Chapter 2
Geometry of Linear Maps

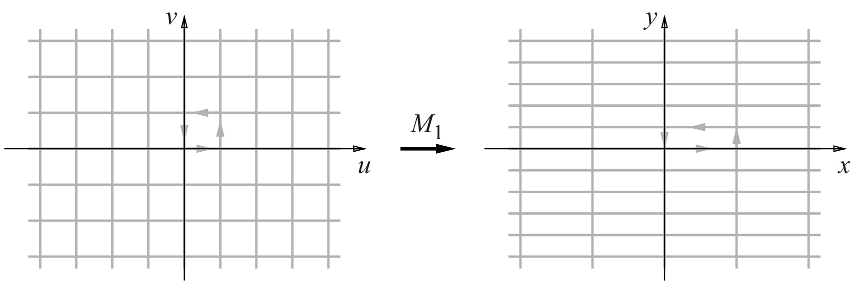
Abstract The geometric meaning of a linear function $x \mapsto y = mx$ is simple and clear: it maps $\mathbb{R}^1$ to itself, multiplying lengths by the factor m. As we show, linear maps $M : \mathbb{R}^n \to \mathbb{R}^n$ also have their multiplication factors of various sorts, for any $n > 1$. In later chapters, these factors play a role in transforming the differentials in multiple integrals that is exactly like the role played by the multiplier $\varphi'(s)$ in the transformation $dx = \varphi'(s)\,ds$ in single-variable integrals. With this in mind, we take up the geometry of linear maps in the simplest case of two variables.

2.1 Maps from $\mathbb{R}^2$ to $\mathbb{R}^2$

Some examples $M : (u,v) \mapsto (x,y)$ illustrate the possibilities that we face.

$M_1 = \begin{pmatrix} 2 & 0 \\ 0 & \frac{3}{5} \end{pmatrix}$

$$M_1 : \begin{cases} x = 2u, \\ y = \frac{3}{5}v, \end{cases} \qquad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & \frac{3}{5} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$



This map carries horizontal lines to horizontal lines and multiplies horizontal lengths by 2. It carries vertical lines to vertical lines and multiplies vertical lengths by $\frac{3}{5}$. These lines are special: they are the only ones whose directions are left unchanged by the map. (For example, the image of a line with slope $\Delta v/\Delta u = 1$ has the different slope $\Delta y/\Delta x = \frac{3}{5}\Delta v/2\Delta u = 3/10$. See the exercises.) A grid of unit

Horizontal and vertical directions are invariant

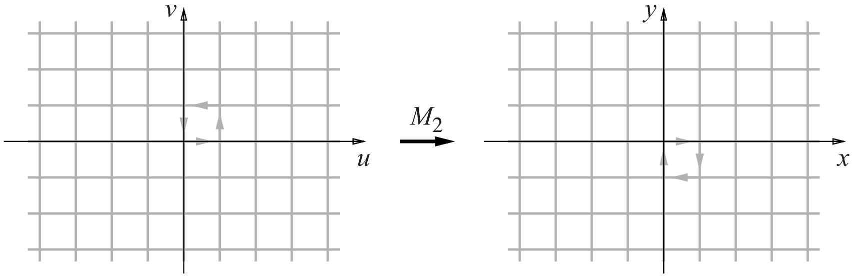
J.J. Callahan, *Advanced Calculus: A Geometric View*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-1-4419-7332-0_2,

squares in the (u,v)-plane is mapped to a grid of rectangles in the (x,y)-plane, and the sides of the rectangles are parallel to the sides of the squares. Finally, orientation is preserved: a counterclockwise circuit around the unit square in the (u,v)-plane maps to a counterclockwise circuit of its image rectangle in the (x,y)-plane.

$M_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

Our second example is also quite simple in form; it is a pure reflection across the horizontal axis:

$$M_2 : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$



Orientation is reversed

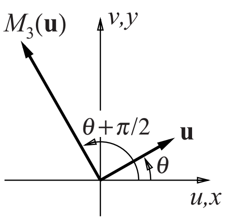
The horizontal and vertical lines are still the invariant ones, and this time even lengths on them are unchanged. Vertical lines are reversed in direction, though, because the vertical multiplier is -1. Orientation of the whole plane is therefore reversed: the counterclockwise circuit in the (u,v)-plane has a clockwise image in the (x,y)-plane. Note that M_1 and M_2 are both diagonal matrices, and the multipliers are their diagonal elements.

$M_3 = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}$

Our third example, although still simple in form, introduces a new action: rotation,

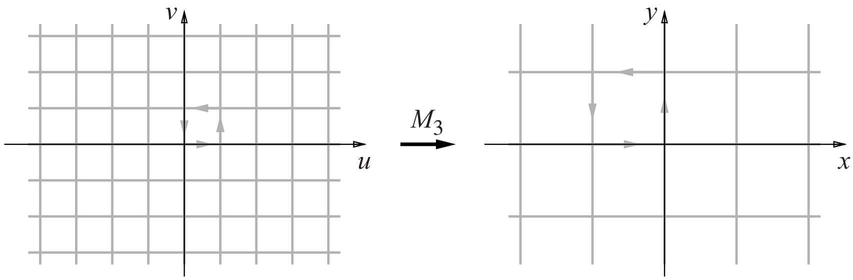
$$M_3 : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Consider the effect M_3 has on a unit vector $\mathbf{u}$ that makes an angle θ with the positive horizontal axis:



$$M_3(\mathbf{u}) = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} -2\sin\theta \\ 2\cos\theta \end{pmatrix} = 2 \begin{pmatrix} \cos(\theta+\pi/2) \\ \sin(\theta+\pi/2) \end{pmatrix}.$$

Thus, $M_3(\mathbf{u})$ is two units long and makes an angle $\theta+\pi/2$ with the horizontal axis. (You should check that $-\sin\theta = \cos(\theta+\pi/2)$ and $\cos\theta = \sin(\theta+\pi/2)$.) Every unit vector, and therefore every nonzero vector, is rotated by $\pi/2$. For this linear map, no line is special in the sense that it is preserved with at most a change in length, so there are no length multipliers. Nevertheless, M_3 doubles the length of every vector and it preserves orientation. It is the combination of a rotation (by $90°$) and a uniform dilation (by a factor of 2), as the following figure shows. Any combination of a rotation with a uniform dilation is a linear map of the plane to itself.



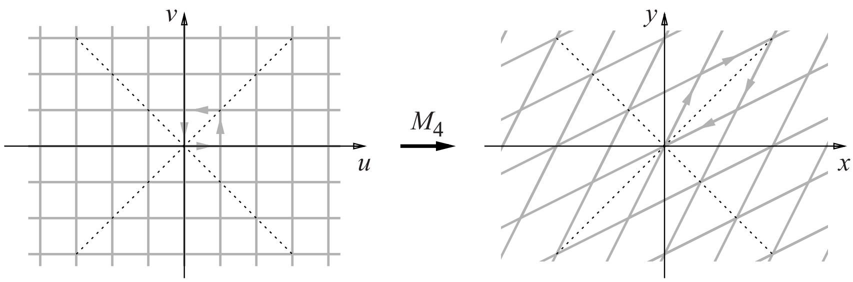
Strains

The maps M_1 and M_2 have similarities not shared with M_3; M_1 and M_2 are what we call **strains**. The next two matrices provide us further examples of strains.

$M_4 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

Example 4 has a more complicated formula than the previous ones, but we ultimately show that it is as simple geometrically as the first two.

$$M_4 : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$



Diagonals are invariant

Neither horizontal nor vertical lines are preserved. The image of the grid of unit squares is a grid of congruent parallelograms, but there is apparently little to connect the two grids geometrically. Notice, however, that the diagonals of the square grid are invariant; they are shown dotted in the figure. The image of a vector that lies on the diagonals is just a multiple of itself:

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1 \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

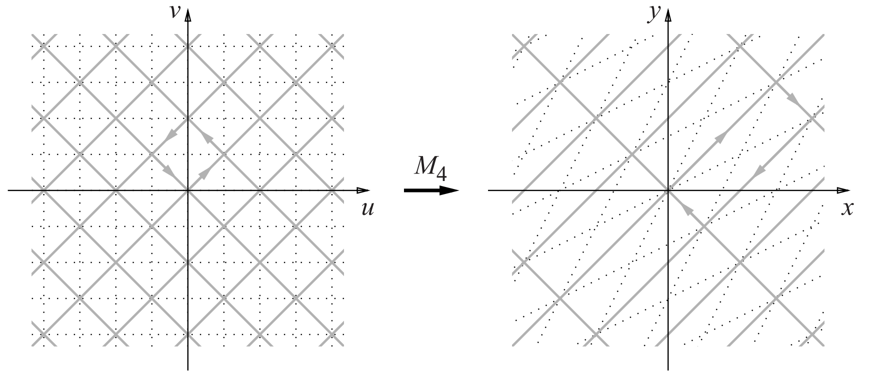
Multipliers are 3 and -1 on diagonals

Specifically, the diagonal in the first and third quadrants is stretched by the factor 3, whereas the diagonal in the second and fourth is simply flipped, with no change in length. The presence of a negative multiplier suggests that orientation is reversed, and that is confirmed by the clockwise circuit in the image.

Geometric clarity with a new basis...

In the figure below, we have switched to a new grid that is parallel to the invariant diagonals in order to see the geometric action of M_4 more clearly. The basis vectors for the new grid (in both source and target) are

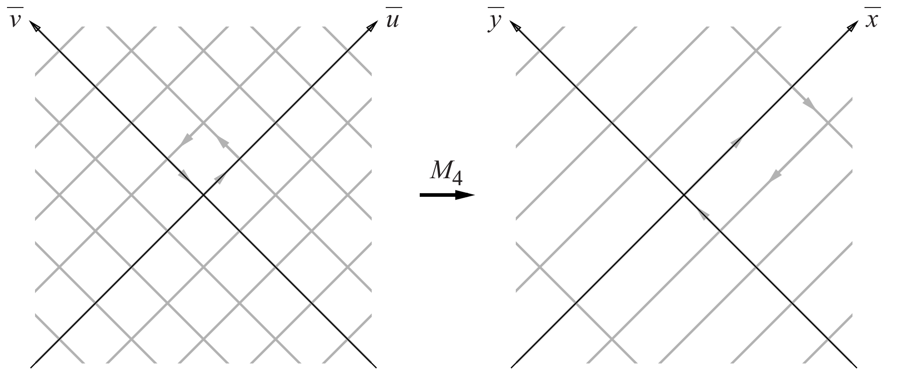
$$\overline{\mathbf{u}} = \overline{\mathbf{x}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \overline{\mathbf{v}} = \overline{\mathbf{y}} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$



. . . and new coordinates

Every point (or vector) in the source now has new coordinates $(\overline{u},\overline{v})$ as well as the original coordinates (u,v), so we must be able to change from one to the other. To see how, suppose vector $\mathbf{p}$ has new coordinates $(\overline{u},\overline{v})$; then

$$\mathbf{p} = \overline{u}\begin{pmatrix}1\\1\end{pmatrix} + \overline{v}\begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}\overline{u}-\overline{v}\\ \overline{u}+\overline{v}\end{pmatrix} = (\overline{u}-\overline{v})\begin{pmatrix}1\\0\end{pmatrix} + (\overline{u}+\overline{v})\begin{pmatrix}0\\1\end{pmatrix}.$$



$\overline{M_4} = \begin{pmatrix}3 & 0\\0 & -1\end{pmatrix}$

The original coordinates of $\mathbf{p}$ are therefore $(\overline{u}-\overline{v},\overline{u}+\overline{v})$, so the formulas for the coordinate change and its inverse are

$$u = \overline{u}-\overline{v} \qquad \overline{u} = \tfrac{1}{2}(u+v)$$
$$v = \overline{u}+\overline{v} \qquad \overline{v} = \tfrac{1}{2}(-u+v)$$

The coordinates $(\overline{x},\overline{y})$ and (x,y) change the same way, of course. Using the coordinate changes we can transform the original formulas for the linear map M_4 into formulas that use the new coordinates:

$$\overline{x} = \tfrac{1}{2}(x+y) = \tfrac{1}{2}(u+2v+2u+v) = \tfrac{1}{2}(3u+3v) = 3\overline{u},$$
$$\overline{y} = \tfrac{1}{2}(-x+y) = \tfrac{1}{2}(-u-2v+2u+v) = \tfrac{1}{2}(u-v) = -\overline{v}.$$

Thus, in the new coordinates, our linear map is described by a new matrix:

$$\begin{pmatrix}\overline{x}\\ \overline{y}\end{pmatrix} = \begin{pmatrix}3 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\overline{u}\\ \overline{v}\end{pmatrix}, \quad \overline{M_4} = \begin{pmatrix}3 & 0\\ 0 & -1\end{pmatrix}.$$

With the new matrix, there is no doubt that M_4 has the same kind of geometric action as M_1 and M_2 (but not M_3!); as a combination of stretches in two different directions, it is a *strain*. Thus, a coordinate change can bring clarity and simplicity to the study of linear maps, just as it can for the study of integration.

Equivalence of matrices

It is worth seeing the connection between M_4 and $\overline{M_4}$ directly in terms of matrices. The coordinate change itself is a matrix multiplication, $U = G\overline{U}$, $\overline{U} = G^{-1}U$, where

$$U = \begin{pmatrix}u\\ v\end{pmatrix}, \quad \overline{U} = \begin{pmatrix}\overline{u}\\ \overline{v}\end{pmatrix}, \qquad G = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}, \quad G^{-1} = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix}.$$

(Notice that the columns of G are the coordinates of the new basis with respect to the old.) The same change $X = G\overline{X}$ and $\overline{X} = G^{-1}X$ happens in the target. In the new coordinates, the map $X = M_4U$ takes the form

$$\overline{X} = G^{-1}X = G^{-1}M_4U = G^{-1}M_4G\overline{U} = \overline{M_4}\,\overline{U}.$$

For us, the object of this string of equalities is the conclusion

$$\overline{M_4} = G^{-1}M_4G,$$

which leads, finally, to the following definition.

Definition 2.1 *Suppose A and B are $n\times n$ matrices; then we say that **B is equivalent to A** if there is an invertible matrix G for which $B = G^{-1}AG$.*

What we have just shown about $\overline{M_4}$ and M_4 implies that *if B is equivalent to A, then there is a basis of $\mathbb{R}^n$ on which A acts in the same way that B acts on the standard basis of $\mathbb{R}^n$*. Alternately, *A and B represent the same linear map in different coordinates*. The matrix G, in $B = G^{-1}AG$, represents the coordinate change.

Equivalence classes of matrices

Note: if $B = G^{-1}AG$, then $A = H^{-1}BH$, where $H = G^{-1}$, so A is equivalent to B when B is equivalent to A. This allows us to say, more symmetrically, that "A and B are equivalent." In the exercises you are asked to show that if C is equivalent to B and B is equivalent to A, then C is also equivalent to A. In other words, equivalent matrices are always *mutually* equivalent. We define an **equivalence class** of $n\times n$ matrices to be the set of all matrices equivalent to some given one. (An example of an equivalence class in a more familiar context is a *rational number*: a rational number is a set of mutually equivalent integer fractions, where two such fractions a/b and c/d are defined to be equivalent if $ad = bc$.) Our aim, which is to identify the different geometric actions of a linear map $M : \mathbb{R}^2 \to \mathbb{R}^2$, is accomplished by determining the equivalence classes of 2×2 matrices.

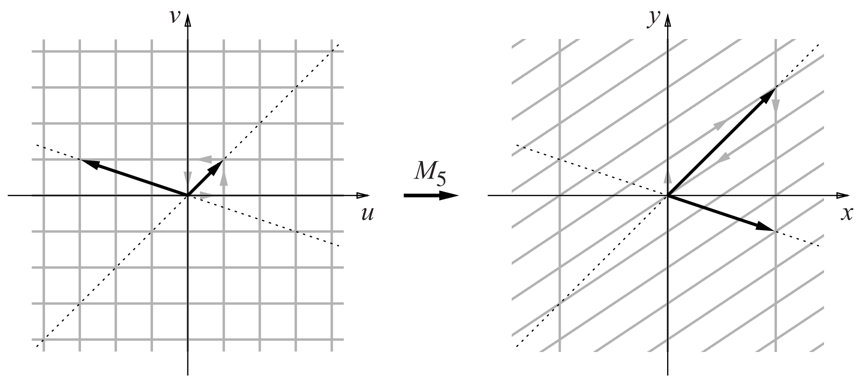
$M_5 = \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix}$

In all our examples where there were invariant lines, those lines were mutually perpendicular. Our next example shows us we cannot expect this to happen in general.

$$M_5 : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

In the figure below, it may appear that M_5 leaves vertical lines invariant. But this is not true: a vertical line in the target is the image of a horizontal line in the source, not a vertical one. In fact, the directions of the invariant lines are indicated by the heavy vectors and the dotted lines in that figure, because

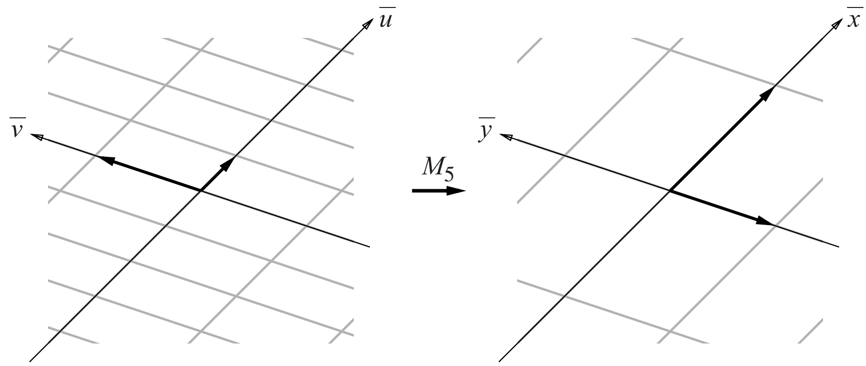
$$\begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix} = -\begin{pmatrix}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



Multipliers are 3 and -1 again

We show how to find invariant lines for an arbitrary linear map immediately below. For the moment, we just observe that M_5 has the same length multipliers as M_4: 3 and -1. The two maps have different invariant lines, though; in particular, the ones for M_5 are not mutually perpendicular. Nevertheless, it is reasonable to call M_5 a strain. With a new coordinate system and grid that is based on vectors along the invariant lines (cf. Exercise 2.2), M_5 has the following form.

$$\begin{pmatrix} \overline{x} \\ \overline{y} \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \overline{u} \\ \overline{v} \end{pmatrix}, \quad \overline{M_5} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}.$$



Characteristics of a linear map

We discover that M_5 is in the same equivalence class as M_4 (because they are both equivalent to $\overline{M_5} = \overline{M_4}$). Indeed, they have the same geometric description: they map the plane to itself by stretching it by a factor of 3 in one direction and simply flipping it—without a stretch—in another. Yet M_5 and M_4 are not the same, because they perform their stretches and flips along different lines (their own invariant lines). It is evident, though, that the invariant lines and the associated multipliers, taken together, characterize each linear map geometrically: two maps with the same multipliers acting on the same lines must be identical.

Eigenvectors and eigenvalues

Multipliers and invariant lines therefore give us an important way to characterize a linear map. They are introduced in the following definition with their usual names. Rather curiously, historical accident has cast those names—*eigen*value and *eigen*vector—half in German and half in English. *Eigen* means "one's own", or "characteristic", and the alternatives *characteristic value* and *characteristic vector* are also used. Furthermore, *eigen* can be translated into French as *propre*, and the terms *proper value* and *proper vector* are likewise in use, but less frequently.

Definition 2.2 *Let $M : \mathbb{R}^n \to \mathbb{R}^n : U \mapsto X$ be a linear map defined by matrix multiplication: $X = MU$. A vector $U \neq 0$ is an* **eigenvector of M with eigenvalue λ** *if $MU = \lambda U$.*

$\lambda = 0$ and the kernel of M

Note that an eigenvector is nonzero by definition because it has to determine an invariant line. An eigenvalue can be 0, though; it just means M has a nonzero *kernel* consisting of the eigenvectors with eigenvalue zero.

Characteristic equation

Let us rewrite the "eigen" condition $MU = \lambda U$ first as $MU = \lambda IU$ (where I is the identity matrix), and then as $(M - \lambda I)U = 0$. This says that U is in the kernel of the newly defined matrix $M - \lambda I$. But $U \neq 0$, so $M - \lambda I$ must be noninvertible, implying $\det(M - \lambda I) = 0$. Because the determinant of a matrix is a polynomial function of the elements of the matrix, the expression $p(\lambda) = \det(M - \lambda I)$ is is a polynomial in λ, called the **characteristic polynomial** of M. The equation $p(\lambda) = 0$ is the **characteristic equation** of M.

Theorem 2.1. *Each eigenvalue of M is a root of its characteristic equation.* □

Complex roots

But real polynomials can have complex roots, too. For example, our rotation matrix M_3 has the characteristic polynomial $p(\lambda) = \lambda^2 + 4$ whose roots are $\lambda = \pm 2\mathrm{i}$. Furthermore,

$$\begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} = \begin{pmatrix} 2\mathrm{i} \\ 2 \end{pmatrix} = 2\mathrm{i}\begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix}; \quad \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix} = \begin{pmatrix} -2\mathrm{i} \\ 2 \end{pmatrix} = -2\mathrm{i}\begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}.$$

In other words, when we allow M_3 to act on ordered pairs of complex numbers, we find that M_3 does have invariant directions. A real polynomial always has complex roots, but need not have any real roots. Thus this example suggests that, for the purpose of getting the simple view of the action of matrix multiplication, we use complex n-tuples instead of real ones ($M : \mathbb{C}^n \to \mathbb{C}^n$) to define eigenvectors and eigenvalues (Definition 2.2). With this understanding, every root of the characteristic equation of M becomes an eigenvalue of M.

Trace and determinant

If M is an 2×2 matrix, we have

$$\begin{aligned} p(\lambda) &= \det\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] = \det\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} \\ &= (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + ad - bc \\ &= \lambda^2 - \operatorname{tr}(M)\lambda + \det(M), \end{aligned}$$

where $\operatorname{tr}(M) = a + d$ is the **trace** of M and $\det(M) = ad - bc$ is, of course, its **determinant**. Thus, the eigenvalues of M are the roots of a quadratic equation that involves the trace and determinant of M. If λ_1 and λ_2 are these roots, then

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2$$

is also the characteristic polynomial, so we have the following proposition.

Theorem 2.2. *The sum of the eigenvalues of a 2×2 matrix is equal to its trace and their product is equal to its determinant.* □

If we write the equation for the eigenvalues of the 2×2 matrix M in the form

$$\lambda_{1,2} = \frac{\operatorname{tr} M \pm \sqrt{\operatorname{tr}^2 M - 4 \det M}}{2},$$

we see these roots will be complex when the *discriminant* is negative:

$$\operatorname{tr}^2 M - 4 \det M = (a+d)^2 - 4(ad - bc) = (a-d)^2 + 4bc < 0.$$

Because $(a-d)^2 \geq 0$, b and c must be of opposite sign and have $bc < -(a-d)^2/4$ for M to have complex eigenvalues.

As we have seen, equivalent matrices describe the same linear map but in terms of different bases. We would expect, then, that such matrices have the same eigenvalues, and their eigenvectors would be mapped to one another by the coordinate change that connects the matrices.

Eigenvalues of equivalent matrices

Theorem 2.3. *Suppose A and $B = G^{-1}AG$ are equivalent matrices, and $\overline{U}$ is an eigenvector of B with eigenvalue λ. Then $U = G\overline{U}$ is an eigenvector of A with the same eigenvalue λ.*

Proof. Suppose $\overline{U}$ is an eigenvector of B with eigenvalue λ: $B\overline{U} = \lambda\overline{U}$. Then

$$G^{-1}AG\overline{U} = \lambda\overline{U}, \text{ so } A(G\overline{U}) = G\lambda\overline{U} = \lambda(G\overline{U}).$$ □

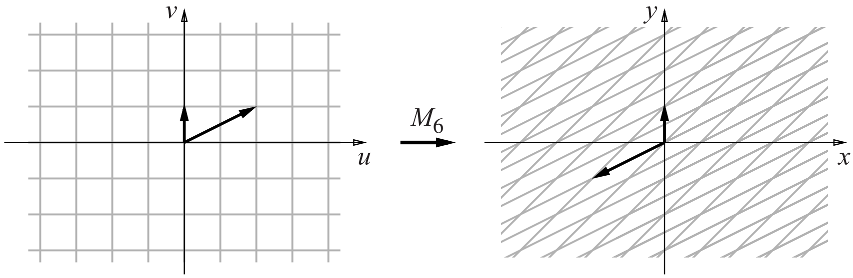
Corollary 2.4 *Equivalent matrices have the same eigenvalues and therefore the same trace, determinant, and characteristic polynomial.*

Proof. According to the theorem, every eigenvalue of $B = G^{-1}AG$ is an eigenvalue of A. But equivalence is symmetric ($A = H^{-1}BH$ with $H = G^{-1}$), so every eigenvalue of A is an eigenvalue of B. □

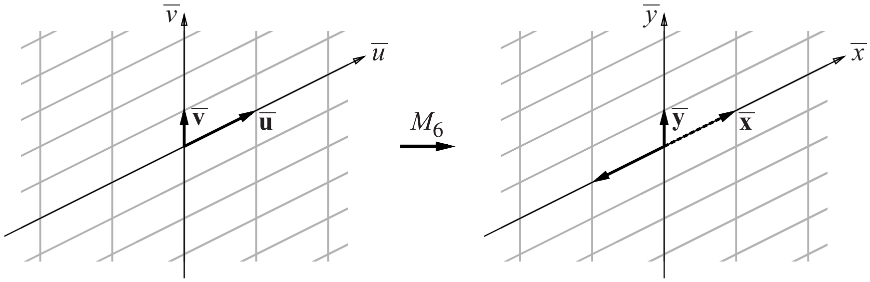
Even when the eigenvalues and eigenvectors of M are complex, they can provide crucial information about the geometric action of M on the real plane $\mathbb{R}^2$. Consider the map

$M_6 = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$

$$M_6 : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$



Here $\operatorname{tr} M_6 = 0$ and $\det M_6 = 1$, so the characteristic polynomial is $\lambda^2 + 1$ and the eigenvalues are $\pm i$. However, M_6 is not a rotation: it turns the coordinate axes by different amounts (so their images are not perpendicular, as they would be under a rotation).



A clearer picture emerges, however, when we consider the action of M_6 on the heavy vectors; note that

$$M_6(\overline{\mathbf{u}}) = \overline{\mathbf{y}}, \quad M_6(\overline{\mathbf{v}}) = -\overline{\mathbf{x}}.$$

Consequently, in terms of the new grid and coordinates (which use $\{\overline{\mathbf{u}}, \overline{\mathbf{v}}\}$ and $\{\overline{\mathbf{x}}, \overline{\mathbf{y}}\}$ as bases in the source and target), our linear map is now described by

$$\overline{M_6} : \begin{pmatrix} \overline{x} \\ \overline{y} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \overline{u} \\ \overline{v} \end{pmatrix}.$$

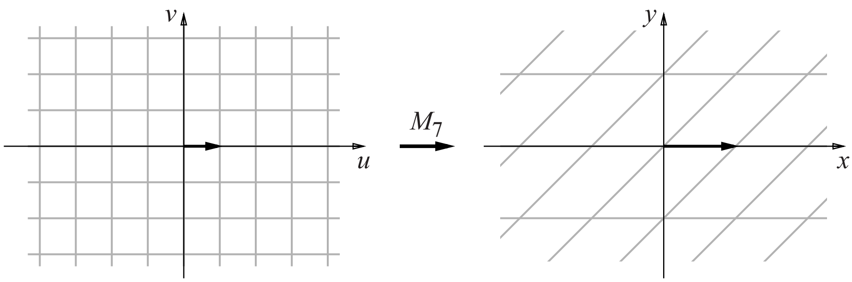
You can check directly that the new matrix $\overline{M_6}$ has the same trace, determinant, characteristic polynomial, and eigenvalues as M_6. But the geometric action of $\overline{M_6}$ is simpler to describe: applied to the standard basis (instead of $\{\overline{\mathbf{u}}, \overline{\mathbf{v}}\}$), $\overline{M_6}$ is just rotation by 90°. Note, however, that $\overline{M_6}$, applied to the basis $\{\overline{\mathbf{u}}, \overline{\mathbf{v}}\}$, is not a rotation, because M_6 is not a rotation. Thus, although M_6 is not a rotation, it is equivalent to a 90° rotation.

M_6 is equivalent to a 90° rotation

$M_7 = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}$

There is essentially only one more type of linear map for us to analyze. Here is an example:

$$M_7 : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 2 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$



M_7 has only one invariant direction

The horizontal lines are invariant, but the vertical ones are not. In fact, there is no second set of invariant lines. We can trace this shortcoming to the fact that M_7 has only one eigenvalue, $\lambda = 2$; it is a *repeated* root of the characteristic polynomial $\lambda^2 - 4\lambda + 4 = (\lambda - 2)^2$. All eigenvectors are therefore the solutions of the single pair of equations

$$\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ giving just } \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ 0 \end{pmatrix} = u \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

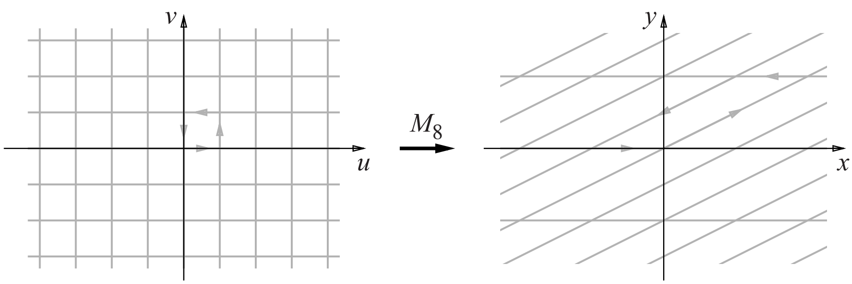
Shears

This eigenvector is horizontal; it implies that horizonal lines are invariant under M_7. But because no vector in any other direction is an eigenvector, no other direction is invariant. The geometric action of M_7 is called a **shear**. Of course we associate eigenvalues with stretches; in this example the shear is combined with a uniform dilation whose magnitude is given by the single eigenvalue 2.

$M_8 = \begin{pmatrix} 4 & -2 \\ 2 & 0 \end{pmatrix}$

A shear can take a less recognizable form, as in the following example.

$$M_8 : \begin{cases} x = 4u - 2v, \\ y = 2u, \end{cases} \qquad M_8 = \begin{pmatrix} 4 & -2 \\ 2 & 0 \end{pmatrix}.$$



Because $\operatorname{tr} M_8 = \det M_8 = 4$, M_8 has the single eigenvalue $\lambda = 2$. There is an eigenvector in only one direction. In the exercises you are asked to find that eigenvector and then to verify that the coordinate change

$$G : \begin{cases} u = \overline{u} + \overline{v}, \\ v = \overline{u}, \end{cases}$$

converts M_8 into M_7: $G^{-1} M_8 \, G = M_7$. This implies M_8 is equivalent to a shear combined with a uniform dilation by the factor 2.

Rotations

Before we proceed to a description of all the different geometric actions of a linear map $M : \mathbb{R}^2 \to \mathbb{R}^2$, it is helpful to comment on a few specific matrices. The matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

rotates the plane by θ radians; it has complex eigenvalues $\lambda_\pm = \cos\theta \pm \mathrm{i}\sin\theta$. The matrix

$$C_{a,b} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

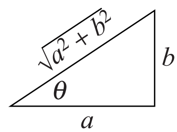
has a similar form, and has the complex eigenvalues, $\lambda_\pm = a \pm \mathrm{i}b$, but is not a simple rotation if $a^2 + b^2 \neq 1$. However, the following theorem connects it to a rotation.

Theorem 2.5. *If* $(a,b) \neq (0,0)$*, then the matrix* $C_{a,b}$ *rotates the plane by the angle* $\theta = \arctan(b/a)$ *and then performs a uniform dilation by the factor* $\sqrt{a^2+b^2}$.

Proof. By hypothesis, $\sqrt{a^2+b^2} \neq 0$, so we can factor this term out of each component of $C_{a,b}$:

$$C_{a,b} = \sqrt{a^2+b^2} \begin{pmatrix} \dfrac{a}{\sqrt{a^2+b^2}} & \dfrac{-b}{\sqrt{a^2+b^2}} \\ \dfrac{b}{\sqrt{a^2+b^2}} & \dfrac{a}{\sqrt{a^2+b^2}} \end{pmatrix} = \sqrt{a^2+b^2} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In the matrix on the right, we have made the replacements



$$\frac{a}{\sqrt{a^2+b^2}} = \cos\theta \quad \text{and} \quad \frac{b}{\sqrt{a^2+b^2}} = \sin\theta$$

by using the angle $\theta = \arctan(b/a)$, as the figure shows. In fact, we can extend $\theta = \arctan(b/a)$ as a function of two variables a and b (cf. Exercise 2.10) to define a unique value of θ in the interval $-\pi < \theta \leq \pi$ for every $(a,b) \neq (0,0)$. That is, we need not require that a and b be positive. Therefore,

$$C_{a,b} = \left(\sqrt{a^2+b^2}\right) R_{\arctan(b/a)}. \qquad \square$$

Collapse and shear–collapse

Suppose M has an eigenvector U with eigenvalue 0; then M collapses $\mathbb{R}^2$ along the direction of U. For this reason, we describe any matrix with a zero eigenvalue as a **collapse**. A rather special example is

$$K = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{with eigenvector} \quad U = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Because $\operatorname{tr} K = \det K = 0$, the characteristic equation of K is $\lambda^2 = 0$. This has only a single root, the repeated eigenvalue $\lambda = 0$. Because there is only a single eigendirection, given by the eigenvector U, above, K behaves like a *shear*. Because its sole eigenvalue is 0, it is also a *collapse*; we call it a **shear–collapse**.

Theorem 2.6. *Every linear map $M : \mathbb{R}^2 \to \mathbb{R}^2$ is equivalent to precisely one of the types listed in the following table; M lies in the equivalence class of matrices that have the same eigenvalues and the same number of eigendirections.*

Equivalence Classes of 2×2 Matrices
and Their Representatives

Name	Matrix	Eigenvalues*	Eigendirections
Zero	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$0, 0$	all
Shear–collapse	$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$	$0, 0$	one
Strain–collapse	$\begin{pmatrix} \lambda & 0 \\ 0 & 0 \end{pmatrix}$	$0, \lambda$	two
Pure dilation	$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$	λ, λ	all
Shear–dilation	$\begin{pmatrix} \lambda & \lambda \\ 0 & \lambda \end{pmatrix}$	λ, λ	one
Strain	$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$	$\lambda_1 \neq \lambda_2$	two
Rotation–dilation	$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$	$a \pm ib$	none

*An eigenvalue not written as 0 is understood to be nonzero.

Proof. You carry out parts of the proof in the exercises. The basic classification is by the eigenvalues of M:

- *Real and equal*: zero, shear–collapse, pure dilation, shear–dilation
- *Real and unequal*: strain–collapse, strain
- *Complex conjugates*: rotation–dilation

Types are then further separated by the number of eigendirections that M has:

- *None*: rotation–dilation

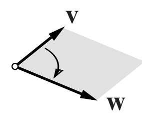
- *One*: all shears
- *Two*: all strains
- *All*: pure dilations, including zero □

Area multiplier

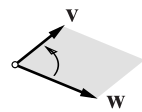
If a matrix of a linear map has real eigenvalues and eigenvectors, the eigenvectors determine the map's invariant lines and the eigenvalues give the length multiplication factors along those lines. But even more is true: the product of those factors then tells us how much the map magnifies *areas*; the sign of the product even indicates how the map affects orientation. Furthermore, the area multiplier is just the determinant of the matrix (because the product of the eigenvalues is the determinant), so the area multiplier can be determined directly from the matrix itself, without first calculating the eigenvalues. (This is particularly useful when the eigenvalues and eigenvectors are complex because then the matrix has no usable length multipliers.)

Orientation and signed area

We need a notation that indicates the orientation of a parallelogram, and this is easily obtained. We use $\mathbf{v}\wedge\mathbf{w}$ to denote the parallelogram spanned by the vectors $\mathbf{v}$ and $\mathbf{w}$, *in that order*. Call this the **wedge product** of $\mathbf{v}$ and $\mathbf{w}$. The order determines the "sense of rotation"—either clockwise or counterclockwise—that carries the first-named vector, $\mathbf{v}$, to the second, $\mathbf{w}$. Reversing the order, to $\mathbf{w}\wedge\mathbf{v}$, reverses the sense of rotation; we write $\mathbf{w}\wedge\mathbf{v} = -\mathbf{v}\wedge\mathbf{w}$. A parallelogram has **positive orientation** if it has the same sense of rotation as the positive coordinate axes, and **negative orientation** if it has the opposite sense. (If $\mathbf{v}$ and $\mathbf{w}$ are linearly dependent, $\mathbf{v}\wedge\mathbf{w}$ collapses to a line segment and has no orientation.) As a rule, we take the positive sense of rotation to be counterclockwise. Thus, in the adjacent figure, $\mathbf{v}\wedge\mathbf{w}$ is negatively oriented and $\mathbf{w}\wedge\mathbf{v}$ is positively oriented.



$\mathbf{v}\wedge\mathbf{w}$



$\mathbf{w}\wedge\mathbf{v} = -\mathbf{v}\wedge\mathbf{w}$

The signed area, area $\mathbf{v}\wedge\mathbf{w}$, will then be determined by the following two stipulations. First, the signed area of the unit square $\mathbf{e}_1\wedge\mathbf{e}_2$ should be $+1$ (rather than -1). Second, $\text{area}(\mathbf{w}\wedge\mathbf{v}) = -\,\text{area}(\mathbf{v}\wedge\mathbf{w})$ for all $\mathbf{v}$, $\mathbf{w}$.

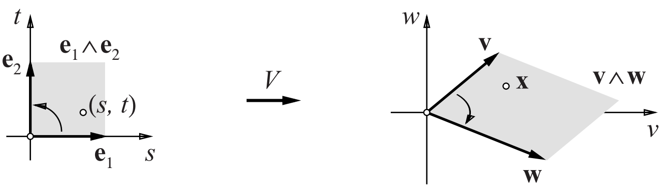
Theorem 2.7. $\text{area}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \wedge \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \det\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}.$

Proof. See Exercise 2.15. □

Parametrizing $\mathbf{v}\wedge\mathbf{w}$

The signed area is the determinant of the matrix V whose columns are the coordinates of $\mathbf{v}$ and $\mathbf{w}$, in that order. The matrix represents a linear map

$$\mathbf{x} = V(s,t) = s\mathbf{v} + t\mathbf{w}$$



that maps the unit square $\mathbf{e}_1\wedge\mathbf{e}_2$ to $\mathbf{v}\wedge\mathbf{w}$. Thus, the orientation and area of $\mathbf{v}\wedge\mathbf{w}$ are determined by a parametrization.

Theorem 2.8. *If $M : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map and $\mathbf{v} \wedge \mathbf{w}$ is an oriented parallelogram in the source, then $M(\mathbf{v} \wedge \mathbf{w}) = M(\mathbf{v}) \wedge M(\mathbf{w})$ is an oriented parallelogram in the target and*

$$\operatorname{area} M(\mathbf{v} \wedge \mathbf{w}) = \det M \times \operatorname{area}(\mathbf{v} \wedge \mathbf{w}).$$

Proof. In Exercise 2.16 you are asked to prove this directly by analyzing the function $\operatorname{area} M(\mathbf{v}) \wedge M(\mathbf{w})$. □

Corollary 2.9 *The **area multiplier** of the linear map $M : \mathbb{R}^2 \to \mathbb{R}^2$ is* $\det M$. *The map M reverses orientation precisely when* $\det M < 0$. □

2.2 Maps from $\mathbb{R}^n$ to $\mathbb{R}^n$

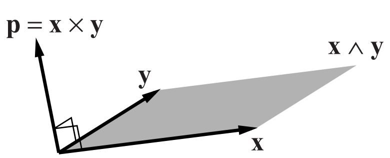
Because we found the area multiplier to be the most salient geometric feature of a linear map of the plane, we can expect that the volume multiplier, and its higher-dimensional analogues, will play a similar role here.

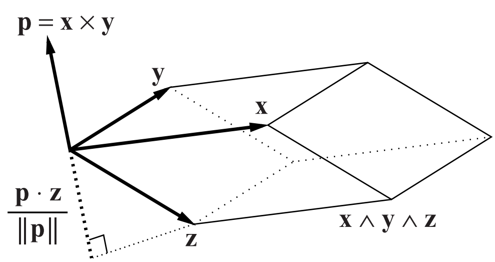
$\mathbf{x} \wedge \mathbf{y}$ in $\mathbb{R}^n$, $n \geq 3$

In $\mathbb{R}^n$, $n \geq 3$, we continue to use $\mathbf{x} \wedge \mathbf{y}$ to denote the oriented parallelogram spanned by the vectors $\mathbf{x}$ and $\mathbf{y}$. When $n = 2$, the orientation of $\mathbf{x} \wedge \mathbf{y}$ is fixed in relation to the orientation of the two coordinate axes. However, when $n \geq 3$, this is not true: an orientation-preserving linear map of $\mathbb{R}^n$ can reverse the orientation of $\mathbf{x} \wedge \mathbf{y}$ (see below, p. 44). Moreover, because the coordinates of $\mathbf{x}$ and $\mathbf{y}$ now make up an $n \times 2$ matrix V—for which the determinant is not even defined—we cannot express $\operatorname{area}(\mathbf{x} \wedge \mathbf{y})$ as the determinant of V. (Let $V^\dagger$ be the *transpose* of the matrix of V; it is a $2 \times n$ matrix. The product $V^\dagger V$ does give a square 2×2 matrix, and $\operatorname{area}^2(\mathbf{x} \wedge \mathbf{y}) = \det V^\dagger V$. See the exercises.)

$\mathbf{x} \times \mathbf{y}$ in $\mathbb{R}^3$

In $\mathbb{R}^3$, the cross-product of two vectors is defined: $\mathbf{p} = \mathbf{x} \times \mathbf{y}$ is the unique vector with length $|\operatorname{area}(\mathbf{x} \wedge \mathbf{y})|$ and with direction orthogonal to both $\mathbf{x}$ and $\mathbf{y}$ so that the three vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{p}$—in that order—have the same orientation as the three coordinate axes. We call this the **positive orientation**, and always take it to be *right-handed*, meaning that the thumb, index finger, and middle finger of the right hand can be lined up with the first, second, and third coordinate axes, respectively.





$\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}$ in $\mathbb{R}^3$

For vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ in $\mathbb{R}^3$, we define $\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}$ to be the **oriented parallelepiped** spanned by $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, in that order. Notice that the parallelepiped shown in the figure on the right, above, has *left*-handed, or *negative*, orientation. To calculate its volume, we take $\mathbf{x}\wedge\mathbf{y}$ as base and measure its height as the length of the projection of $\mathbf{z}$ on $\mathbf{p}=\mathbf{x}\times\mathbf{y}$:

$$\begin{aligned}\mathrm{vol}(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}) &= \text{area of base}\cdot\text{height}\\ &= \|\mathbf{x}\times\mathbf{y}\|\,\frac{(\mathbf{x}\times\mathbf{y})\cdot\mathbf{z}}{\|\mathbf{x}\times\mathbf{y}\|} = (\mathbf{x}\times\mathbf{y})\cdot\mathbf{z}.\end{aligned}$$

The quantity $(\mathbf{x}\times\mathbf{y})\cdot\mathbf{z}$ is called the **scalar triple product** of $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. Note that the parentheses can be removed, because $\mathbf{x}\times(\mathbf{y}\cdot\mathbf{z})$ is meaningless. The *order* of the three vectors in $\mathbf{x}\times\mathbf{y}\cdot\mathbf{z}$ is still important, though. Scalar triple product

Theorem 2.10. *The signed volume of the oriented parallelepiped* $\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}$ *is the scalar triple product* $\mathbf{x}\times\mathbf{y}\cdot\mathbf{z}$*. The volume is negative precisely when the parallelepiped has negative orientation.*

Proof. The first statement has been proven above. To prove the second, note that the parallelepiped $\mathbf{x}\wedge\mathbf{y}\wedge(\mathbf{x}\times\mathbf{y})$ has positive orientation by definition, at least when $\mathbf{x}\times\mathbf{y}\neq 0$. Therefore, $\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}$ has negative orientation precisely when $\mathbf{z}$ and $\mathbf{x}\times\mathbf{y}$ lie on opposite sides of the plane determined by $\mathbf{x}\wedge\mathbf{y}$. But $\mathbf{x}\times\mathbf{y}$ is perpendicular to $\mathbf{x}\wedge\mathbf{y}$, so $\mathbf{z}$ and $\mathbf{x}\times\mathbf{y}$ are on opposite sides of $\mathbf{x}\wedge\mathbf{y}$ when $\mathbf{z}$ makes an obtuse angle with $\mathbf{x}\times\mathbf{y}$, and that is precisely the condition that

$$(\mathbf{x}\times\mathbf{y})\cdot\mathbf{z}=\mathrm{vol}(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z})<0. \qquad \square$$

Theorem 2.11. *Let V be the matrix whose columns are the coordinates of* $\mathbf{x}$, $\mathbf{y}$, *and* $\mathbf{z}$, *in that order. Then* $\mathrm{vol}(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z})=\det V$. Volumes and determinants

Proof. Let $\mathbf{x}=(x_1,x_2,x_3)^\dagger$, $\mathbf{y}=(y_1,y_2,y_3)^\dagger$, $\mathbf{z}=(z_1,z_2,z_3)^\dagger$. Then

$$V=\begin{pmatrix}x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{pmatrix},$$

and if we calculate the determinant of V along the third column, we get

$$\det V=\begin{vmatrix}x_2 & y_2\\ x_3 & y_3\end{vmatrix}z_1+\begin{vmatrix}x_3 & y_3\\ x_1 & y_1\end{vmatrix}z_2+\begin{vmatrix}x_1 & y_1\\ x_2 & y_2\end{vmatrix}z_3;$$

note the order of the rows in the second determinant. On the other hand,

$$\mathbf{x}\times\mathbf{y}=\left(\begin{vmatrix}x_2 & y_2\\ x_3 & y_3\end{vmatrix},\begin{vmatrix}x_3 & y_3\\ x_1 & y_1\end{vmatrix},\begin{vmatrix}x_1 & y_1\\ x_2 & y_2\end{vmatrix}\right)^\dagger,$$

so

$$\mathrm{vol}(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z})=\mathbf{x}\times\mathbf{y}\cdot\mathbf{z}=\begin{vmatrix}x_2 & y_2\\ x_3 & y_3\end{vmatrix}z_1+\begin{vmatrix}x_3 & y_3\\ x_1 & y_1\end{vmatrix}z_2+\begin{vmatrix}x_1 & y_1\\ x_2 & y_2\end{vmatrix}z_3=\det V. \qquad \square$$

Orientation

Corollary 2.12 *The parallelepiped* $\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}$ *has positive orientation if and only if the linear map* $V:\mathbb{R}^3\to\mathbb{R}^3$ *that maps the standard basis* $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$ *to* $\mathbf{x},\mathbf{y},\mathbf{z}$*, in that order, has* $\det V>0$.

Proof. We use two fundamental results of linear algebra: (a) a linear map is uniquely defined by its action on a basis; and (b) the matrix V whose columns are the coordinates of $\mathbf{x},\mathbf{y},\mathbf{z}$, in that order, has $V(\mathbf{e}_1)=\mathbf{x}$, $V(\mathbf{e}_2)=\mathbf{y}$, $V(\mathbf{e}_3)=\mathbf{z}$. By the preceding theorems, $\det V>0$ if and only if $\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}$ has positive orientation. □

Corollary 2.13 *An ordered set of vectors* $\{\mathbf{x},\mathbf{y},\mathbf{z}\}$ *has positive orientation if and only if it is the image of the standard basis* $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ *under a linear map* M *with* $\det M>0$. □

Corollary 2.14 *If* $M:\mathbb{R}^3\to\mathbb{R}^3$ *is a linear map and* $\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}$ *is an oriented parallelepiped in the source, then* $M(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z})=M(\mathbf{x})\wedge M(\mathbf{y})\wedge M(\mathbf{z})$ *is an oriented parallelepiped in the target and*

$$\operatorname{vol} M(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z})=\det M\times\operatorname{vol}(\mathbf{x}\wedge\mathbf{y}\wedge\mathbf{z}).$$

Proof. This is analogous to Theorem 2.8 (p. 42) and is proven the same way. □

Volume multiplier

Corollary 2.15 *The* ***volume multiplier*** *for the linear map* $M:\mathbb{R}^3\to\mathbb{R}^3$ *is* $\det M$. *The map* M *reverses orientation precisely when* $\det M<0$. □

Parallelograms in $\mathbb{R}^3$

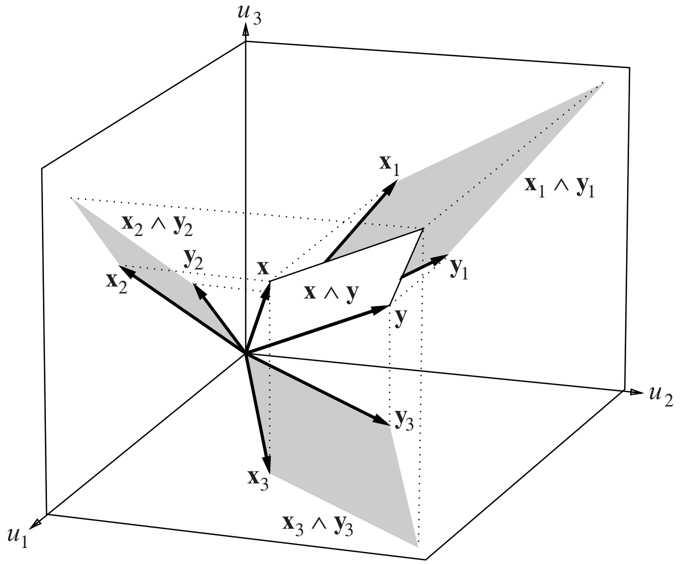
Suppose $\mathbf{x}\times\mathbf{y}\neq\mathbf{0}$. Then there is a linear map $L:\mathbb{R}^3\to\mathbb{R}^3$ with positive determinant for which $L(\mathbf{x})=\mathbf{y}$, $L(\mathbf{y})=\mathbf{x}$ (see Exercise 2.21). Consequently, orientation-preserving linear maps of $\mathbb{R}^3$ need not preserve the orientation of 2-dimensional parallelograms that lie in $\mathbb{R}^3$. (We still orient such objects; see pp. 388ff.) The sign of $\operatorname{area}(\mathbf{x}\wedge\mathbf{y})$ has no intrinsic geometric significance in $\mathbb{R}^3$; thus we always take $\operatorname{area}(\mathbf{x}\wedge\mathbf{y})\geq 0$.

There is a remarkable connection between $\mathbf{x}\wedge\mathbf{y}$ and its projections onto the three coordinate planes. First of all, if $\mathbf{x}=(x_1,x_2,x_3)^\dagger$, $\mathbf{y}=(y_1,y_2,y_3)^\dagger$, then

$$\begin{aligned}\operatorname{area}(\mathbf{x}\wedge\mathbf{y})=\|\mathbf{x}\times\mathbf{y}\| &= \left\|\left(\begin{vmatrix}x_2&y_2\\x_3&y_3\end{vmatrix},\begin{vmatrix}x_3&y_3\\x_1&y_1\end{vmatrix},\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}\right)^\dagger\right\|\\ &=\sqrt{\begin{vmatrix}x_2&y_2\\x_3&y_3\end{vmatrix}^2+\begin{vmatrix}x_3&y_3\\x_1&y_1\end{vmatrix}^2+\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}^2}.\end{aligned}$$

Let $\mathbf{x}_i$ denote the projection of $\mathbf{x}$ onto the coordinate plane $u_i=0$, $i=1,2,3$, and similarly for $\mathbf{y}_i$. Then $\mathbf{x}_i\wedge\mathbf{y}_i$ is a parallelogram in a 2-dimensional plane whose area is therefore a simple 2×2 determinant:

$$\operatorname{area}(\mathbf{x}_1\wedge\mathbf{y}_1)=\begin{vmatrix}x_2&y_2\\x_3&y_3\end{vmatrix},\quad \operatorname{area}(\mathbf{x}_2\wedge\mathbf{y}_2)=\begin{vmatrix}x_3&y_3\\x_1&y_1\end{vmatrix},\quad \operatorname{area}(\mathbf{x}_3\wedge\mathbf{y}_3)=\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}.$$



A "Pythagorean" theorem

We can therefore rewrite our expression for area$(\mathbf{x} \wedge \mathbf{y})$ as

$$\text{area}(\mathbf{x} \wedge \mathbf{y}) = \sqrt{\text{area}^2(\mathbf{x}_1 \wedge \mathbf{y}_1) + \text{area}^2(\mathbf{x}_2 \wedge \mathbf{y}_2) + \text{area}^2(\mathbf{x}_3 \wedge \mathbf{y}_3)};$$

in other words, the square of the area of a parallelogram is equal to the sum of the squares of the areas of its projections onto the three coordinate planes. We can think of this as a "Pythagorean" theorem whose more usual form deals with lengths rather than areas, but relates, in the same way, the length of a vector to the lengths of its projections to the three coordinate axes.

Geometry in $\mathbb{R}^n$, $n > 3$

Although we cannot visualize $\mathbb{R}^n$ directly when $n > 3$, we do carry over geometric concepts by analogy. For example, we continue to say the vector $\mathbf{x}$ has length $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ and the angle θ between the vectors $\mathbf{x}$ and $\mathbf{y}$ (assuming $\mathbf{x} \neq 0 \neq \mathbf{y}$) is

$$\theta = \arccos\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right).$$

Of course, $\arccos q$ is defined only when $|q| \leq 1$, so our definition of θ makes sense only if $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$ for all vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$. This fact is established in the exercises. In what follows, $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$. As you can see, the definitions relate the orientation and volume of an n-dimensional parallelepiped to the determinant of a certain $n \times n$ matrix. We review the definition of an $n \times n$ determinant in the exercises.

Definition 2.3 *An ordered set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ has **positive orientation** if $\det V > 0$, where $V : \mathbb{R}^n \to \mathbb{R}^n$ is the linear map defined by the conditions $V(\mathbf{e}_i) = \mathbf{v}_i$ and $i = 1, 2, \ldots, n$. The set has **negative orientation** if $\det V < 0$.*

We can now construct the analogue of a parallelepiped, and define its volume and orientation, by extending the wedge product as follows.

n-parallelepipeds

Definition 2.4 *Let* $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ *be an ordered set of vectors in* $\mathbb{R}^n$*; the* ***oriented n-dimensional parallelepiped*** $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_n$ *is the set of vectors*

$$\mathbf{w} = \sum_{i=1}^{n} t_i \mathbf{v}_i, \quad 0 \le t_i \le 1, \quad i = 1, \dots, n.$$

Orientation and volume

Definition 2.5 *The* ***orientation*** *of* $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_n$ *is the orientation of the ordered set* $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$*; its* ***n-volume*** *(or just* ***volume****) is*

$$\mathrm{vol}(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_n) = \det V,$$

where V *is the matrix whose* j*th column consists of the coordinates of* $\mathbf{v}_j$*; that is,* $V(\mathbf{e}_j) = \mathbf{v}_j$*.*

$n \times n$ determinants

The volume of an n-parallelepiped can be either positive, negative, or zero. If the volume is zero, then $\det V = 0$, so the columns of V (which are the coordinates of the edges of the parallelepiped) are linearly dependent. The parallelepiped does not fill out an n-dimensional region in $\mathbb{R}^n$. Our final statement about volumes is the analogue of similar results in $\mathbb{R}^2$ and $\mathbb{R}^3$, and is proven the same way.

Theorem 2.16. *The* ***volume multiplier*** *of the linear map* $M : \mathbb{R}^n \to \mathbb{R}^n$ *is* $\det M$*; that is,*

$$\mathrm{vol}\, M(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n) = \mathrm{vol}\, M(\mathbf{v}_1) \wedge \cdots \wedge M(\mathbf{v}_n) = \det M \times \mathrm{vol}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n)$$

for every oriented n-parallelepiped $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n$*. The map* M *reverses orientation precisely when* $\det M < 0$*.* □

2.3 Maps from $\mathbb{R}^n$ to $\mathbb{R}^p$, $n \neq p$

Image and kernel subspaces are graphs

A good example of a map between spaces of the same dimension is a coordinate change. Of course, a coordinate change has to work both ways; that is, the map must be invertible. When the source and target have different dimensions, invertibility is out of the question, but the geometric action of such linear maps still has a simple description. When the source is larger, the map cannot be one-to-one: the kernel of the map must be a linear subspace of positive dimension in the source. When the target is larger, the map cannot be onto: the image must be a linear subspace of strictly smaller dimension than the target. As we show, each of these subspaces is the graph of a new linear map that is defined implicitly by the original one.

Rank–nullity theorem

When $L : \mathbb{R}^n \to \mathbb{R}^p$ is a linear map, the **kernel**, or **null space**, of L is the linear subspace $\ker L$ of the source $\mathbb{R}^n$ that consists of all vectors $\mathbf{v}$ for which $L(\mathbf{v}) = 0$. The **image** of L is the linear subspace $\mathrm{im}\, L$ of the target consisting of all vectors $\mathbf{x}$

of the form $\mathbf{x} = L(\mathbf{v})$, for some $\mathbf{v}$ in the source. We call $r = \dim \operatorname{im} L$ the **rank** of L and $k = \dim \ker L$ its **nullity**. The **rank–nullity theorem** of linear algebra says that

$$r + k = \text{rank of } L + \text{nullity of } L = \dim \text{source of } L = n.$$

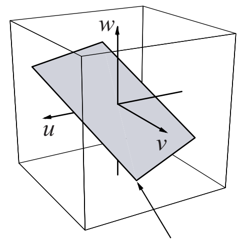
Kernel L has positive dimension when $n > p$

To analyze linear maps $L : \mathbb{R}^n \to \mathbb{R}^p$ for which $n \neq p$, let us first assume $n > p$. Because the image is a linear subspace of the target, we always have $r \leq p$. Because $n - p > 0$, the rank–nullity theorem implies $k = n - r \geq n - p > 0$. In other words, the kernel of L has positive dimension, at least as large as $n - p$. Let us now look more closely at $\ker L$.

Example: one equation in three variables

We begin with an example in which $n = 3$ and $p = 1$, so L has the general form $x = L(u, v, w) = au + bv + cw$. How can we describe $\ker L$? To illustrate, suppose $L(u, v, w) = u - 2v - 3w$. The kernel of L is the locus of points in (u, v, w)-space that satisfy the equation

$$u - 2v - 3w = 0.$$



$u - 2v - 3w = 0$

The figure shows this locus is a (2-dimensional) plane through the origin. We can solve the equation for w, for example, and get

$$w = \frac{u - 2v}{3}.$$

The original equation $u - 2v - 3w = 0$ therefore implies that w is a (linear) function of u and v. Thus, we can view the plane in the figure as either a *locus* (i.e., the locus of zeros of $L(u, v, w)$) or a *graph* (e.g., the graph of $w = (u - 2v)/3$).

Implicit functions

Of course, this is not the only functional relation that is implicit in the equation $u - 2v - 3w = 0$; we get others by solving for u or v:

$$u = 2v + 3w \quad \text{or} \quad v = \frac{u - 3w}{2}.$$

In each case, however, precisely one of the variables is expressed in terms of the other two.

$\ker L = \operatorname{graph} M$

We can say the same about an arbitrary linear function of three variables:

$$x = L(u, v, w) = au + bv + cw.$$

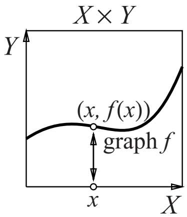
We have already seen that $\dim \ker L$ is at least $n - p = 2$. If $a = b = c = 0$, then every point satisfies the kernel equation $au + bv + cw = 0$, and $\dim \ker L = 3$. Otherwise, at least one of the coefficients is different from zero. Suppose $c \neq 0$; then we can solve the kernel equation for w in terms of u and v:

$$w = -\frac{a}{c}u - \frac{b}{c}v = M(u, v).$$

The linear function $M : \mathbb{R}^2 \to \mathbb{R}^1$ that expresses w in terms of u and v is implicitly defined by the equation $L(u, v, w) = 0$. In fact, the graph of M is the locus of the equation $L(u,v,w)=0$. The dimension of a graph is equal to the dimension of its source (see below); therefore $\dim\ker L = \dim \operatorname{graph} M = 2$.

Graphs in general

Because we have found that the kernel of one linear map can be the graph of another, we pause to state some facts about graphs generally.



Definition 2.6 *The* ***graph*** *of an arbitrary map $f : X \to Y$ is the subset of the product $X \times Y$ that is defined by*

$$\operatorname{graph} f = \{(x, f(x)) \,|\, x \in X\}.$$

The definition makes it clear that there is a 1–1 correspondence between the source X and graph f. If the map is linear, we can say more.

Theorem 2.17. *If $L : \mathbb{R}^n \to \mathbb{R}^p$ is linear, then* graph$L$ *is a linear subspace of the product $\mathbb{R}^n \times \mathbb{R}^p = \mathbb{R}^{n+p}$ and* $\dim \operatorname{graph} L = n$.

Proof. Do this as an exercise. □

Example: p equations in $p+k$ variables

Now consider a general linear map $L : \mathbb{R}^n \to \mathbb{R}^p : \mathbf{v} \mapsto \mathbf{x}$ with $n > p$. To begin, assume that L has maximal rank, so $r = p$ and $k = \dim\ker L = n - p$. If $A = (a_{ij})$ is the $p \times n$ matrix representing L, then the vector kernel equation $L(\mathbf{v}) = \mathbf{0}$ translates into a system of p ordinary equations in the $n = p + k$ unknowns $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and the coefficients a_{ij}:

$$\begin{aligned}
a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n &= 0,\\
a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n &= 0,\\
&\ \vdots\\
a_{p1}v_1 + a_{p2}v_2 + \cdots + a_{pn}v_n &= 0.
\end{aligned}$$

Our previous example suggests that we should be able to solve these equations so as to express p of the unknowns as linear functions of the remaining $k = n - p$.

Solving the equations

To solve the equations, note first that the rank of L is the number of linearly independent columns of A. By our assumption, A must have p linearly independent columns. By rearranging them (and with them the variables v_j), if necessary, we can assume that the final p columns of A are linearly independent. They form an invertible $p \times p$ submatrix, C. The initial $n - p = k$ columns form a $p \times k$ submatrix B. With these identifications, the kernel equations take the form

$$\begin{aligned}
\mathbf{0} &= A\mathbf{v}\\
&= (B\ C)\begin{pmatrix}\mathbf{u}\\ \mathbf{w}\end{pmatrix}\\
&= B\mathbf{u} + C\mathbf{w}
\end{aligned}$$

$$\begin{pmatrix}0\\0\\ \vdots\\0\end{pmatrix} = \left(\overbrace{\begin{matrix}a_{11} & \cdots & a_{1k}\\ a_{21} & \cdots & a_{2k}\\ \vdots & \ddots & \vdots\\ a_{p1} & \cdots & a_{pk}\end{matrix}}^{B_{p\times k}}\ \overbrace{\begin{matrix}a_{1,k+1} & \cdots & a_{1,k+p}\\ a_{2,k+1} & \cdots & a_{2,k+p}\\ \vdots & \ddots & \vdots\\ a_{p,k+1} & \cdots & a_{p,k+p}\end{matrix}}^{C_{p\times p}}\right)\begin{pmatrix}u_1\\ \vdots\\ u_k\\ w_1\\ \vdots\\ w_p\end{pmatrix}\begin{matrix}\left.\vphantom{\begin{matrix}u_1\\ \vdots\\ u_k\end{matrix}}\right\}\mathbf{u}\\ \left.\vphantom{\begin{matrix}w_1\\ \vdots\\ w_p\end{matrix}}\right\}\mathbf{w}\end{matrix} = (B\ C)\begin{pmatrix}\mathbf{u}\\ \mathbf{w}\end{pmatrix},$$

The implicit functions

or just $B\mathbf{u} + C\mathbf{w} = 0$ in terms of matrices. Because C is invertible, we can solve for $\mathbf{w}$:

$$\begin{aligned}
\mathbf{w} &= -C^{-1}B\,\mathbf{u} = M(\mathbf{u}),\\
w_1 &= \beta_{11}u_1 + \cdots + \beta_{1k}u_k,\\
w_2 &= \beta_{21}u_1 + \cdots + \beta_{2k}u_k,\\
&\ \ \vdots\\
w_p &= \beta_{p1}u_1 + \cdots + \beta_{pk}u_k.
\end{aligned}$$

ker L = graph M

This is what we want: the equation $\mathbf{w} = M(\mathbf{u})$ expresses p of the variables as linear functions of the remaining k. The $p \times k$ matrix $-C^{-1}B = (\beta_{ij})$ that represents M is constructed from certain submatrices of the matrix A that represents L. Finally, the argument shows that the graph of M is precisely the kernel of L, because

$$\mathbf{w} = M(\mathbf{u}) \ \text{ if and only if } \ L(\mathbf{u},\mathbf{w}) = B\mathbf{u} + C\mathbf{w} = \mathbf{0}.$$

That is, a point $(\mathbf{u},\mathbf{w})$ is in the graph of M (so $\mathbf{w} = M(\mathbf{u})$) if and only if it is in the kernel of L.

Notice, incidentally, how our general result echoes what we found in the first example, in which $L(u,v,w) = au + bv + cw$. We had $A = (a\ b\ c)$ (a 1×3 matrix), $B = (a\ b)$, and $C = (c)$, implying

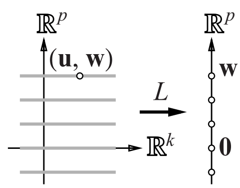
$$C^{-1} = \left(\frac{1}{c}\right) \quad \text{and} \quad -C^{-1}B = \left(-\frac{a}{c} \quad -\frac{b}{c}\right).$$

The following theorem summarizes our result in both an algebraic form involving equations, and a geometric form involving maps. The condition that the original equations are linearly independent implies that the rank of the associated linear map is p.

Theorem 2.18. Algebraically: *A set of p linearly independent linear equations in $k+p$ variables implicitly defines p of the variables as linear functions of the remaining k variables.* Geometrically: *The kernel of a linear map $L : \mathbb{R}^{k+p} \to \mathbb{R}^p$ of maximal rank p is the graph of another linear map $M : \mathbb{R}^k \to \mathbb{R}^p$.* □

One such map $L : \mathbb{R}^{k+p} \to \mathbb{R}^p$ of maximal rank p is just the identity on the last p variables:

$$\begin{aligned}
w_1 &= v_{k+1} = x_1,\\
w_2 &= v_{k+2} = x_2,\\
&\ \ \vdots\\
w_p &= v_{k+p} = x_p,
\end{aligned} \qquad (O_{p\times k}\ I_{p\times p})\begin{pmatrix}\mathbf{u}\\ \mathbf{w}\end{pmatrix} = \mathbf{x}.$$



The $p \times k$ zero matrix $O_{p\times k}$ eliminates the $\mathbf{u}$ variables $u_1 = v_1, \ldots, u_k = v_k$ from the formulas. Geometrically, L projects $\mathbb{R}^{k+p} = \mathbb{R}^k \times \mathbb{R}^p$ onto its second factor. It projects the first factor $\mathbb{R}^k$ (the kernel of L) to $\mathbf{0}$, and it projects the parallel translate of $\mathbb{R}^k$ by the vector $(\mathbf{u},\mathbf{w})$ (for an arbitrary $\mathbf{u}$) to the point $\mathbf{w}=\mathbf{x}$ in the target. Although this example may seem special, the following theorem shows that, in a sense, it is the only possibility. Because the theorem is geometric, it is helpful to write the linear map in its original form $L:\mathbb{R}^n \to \mathbb{R}^p$, keeping in mind it is an onto map with $n>p$.

Every *onto* map is a projection

Theorem 2.19. *If the linear map $L:\mathbb{R}^n \to \mathbb{R}^p$ is onto, then there is a linear coordinate change H in the source $\mathbb{R}^n$ that transforms L into the projection that is the identity on the final p variables.*

Proof. We assume, as in the proof of Theorem 2.18, that variables in the source have been permuted so that L has the form

$$L=\begin{pmatrix}B_{p\times k} & C_{p\times p}\end{pmatrix}, \quad \mathbf{x}=L(\mathbf{u},\mathbf{w})=\begin{pmatrix}B & C\end{pmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{w}\end{pmatrix}=B\mathbf{u}+C\mathbf{w},$$

where $k=n-p\geq 0$ and the square submatrix C is invertible. Define $H:\mathbb{R}^n \to \mathbb{R}^n$: $(\overline{\mathbf{u}},\overline{\mathbf{w}})\mapsto(\mathbf{u},\mathbf{w})$ as the pullback

$$H:\begin{cases}\mathbf{u}=\overline{\mathbf{u}},\\ \mathbf{w}=-C^{-1}B\overline{\mathbf{u}}+C^{-1}\overline{\mathbf{w}},\end{cases} \qquad H=\begin{pmatrix}I_{k\times k} & O_{k\times p}\\ -(C^{-1}B)_{p\times k} & C^{-1}_{p\times p}\end{pmatrix}.$$

Now H is invertible,

$$H^{-1}=\begin{pmatrix}I & O\\ B & C\end{pmatrix},$$

so H is a valid coordinate change. Applying H to $\mathbf{x}=B\mathbf{u}+C\mathbf{w}$ gives

$$\mathbf{x}=B\overline{\mathbf{u}}+C\left[-C^{-1}B\overline{\mathbf{u}}+C^{-1}\overline{\mathbf{w}}\right]=B\overline{\mathbf{u}}-B\overline{\mathbf{u}}+\overline{\mathbf{w}}=\overline{\mathbf{w}}$$

(i.e., $\mathbf{x}=\overline{\mathbf{w}}$). The matrix for L in the new coordinates is

$$\overline{L}=LH=\begin{pmatrix}B & C\end{pmatrix}\begin{pmatrix}I & O\\ -C^{-1}B & C^{-1}\end{pmatrix}=\begin{pmatrix}O & I\end{pmatrix}.$$

Thus $\overline{L}$ is the identity on the second component of $(\overline{\mathbf{u}},\overline{\mathbf{w}})$; it projects $(\overline{\mathbf{u}},\overline{\mathbf{w}})$ to $\overline{\mathbf{w}}$. □

Corollary 2.20 *If the linear map $L:\mathbb{R}^n \to \mathbb{R}^p$ is onto, and Y is a linear subspace of the target of dimension q, then its preimage*

$$L^{-1}(Y)=\{\mathbf{v}\ \textit{in}\ \mathbb{R}^n : L(\mathbf{v})\ \textit{is in}\ Y\}$$

is a linear subspace of dimension $q+k$, where $k=n-p\geq 0$.

Proof. The theorem provides new coordinates $(\overline{\mathbf{u}},\overline{\mathbf{w}})$ in $\mathbb{R}^k\times\mathbb{R}^p=\mathbb{R}^n$ in which L becomes a projection onto the second factor. Then

$$L^{-1}(Y)=\{(\overline{\mathbf{u}},\overline{\mathbf{w}}) : \overline{\mathbf{w}}\text{ is in }Y\}=\mathbb{R}^k\times Y,$$

because $\overline{\mathbf{u}}$ is an arbitrary point in $\mathbb{R}^k$. Hence $\dim L^{-1}(Y) = k + \dim Y$. Standard arguments in linear algebra show that $L^{-1}(Y)$ is a linear subspace. □

Codimension

It is intuitively clear that, if L projects a larger space onto a smaller one, the pullback $L^{-1}(Y)$ will always be larger than the original Y. The corollary says that the difference is equal to the difference in dimension of the spaces themselves:

$$\dim L^{-1}(Y) - \dim Y = k = \dim \mathbb{R}^n - \dim \mathbb{R}^p.$$

But this means

$$\dim \mathbb{R}^n - \dim L^{-1}(Y) = \dim \mathbb{R}^p - \dim Y.$$

Definition 2.7 *The* ***codimension*** *of a linear subspace Y of a vector space V is*

$$\operatorname{codim} Y = \dim V - \dim Y.$$

Think of *codimension* as "dimension of the complement" or "complement's dimension." A complement of Y in V is a linear subspace Z for which $V = Y \times Z$. Obviously, $\dim Z = \dim V - \dim Y = \operatorname{codim} Y$. In the previous corollary, it is thus more useful to work with the *codimension* of a linear subspace than its *dimension*, because pullback alters dimension but preserves codimension.

Corollary 2.21 *If the linear map $L : \mathbb{R}^n \to \mathbb{R}^p$ is onto, and Y is a linear subspace of $\mathbb{R}^p$, then* $\operatorname{codim} L^{-1}(Y) = \operatorname{codim} Y$. □

Classifying linear maps that are *onto*

Theorem 2.19 provides a classification of onto linear maps that is similar to the classification of linear maps of the plane by Theorem 2.6 (p. 40). There we found several classes, each with a typical representative (that shares eigenvalues with all members of the class). Here there is only a single class, and projection is chosen as the typical representative of that class.

Assume rank is not maximal

It remains to consider the kernel of $L : \mathbb{R}^n \to \mathbb{R}^p$ when the rank of L is no longer maximal. In this case it turns out not to matter that $n > p$. We can illustrate this with a simple example.

$$\begin{aligned} u \qquad - \ \ w &= 0, \\ v - 2w &= 0, \\ u + v - 3w &= 0, \\ u - v + \ \ w &= 0. \end{aligned}$$

These four equations in three variables describe the kernel of a particular linear map $L : \mathbb{R}^3 \to \mathbb{R}^4$. The maximum possible rank of L is 3. However, the actual rank is only $r = 2$: only two of the four equations are linearly independent; the other two are linear combinations of those. By the rank–nullity theorem, the dimension of the kernel of L is $k = n - r = 1$. We expect, therefore, that the kernel of L is the graph of some other linear map $M : \mathbb{R}^1 \to \mathbb{R}^2$. Indeed, the kernel equations imply

$$M : \begin{cases} u = w, \\ v = 2w. \end{cases}$$

This is the linear map we seek. The following theorem generalizes this example; it makes no assumption about the relative sizes of n and p.

$\ker L$ is a graph when $r < n$

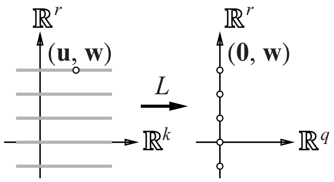
Theorem 2.22 (Linear implicit function theorem). *If $L : \mathbb{R}^n \to \mathbb{R}^p$ has rank $r < n$, then $\ker L$ in $\mathbb{R}^n$ is the graph of a linear map $M : \mathbb{R}^{n-r} \to \mathbb{R}^r$.*

Proof. We consider cases; in every case, $r \le p$. If $r = p$, Theorem 2.18 applies. If $r = 0$, then L is identically zero and $\ker L = \mathbb{R}^n$. Therefore, the "zero" linear map $M : \mathbb{R}^n \to \mathbb{R}^0 : \mathbf{v} \mapsto 0$ has $\operatorname{graph} M = \ker L$.

The only remaining possibility is $0 < r < p$. Then $q = p - r > 0$ of the equations that determine $\ker L$ depend linearly on the remaining r equations. Select q dependent equations and discard them. Then use the remaining $p - q = r > 0$ equations to define a new linear map $L^* : \mathbb{R}^n \to \mathbb{R}^r$. Because the discarded kernel equations add no information, $\ker L^* = \ker L$.

By construction, L^* does have maximal rank r, so Theorem 2.18 applies again: the kernel equations for L^* implicitly define r of the variables $v_1, \ldots, v_n$ as linear functions of the remaining $n - r$ variables. In other words, $\ker L^* = \ker L$ is the graph of a linear map $M : \mathbb{R}^{n-r} \to \mathbb{R}^r$. □

One linear map $L : (u_1, \ldots, u_k, w_1, \ldots, w_r) \mapsto (x_1, \ldots, x_q, y_1, \ldots, y_r)$ of rank r is given by



$$L : \begin{cases} x_1 = 0, \\ \quad \vdots \\ x_q = 0, \\ y_1 = w_1, \\ \quad \vdots \\ y_r = w_r, \end{cases} \qquad L = \begin{pmatrix} O_{q\times k} & O_{q\times r} \\ O_{r\times k} & I_{r\times r} \end{pmatrix}.$$

To make this different from the previous example (p. 49), it is sufficient to require $q > 0$. The kernel of L consists of the points $(\mathbf{u}, \mathbf{0})$, and the image of L consists of the points $(\mathbf{0}, \mathbf{y})$. If we write the source as $\mathbb{R}^k \times \mathbb{R}^r$ and the target as $\mathbb{R}^q \times \mathbb{R}^r$, then L is the projection that is the identity on the second components:

$$L : (\mathbf{u}, \mathbf{w}) \mapsto (\mathbf{0}, \mathbf{w}).$$

The next theorem shows that this example is essentially the only possibility.

Theorem 2.23. *Let $L : \mathbb{R}^n \to \mathbb{R}^p$ be a linear map of rank r, and let $k = n - r$, $q = p - r$. Then there are coordinates in the source and target for which the matrix representing L is*

$$\Pi = \begin{pmatrix} O_{q\times k} & O_{q\times r} \\ O_{r\times k} & I_{r\times r} \end{pmatrix}.$$

Proof. We obtain coordinates for the source and for the target in which the map L is represented by multiplication by the given matrix Π. According to the rank–nullity theorem, $k = n - r$ is the dimension of the kernel of L. Let $\{U_1, \ldots, U_k\}$ be any basis for the kernel; add additional vectors $W_1, \ldots, W_r$ so that

$$\{U_1, \ldots, U_k, W_1, \ldots, W_r\}$$

is a basis for the entire source $\mathbb{R}^{k+r}$. Then any vector $\mathbf{v}$ in $\mathbb{R}^{k+R}$ can be written as

$$\mathbf{v} = u_1 U_1 + \cdots + u_k U_k + w_1 W_1 + \cdots + w_r W_r,$$

and (inasmuch as every $L(U_i) = 0$)

$$L(\mathbf{v}) = w_1 L(W_1) + \cdots + w_r L(W_r).$$

Because $\mathbf{v}$ is an arbitrary vector in the source, the vectors $L(\mathbf{v})$ constitute the entire image of L, and the equation for $L(\mathbf{v})$ shows that the vectors $L(W_j)$ span the image. Because the image has dimension r, the vectors $Y_j = L(W_j)$, $j = 1, \ldots, r$ must, in fact, form a basis for the image. Add additional vectors $X_1, \ldots, X_q$ so that

$$\{X_1, \ldots, X_q, Y_1, \ldots, Y_r\}$$

is a basis for the entire target $\mathbb{R}^{q+r}$. Then, in terms of these two bases, the coordinates of $\mathbf{v}$ and $L(\mathbf{v})$ are

$$\begin{aligned} \mathbf{v} &\leftrightarrow (u_1, \ldots, u_k, w_1, \ldots, w_r) = (\mathbf{u}, \mathbf{w}), \\ L(\mathbf{v}) &\leftrightarrow (0, \ldots, 0, w_1, \ldots, w_r) = (\mathbf{0}, \mathbf{w}). \end{aligned}$$

Multiplication by Π gives

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} O & O \\ O & I \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{w} \end{pmatrix};$$

thus Π does indeed represent L in terms of these coordinates. □

When is $\mathrm{im}\, L$ a graph?

We now switch our attention from kernels to images. We show that when $\mathrm{im}\, L$ is a proper subspace of the target of L, it too can be thought of as the graph of a linear map implicitly defined by L. As we did for kernels, we assume first that L has maximal rank.

Example: three equations in two variables

We begin with an example. Consider the linear map $L : \mathbb{R}^2 \to \mathbb{R}^3 : (u, v) \mapsto (x, y, z)$ given by

$$\begin{aligned} au + bv &= x, \\ cu + dv &= y, \\ eu + fv &= z. \end{aligned}$$

If L has maximum rank, namely 2, then precisely two of the equations are linearly independent. By rearranging them, if necessary, we may assume the first two equations are. Then we can solve these two equations for u and v in terms of x and y. It is perhaps easiest to see this if we work with matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{D}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

The implicit function

Here $D = ad - bc$, the determinant; linear independence of the first two equations implies $D \neq 0$. We can now express z directly in terms of x and y:

$$z = e\frac{dx - by}{D} + f\frac{-cx + ay}{D} = \frac{de - cf}{D}x + \frac{af - be}{D}y = M(x,y).$$

In geometric terms, the image of L is a plane in the target, $\mathbb{R}^3$. We have shown that if (x,y,z) are the coordinates of a point in this plane, then z is not independent of x and y, but depends linearly on them. The points of the image plane are of the form $(x,y,M(x,y))$; in other words, the image of L is the graph of M.

Example: $n+q$ equations in n variables

To generalize our example, take a linear map $L : \mathbb{R}^n \to \mathbb{R}^{n+q}$, $q > 0$, with maximal rank $r = n$. Let $A = (a_{ij})$ be the matrix representing L. Assume, by rearranging the rows of A, if neccesary, that the *first* n rows are linearly independent. Then, if we write the equation $L(\mathbf{v}) = A\mathbf{x} = \mathbf{x}$ in the form,

$$\begin{array}{c} B\left\{\begin{array}{c}\\ \\ \\ \end{array}\right. \\ C\left\{\begin{array}{c}\\ \\ \\ \end{array}\right. \end{array}\!\!\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \\ a_{n+1,1} & \cdots & a_{n+1,n} \\ \vdots & \ddots & \vdots \\ a_{n+q,1} & \cdots & a_{n+q,n} \end{pmatrix}\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ z_{n+1} \\ \vdots \\ z_{n+q} \end{pmatrix}\!\!\begin{array}{c} \left.\begin{array}{c}\\ \\ \\ \end{array}\right\}\mathbf{y} \\ \left.\begin{array}{c}\\ \\ \\ \end{array}\right\}\mathbf{z} \end{array} \qquad \begin{pmatrix} B \\ C \end{pmatrix}\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix},$$

we can express this in terms of the submatrices B and C as a pair of matrix equations:

$$B\mathbf{v} = \mathbf{y}, \quad C\mathbf{v} = \mathbf{z}.$$

The implicit functions

We rearranged the rows of A to guarantee that the $n \times n$ submatrix B is invertible. Hence we can solve for $\mathbf{v}$ in terms of $\mathbf{y}$: $\mathbf{v} = B^{-1}\mathbf{y}$ and we can then express $\mathbf{z}$ in terms of $\mathbf{y}$:

$$\mathbf{z} = CB^{-1}\mathbf{y}.$$

In terms of the components $\mathbf{y}$, $\mathbf{z}$ of $\mathbf{x}$, the equation $\mathbf{z} = CB^{-1}\mathbf{y}$ takes the form

$$\begin{aligned} z_{n+1} &= \gamma_{11}y_1 + \cdots + \gamma_{1n}y_n, \\ z_{n+2} &= \gamma_{21}y_1 + \cdots + \gamma_{2n}y_n, \\ &\;\;\vdots \\ z_{n+q} &= \gamma_{q1}y_1 + \cdots + \gamma_{qn}y_n. \end{aligned}$$

In this way we have expressed the image of L as a system of q linear equations in n variables. The coefficients γ_{ij} in these equations are the components of the $q \times n$ matrix CB^{-1} that represents a certain linear map $M : \mathbb{R}^n \to \mathbb{R}^q$. We see that a point $(\mathbf{y}, \mathbf{z})$ is in the image of L if and only if $\mathbf{z} = M(\mathbf{y})$, that is, if and only if it is of the form $(\mathbf{y}, M(\mathbf{y}))$. These are precisely the points of the graph of M.

Theorem 2.24. *Suppose the linear map $L : \mathbb{R}^n \to \mathbb{R}^{n+q}$ has maximal rank n. Then the image of L in $\mathbb{R}^{n+q}$ is the graph of a linear map $M : \mathbb{R}^n \to \mathbb{R}^q$.* □

Volume multiplier

In these circumstances, the image $L(\mathbb{R}^n)$ is n-dimensional. Therefore, if P is a parallelpiped in the source and $L(P)$ is its image, both are n-dimensional, and both have n-volume. It is natural, then, to ask what is the volume multiplier for L. Of course, because the target dimension is larger than n, the image parallelepiped $L(P)$ cannot be oriented and the sign of its n-volume will have no meaning. Thus, we always understand the volume multiplier for L to be nonnegative. Let us consider first the special case $L : \mathbb{R}^2 \to \mathbb{R}^3$ when the image is a 2-dimensional plane.

$$L = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

Theorem 2.25. *Suppose $L : \mathbb{R}^2 \to \mathbb{R}^3$ has maximal rank 2 and is represented by the 3×2 matrix (a_{ij}), $i = 1,2,3$, $j = 1,2$. Then*

$$\textit{area multiplier of } L = \sqrt{\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}^2 + \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}^2 + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}^2}.$$

Proof. By linearity, the area multiplier will equal the area of the image of a parallelogram of unit area. In particular, take P to be the unit square

$$P = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \text{ then } L(P) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} \wedge \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}.$$

It follows from page 44 that the area of the parallelogram $L(P)$ is

$$\sqrt{\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}^2 + \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}^2 + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}^2}.$$ □

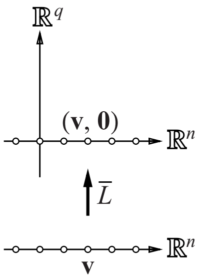
According to Exercise 2.26.c. (p. 64), the area of the same parallelogram $L(P)$ can also be written as $\sqrt{\det L^{\dagger} L}$. More generally, then, Exercise 2.36 and the discussion leading up to it establish the following ("Pythagorean") theorem.

Theorem 2.26. *Suppose $L : \mathbb{R}^n \to \mathbb{R}^{n+q}$ has maximal rank n; then the n-volume multiplier for L is $\sqrt{\det L^{\dagger} L}$.* □

Splitting the target of a 1–1 map

Because the rank of L equals the dimension of the source, the kernel of L is zero so L is a 1–1 map. The next theorem says we can split the target into two factors so that L is just the identity mapping onto the first factor. This is a special case of Theorem 2.23, but it is worth having another proof that follows different lines.

Theorem 2.27. *Suppose* $L : \mathbb{R}^n \to \mathbb{R}^{n+q}$ *is* 1–1*; then a coordinate change H in the target transforms L into the matrix*



$$\overline{L} = \begin{pmatrix} I_{n\times n} \\ O_{q\times n} \end{pmatrix}.$$

Proof. To begin, we assume (as in the proof of Theorem 2.24) that L has the matrix form

$$\begin{pmatrix} B \\ C \end{pmatrix} \mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}, \quad \text{that is,} \quad \begin{matrix} B\mathbf{v} = \mathbf{y}, \\ C\mathbf{v} = \mathbf{z}, \end{matrix}$$

with B an $n \times n$ invertible matrix. Define $H : \mathbb{R}^{n+q} \to \mathbb{R}^{n+q}$ by

$$H : \begin{cases} \overline{\mathbf{y}} = B^{-1}\mathbf{y}, \\ \overline{\mathbf{z}} = -CB^{-1}\mathbf{y} + \mathbf{z}, \end{cases} \qquad H = \begin{pmatrix} B^{-1} & O \\ -CB^{-1} & I \end{pmatrix}$$

Because $\det H = \det(B^{-1}) \neq 0$, H is a valid coordinate change. Moreover,

$$\overline{L} = HL = \begin{pmatrix} B^{-1} & O \\ -CB^{-1} & I \end{pmatrix} \begin{pmatrix} B \\ C \end{pmatrix} = \begin{pmatrix} I \\ O \end{pmatrix}, \quad \begin{matrix} \overline{\mathbf{y}} = B^{-1}(B\mathbf{v}) = \mathbf{v}, \\ \overline{\mathbf{z}} = -CB^{-1}(B\mathbf{v}) + C\mathbf{v} = \mathbf{0}. \end{matrix}$$ □

Injections

By analogy with projections, a linear map with the special form of $\overline{L}$ is sometimes called an **injection**. The theorem thus says that every 1–1 linear map is (equivalent to) an injection. The following corollary stands in the same relation to Theorem 2.27 that Corollaries 2.20 and 2.21 do to Theorem 2.19. It says that a 1–1 linear map preserves dimension under push-forward.

Corollary 2.28 *If the linear map* $L : \mathbb{R}^n \to \mathbb{R}^p$ *is* 1–1*, and X is a linear subspace of dimension k in the source, then* $L(X)$ *is a linear subspace of the same dimension k in the target.*

Proof. In terms of the $(\overline{\mathbf{y}}, \overline{\mathbf{z}})$ coordinates in the proof of Theorem 2.27, the linear map L becomes the injection $(\overline{\mathbf{y}}, \overline{\mathbf{z}}) = \overline{L}(\mathbf{v}) = (\mathbf{v}, \mathbf{0})$. Thus, for any subspace X, we have $\overline{L}(X) = X \times \mathbf{0}$, implying $\dim L(X) = \dim \overline{L}(X) = \dim X$. □

Assume rank is not maximal

The final possibility to consider is the image of a linear map $L : \mathbb{R}^n \to \mathbb{R}^p$ whose rank may not be maximal. In this case we need make no assumption about the relative sizes of n and p.

Theorem 2.29. *If* $L : \mathbb{R}^n \to \mathbb{R}^p$ *has rank* $r < p$*, then* $\operatorname{im} L$ *in* $\mathbb{R}^p$ *is the graph of a linear map* $M : \mathbb{R}^r \to \mathbb{R}^{p-r}$.

Proof. As a preliminary step, write $L : \mathbb{R}^n \to \mathbb{R}^{r+q}$, where $q = p - r$ and $q > 0$ by hypothesis. The rest of the proof now proceeds by analogy with the proof of Theorem 2.22.

In every case, $r \leq n$. If $r = n$, the previous theorem applies. If $r = 0$, then L is identically 0 and $\operatorname{im} L = \mathbb{R}^0$. In this case, $\operatorname{im} L$ is the graph of the linear map $M : \mathbb{R}^0 \to \mathbb{R}^p : 0 \mapsto (0, 0, \ldots, 0)$.

The only remaining possibility is $0 < r < n$. Let A be the $p \times n$ matrix that represents L. The n columns of A are elements of $\mathbb{R}^p$ that span $\operatorname{im} L$ in $\mathbb{R}^p$. Our assumption implies that only r of the n columns are linearly independent; the remaining $n - r > 0$ columns depend linearly upon these. Delete the dependent columns from A to create a new matrix A^* of size $p \times r = (r + q) \times r$, and let $L^* : \mathbb{R}^r \to \mathbb{R}^{r+q}$ be the linear map defined by A^*. Because the columns of A^* and A span the same r-dimensional linear subspace of $\mathbb{R}^p = \mathbb{R}^{r+q}$, we have $\operatorname{im} L^* = \operatorname{im} L$.

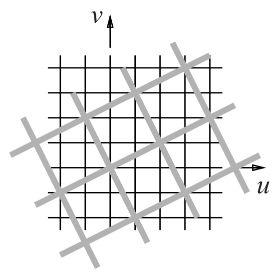
By construction, L^* has maximal rank r, so Theorem 2.24 applies: $\operatorname{im} L^* = \operatorname{im} L$ is the graph of a linear map $M : \mathbb{R}^r \to \mathbb{R}^q = \mathbb{R}^{p-r}$. □

Exercises

2.1. Show that the linear map M_1 (p. 29) alters the slope of any line that is neither horizontal nor vertical. Specifically, show that if a line has slope $\Delta v / \Delta u = m$, its image has slope $\Delta y / \Delta x = 3m/10 \neq m$ if $m \neq 0, \infty$.

2.2. Determine the coordinate change from (u, v) to $(\overline{u}, \overline{v})$ (and from (x, y) to $(\overline{x}, \overline{y})$) that converts the matrix M_5 into $\overline{M_5}$ (cf. p. 34)).

2.3. Let A, B, and C be $n \times n$ matrices. Suppose C is equivalent to B (cf. Definition 2.1, p. 33) and B is equivalent to A; show that C is equivalent to A.



2.4. This question concerns the map $M : \mathbb{R}^2 \to \mathbb{R}^2 : (u, v) \mapsto (x, y)$ defined by the matrix

$$M = \begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}.$$

a. Determine the area multiplier for M.
b. Sketch in the (x, y)-plane the image of the standard unit grid from the (u, v)-plane.
c. Show that the *image* of the line in the direction of the vector $(u, v) = (2, 1)$ is the line in the direction of the vector $(x, y) = (2, 1)$ (in the (x, y)-plane). In other words, show that this line is invariant under M.
d. Show that the same is true for the line in the direction of the vector $(u, v) = (-1, 2)$.
e. Sketch in the (x, y)-plane the image of the solid gray grid shown in the (u, v)-plane. (Notice that the lines in this grid are parallel to the lines from parts (c) and (d).)
f. What is the shape of the image of a single square from the new grid? What are the dimensions of that shape if you use that solid gray grid to define new unit lengths? What is the area of that shape in terms of these new units? What is the area multiplier in terms of these new units?

2.5. Determine the eigenvalues and eigenvectors of each of the following matrices/maps.

a. $M = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}$ b. $M = \begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}$ c. $M = \begin{pmatrix} 0 & \sqrt{6} \\ \sqrt{6} & -1 \end{pmatrix}$

2.6. Carry out an analysis similar to what you did in Exercise 2.4 for the linear map defined by each of the the following matrices.

a. $M = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}$ b. $M = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}$

In particular, construct a grid whose image is "parallel to itself." Note that, in the second case, the grid consists of parallelograms rather than rectangles (or squares). Determine the **linear multipliers** for the map and show that the sides of the grid are stretched by these factors. Determine the **area multiplier** for the map and indicate how your diagram confirms that value. Comment on how the map affects orientation.

2.7. Consider the following from Example 8 (p. 38):

$$M_8 = \begin{pmatrix} 4 & -2 \\ 2 & 0 \end{pmatrix}, \quad \overline{\mathbf{u}} = \overline{\mathbf{x}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \overline{\mathbf{v}} = \overline{\mathbf{y}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

a. You know M_8 has the repeated eigenvalue 2, and that therefore the kernel of the matrix $M_8 - 2I$ contains all the eigenvectors of M_8. Show that the dimension of the image of $M_8 - 2I$ is 1; by the rank–nullity theorem, the dimension of the set of eigenvectors of M_8 is only 1, not 2.

b. Show that $M_8(\overline{\mathbf{u}}) = 2\overline{\mathbf{x}}$, $M_8(\overline{\mathbf{v}}) = 2\overline{\mathbf{x}} + 2\overline{\mathbf{y}}$. In particular, identify an eigenvector of M_8.

c. Let $(\overline{u}, \overline{v})$ be coordinates based on $\overline{\mathbf{u}}$ and $\overline{\mathbf{v}}$, and let $(\overline{x}, \overline{y})$ be coordinates based, in the same way, on $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$. Show that the coordinate change from the standard basis to the new one is

$$G : \begin{cases} u = \overline{u} + \overline{v}, & x = \overline{x} + \overline{y}, \\ v = \overline{u}, & y = \overline{x}. \end{cases}$$

d. Show that, in terms of the new coordinates, M_8 becomes

$$\overline{M_8} : \begin{cases} \overline{x} = 2\overline{u} + 2\overline{v}, \\ \overline{y} = 2\overline{v}, \end{cases} \qquad \overline{M_8} = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}.$$

2.8. a. Show that the complex numbers $\mathrm{e}^{\pm \mathrm{i}\theta} = \cos\theta \pm \mathrm{i}\sin\theta$ are the eigenvalues of the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

b. Show that R_θ rotates the plane by θ radians by showing that

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} r\cos\alpha \\ r\sin\alpha \end{pmatrix} = \begin{pmatrix} r\cos(\alpha+\theta) \\ r\sin(\alpha+\theta) \end{pmatrix}.$$

In other words, R_θ maps the point with polar coordinates (r,α) to the point with polar coordinates $(r,\alpha+\theta)$. Explain why this implies R_θ has no real eigenvectors when $\theta \neq n\pi$, n integer.

c. Find a (complex) eigenvector for each of the eigenvalues of R_θ, $\theta \neq n\pi$.

2.9. Show that the only matrix equivalent (cf. Definition 2.1, p. 33) to the uniform dilation $D = \lambda I$ is D itself.

2.10. Let $\arctan(y/x)$, viewed as a function of two variables, be defined in terms of the usual arctangent function for all $(x,y) \neq (0,0)$ as follows:

$$\arctan(y/x) = \begin{cases} \arctan(y/x), & 0 < x, \\ \pi/2, & x = 0,\ 0 < y, \\ -\pi/2, & x = 0,\ y < 0, \\ \arctan(y/x) - \pi, & x < 0,\ y < 0, \\ \arctan(y/x) + \pi, & x < 0,\ 0 \leq y. \end{cases}$$

a. Show that $\arctan(y/x)$ is continuous across the y-axis, and is thus continuous on $\mathbb{R}^2 \setminus \{(x,0) | x \leq 0\}$; this is the plane with the origin and the negative x-axis deleted.

b. The graph $z = \arctan(y/x)$ is a *spiral ramp*; sketch it.

2.11. Suppose the 2×2 matrix M has real unequal eigenvalues λ_1 and λ_2, with corresponding eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$.

a. Explain why $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent.

b. Let G be the matrix whose columns are $\mathbf{u}_1$ and $\mathbf{u}_2$, in that order. Explain why G is invertible, and then show

$$G^{-1}MG = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

2.12. Suppose the 2×2 matrix M has the repeated eigenvalue $\lambda \neq 0$, but has only a single eigendirection, along the eigenvector $\mathbf{u}$. The purpose of this exercise is to show that M is equivalent to the standard shear–dilation matrix

$$S_\lambda = \begin{pmatrix} \lambda & \lambda \\ 0 & \lambda \end{pmatrix}.$$

a. Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the standard basis vectors. Show by direct computation that the vectors $(M - \lambda I)\mathbf{e}_1$ and $(M - \lambda I)\mathbf{e}_2$ are both eigenvectors of M. Conclude that $(M - \lambda I)\mathbf{w}$ is an eigenvector of M for every $\mathbf{w}$ in $\mathbb{R}^2$. Suggestion: The vectors $(M - \lambda I)\mathbf{e}_1$ and $(M - \lambda I)\mathbf{e}_2$ are the columns of the matrix

$$M-\lambda I=\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}.$$

b. The image of $M-\lambda I$ is 1-dimensional and contains $\lambda\mathbf{u}$; why? Conclude that there is a vector $\mathbf{v}$ for which $(M-\lambda I)\mathbf{v}=\lambda\mathbf{u}$. Explain why $\mathbf{u}$ and $\mathbf{v}$ are linearly independent.

c. Let G be the matrix whose columns are $\mathbf{u}$ and $\mathbf{v}$. Explain why G is invertible, and show that $G^{-1}MG=S_\lambda$.

2.13. Suppose the real 2×2 matrix M has complex eigenvalues $a\pm b\mathrm{i}$, $b\neq 0$, and the real vectors $\mathbf{u}$ and $\mathbf{v}$ form the complex eigenvector $\mathbf{u}+\mathrm{i}\mathbf{v}$ for M with eigenvalue $a-b\mathrm{i}$ (note the difference in signs). The purpose of this exercise is to show that M is equivalent to the standard rotation–dilation matrix $C_{a,b}$ (cf. p. 39).

a. Show that the following *real* matrix equations are true:

$$M\mathbf{u}=a\mathbf{u}+b\mathbf{v},\quad M\mathbf{v}=-b\mathbf{u}+a\mathbf{v}.$$

b. Let G be the matrix whose columns are $\mathbf{u}$ and $\mathbf{v}$, in that order. Show that $MG=GC_{a,b}$.

c. Show that the real vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly independent in $\mathbb{R}^2$. Suggestion: first show $\mathbf{u}\neq\mathbf{0}$, $\mathbf{v}\neq\mathbf{0}$. Then suppose there are real numbers r, s for which $r\mathbf{u}+s\mathbf{v}=\mathbf{0}$. Show that $\mathbf{0}=M(r\mathbf{u}+s\mathbf{v})$ implies that $-s\mathbf{u}+r\mathbf{v}=\mathbf{0}$, and hence that $r=s=0$.

d. Conclude that G is invertible and $G^{-1}MG=C_{a,b}$.

2.14. Notice that, in Exercise 2.6, the map whose invariant grid was rectangular (and hence whose eigenvectors were orthogonal) was the one whose matrix was *symmetric*. A matrix is symmetric if it is equal to its own transpose; the 2×2 symmetric matrices are

$$M=\begin{pmatrix} a & b \\ b & c \end{pmatrix}\qquad a,\ b,\text{ and } c \text{ real.}$$

The purpose of this exercise is to show that a symmetric matrix always has linearly independent orthogonal eigenvectors that define an invariant rectangular grid.

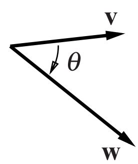
a. Show that the eigenvalues of a 2×2 symmetric matrix M are real. (Suggestion: Look at the formula for the roots of the quadratic equation that gives the eigenvalues and focus on the part of the formula that leads to complex values.)

b. Suppose a 2×2 symmetric matrix M has unequal eigenvalues. Show that the eigenvectors that correspond to these eigenvalues are orthogonal (e.g., their dot product is zero).

c. Suppose a 2×2 symmetric matrix M has equal eigenvalues. Show that this can happen only if $b=0$ and $a=c$. (Again, look at the formula for the roots.) This implies M must reduce to a multiple of the identity matrix.

In the last case (c), every nonzero vector is an eigenvector. Therefore, for every symmetric 2×2 matrix M, $\mathbb{R}^2$ has a basis consisting of orthogonal eigenvectors of M. Thus if a linear map $M:\mathbb{R}^2\to\mathbb{R}^2$ is represented by a symmetric matrix, there is a grid of squares that is stretched into a parallel grid of rectangles (as happened in Exercises 2.4 and 2.6a).

2.15. The purpose of this exercise is to show that the area of $\mathbf{v}\wedge\mathbf{w}$ is the determinant of the matrix whose columns are the coordinates of $\mathbf{v}$ and $\mathbf{w}$.



a. Let θ be the angle between $\mathbf{v}$ and $\mathbf{w}$, taken from $\mathbf{v}$ to $\mathbf{w}$ and chosen so that $-\pi<\theta\le\pi$. Thus, θ is negative precisely when $\mathbf{v}\wedge\mathbf{w}$ has negative orientation. Show that $\mathrm{area}(\mathbf{v}\wedge\mathbf{w})=\|\mathbf{v}\|\|\mathbf{w}\|\sin\theta$. (Notice that this has a negative value when $\mathbf{v}\wedge\mathbf{w}$ has negative orientation.)

b. Show that $\mathrm{area}^2(\mathbf{v}\wedge\mathbf{w})=\|\mathbf{v}\|^2\|\mathbf{w}\|^2-(\mathbf{v}\cdot\mathbf{w})^2$.

c. Let $\mathbf{v}=(v_1,v_2)$, $\mathbf{w}=(w_1,w_2)$. Show that

$$\|\mathbf{v}\|^2\|\mathbf{w}\|^2-(\mathbf{v}\cdot\mathbf{w})^2=(v_1w_2-v_2w_1)^2,$$

and hence that $\mathrm{area}(\mathbf{v}\wedge\mathbf{w})=\pm(v_1w_2-v_2w_1)$.

d. Show that requiring the area of the unit square $(1,0)\wedge(0,1)$ be positive implies we should choose the positive root in part (c):

$$\mathrm{area}(\mathbf{v}\wedge\mathbf{w})=v_1w_2-v_2w_1=\det\begin{pmatrix}v_1 & w_1\\ v_2 & w_2\end{pmatrix}.$$

2.16. Suppose $M=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$, $\mathbf{v}=\begin{pmatrix}v_1\\ v_2\end{pmatrix}$, $\mathbf{w}=\begin{pmatrix}w_1\\ w_2\end{pmatrix}$.

a. Let V be the matrix whose columns are $\mathbf{v}$ and $\mathbf{w}$, and let $\overline{V}$ be the matrix whose columns are $M(\mathbf{v})$ and $M(\mathbf{w})$. Show that $\overline{V}=MV$ and hence that $\det\overline{V}=\det M\times\det V$.

b. Conclude that

$$\mathrm{area}\,M(\mathbf{v}\wedge\mathbf{w})=\mathrm{area}\,M(\mathbf{v})\wedge M(\mathbf{w})=\det M\times\mathrm{area}(\mathbf{v}\wedge\mathbf{w}).$$

2.17. The aim of this exercise is to show that the determinant of a 2×2 matrix can be viewed as a certain function $D(\mathbf{v},\mathbf{w})$ of its columns $\mathbf{v}$ and $\mathbf{w}$ that is uniquely characterized by the following three properties. Defining properties of the determinant

- $D(\mathbf{e}_1,\mathbf{e}_2)=1$, where $\mathbf{e}_i$ is the ith column of the identity matrix.
- $D(\mathbf{v},\mathbf{w})=0$ if $\mathbf{v}=\mathbf{w}$.
- $D(\mathbf{v},\mathbf{w})$ is a linear function of each of its arguments $\mathbf{v}$ and $\mathbf{w}$. That is,

$$D(t\mathbf{v},\mathbf{w}) = tD(\mathbf{v},\mathbf{w}),$$
$$D(\mathbf{v}_1+\mathbf{v}_2,\mathbf{w}) = D(\mathbf{v}_1,\mathbf{w})+D(\mathbf{v}_2,\mathbf{w}),$$

and similarly for the second argument. We say D is **bilinear**.

a. Show that D is **antisymmetric**; that is, $D(\mathbf{y},\mathbf{x}) = -D(\mathbf{x},\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$. Suggestion: First expand $D(\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y})$ to four terms using the bilinearity of D (this is a kind of "FOIL"). The second property will guarantee two of those terms are zero; the remaining two terms then give the result.
b. Show that $D(\mathbf{e}_2,\mathbf{e}_1) = -1$.
c. Let $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2$ and $\mathbf{w} = w_1\mathbf{e}_1 + w_2\mathbf{e}_2$. Using the bilinearity of D ("FOIL" again!), show that

$$D(\mathbf{v},\mathbf{w}) = v_1w_2D(\mathbf{e}_1,\mathbf{e}_2)+v_2w_1D(\mathbf{e}_2,\mathbf{e}_1) = v_1w_2 - v_2w_1,$$

proving that $D(\mathbf{v},\mathbf{w}) = \det\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}$.

The determinant in higher dimensions

The reason we have paused here to characterize the determinant of a 2×2 matrix by these three properties is that we later use (suitable modifications of) the same three properties to define the determinant of an $n\times n$ matrix. (See the exercises below). The properties of $D(\mathbf{v},\mathbf{w})$ are the properties of area$(\mathbf{v}\wedge\mathbf{w})$; thus the determinant of an $n\times n$ matrix will be connected with the volume of an n-dimensional parallelepiped, the analogue of a 2-dimensional parallelogram.

2.18. Determine $(5,2,-1)\times(3,4,2)$ and $(1,1,1)\times(1,1,-1)$.

2.19. Determine the volume of the parallelepiped:

a. $(5,2,-1)\wedge(3,4,2)\wedge(1,0,-1)$.
b. $(1,1,1)\wedge(1,1,-1)\wedge(1,-1,-1)$.

2.20. Consider the linear map $M:\mathbb{R}^3\to\mathbb{R}^3$ and the parallelepiped P defined by

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ -1 & 0 & 0 \end{pmatrix}, \quad P = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\wedge\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\wedge\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

a. What is the orientation of P? What is the volume of P?
b. Describe the parallelepiped $M(P)$ by listing its edges, in proper order.
c. Determine directly from your answer in (b) the orientation and volume of $M(P)$.
d. What is the volume multiplier of M? Does this value account for the orientation and volume of $M(P)$ you found in part (c)?

2.21. Suppose $\mathbf{z} = \mathbf{x}\times\mathbf{y} \neq \mathbf{0}$. Show that the linear map $L:\mathbb{R}^3\to\mathbb{R}^3$ defined by

$$L(\mathbf{x}) = \mathbf{y}, \quad L(\mathbf{y}) = \mathbf{x}, \quad L(\mathbf{z}) = -\mathbf{z},$$

has positive determinant and maps $\mathbf{x} \wedge \mathbf{y}$ to $\mathbf{y} \wedge \mathbf{x}$.

2.22. Show that $\mathrm{vol}(\mathbf{x} \wedge \mathbf{y} \wedge \overline{\mathbf{z}}) = \mathrm{vol}(\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z})$, where $\overline{\mathbf{z}} = \mathbf{z} + \alpha\mathbf{x} + \beta\mathbf{y}$ and α and β are arbitrary. One way to do this is to note that the two parallelepipeds $\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}$ and $\mathbf{x} \wedge \mathbf{y} \wedge \overline{\mathbf{z}}$ have the same base $\mathbf{x} \wedge \mathbf{y}$, and their third edges $\mathbf{z}$ and $\overline{\mathbf{z}}$ lie the same distance above that base. Draw a picture.

2.23. The parallelpipeds obtained from $\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}$ by an arbitrary permutation of $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are all equal as sets. However, they differ in orientation. Show that the cyclic permutations $\mathbf{y} \wedge \mathbf{z} \wedge \mathbf{x}$ and $\mathbf{z} \wedge \mathbf{x} \wedge \mathbf{y}$ have the same orientation as $\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}$, but the permutations that simply transpose a pair of edges, namely $\mathbf{y} \wedge \mathbf{x} \wedge \mathbf{z}$, $\mathbf{x} \wedge \mathbf{z} \wedge \mathbf{y}$, and $\mathbf{z} \wedge \mathbf{y} \wedge \mathbf{x}$, all have the opposite orientation. We can express this in the following way.

$$\mathbf{y} \wedge \mathbf{z} \wedge \mathbf{x} = \mathbf{z} \wedge \mathbf{x} \wedge \mathbf{y} = \mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z},$$
$$\mathbf{x} \wedge \mathbf{z} \wedge \mathbf{y} = \mathbf{y} \wedge \mathbf{x} \wedge \mathbf{z} = \mathbf{z} \wedge \mathbf{y} \wedge \mathbf{x} = -\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z}.$$

2.24. The purpose of this exercise is to show that the determinant of a 3×3 matrix is a certain function D of its columns that is uniquely defined by the following three properties. This is exactly analogous to the 2×2 case as dealt with in Exercise 2.17 (p. 61), and is preparation for the $n \times n$ case addressed below (Exercise 2.28).

- $D(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = 1$, where $\mathbf{e}_i$ is the ith column of the identity matrix.
- $D(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0$ if any two of the columns $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ are equal.
- $D(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is a linear function of each of its arguments $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. That is,

$$D(t\mathbf{x}, \mathbf{y}, \mathbf{z}) = tD(\mathbf{x}, \mathbf{y}, \mathbf{z}),$$
$$D(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}, \mathbf{z}) = D(\mathbf{x}_1, \mathbf{y}, \mathbf{z}) + D(\mathbf{x}_2, \mathbf{y}, \mathbf{z}),$$

and similarly for the second and third arguments.) We say D is **multilinear**.

a. Show that D is **antisymmetric**; that is, D changes sign when any two columns are interchanged:

$$D(\mathbf{x}, \mathbf{z}, \mathbf{y}) = D(\mathbf{z}, \mathbf{y}, \mathbf{x}) = D(\mathbf{y}, \mathbf{x}, \mathbf{z}) = -D(\mathbf{x}, \mathbf{y}, \mathbf{z}).$$

(Compare this with the previous exercise.)

b. Show that

$$D(\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2) = D(\mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_1) = D(\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3) = -1,$$
$$D(\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1) = D(\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2) = D(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = +1.$$

c. Write $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in terms of their coordinates with respect to the standard basis:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3;$$

similarly, $\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + y_3\mathbf{e}_3$; $\mathbf{z} = z_1\mathbf{e}_1 + z_2\mathbf{e}_2 + z_3\mathbf{e}_3$. Using the multilinearity of D, expand $D(\mathbf{x},\mathbf{y},\mathbf{z})$ as a sum of 27 terms of the form $x_i y_j z_k D(\mathbf{e}_i,\mathbf{e}_j,\mathbf{e}_k)$. Note that, in a given one of these expressions, the indices i, j, k need not be distinct.

d. Precisely 21 of the 27 terms you just obtained are automatically zero. Which ones, and why?

e. Show that the remaining six terms yield

$$D(\mathbf{x},\mathbf{y},\mathbf{z}) = \det\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix},$$

the familiar 3×3 determinant. Hence, the 3×3 determinant is uniquely determined by the three properties defining D.

2.25. Show that $D(\mathbf{x},\mathbf{y},\bar{\mathbf{z}}) = D(\mathbf{x},\mathbf{y},\mathbf{z})$, where $\bar{\mathbf{z}} = \mathbf{z} + \alpha\mathbf{x} + \beta\mathbf{y}$ and α and β are arbitrary. This is obviously the same result as Exercise 2.22 above; however, you should prove it here using only the properties of D defined and deduced in the previous exercise.

2.26. Let $P = \mathbf{v}\wedge\mathbf{w}$ be the parallelogram in $\mathbb{R}^3 : (u_1,u_2,u_3)$ spanned by $\mathbf{v} = (1,1,2)$ and $\mathbf{w} = (1,0,-1)$.

a. For $i = 1,2,3$, describe in the form $P_i = \mathbf{v}_i \wedge \mathbf{w}_i$ the projection of P onto the coordinate plane $u_i = 0$. For clarity, write $\mathbf{v}_i$ and $\mathbf{w}_i$ as elements of $\mathbb{R}^2$ rather than $\mathbb{R}^3$.

b. Determine the areas of the three projections P_i; then use the "Pythagorean" theorem to calculate

$$\operatorname{area} P = \sqrt{\operatorname{area}^2 P_1 + \operatorname{area}^2 P_2 + \operatorname{area}^2 P_3}.$$

c. Let V be the 3×2 matrix whose columns are the components of $\mathbf{v}$ and $\mathbf{w}$, in that order, and let $V^\dagger$ be its transpose. Show that

$$V^\dagger V = \begin{pmatrix} \mathbf{v}\cdot\mathbf{v} & \mathbf{v}\cdot\mathbf{w} \\ \mathbf{w}\cdot\mathbf{v} & \mathbf{w}\cdot\mathbf{w} \end{pmatrix},$$

implying $\det V^\dagger V = \|\mathbf{v}\|^2\|\mathbf{w}\|^2 - (\mathbf{v}\cdot\mathbf{w})^2$. Show that this is $\operatorname{area}^2 P$ (cf. Exercise 2.15, p. 61). Confirm that this value of $\operatorname{area} P$ agrees with the value you found in part (b).

2.27. Show that $(\mathbf{x}\cdot\mathbf{y})^2 \le \|\mathbf{x}\|^2\|\mathbf{y}\|^2$ for all vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. (Suggestion: both sides are zero if $\mathbf{y} = 0$, so assume $\mathbf{y} \ne 0$ and let

$$\mathbf{z} = \mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}.$$

Now consider the implications of $0 \le \mathbf{z}\cdot\mathbf{z}$.)

The next exercise defines the determinant of an $n \times n$ matrix V as a function D of its columns that satisfies certain properties. We have already seen how this approach works with a 2×2 matrix (Exercise 2.17, p. 61) and with a 3×3 matrix (Exercise 2.24).

As we saw in the earlier exercises, and see again in the $n \times n$ case, an essential property of D is antisymmetry. A description of this property involves rearranging, or permuting, the columns. When there were only two or three columns, this was simple to follow. However, it is useful here to pause and introduce some facts about permutations of an arbitrary number, n, of objects.

Permutations and transpositions

A **permutation on *n* elements** is an invertible map π of the set $\{1,2,\ldots,n\}$ to itself. A **transposition** is a permutation $\tau_{i,j}$ that interchanges the elements i and j and leaves all the others unchanged: $\tau_{i,j}(i) = j$; $\tau_{i,j}(j) = i$; $\tau_{i,j}(k) = k, k \neq i, j$. The **product** of two permutations is the permutation that results from their composition: $(\pi_1 \cdot \pi_2)(i) = (\pi_1 \pi_2)(i) = \pi_1(\pi_2(i))$. The identity map is a permutation and it is the identity element in the product. (With this product, the set S_n of permutations on n elements is a **group** with $n!$ elements.)

Even and odd permutations

Every permutation can be written as a product of transpositions. The number of transpositions in such a product is not unique, but its *parity* is. Therefore, we say a permutation π is **even**, and write $\operatorname{sgn}\pi = +1$, if π is always the product of an *even* number of transpositions; we say it is **odd**, and write $\operatorname{sgn}\pi = -1$, if it is always the product of an *odd* number of transpositions.

The defining properties of D

If $\mathbf{v}_1$, $\mathbf{v}_2$, $\ldots \mathbf{v}_n$, are the columns of the matrix V, then $D(\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n)$ is a function that satisfies the following conditions:

- $D(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 1$, where $\mathbf{e}_i$ is the ith column of the $n \times n$ identity matrix.
- $D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0$ if any two of the columns $\mathbf{v}_i$ are equal.
- D is multilinear; that is, $D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is a linear function of each of its arguments $\mathbf{v}_1$, $\mathbf{v}_2$, $\ldots$, $\mathbf{v}_n$.

Note: It is not yet evident that there is such a function D; the exercise shows that D exists.

2.28. a. Show that D is **antisymmetric**; that is, D changes signs when any two columns are interchanged. In terms of permutations: if π is a transposition, then

$$D(\mathbf{v}_{\pi(1)}, \mathbf{v}_{\pi(2)}, \ldots, \mathbf{v}_{\pi(n)}) = -D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{e}_n).$$

b. Suppose π is a permutation of $\{1,2,\ldots,n\}$. Show that

$$D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)}) = \operatorname{sgn}\pi = \pm 1.$$

c. Suppose that the map $\pi : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ is not a permutation, that is, π is not onto or one-to-one. Explain why

$$D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)}) = 0.$$

d. First write each column $\mathbf{v}_i$ as a linear combination of the standard basis elements $\mathbf{e}_k$:

$$\begin{aligned}
\mathbf{v}_1 &= v_{11}\mathbf{e}_1 + v_{21}\mathbf{e}_2 + \cdots + v_{n1}\mathbf{e}_n,\\
\mathbf{v}_2 &= v_{12}\mathbf{e}_1 + v_{22}\mathbf{e}_2 + \cdots + v_{n2}\mathbf{e}_n,\\
&\vdots\\
\mathbf{v}_n &= v_{1n}\mathbf{e}_1 + v_{2n}\mathbf{e}_2 + \cdots + v_{nn}\mathbf{e}_n.
\end{aligned}$$

Then, using the multilinearity of D, expand $D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v_n})$ as a sum of n^n terms of the form

$$v_{\pi(1),1} v_{\pi(2),2} \cdots v_{\pi(n),n} D(\mathbf{e}_{\pi(1)}, \mathbf{e}_{\pi(2)}, \ldots, \mathbf{e}_{\pi(n)}).$$

e. Most of the n^n terms you just obtained in the previous part are automatically zero; why? Which ones are not?

f. Conclude that

$$D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{\pi \text{ in } S_n} (\operatorname{sgn}\pi)\, v_{\pi(1),1}\, v_{\pi(2),2} \cdots v_{\pi(n),n}.$$

This formula for D shows that there is one and only one function D that satisfies the three given properties.

This exercise gives us a way to define the determinant of an $n \times n$ matrix.

Definition 2.8 *Suppose $V = (v_{ij})$ is the $n \times n$ matrix that has the element v_{ij} in the ith row and jth column. Let $\mathbf{v}_j = (v_{ij})$, $i = 1, \ldots, n$, denote the jth column of V, so $V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. The **determinant of V** is*

$$\det V = D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \sum_{\pi \text{ in } S_n} (\operatorname{sgn}\pi)\, v_{\pi(1),1}\, v_{\pi(2),2} \cdots v_{\pi(n),n}.$$

Thus, the determinant of V is the sum of all possible products of n elements of V, one taken from each row and each column, switching the sign of a particular product if the row indices represent an odd permutation of the columns indices.

2.29. Write out the 24 terms of the determinant of a 4×4 matrix.

2.30. Suppose that $V = (v_{ij})$ and $v_{ij} = 0$ if $i > j$. This is called an *upper triangular* matrix, because all entries below the main diagonal are zero. Show that $\det V = v_{11}v_{22}\cdots v_{nn}$.

2.31. Let A and B be 2×2 matrices, and let O denote the 2×2 matrix whose elements are all 0. Find the determinant of each of the following 4×4 matrices in terms of $\det A$ and $\det B$.

$$M_1 = \begin{pmatrix} A & O \\ O & B \end{pmatrix}, \quad M_2 = \begin{pmatrix} O & A \\ B & O \end{pmatrix}, \quad M_3 = \begin{pmatrix} A & B \\ O & O \end{pmatrix}.$$

2.32. Show that a square matrix A with a row or a column of zeros has $\det A = 0$.

2.33. Show that A and $A^\dagger$ have the same determinant.

2.34. The *minor* M_{ij} of A is the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and jth column of A. Show that we can "expand $\det A$ by minors along the ith row" in the following way.

$$\det A = (-1)^{i+1} a_{i1} \det M_{i1} + (-1)^{i+2} a_{i2} \det M_{i2} + \cdots + (-1)^{i+n} a_{in} \det M_{in}.$$

Write the analogous formula to "expand by minors along the jth column."

k-parallelepipeds

The definition of a parallelogram in $\mathbb{R}^3$ suggests we can define similar objects in $\mathbb{R}^n$, $n > 3$; in fact, we can generalize Definition 2.4 to define a parallelepiped of any dimension $k \le n$ in $\mathbb{R}^n$.

The ***k*-dimensional parallelepiped** $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_k$ is the set of vectors

$$\mathbf{w} = \sum_{i=1}^{k} t_i \mathbf{v}_i,$$

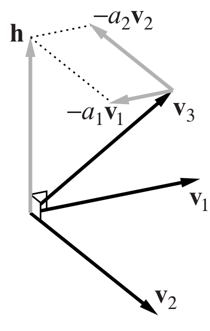
where $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are vectors in $\mathbb{R}^n$ and $0 \le t_i \le 1$, $i = 1, \ldots, k$. We take $k \le n$, but if $k \ne n$, $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_k$ is not oriented. We continue to call a 2-parallelepiped $\mathbf{v} \wedge \mathbf{w}$ a *parallelogram*.

2.35. Let $\mathbf{v} \wedge \mathbf{w}$ be a parallelogram in $\mathbb{R}^n$, and let V be the $n \times 2$ matrix whose columns are the coordinates of the vectors $\mathbf{v}$ and $\mathbf{w}$, in that order. Show that $\det V^\dagger V = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - (\mathbf{v} \cdot \mathbf{w})^2$. By Exercise 2.26, we can take this to be $\text{area}^2(\mathbf{v} \wedge \mathbf{w})$.

3-volume

Now that we have the area of a parallelogram, we can define the k-volume of a k-parallelepiped inductively on k. We start with a 3-parallelepiped $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3$. Think of this as having base $\mathbf{v}_1 \wedge \mathbf{v}_2$; then we want the "3-volume" to be the area of the base times the perpendicular height:

$$\text{vol}(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3) = \text{area}(\mathbf{v}_1 \wedge \mathbf{v}_2)\,\|\mathbf{h}\|.$$



Here $\mathbf{h}$ is the vector that is orthogonal to the base $\mathbf{v}_1 \wedge \mathbf{v}_2$ and in the plane that contains $\mathbf{v}_3$ and is parallel to the base. A vector in that parallel plane is of the form

$$\mathbf{h} = \mathbf{v}_3 - a_1\mathbf{v}_1 - a_2\mathbf{v}_2,$$

for some real numbers a_1 and a_2. The orthogonality condition on $\mathbf{h}$ gives us two equations,

$$0 = \mathbf{v}_1 \cdot \mathbf{h} = \mathbf{v}_1 \cdot \mathbf{v}_3 - a_1 \mathbf{v}_1 \cdot \mathbf{v}_1 - a_2 \mathbf{v}_1 \cdot \mathbf{v}_2,$$
$$0 = \mathbf{v}_2 \cdot \mathbf{h} = \mathbf{v}_2 \cdot \mathbf{v}_3 - a_1 \mathbf{v}_2 \cdot \mathbf{v}_1 - a_2 \mathbf{v}_2 \cdot \mathbf{v}_2,$$

that we can convert into a matrix equation for the unknowns a_1 and a_2:

$$\begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{v}_3 \\ \mathbf{v}_2 \cdot \mathbf{v}_3 \end{pmatrix}.$$

There are unique values for a_1 and a_2—that is, $\mathbf{h}$ is uniquely defined—precisely when the matrix on the left is invertible. But notice that the determinant of that matrix is, by Exercise 2.26, the square of the area of the base parallelogram $\mathbf{v}_1 \wedge \mathbf{v}_2$. If $\mathrm{area}(\mathbf{v}_1 \wedge \mathbf{v}_2) = 0$, the 3-volume of $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3$ is zero; otherwise, we can find $\mathbf{h}$, as above, and obtain $\mathrm{vol}(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3)$. In fact, the volume is a determinant:

$$\begin{aligned} \begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & 0 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & 0 \\ 0 & 0 & \mathbf{h} \cdot \mathbf{h} \end{vmatrix} &= \begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 \end{vmatrix} \mathbf{h} \cdot \mathbf{h} \\ &= \mathrm{area}^2(\mathbf{v}_1 \wedge \mathbf{v}_2)\, \|\mathbf{h}\|^2 = \mathrm{vol}^2(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3). \end{aligned}$$

Now replace the zeros in the 3×3 determinant with $\mathbf{v}_1 \cdot \mathbf{h}$ and $\mathbf{v}_2 \cdot \mathbf{h}$, as appropriate. Then substitute for the right-hand factor $\mathbf{h}$ in each entry in the third column its expression as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$:

$$\begin{aligned} \begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \mathbf{v}_1 \cdot \mathbf{h} \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \mathbf{v}_2 \cdot \mathbf{h} \\ \mathbf{h} \cdot \mathbf{v}_1 & \mathbf{h} \cdot \mathbf{v}_2 & \mathbf{h} \cdot \mathbf{h} \end{vmatrix} &= \begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \mathbf{v}_1 \cdot \mathbf{v}_3 - a_1 \mathbf{v}_1 \cdot \mathbf{v}_1 - a_2 \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \mathbf{v}_2 \cdot \mathbf{v}_3 - a_1 \mathbf{v}_2 \cdot \mathbf{v}_1 - a_2 \mathbf{v}_2 \cdot \mathbf{v}_2 \\ \mathbf{h} \cdot \mathbf{v}_1 & \mathbf{h} \cdot \mathbf{v}_2 & \mathbf{h} \cdot \mathbf{v}_3 - a_1 \mathbf{h} \cdot \mathbf{v}_1 - a_2 \mathbf{h} \cdot \mathbf{v}_2 \end{vmatrix} \\ &= \begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \mathbf{v}_1 \cdot \mathbf{v}_3 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \mathbf{v}_2 \cdot \mathbf{v}_3 \\ \mathbf{h} \cdot \mathbf{v}_1 & \mathbf{h} \cdot \mathbf{v}_2 & \mathbf{h} \cdot \mathbf{v}_3 \end{vmatrix}. \end{aligned}$$

To get the simpler expression in the last step, above, we have used the multilinearity of the determinant and the fact that the determinant of a matrix with two equal columns is zero; cf. Exercise 2.25. The net result is that $\mathbf{h}$ is replaced by $\mathbf{v}_3$ in the third column. The same substitution for the factor $\mathbf{h}$ in the third row, followed by a similar addition of rows, will leave us $\mathbf{h}$ replaced by $\mathbf{v}_3$ in the third row. We discover that the volume is a determinant involving the vectors $\mathbf{v}_i$ in a symmetric way.

$$\begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \mathbf{v}_1 \cdot \mathbf{v}_3 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \mathbf{v}_2 \cdot \mathbf{v}_3 \\ \mathbf{v}_3 \cdot \mathbf{v}_1 & \mathbf{v}_3 \cdot \mathbf{v}_2 & \mathbf{v}_3 \cdot \mathbf{v}_3 \end{vmatrix} = \mathrm{vol}^2(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3)$$

$\det V^\dagger V = \mathrm{vol}^2$

Finally, if V is the $n \times 3$ matrix whose columns are $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, in that order, then

$$V^\dagger V = \begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \mathbf{v}_1 \cdot \mathbf{v}_3 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \mathbf{v}_2 \cdot \mathbf{v}_3 \\ \mathbf{v}_3 \cdot \mathbf{v}_1 & \mathbf{v}_3 \cdot \mathbf{v}_2 & \mathbf{v}_3 \cdot \mathbf{v}_3 \end{pmatrix},$$

so $\det V^\dagger V = \mathrm{vol}^2(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3)$.

We consider next the 4-parallelepiped $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3 \wedge \mathbf{v}_4$; it has

$$\text{base, a 3-parallelepiped}: \mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \mathbf{v}_3,$$
$$\text{height vector, orthogonal to base}: \mathbf{h} = \mathbf{v}_4 - a_1\mathbf{v}_1 - a_2\mathbf{v}_2 - a_3\mathbf{v}_3,$$

and we can define its 4-volume as the product of the 3-volume of its base with the length of its height vector. The same, rather lengthy, argument we have just carried out allows us to determine the square of the 4-volume as $\det V^\dagger V$, where V is the matrix whose columns are the vectors $\mathbf{v}_i$. This process of establishing the volume of a k-parallelepiped from the volume of a $(k-1)$-parallelepiped is an example of *mathematical induction*, and shows for every $k \le n$ that the squared k-volume is $\det V^\dagger V$.

2.36. Show that the square of the n-volume of an n-parallelepiped, as defined in the text, equals $\det V^\dagger V$, as derived in these exercises.

2.37. Let V be the $n \times 3$ matrix whose columns are $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$; verify that

$$V^\dagger V = \begin{pmatrix} \mathbf{v}_1\cdot\mathbf{v}_1 & \mathbf{v}_1\cdot\mathbf{v}_2 & \mathbf{v}_1\cdot\mathbf{v}_3 \\ \mathbf{v}_2\cdot\mathbf{v}_1 & \mathbf{v}_2\cdot\mathbf{v}_2 & \mathbf{v}_2\cdot\mathbf{v}_3 \\ \mathbf{v}_3\cdot\mathbf{v}_1 & \mathbf{v}_3\cdot\mathbf{v}_2 & \mathbf{v}_3\cdot\mathbf{v}_3 \end{pmatrix}.$$

2.38. Let $\mathbf{h} = \mathbf{v}_k - a_1\mathbf{v}_1 - \cdots - a_{k-1}\mathbf{v}_{k-1}$, where $a_1, \ldots, a_{k-1}$ are arbitrary real numbers. Show that $\mathbf{h}$ is in the (hyper)plane that contains $\mathbf{v}_k$ and is parallel to the linear subspace of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$.

2.39. Find a vector $\mathbf{h}$ in the plane that contains $\mathbf{v}_3$ and is parallel to $\mathbf{v}_1 \wedge \mathbf{v}_2$, when

a. $\mathbf{v}_1 = (1,0,1,0)$, $\mathbf{v}_2 = (2,1,1,1)$, $\mathbf{v}_3 = (0,1,2,0)$.
b. $\mathbf{v}_1 = (1,0,-1,0)$, $\mathbf{v}_2 = (2,1,1,1)$, $\mathbf{v}_3 = (0,1,2,0)$.

2.40. Determine the rank and nullity of each of the following matrices, viewing each as a linear map.

a. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ b. $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$ c. $\begin{pmatrix} 1 & -2 \\ -2 & 4 \\ 3 & -6 \end{pmatrix}$ d. $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$

2.41. a. Solve the equations

$$5u + 3v - 3w + x = 0,$$
$$3u + 2v + 6w - 2x = 0,$$

for u and v in terms of w and x.

b. Can you solve for w and x in terms of u and v? What happens?
c. Can you solve for u and x in terms of v and w? What is the result?

2.42. This question concerns the linear map $L : \mathbb{R}^3 \to \mathbb{R}^3$ defined by the equations

$$L : \begin{cases} x = u + v, \\ y = v + w, \\ z = u - w. \end{cases}$$

a. What is the dimension of $\ker L$? Give a basis for $\ker L$.
b. Find the matrix for a linear map $M : \mathbb{R}^p \to \mathbb{R}^q$ whose graph is the kernel of L. What are the values of p and q?
c. Write M as a set of q equations in p variables.
d. What is the dimension of $\operatorname{im} L$? Give a basis for $\operatorname{im} L$.
e. Find the matrix for a linear map $A : \mathbb{R}^j \to \mathbb{R}^k$ whose graph is the image of L. What are the values of j and k?
f. Write A as a set of k equations in j variables.

2.43. a. Find all solutions (u, v) to the equations

$$\begin{aligned} u - 2v &= 5, \\ 4v - 2u &= -10, \end{aligned}$$

and sketch the solution set in the (u, v)-plane.
b. Describe the solution set in (a) as the graph of a suitable function. Is your sketch in (a) the graph of that function?
c. Describe the relation between the solution set in part (a) to the set of solutions to the equations

$$\begin{aligned} u - 2v &= 0, \\ 4v - 2u &= 0. \end{aligned}$$

2.44. Let $L : \mathbb{R}^n \to \mathbb{R}^p$ be an arbitrary linear map, and let

$$V = \{(\mathbf{u}, L(\mathbf{u})) \mid \mathbf{u} \in \mathbb{R}^n\} \subset \mathbb{R}^n \times \mathbb{R}^p = \mathbb{R}^{n+p}$$

be the graph of L. The purpose of this exercise is to show that V is a linear subspace of $\mathbb{R}^{n+p}$ of dimension n.

a. Show that the sum of two vectors in V is also in V. That is, given $\mathbf{v}_1 = (\mathbf{u}_1, L(\mathbf{u}_1))$ and $\mathbf{v}_2 = (\mathbf{u}_2, L(\mathbf{u}_2))$ with $\mathbf{u_1}$ and $\mathbf{u}_2$ in $\mathbb{R}^n$, show that $\mathbf{v}_1 + \mathbf{v}_2$ also has the form $(\mathbf{w}, L(\mathbf{w}))$ for some suitable $\mathbf{w}$ in $\mathbb{R}^n$.
b. Show that any scalar multiple of a vector in V is also in V.
c. Suppose $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$. Let $\mathbf{v}_j = (\mathbf{u}_j, L(\mathbf{u}_j))$ for $j = 1, 2, \ldots, n$. Show that
   i. $\mathcal{G} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a linearly independent set of vectors in $\mathbb{R}^{n+p}$;
   ii. $\mathcal{G}$ spans V; that is, any vector in V can be written as a linear combination of the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ that span V.
d. Explain why $\dim V = \dim \operatorname{graph} L = n$.