a_{1,1} & + & \cdots & + & v_m a_{1,m} \\ v_1 a_{2,1} & + & \cdots & + & v_m a_{2,m} \\ \cdots & + & \cdots & + & \cdots \\ \cdots & + & \cdots & + & \cdots \\ v_1 a_{n,1} & + & \cdots & + & v_m a_{n,m} \end{pmatrix} = v_1 \begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ \vdots \\ a_{n,1} \end{pmatrix} + \cdots + v_m \begin{pmatrix} a_{1,m} \\ a_{2,m} \\ \vdots \\ \vdots \\ a_{n,m} \end{pmatrix}$$

If we denote by Col_i the i'th column of the matrix on the left, this can be rewritten as

$$\begin{pmatrix} \mathrm{Col}_1 & \mathrm{Col}_2 & \cdots & \mathrm{Col}_m \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ \vdots \\ v_m \end{pmatrix} = v_1 \mathrm{Col}_1 + \cdots + v_m \mathrm{Col}_m. \tag{4.3}$$

Exercise Prove the corresponding statement for the rows of a matrix: denoting by

$$\begin{pmatrix} \mathrm{Row}_1 \\ \vdots \\ \mathrm{Row}_n \end{pmatrix}$$

the matrix with rows $\mathrm{Row}_1, \ldots, \mathrm{Row}_n$, and by $\begin{pmatrix} v_1 & \cdots & \cdots & v_n \end{pmatrix}$ the row vector with entries $v_1, \ldots, v_n$, we have

$$\begin{pmatrix} v_1 & \cdots & \cdots & v_n \end{pmatrix} \begin{pmatrix} \mathrm{Row}_1 \\ \vdots \\ \mathrm{Row}_n \end{pmatrix} = v_1 \mathrm{Row}_1 + \cdots + v_n \mathrm{Row}_n \tag{4.4}$$

Let us now return to the map $T : V \to W$. Let A denote the $n \times n$ matrix with columns $[T(e_1)]_F, \ldots, [T(e_m)]_F$. Suppose that $v \in V$ and $v = v_1 e_1 + \cdots + v_m e_m$. Then

$$\begin{aligned} A\,[v]_E &= \begin{pmatrix} [T(e_1)]_F, \ldots, [T(e_m)]_F \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} = v_1 [T(e_1)]_F + \cdots + v_m [T(e_m)]_F \\ &= [v_1 T(e_1)]_E + \cdots + [v_m T(e_m)]_E \\ &= [T(v_1 e_1)]_E + \cdots + [T(v_m e_m)]_E \\ &= [T(v_1 e_1) + \cdots + T(v_m e_m)]_E \qquad = [T(v)]_E \end{aligned} \tag{4.5}$$

where the equalities are dues, respectively, to (4.3), to (4.2), to the linearity of T, to (4.1), and, finally, to the linearity of T once again. To summarise, the chain of equalities says that if A is the matrix with columns $[T(e_1)]_F, \ldots, [T(e_m)]_F$ then

$$A\,[v]_E = [T(v)]_F.$$

This equation is crucial. It says that left-multiplication by the matrix $\begin{pmatrix} [T(e_1)]_F, \ldots, [T(e_m)]_F \end{pmatrix}$ converts the expression for the vector V in the basis E to the expression for the vector $T(v)$ in the basis F. From now on we denote the matrix $\begin{pmatrix} [T(e_1)]_F, \ldots, [T(e_m)]_F \end{pmatrix}$ by $[T]_F^E$, so that the key equation relating $[v]_E$ and $[T(v)]_F$ becomes:

$$\text{Defining } \Big[\ T\ \Big]_F^E \text{ as } \begin{pmatrix} [T(e_1)]_F, \ldots, [T(e_m)]_F \end{pmatrix}, \text{ we have } \Big[\ T\ \Big]_F^E [v]_E = [T(v)]_F. \tag{4.6}$$

Example 4.1. 1. Let V be the plane $\{(x,y,z) : 2x - 4y + z = 0\}$ in $\mathbb{R}^3$. It is a 2-dimensional space. Let W be the xy-plane in $\mathbb{R}^3$, and let $T : V \to W$ be the orthogonal projection,

$$T(x,y,z) = (x,y).$$

To represent T as a matrix we need to choose a basis for V and a basis for W. Let us take, as basis for V, the two vectors $e_1 := (2,1,0), e_2 := (0,1,4)$; we will denote this basis by E. As basis for W we take the "standard" basis F consisting of the two vectors $(1,0)$ and $(0,1)$. Then

$$[T]_F^E = \begin{pmatrix}[T(e_1)]_F & [T(e_2)]_F\end{pmatrix} = \begin{pmatrix}2 & 0\\ 1 & 1\end{pmatrix}.$$

2. Let V be the space of polynomials in one variable X of degree ≤ 3, and let $T : V \to V$ be differentiation d/dx. As basis for V we take the four vectors $e_0 = 1, e_1 = X, e_2 = X^2, e_3 = X^3$. We have

$$\begin{aligned}
T(e_0) &= \tfrac{d}{dx}1 &&= 0\\
T(e_1) &= \tfrac{d}{dx}(x) &&= 1 &&= e_0\\
T(e_2) &= \tfrac{d}{dx}(x^2) &&= 2x &&= 2e_1\\
T(e_3) &= \tfrac{d}{dx}(x^3) &&= 3x^2 &&= 3e_2
\end{aligned}$$

Therefore

$$\left[\frac{d}{dx}\right]_E^E = \begin{pmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 2 & 0\\ 0 & 0 & 0 & 3\\ 0 & 0 & 0 & 0\end{pmatrix}$$

Exercise With the same space V, let $a \in \mathbb{R}$ and let $T_a : V \to V$ be the map defined by composition with a translation by a: for each polynomial $P \in V$, $T_a(P)$ is the polynomial defined by

$$T_a(P)(x) = P(x+a).$$

Write down $\left[\ T_a\ \right]_E^E$.

Exercise Let $V \subset C^\infty(\mathbb{R})$ be the two-dimensional subspace spanned by the two "vectors" cos and sin, and let E be the basis of V consisting of these two vectors. It is well-known that differentiation d/dx maps V to itself. Write down the matrices $\left[\ \frac{d}{dx}\ \right]_E^E$ and $\left[\ \frac{d^2}{dx^2}\ \right]_E^E$. What is the relation between these two matrices?

3. Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map given by anticlockwise rotation about the origin through an angle θ. Let E be the basis $e_1 := (1,0), e_2 := (0,1)$ for $\mathbb{R}^2$. Elementary trigonometry tells us that

$$R_\theta(e_1) = (\cos\theta, \sin\theta), \qquad R_\theta(e_2) = (-\sin\theta, \cos\theta) \tag{4.7}$$

so

$$[R_\theta(e_1)]_E = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \qquad [R_\theta(e_2)]_E = \begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}$$

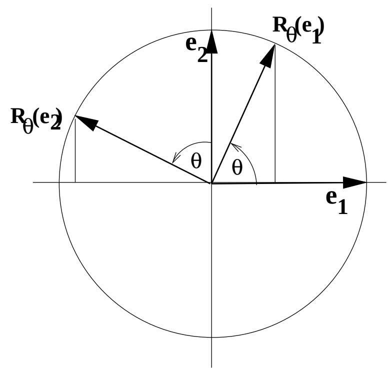


Figure 2: Effect on basis vectors of rotation through angle θ

and thus

$$\Big[R_\theta\Big]_E^E = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Exercise Find $[R_\theta]_F^F$ in case

(a) F is the basis for $\mathbb{R}^2$ consisting of the vectors $f_1 = (0, 1)$ and $f_2 = (1, 0)$.

(b) F is the basis for $\mathbb{R}^2$ consisting of the vectors $f_1 = (2, 0)$ and $f_2 = (0, 1)$.

(c) F is the basis for $\mathbb{R}^2$ consisting of the vectors $f_1 = (2, 0)$ and $f_2 = (0, 2)$.

(d) F is the orthonormal basis for $\mathbb{R}^2$ consisting of the two vectors $g_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and $g_2 = (-1/\sqrt{2}, 1/\sqrt{2})$.

4. Let $r_V : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map defined by reflection in the plane $V := \{(x, y, z) \in \mathbb{R}^3 : 2x - 4y + z = 0\}$. Then r_V leaves vectors *in* the plane unchanged, and sends vectors orthogonal to the plane to their opposite – i.e. multiplies them by -1. If E is the standard basis of $\mathbb{R}^3$ consisting of the vectors $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$, It is not at all obvious how to write down the matrix $\Big[\ r_V\ \Big]_E^E$. However,we can choose as basis for $\mathbb{R}^3$ one better adapted to the situation: take two vectors in V, for example $f_1 = (2, 1, 0)$ and $f_2 = (0, 1, 4)$, and one vector orthogonal to the plane, such as $f_3 = (2, -4, 1)$. Then the collection $F := \{f_1, f_2, f_3\}$ is a basis for $\mathbb{R}^3$. Since

$$r_V(f_1) = f_1, \quad r_V(f_2) = f_2 \quad r_V(f_3) = -f_3,$$

we have

$$\Big[\ r_V\ \Big]_F^F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

5. Let $R_{L,\theta} : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map given by rotation through an angle θ about the line L spanned by the vector $(2, -4, 1)$. To remove the ambiguity of the direction in which the rotation should take place, we specify that the rotation should be anticlockwise if viewed from the tip of the vector $(2, -4, 1)$ looking towards the origin. Once again, it is hard at this stage to write down the matrix of $R_{L,\theta}$ with respect to the standard basis of $\mathbb{R}^3$. But if we choose a basis adapted to the situation, it is easy. As before, let V be the plane orthogonal to the vector $(2, -4, 1)$. Take an *orthonormal* basis for V consisting of two unit vectors f_1, f_2, and let $f_3 = (2, 4, -1)$. Make sure

that viewed from the tip of $(2,-4,1)$, the shortest rotation from f_1 to f_2 should be anticlockwise.

Exercise How can we tell if this is the case simply by looking at the three vectors f_1, f_2, f_3?

Then $R_{L,\theta}(f_1) = \cos\theta f_1 + \sin\theta f_2$ and $R_{L,\theta}(f_2) = -\sin\theta f_1 + \cos\theta f_2$, exactly as in (4.7), while $R_{L,\theta}(f_3) = f_3$. Thus, taking F to be the basis f_1, f_2, f_3,

$$\Big[\ R_{L,\theta}\ \Big]_F^F = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

4.1 Composition of linear maps and multiplication of matrices

Now we consider the effect of carrying out two linear maps in succession. Suppose we are given spaces U, V and W, linear maps $S : U \to V$ and $T : V \to W$, and bases E, F, G for U, V, W respectively. What is the relation between $\Big[\ S\ \Big]_F^E$, $\Big[\ T\ \Big]_G^F$ and $\Big[\ T \circ S\ \Big]_G^E$? The answer is that

$$\Big[\ T \circ S\ \Big]_G^E = \Big[\ T\ \Big]_G^F \Big[\ S\ \Big]_F^E, \tag{4.8}$$

an equation that will be very useful. To prove it, suppose that $u \in U$. We have

$$\begin{aligned} \Big(\Big[\ T\ \Big]_G^F \Big[\ S\ \Big]_F^E\Big)\,[u]_E &= \Big[\ T\ \Big]_G^F \Big(\Big[\ S\ \Big]_F^E [u]_E\Big) \\ &= \Big[\ T\ \Big]_G^F \big([S(u)]_F\big) \\ &= \big[T(S(u))\big]_G \end{aligned} \tag{4.9}$$

where the three equalities are due, respectively, to the associativity of matrix multiplication, and to (4.6) applied twice in succession. Now applying (4.6) directly to the linear map $T \circ S$, we have

$$\Big[\ T \circ S\ \Big]_G^E [u]_E = [T \circ S(u)]_G. \tag{4.10}$$

Comparing (4.9) and (4.10) we see that the product $\Big[\ T\ \Big]_G^F \Big[\ S\ \Big]_F^E$ does exactly the same thing to every column vector as does the matrix $\Big[\ T \circ S\ \Big]_G^E$, and conclude that $\Big[\ T\ \Big]_G^F \Big[\ S\ \Big]_F^E = \Big[\ T \circ S\ \Big]_G^E$. Thus, (4.8) is proved.

Example 4.2. Let R_θ be rotation through an angle θ about the origin in $\mathbb{R}^2$. We saw in Example 4.1(3) that

$$\Big[\ R_\theta\ \Big]_E^E = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where E is the basis $E = (1,0), (0,)$ for $\mathbb{R}^2$. If α and β are any two angles, it is clear that $R_\alpha \circ R_\beta = R_{\alpha+\beta}$. By (4.8) we know that

$$\Big[\ R_\alpha \circ R_\beta\ \Big]_E^E = \Big[\ R_\alpha\ \Big]_E^E \Big[\ R_\beta\ \Big]_E^E.$$

It follows that

$$\begin{pmatrix}\cos\,\alpha+\beta & -\sin\,\alpha+\beta\\ \sin\,\alpha+\beta & \cos\,\alpha+\beta\end{pmatrix} = \begin{pmatrix}\cos\,\alpha & -\sin\,\alpha\\ \sin\,\alpha & \cos\,\alpha\end{pmatrix}\begin{pmatrix}\cos\,\beta & -\sin\,\beta\\ \sin\,\beta & \cos\,\beta\end{pmatrix}$$

$$= \begin{pmatrix}\cos\,\alpha\cos\,\beta-\sin\,\alpha\sin\,\beta & -\big(\cos\,\alpha\sin\,\beta+\sin\,\alpha\cos\,\beta\big)\\ \sin\,\alpha\cos\,\beta+\cos\,\alpha\sin\,\beta & \cos\,\alpha\cos\,\beta-\sin\,\alpha\sin\,\beta\end{pmatrix}$$

Of course we all recognise the trigonometric equalities

$$\cos\,\alpha+\beta=\cos\,\alpha\cos\,\beta-\sin\,\alpha\sin\,\beta \qquad \text{and} \qquad \sin\,\alpha+\beta=\sin\,\alpha\cos\,\beta+\cos\,\alpha\sin\,\beta.$$

Here we have deduced them from our key formula (4.8).

4.2 Using the key formula (4.8) to change basis

Suppose that E and F are two bases for the same vector space V, and that $v \in V$. What is the relation between $[v]_E$ and $[v]_F$? We can use (4.6) to answer this question easily. Let us guess that the two column vectors are related by a matrix:

$$X[v]_E = [v]_F \tag{4.11}$$

for some unknown matrix X. What could X be? Compare (4.11) with (4.6): if T is any linear map, then to convert $[v]_E$ to $[T(v)]_F$, I must multiply it on the left by $\Big[\ T\ \Big]_F^E$. In (4.11) there is apparently no linear map T; but as mathematicians, we know that this need not stop us. Let us simply take, as T in (4.6), the identity map $I : V \to V$ – that is, the map which leaves every vector unchanged. Then (4.6) becomes

$$\Big[\ I\ \Big]_F^E [v]_E = [I(v)]_F \tag{4.12}$$

– in other words, since $I(v) = v$, we get

$$\Big[\ I\ \Big]_F^E [v]_E = [v]_F \tag{4.13}$$

So we have found the unknown matrix X of (4.11), and we have a simple and foolproof formula, (4.13), for changing basis.

Notice that by (4.6), if V is n-dimensional then

$$\Big[\ I\ \Big]_F^E = \Big([I(e_1)]_F \quad \cdots \quad [I(e_n)]_F\Big)$$

$$= \Big([e_1]_F \quad \cdots \quad [e_n]_F\Big)$$

Although $\Big[\ I\ \Big]_F^E$ is the matrix of the identity *map*, it will not be the identity *matrix* unless the two bases E and F are equal. If $E = F$ then indeed

$$[e_i]_F = [e_i]_E = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{(1 in the } i\text{'th place)}$$

and so

$$\Big[\ I\ \Big]_E^E = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

is the identity matrix. It follows that

$$\Big[\ I\ \Big]_F^E \Big[\ I\ \Big]_E^F = \Big[\ I\ \Big]_F^F = \text{identity matrix} \tag{4.14}$$

and thus

$$\Big[\ I\ \Big]_E^F = \left(\Big[\ I\ \Big]_F^E\right)^{-1} \tag{4.15}$$

Two more equalities are now easily deduced: first, if $T : V \to W$ is a linear map, E, E' are two distinct bases for V, and F, F' are distinct bases for W, then

$$\Big[\ T\ \Big]_{F'}^{E'} = \Big[\ I\ \Big]_{F'}^F \Big[\ T\ \Big]_F^E \Big[\ I\ \Big]_E^{E'} \tag{4.16}$$

and, as a special case, if $V = W$, $E = F$ and $E' = F'$, then

$$\Big[\ T\ \Big]_{E'}^{E'} = \Big[\ I\ \Big]_{E'}^E \Big[\ T\ \Big]_E^E \Big[\ I\ \Big]_E^{E'} = \left(\Big[\ I\ \Big]_E^{E'}\right)^{-1} \Big[\ T\ \Big]_E^E \Big[\ I\ \Big]_E^{E'}. \tag{4.17}$$

Formula (4.17) is especially important.

Theorem 4.3. *If V is an n-dimensional vector space, $T : V \to V$ is a linear map, and E is a basis for V, then* $\det\Big[\ T\ \Big]_E^E$ *is independent of the choice of E, and thus is a property of the linear map T itself.*

Proof. The determinant of a product of matrices is the product of their determinants (this will be proved in Theorem 5.1 in the next section). From this it follows that if E and E' are any two bases then

$$\det\Big[\ T\ \Big]_{E'}^{E'} = \det\Big[\ I\ \Big]_{E'}^E \det\Big[\ T\ \Big]_E^E \det\Big[\ I\ \Big]_E^{E'}. \tag{4.18}$$

Because $\Big[\ I\ \Big]_{E'}^{E}$ is the inverse of $\Big[\ I\ \Big]_{E}^{E'}$, it follows that

$$\det\Big[\ I\ \Big]_{E'}^{E} = \left(\det\Big[\ I\ \Big]_{E'}^{E'}\right)^{-1}$$

and thus, from (4.18), that

$$\det\Big[\ T\ \Big]_{E'}^{E'} = \det\Big[\ T\ \Big]_{E}^{E}.$$

□

If this determinant is a property of T itself, independent of the choice of basis, do we really need a basis to compute it? Does it have a meaning independently of reference to a basis? We will answer these questions in the next section.

Example 4.4. We use (4.17) to solve a problem that was left unresolved in Example 4.1(4). Given a plane V in $\mathbb{R}^3$, what is the matrix of reflection in V with respect to the basis $E = (1, 0,), (0, 1, 0), (0, 0, 1)$? Recall that by definition, reflection in V, r_V, is the map that leaves all vectors of V fixed, and sends each vector orthogonal to V to -1 times itself. We saw that if we choose a basis F adapted to the situation, whose first two vectors lie in V and whose third vector is orthogonal to V, then

$$\Big[\ r_V\ \Big]_{F}^{F} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Now using (4.17), we have

$$\Big[\ r_V\ \Big]_{E}^{E} = \Big[\ I\ \Big]_{E}^{F}\Big[\ r_V\ \Big]_{F}^{F}\Big[\ I\ \Big]_{F}^{E}.$$

Which of the two matrices $\Big[\ I\ \Big]_{F}^{E}$ and $\Big[\ I\ \Big]_{E}^{F}$ is easier to find? The answer is that if we have specified the three vectors of the basis F in terms of the standard basis of $\mathbb{R}^3$, as in Example 4.1(4), then we already know the matrix $\Big[\ I\ \Big]_{E}^{F}$. In Example 4.1(4), $f_1 = (2, 1, 0)$, $f_2 = (0, 1, 4)$, and $f_3 = (2, -4, 1)$, so that

$$\Big[\ I\ \Big]_{E}^{F} = \begin{pmatrix} 2 & 0 & 2 \\ 1 & 1 & -4 \\ 0 & 4 & 1 \end{pmatrix}.$$

It follows (by a calculation which we do not go through here) that

$$\Big[\ I\ \Big]_{F}^{E} = \left(\Big[\ I\ \Big]_{E}^{F}\right)^{-1} = \frac{1}{42}\begin{pmatrix} 17 & 8 & -2 \\ -1 & 2 & 10 \\ 4 & -8 & 2 \end{pmatrix}$$

Thus, by (4.17),

$$\Big[\ r_V\ \Big]_{E}^{E} = \begin{pmatrix} 2 & 0 & 2 \\ 1 & 1 & -4 \\ 0 & 4 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} 17/42 & 8/42 & -2/42 \\ -1/42 & 2/42 & 10 \\ 4/42 & -8/42 & 2/42 \end{pmatrix} = \begin{pmatrix} 13/21 & 16/21 & -4/21 \\ 16/21 & -11/21 & 8/21 \\ -4/21 & 8/21 & 19/21 \end{pmatrix}$$

Example 4.5. Another problem left unresolved in Example 4.1(5): what is the matrix, with respect to the standard basis of $\mathbb{R}^3$, of rotation through an angle θ about the axis spanned by the vector $(2,-4,1)$? In the example, we saw that if we chose an orthonormal basis f_1, f_2 for the plane V perpendicular to $(2,-4,1)$, and then took $f_3 = (2,-4,1)$, and let F be the basis f_1, f_2, f_3 of $\mathbb{R}^3$, then

$$\Big[\ R_{L,\theta}\ \Big]_F^F = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let us choose $f_1 = \frac{1}{\sqrt{41}}(1,-2,6), f_2 = \frac{1}{\sqrt{5}}(2,1,0)$. Then

$$\begin{aligned} \Big[\ R_{L,\theta}\ \Big]_E^E &= \Big[\ I\ \Big]_E^F \Big[\ R_{L,\theta}\ \Big]_F^F \Big[\ I\ \Big]_F^E \\ &= \begin{pmatrix} \frac{1}{\sqrt{105}} & \frac{2}{\sqrt{5}} & 2 \\ \frac{-2}{\sqrt{105}} & \frac{1}{\sqrt{5}} & -4 \\ \frac{-10}{\sqrt{105}} & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{105}} & \frac{2}{\sqrt{5}} & 2 \\ \frac{-2}{\sqrt{105}} & \frac{1}{\sqrt{5}} & -4 \\ \frac{-10}{\sqrt{105}} & 0 & 1 \end{pmatrix}^{-1} \end{aligned}$$

At this point we could compute the inverse matrix that we need, as we did (without spelling out he calculation) in the previous example. But finding the inverse of a matrix is usually rather tedious. It turns out that by making a very minor modification to the basis F, we can avoid all this hard work. The reason is this: the three vectors of the basis F are orthogonal to one another. The first two have length 1, but the third does not. If the third also had length 1, then the inverse of the matrix $\Big[\ I\ \Big]_E^F$ would just be the transpose of $\Big[\ I\ \Big]_E^F$. This is explained below. So let us replace $(2,-4,1)$ by $\frac{1}{\sqrt{21}}(2,-4,1)$. Then

$$\Big[\ R_{L,\theta}\ \Big]_E^E = \begin{pmatrix} \frac{1}{\sqrt{105}} & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{21}} \\ \frac{-2}{\sqrt{105}} & \frac{1}{\sqrt{5}} & \frac{-4}{\sqrt{21}} \\ \frac{-10}{\sqrt{105}} & 0 & \frac{1}{\sqrt{21}} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{105}} & \frac{-2}{\sqrt{105}} & \frac{-10}{\sqrt{105}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{21}} & \frac{-4}{\sqrt{21}} & \frac{1}{\sqrt{21}} \end{pmatrix}. \tag{4.19}$$

We say a matrix A is *orthogonal* if its columns all have length 1, and are mutually orthogonal.

Proposition 4.6. *A is orthogonal if and only if A^{-1} is equal to the transpose of A, A^t.*

Proof. Suppose that A is orthogonal. Write A in the form $A = (\mathrm{Col}_1 \cdots \mathrm{Col}_n)$. Then

$$A^t\ A = \begin{pmatrix} (\mathrm{Col}_1)^t \\ \vdots \\ (\mathrm{Col}_n)^t \end{pmatrix} (\mathrm{Col}_1 \cdots \mathrm{Col}_n)$$

The entry in row i and column j of the product is equal to

$$(\mathrm{Col}_i)^t\ \mathrm{Col}_j.$$

This is the same as the inner product

$$\mathrm{Col}_i \cdot \mathrm{Col}_j$$

and thus is equal to 1 if $i = j$ and to 0 otherwise. So $A^t\ A$ is the identity matrix.

The converse is proved similarly: if A^tA is the identity matrix then for all i, j, $C_i \cdot C_j = 0$ if $i \neq j$ and $C_i \cdot C_i = 1$. So A is orthogonal. $\square$

The set of all orthogonal $n \times n$ matrices with real entries is a group under multiplication, the *orthogonal group*. It is denoted $O(n)$. It plays a central role in a great deal of mathematics! It is not only a group, but a smooth manifold, for which moreover the group operations of multiplication and inversion are smooth maps; groups with this property are known as *Lie groups*.

5 Determinant and Volume

The determinant of a square matrix

$$A := \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & \cdots & \alpha_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & \cdots & a_{n,n} \end{pmatrix}$$

is defined by a rather indigestible formula involving permutations of the numbers $1, \ldots, n$:

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}, \tag{5.1}$$

where S_n is the group of permutations of n objects. From the definition it is easy to deduce the following properties, which will be our starting point:

1. det is additive on each column individually:

$$\det \begin{pmatrix} \mathrm{Col}_1 & \cdots & \mathrm{Col}_i + \mathrm{Col}'_i & \cdots & \mathrm{Col}_n \end{pmatrix} =$$
$$\det \begin{pmatrix} \mathrm{Col}_1 & \cdots & \mathrm{Col}_i & \cdots & \mathrm{Col}_n \end{pmatrix} + \det \begin{pmatrix} \mathrm{Col}_1 & \cdots & \mathrm{Col}'_i & \cdots & \mathrm{Col}_n \end{pmatrix}$$

2. If all the elements in one column are multiplied by the scalar λ, the value of the determinant is multiplied by λ.

3. If two columns are equal then the determinant is zero.

4. det is additive on each row individually:

$$\det \begin{pmatrix} \mathrm{Row}_1 \\ \vdots \\ \mathrm{Row}_i + \mathrm{Row}'_i \\ \vdots \\ \mathrm{Row}_n \end{pmatrix} = \det \begin{pmatrix} \mathrm{Row}_1 \\ \vdots \\ \mathrm{Row}_i \\ \vdots \\ \mathrm{Row}_n \end{pmatrix} + \det \begin{pmatrix} \mathrm{Row}_1 \\ \vdots \\ \mathrm{Row}'_i \\ \vdots \\ \mathrm{Row}_n \end{pmatrix}$$

5. If all the elements in one row are multiplied by the scalar λ, the value of the determinant is multiplied by λ.

6. If two rows are equal the the determinant is zero.

7. If any column is added to any other, the value of the determinant is unchanged.

8. If any row is added to any other, the value of the determinant is unchanged.

9. The determinant of the identity matrix is equal to 1.

In fact these properties completely characterise the determinant.

Exercises

1. From property 4, prove that the determinant of a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$ along the diagonal is equal to $\lambda_1 \cdots \lambda_n$.

2. Prove that Properties 1, 2 and 3 imply Property 7.

3. Recall that by means of column operations, any $n \times n$ matrix can be transformed either into a diagonal matrix or into a matrix with a column of zeros. Use this to show that given any function $D : \{n \times n \text{ matrices }\} \to \mathbb{R}$ with properties 1, 3, 5 and 9, $D(A) = \det(A)$ for every $n \times n$ matrix A.

4. Prove that if $D : \{n \times n \text{ matrices }\} \to \mathbb{R}$ has properties 1, 3, and 5, but not necessarily property 9, then for all $n \times n$ matrices A,

$$D(A) = \det(A)D(I_n)$$

   where I_n is the identity matrix.

Many textbooks on Linear Algebra introduce the determinant "axiomatically": they show that any two functions $\{n \times n \text{ matrices }\} \to \mathbb{R}$ with properties 1,2,3 and 9 must be equal to one another ("the determinant function, if it exists, is unique"), and then show that the function defined by formula (5.1) has these properties ("the determinant function exists"). This approach has many advantages. In particular, it avoids beginning the topic by confronting the student with the rather complicated and off-putting formula (5.1), which seems to have appeared from nowhere. It is certainly useful to be able to characterise the determinant by some of its properties, as the following proof shows.

Theorem 5.1. *If A and B are square matrices then* $\det(AB) = \det(A)\det(B)$.

Proof. Let A be a fixed $n \times n$ matrix, and consider the function

$$D_A : \{n \times n \text{ matrices }\} \to \mathbb{R}$$

defined by

$$D_A(B) = \det(AB).$$

Let $B_1, \ldots, B_n$ be the columns of B. The columns of AB are equal to $AB_1, \ldots, AB_n$. It follows that

1. If B_i is a sum of columns B_i' and B_i'' then the i'th column of AB is equal to $AB_i' + AB_i''$.

2. If two columns of B are equal then two columns of AB are equal.

3. If we multiply a column of B by a scalar λ then we multiply by λ the corresponding column of AB.

Therefore:

1. Because det is additive on each column individually (Property 1 of det),

$$\begin{aligned} &D_A \begin{pmatrix} B_1 & \cdots & B_i + B_i' & \cdots & B_n \end{pmatrix} \\ &\quad = \det \begin{pmatrix} AB_1 & \cdots & AB_i + AB_i' & \cdots & AB_n \end{pmatrix} \\ &\quad\quad = \det \begin{pmatrix} AB_1 & \cdots & AB_i & \cdots & AB_n \end{pmatrix} + \det \begin{pmatrix} AB_1 & \cdots & AB_i' & \cdots & AB_n \end{pmatrix} \\ &\quad\quad\quad = D_A \left(\begin{pmatrix} B_1 & \cdots & B_i & \cdots & B_n \end{pmatrix}\right) + D_A \begin{pmatrix} B_1 & \cdots & B_i' & \cdots & B_n \end{pmatrix} \end{aligned}$$

   So D_A has Property 1 also.

2. Because the det of a matrix with two equal columns is equal to 0 (Property 3 of det), it follows that if B has two equal columns then $D_A(B) = 0$. So D_A has Property 3 also.

3. If all the elements in one column of B are multiplied by the scalar λ, then the value of $D_A(B)$ is multiplied by λ. So D_A has Property 5 also.

By Exercise 4 above, it follows that

$$D_A(B) = \det(B) D_A(I_n)$$

where I_n is the identity matrix. Since $D_A(I_n)$ is equal to $\det A$, the theorem is proved. □

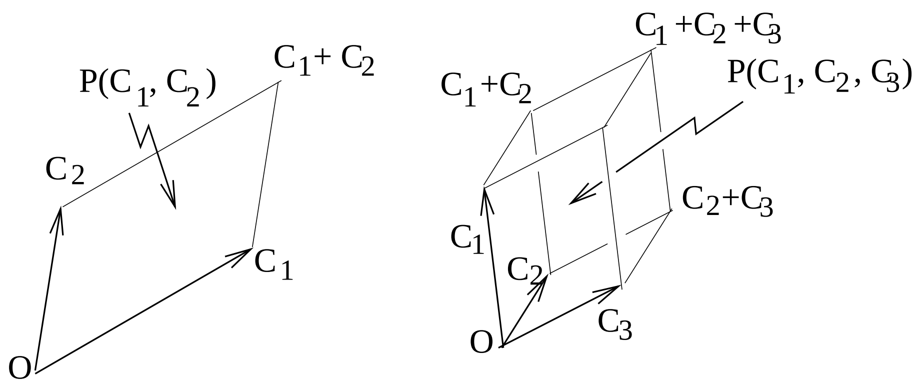
But the most important thing missing from most accounts of the determinant is any geometrical meaning.

Theorem 5.2. *The absolute value of the determinant of a matrix with columns $C_1, \ldots, C_n$ is equal to the n-dimensional volume, with respect to the standard Euclidean metric on $\mathbb{R}^n$, of the parallelipiped $P(C_1, \ldots, C_n)$ in $\mathbb{R}^n$ spanned by the vectors $C_1, \ldots, C_n$.*

Two things need clarifying before we prove the theorem. The first is the meaning of the term "parallelipiped", and the second is the meaning of the term "volume".

Parallelipipeds

The following two drawings show the parallelipiped $P(C_1, \ldots, C_n)$ in the cases $n = 2$ and $n = 3$



In general, $P(C_1, \ldots, C_n)$ is the set of points in $\mathbb{R}^n$ which can be written as a sum $\lambda_1 C_1 + \cdots + \lambda_n C_n$, where $0 \leq \lambda_i \leq 1$ for all i.

Volume

When we say "volume with respect to the standard Euclidean metric", we mean that the usual Euclidean metric gives us our notion of the length of a line-segment (or vector), the area of a plane region, the volume of a solid, etc. Developing this notion from scratch would take us too far afield, but let us just point out that we are using the Euclidean metric when we say that the length of the vector $(1, 0 \ldots, 0)$ is 1, the length of $(v_1, \ldots, v_n)$ is $\sqrt{v_1^2 + \cdots + v_n^2}$, or the area of a triangle is half base times height. In case this is not clear, think for a moment of one of the non-standard examples of vector space described in Section 3. For instance, consider the function 'sin' in the vector space $C^\infty(\mathbb{R})$ (whose members are infinitely differentiable functions on $\mathbb{R}$). What is the length of this vector? Or consider the parallelipiped in $C^\infty(\mathbb{R})$ spanned by the two functions 'cos' and 'sin'. What is its area? And what is the angle between the vectors cos and sin? The questions have no meaning, *because we have not agreed a means of measuring length, area or angle in* $C^\infty(\mathbb{R})$. In our proof of Theorem 5.2, we implicitly use the Euclidean volume. This begins by fixing the n-dimensional volume of the parallelipiped spanned by the vectors $(\lambda_1, 0, \ldots, 0), (0, \lambda_2, 0, \ldots, 0), \ldots, (0, \ldots, 0, \lambda_n)$ as $\lambda_1 \cdots \lambda_n$, and continues by agreeing that the volume of any solid is invariant under Euclidean motions, that volume is additive over disjoint unions, and so on. Our intuition takes care of all of this when we look at objects in $\mathbb{R}^2$ or $\mathbb{R}^3$. In higher dimensions some care is needed!

Incidentally, one of the benefits of giving students non-standard examples of vector spaces, such as $C^\infty(\mathbb{R})$, is that they can be used to disentangle metric properties of $\mathbb{R}^n$ from its purely algebraic properties as vector space, as we tried to do just now. In mathematics, abstraction is the process of separating things that are tangled together in examples. Abstraction is helped by giving a wide range of examples, and encouraging students to make up their own. By seeing what is common to all the examples, and separating it from what is present only in some of the examples, students can focus on the key ideas we are trying to emphasise. And they will appreciate the importance and significance of the ideas, once they have seen them at work in a wide range of examples.

We will denote the n-dimensional volume of the parallelipiped $P(C_1, \ldots, C_n)$ by $\|P(C_1, \ldots, C_n)\|$.

Theorem 5.2 can be proved in a number of different ways. The simplest that I know relies on the generalisation, for an n-dimensional parallelipiped, of the formula for the area of a plane parallelogram:

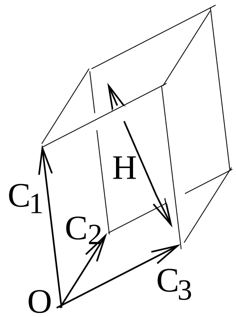
$$\text{Area } = \text{Length of Base } \times \text{ Height.} \tag{5.2}$$

The generalisation is

$$\text{Volume} = (n-1) - \text{dimensional volume of base} \times \text{Height} \tag{5.3}$$

For example, in the following picture we take, as the base, the parallelogram $P(C_2, C_3)$. The height is then the perpendicular distance from the base to the opposite face of the parallelipiped. In the diagram it is the length of the arrow marked H. Equation (5.3) says that the *volume* of the parallelipiped $P(C_1, C_2, C_3)$ is the *area* of $P(C_2, C_3)$, multiplied by the length of H. We could of course take any other of the faces of $P(C_1, C_2, C_3)$ as "base", and redefine its "height" accordingly.

Notice the following important fact: if we add to C_1 a multiple of C_2 or C_3 then the height of the parallelipiped does not change. In the diagram, the height is the perpendicular distance between the base $P(C_2, C_3)$ and the opposite face, which for a moment we will call the "roof". If we add, to C_1, any multiple of C_2 or C_3, then the roof slides in a direction parallel to the added vector, but its perpendicular distance from the base does not change.



Proof of 5.2 using (5.3)*:* We first deal with the case where the n vectors $C_1, \ldots, C_n$ are linearly dependent: there are scalars $\lambda_1, \ldots, \lambda_n$ not all zero, such that

$$\lambda_1 C_1 + \cdots + \lambda_n C_n = 0.$$

If $\lambda_i \neq 0$ we can divide through by λ_i and express C_i as a linear combination of the remaining columns:

$$C_i = -\frac{\lambda_1}{\lambda_i} C_1 - \cdots - \frac{\lambda_{i-1}}{\lambda_i} C_{i-1} - \frac{\lambda_{i+1}}{\lambda_i} C_{i+1} - \cdots - \frac{\lambda_n}{\lambda_i} C_n.$$

Subtracting from C_i these multiples of the remaining columns therefore gives us a matrix with a column of zeros, whose determinant is evidently equal to zero. As these column operations do not change the value of det, we conclude that also $\det(C_1, \ldots, C_n) = 0$. Now take as "base" of the parallelipiped, the face $P(C_1, \ldots, \widehat{C_i}, \ldots, C_n)$ (here the $\widehat{\ }$ over C_i means "omit from the list"). The "height" is the component of C_i orthogonal to this face. But C_i is a linear combination of $C_1, \ldots \widehat{C_i}, \ldots, C_n$, so this component is equal to 0.

We have seen that if $C_1, \ldots, C_n$ are linearly dependent then $\det(C_1 \cdots Cn) = 0 = \|P(C_1, \ldots, C_n)\|$. So in this case the determinant and the volume are equal.

Now assume that $C_1, \ldots, C_n$ are linearly independent.

1. We can transform $P(C_1, \ldots, C_n)$ into a rectanguloid $P(C_1', \ldots, C_n')$ with edges along the coordinate axes, by repeatedly adding multiples of one column to another. This is just the statement that we can transform a square matrix of maximal rank to a diagonal matrix by these column operations, which is well known, and easy to see.

2. These operations do not change the value of the determinant, by Property 7 of determinants.

3. Neither do they change the volume of the parallelipiped. For suppose that we have chosen, as base of $P(C_1, \ldots, C_n)$, the $(n-1)$-dimensional parallelipiped $P(C_1, \ldots, \widehat{C_i}, \ldots, C_n)$. Then the "height" is the component of C_i orthogonal to the face $P(C_1, \ldots, \widehat{C_i}, \ldots, C_n)$. Adding a multiple of C_j to C_i does not change this component. Thus, by (5.3) the operation "add a multiple of C_j to C_i" does not change the volume of the parallelipiped.

4. The n-dimensional volume of a rectanguloid is equal to the product of the lengths of its edges. So if $C_1' = (\lambda_1, 0, \ldots, 0), C_2' = (0, \lambda_2, 0, \ldots, 0), \ldots, C_n' = (0, \ldots, 0, \lambda_n)$ then

$$\|P(C_1', \ldots, C_n')\| = |\lambda_1||\lambda_2| \cdots |\lambda_n| = |\det(C_1' \cdots C_n')|.$$

Now we have

$$\|P(C_1, \ldots, C_n)\| = \|P(C_1', \ldots, C_n')\| = |\det(C_1' \cdots C_n')| = |\det(C_1 \cdots C_n)|$$

and the proof is complete.

The formula (5.3) can be easily proved by multiple integration. This is, perhaps, why Theorem 5.2 is rarely proved in an undergraduate linear algebra course – its simplest proof relies on a technique that is not developed until the later part of the degree. This is a pity, because it gives geometrical meaning to what is otherwise likely to be seen only an abstract, algebraic notion.

Recall from theorem 4.3 that the determinant $\det\Big[\,T\,\Big]_E^E$ of a linear map is independent of the choice of basis E. We will refer to it just as $\det T$.

Theorem 5.3. *If $W \subset \mathbb{R}^n$ is measurable and $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map then*

$$vol\, T(W) = \det T \;\; vol\, W.$$

The theorem is proved by showing successively:

1. It holds if W is the unit cube $P\big((1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\big)$. For if E is the standard basis for $\mathbb{R}^n$ and $\Big[\,T\,\Big]_E^E = \big(C_1 \cdots C_n\big)$, then $T(W) = P(C_1, \ldots, C_n)$.
2. It holds if W is any cube with edges along the coordinate axes,.
3. It holds if W is a translation of any cube with edges along the coordinate axes.
4. It holds for any set which can be well approximated by unions of cubes (this is what we mean by "measurable").

This, then, is the geometrical meaning of the determinant (or, more precisely, the meaning of the absolute value of the determinant): it is the scale factor by which a linear map multiplies volume. This is the key ingredient in the proof of the theorem of change of variable in multiple integration. Since this proof is quite easy, provided one is willing to regard the integral as the limit of a sum and not ask for too much detail about the precision of approximations, we explain it in the next section.

What we have not done is find a meaning for the *sign* of the determinant, and we will not try to do this here. We remark only that in it is the sign of the determinant of the change of basis matrix

6 Change of variable in multiple integration

As an application of Theorem 5.2, we give a sketch proof of the formula for change of variable in multiple integration. We want to show that ideas from linear algebra play a key role in many branches of mathematics.

Theorem 6.1. *Suppose that U is a region in $\mathbb{R}^m$, $h : U \to h(U) \subset \mathbb{R}^m$ is a diffeomorphism, and $f : U \to \mathbb{R}$ is an integrable function. Then*

$$\int_{h(U)} f = \int_U (f \circ h)|J(h)|$$

where $J(h) = \det[dh]$.

The following sketch proof assumes familiarity with the notion of the differential of a map h of class C^1 from an open set $U \subset \mathbb{R}^n$ to $\mathbb{R}^p$. Recall that for each $x \in U$, the differential $d_x h$ of h at x gives the best linear approximation to h in the neighbourhood of x, in the sense that

1. $d_x h$ is a linear map $\mathbb{R}^n \to \mathbb{R}^p$, and
2. For $v \in \mathbb{R}^n$ small enough so that $x + v$ is still in U,

   $$h(x+v) = h(x) + d_x h(v) + E_x(v)$$

   where the error term $E_x(v)$ tends to 0 more quickly than v:

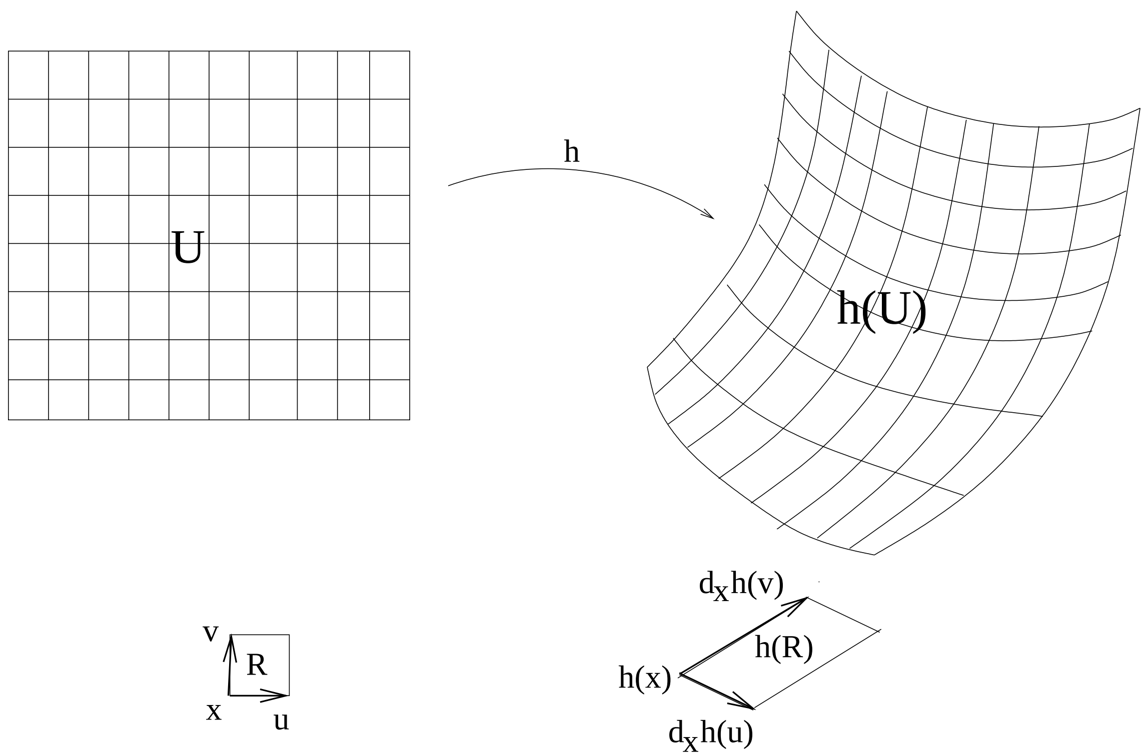
   $$\lim_{\|v\| \to 0} \frac{E_x(v)}{\|v\|} = 0.$$

The next section gives some background on this. In particular, it shows that the matrix of $d_x h$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^p$ is

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial h_p}{\partial x_1} & \cdots & \frac{\partial h_p}{\partial x_n} \end{pmatrix} \tag{6.1}$$

where all partial derivatives are calculated at x.

Sketch proof of Theorem 6.1.



The diagram shows a partition of U (drawn as a rectangle, though this is not necessary) into rectangles R_i. For any function g on U, the integral $\int_U g$ is defined by means of such partitions. If the supremum of g on R_i is achieved at x_i,

$$\sum_i g(x_i) \text{ vol } (R_i) \tag{6.2}$$

is an "approximation from above" to the value of $\int_U g$; the integral itself is the infimum, over all partitions, of these approximations from above. (Of course this is not how one *calculates* it, in general.) Now consider the function $g = f \circ h$. Given any partition of U, by means of the diffeomorphism h we obtain a partition of $h(U)$ (though not into rectangles). Suppose that on R_i, $f \circ h$ achieves its supremum at x_i. Let $y_i = h(x_i)$. Then on $h(R_i)$, f achieves its supremum at y_i. Consider the sum

$$\sum_i f(y_i) \text{ vol } (h(R_i)) \tag{6.3}$$

where $y_i = h(x_i)$. This is an approximation from above of $\int_{h(U)} f$. Suppose $R_i = x_i + P(v_1, \ldots, v_m)$. If R_i is very small,

$$h(R_i) \overset{\text{approx}}{=} y_i + P\big(d_{x_i}h(v_1), \ldots, d_{x_i}h(v_m)\big). \tag{6.4}$$

The parallelipiped on the right has volume $|\det\big(d_{x_i}h(v_1) \quad \cdots \quad d_{x_i}h(v_m)\big)|$. We have

$$\big(d_{x_i}h(v_1) \quad \cdots \quad d_{x_i}h(v_m)\big) = \Big[\ d_{x_i}h\ \Big]_E^E \big(v_1 \quad \cdots \quad v_m\big)$$

By Theorem 5.1, the determinant of the matrix on the left is the product of the determinants of the two matrices on the right. We have

$$\left|\det\Big[\ d_{x_i}h\ \Big]_E^E \det\big(v_1 \quad \cdots \quad v_m\big)\right| = \left|\det\Big[\ d_{x_i}h\ \Big]_E^E\right| \times \|P(v_1, \ldots, v_m)\| = \left|\det\Big[\ d_{x_i}h\ \Big]_E^E\right| \text{vol}(R_i).$$

Therefore

$$\text{vol } (h(R_i)) \overset{\text{approx}}{=} \left|\det\Big[\ d_{x_i}h\ \Big]_E^E\right| \times \text{ vol } (R_i)$$

and

$$\sum_i f(y_i) \text{ vol } (h(R_i)) \overset{\text{approx}}{=} \sum_i f(h(x_i)) \text{ vol } (R_i) \left|\det\Big[\ d_{x_i}h\ \Big]_E^E\right|. \tag{6.5}$$

It is usual to refer to $\det\Big[\ d_x h\ \Big]_E^E$ as the jacobian determinant of h, and denote it by $J(h)$. So our formula becomes

$$\sum_i f(y_i) \text{ vol } (h(R_i)) \overset{\text{approx}}{=} \sum_i f(h(x_i)) \text{ vol } (R_i) \left|J(h)\right|. \tag{6.6}$$

As we take finer and finer partitions, the left hand side tends to $\int_{h(U)} f$, and the right hand side tends to $\int_U (f \circ h)|J(h)|$. The theorem is proved. □

7 The derivative of a C^1 map

This section is not really about linear algebra, but about the way that interesting linear maps arise in multi-variable calculus, as the derivatives of differentiable maps.

Let $f : U \to \mathbb{R}^p$ be a map. For $v \in \mathbb{R}^n$, the *derivative of f at x in the direction v* is defined to be

$$\lim_{t \to 0} \frac{f(x+tv) - f(x)}{t} \tag{7.1}$$

provided this limit exists. We will denote it by $d_x f(v)$. It is obvious that

$$d_x f(\lambda v) = \lambda d_x f(v) \tag{7.2}$$

for $\lambda \in \mathbb{R}$, but completely non-obvious that

$$d_x f(v_1 + v_2) = d_x f(v_1) + d_x f(v_2). \tag{7.3}$$

Indeed, (7.3) is in general false.

Definition 7.1. *Let $U \subset \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^p$ be a map. We say f is differentiable at the point x if the limit* (7.1) *exists for every $v \in \mathbb{R}^n$, and if*

1. *$d_x f(v)$ is linear in v – that is,* (7.3) *holds, as well as* (7.2)*, and if moreover*
2. $$\lim_{\|v\| \to 0} \frac{\|f(x+v) - d_x f(v) - f(x)\|}{\|v\|} = 0.$$

Theorem 7.2. *If all first order partial derivatives of f exist and are continuous at x then f is differentiable at x.*

This is of course a theorem of Analysis, and we make no effort to prove it here. When f is differentiable at x, the linear map $d_x f$ is called the *derivative* of f at x, or, sometimes, the *Fréchet derivative* of f at x.

Let E_n and E_p be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^p$ respectively.

Proposition 7.3. *If f is differentiable at x then*

$$\left[\, d_x f \,\right]_{E_p}^{E_n} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_p}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_n} \end{pmatrix} \tag{7.4}$$

where all partial derivatives are calculated at x.

Proof. **Exercise** (Apply (4.6)). □

Theorem 7.4. Chain Rule *Let $f : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^p$ and $g : V \to \mathbb{R}^q$ be maps. If f is differentiable at x and g is differentiable at $f(x)$ then $g \circ f$ is differentiable at x, and its derivative at x is the composite of the derivatives of f and of g:*

$$d_x(g \circ f) = \left(d_{f(x)} g\right) \circ d_x f.$$

□

Using (4.8) we obtain from this the following equality concerning partial derivatives of a composite:

Corollary 7.5. *Denote by h the composite $g \circ f$. Then for each component function h_k of h, we have*

$$\frac{\partial h_k}{\partial x_i} = \sum_{j=1}^{p} \frac{\partial g_k}{\partial y_j} \frac{\partial f_j}{\partial x_i}$$

where the partial derivatives of g are calculated at $f(x)$ and those of f are calculated at x.

Proof. Follows from Proposition 7.3 by (4.8) □

Exercises Many interesting maps appear in connection with the space $M_{n\times n}(\mathbb{R})$ of $n \times n$ matrices, which we identify with $\mathbb{R}^{n^2}$.

1. Consider the map $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$, and let $A \in M_{n\times n}(\mathbb{R})$. What is $d_A \det$? Notice that $\det(A)$ is a polynomial in the entries of A, and so is certainly a differentiable function. Hint: the vectors in $M_{n\times n}(\mathbb{R})$ should still be thought of as $n \times n$ matrices. If $B \in M_{n\times n}(\mathbb{R})$ then let $B[i]$ denote the matrix whose i'th column is the same as the i'th column of B, and whose remaining columns are equal to 0. Then $B = B[1] + \cdots + B[n]$, so

$$d_A \det(B) = d_A \det(B[1] + \cdots + B[n]) = d_A \det(B[1]) + \cdots + d_A \det(B[n]).$$

   Now you can use well-known properties of det to compute each $d_A \det(B[i])$ as a limit, as in (7.1).

2. Let $\mathrm{Gl}_n(\mathbb{R})$ be the subset of $M_{n\times n}(\mathbb{R})$ consisting of invertible $n \times n$ matrices.

   (a) Let $p : M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ be the product map $p(A, B) = AB$. What is the derivative of p at the point (A, B)? Rather than trying to write down a matrix of this linear map, simply determine $d_{(A,B)}(\hat{A}, \hat{B})$.

   (b) Let $i : \mathrm{Gl}_n(\mathbb{R}) \to \mathrm{Gl}_n(\mathbb{R})$ be the map $i(A) = A^{-1}$. Let I_n be the $n \times n$ identity matrix. What is $d_{I_n} i$? That is, for $B \in M_{n\times n}(\mathbb{R})$, what is $d_{I_n} i(B)$? Let A be any invertible matrix and let B be any $n \times n$ matrix. What is $d_A i(B)$? These questions are quite hard without a good strategy. But there is a good strategy, using the chain rule, and the fact that $p(A, i(A)) = I_n$ for all $A \in \mathrm{Gl}_n(\mathbb{R})$.