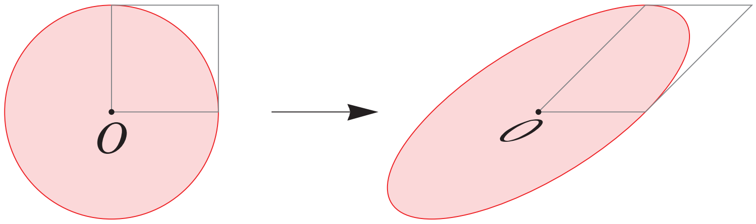
Linear algebra

The second chapter of this book is concerned with vectors, matrices and linear transformations. Determinants are introduced, together with ways in which to calculate them. These concepts are particularly relevant in analytic geometry, where we use them to describe projective transformations.

Linear transformations

Linear transformations are transformations of n-dimensional Euclidean space $\mathbb{R}^n$ expressible as $\underline{x} \to M\,\underline{x}$, where $\underline{x} = (x_1, x_2, \ldots, x_n)$ is the position vector of a point X. M is known as the *transformation matrix*. For example, the linear transformation with matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is shown below.

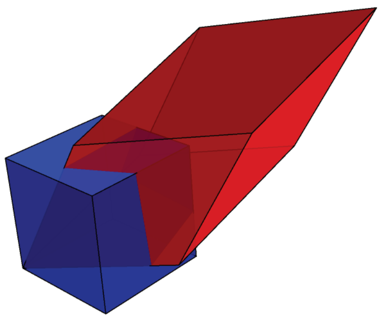


The position of the origin, O, is left unchanged by a linear transformation. Degree-d algebraic curves remain as degree-d algebraic curves; in particular, lines map to lines and conics map to conics. In the shear shown above, a circle is transformed into an ellipse. Parallel lines remain parallel when linear transformations are applied. Finally, the (signed) area of any shape is multiplied by $\det(M)$ when the transformation is applied, where $\det(M)$ is the *determinant* of the transformation matrix. Hence, ratios of areas remain unchanged.

Common linear transformations include rotations (about the origin), reflections (in lines through the origin), dilations (where the origin is the centre of homothety) and stretches (again, preserving the origin). One can combine transformations by multiplying their matrices.

1. Let $A = (1, 0, 0)$, $B = (0, 1, 0)$ and $C = (0, 0, 1)$ be three points in $\mathbb{R}^3$. After applying the transformation with matrix $M = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, find the new locations of A, B and C.

Consider the unit cube $[0, 1] \times [0, 1] \times [0, 1]$, where $\times$ denotes Cartesian product. It is transformed into a parallelepiped with volume $V = \det(M)$.



In the diagram above, the blue cube is transformed into the red parallelepiped. The origin (the common vertex of the cube and parallelepiped) remains fixed.

Determinants

The determinant of a square matrix M is a positive real number $\det(M)$ associated with that matrix. It behaves like the norm of a complex number, in that it is multiplicative.

- For two square matrices A and B of equal dimension, $\det(A\,B) = \det(A)\det(B)$. **[Multiplicativity of determinants]**

If a matrix A has an inverse matrix A^{-1} such that $A\,A^{-1} = A^{-1}\,A = I$, then $\det(A)\det\left(A^{-1}\right) = \det(I) = 1$. Hence, it is clear that a matrix with a determinant of zero has no inverse. Indeed, the converse is also true: all square matrices with non-zero determinants possess unique well-defined inverses. If a matrix is one-dimensional, then its determinant is equal to its only element. Otherwise, we compute it recursively.

$$M = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix}$$

Consider the matrix above. We compute the determinant using the following process:

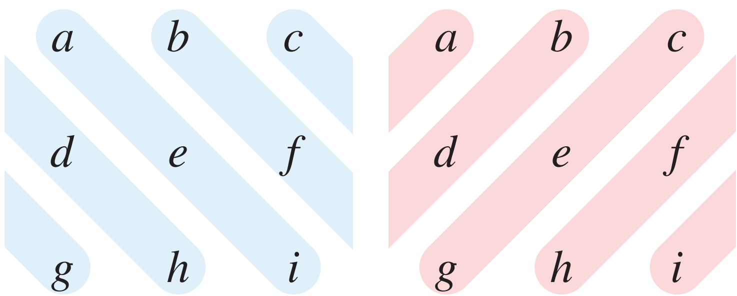
- For some $1 \le i \le n$, consider the element $a_{i,1}$ in the first column of M.
- The $(n-1)$-dimensional matrix M_i is obtained by removing everything in the same row or column as $a_{i,1}$.
- Compute the value $S_i = a_{i,1}\det(M_i)$.
- Then, we have $\det(M) = S_1 - S_2 + S_3 - S_4 + \ldots + (-1)^{n+1} S_n$.

This recursion results in the determinant equating to a sum of $n!$ terms, each of which is a product of n elements of M. After expanding this somewhat complicated recursive definition, we reach a more elegant formulation.

- $\det(M) = \sum_{\text{sym}} (-1)^{P(\sigma)} \left(a_{1,\sigma(1)}\, a_{2,\sigma(2)} \ldots a_{n,\sigma(n)}\right)$, where the sum is taken over all permutations σ of $\{1, 2, 3, \ldots, n\}$. We define $P(\sigma)$ to be even if σ is an even permutation, and odd otherwise. **[Leibniz formula for determinants]**

2. Express $\det\begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix}$ as a polynomial in x, y, z.

You may have noticed that for 3×3 determinants, the even permutations correspond to the three NW-SE 'diagonals' and the odd permutations correspond to the three NE-SW 'diagonals'. The diagonals are considered to wrap around the edges of the matrix as though it were a cylinder. This trick is known as the *Rule of Sarrus*.



$$a\,e\,i + b\,f\,g + c\,d\,h \;-a\,f\,h - b\,d\,i - c\,e\,g$$

Leibniz's formula requires $n(n!)$ elementary operations, so is rather time-consuming for large matrices, taking exponential time. Instead, it helps to simplify the calculation by performing operations on the matrix.

- Applying elementary operations to the rows or columns of M cause its determinant to behave in a predictable manner:
  - Multiplying any row or column of M by x causes the determinant of M to be multiplied by x;
  - Adding (or subtracting) any multiple of one row to another row does not affect the determinant of M;
  - Swapping any two rows causes $\det(M)$ to be multiplied by -1;
  - The transpose of M has the same determinant as M.

This can also be used to easily factorise the determinants of matrices.

3. Factorise $\det\begin{pmatrix} 1 & 1 & 1 \\ x & y & z \\ x^3 & y^3 & z^3 \end{pmatrix}$ into four linear factors.

Interpolating curves

The determinant of a matrix is zero if and only if one row can be expressed as a linear combination of the others. This is known as *linear dependence*. This enables one to create a curve of some type (*e.g.* a polynomial, circle or conic) interpolating between various points. For example, if we have a sequence of n points (x_i, y_i), then the following curve is a degree-$(n-1)$ polynomial passing through all n points.

- The curve $\det\begin{pmatrix} 1 & y & x & x^2 & x^3 & \cdots & x^{n-1} \\ 1 & y_1 & x_1 & {x_1}^2 & {x_1}^3 & \cdots & {x_1}^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y_n & x_n & {x_n}^2 & {x_n}^3 & \cdots & {x_n}^{n-1} \end{pmatrix} = 0$ passes through all points (x_i, y_i). **[Lagrange interpolating polynomial]**

This is obvious, as the determinant equals zero if two rows are identical. It is also a degree-$(n-1)$ polynomial, as we can use the recursive determinant formula to express it as $A_1 + A_2\, y + A_3\, x + A_4\, x^2 + A_5\, x^3 + \ldots + A_{n+1}\, x^{n-1}$ and rearrange it. If $A_2 = 0$ then this method will fail, but that only occurs if two points have the same abscissa.

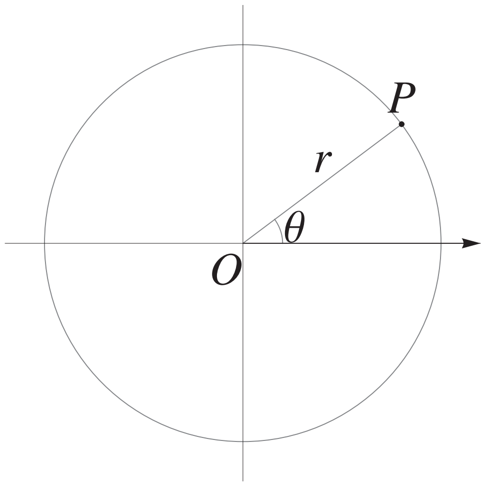
Using this idea, we can create a unique conic passing through any 5 points in general position, a cubic passing through 9 points *et cetera*. If the points are not in general position, then seemingly paradoxical things can occur. This forms the basis of the powerful *Cayley-Bacharach theorem* explored in the projective geometry chapter. The general equation of a conic is $A + B\, x + C\, y + D\, x^2 + E\, y^2 + F\, x\, y = 0$, so we can determine the equation of the conic passing through five given points.

- The conic $\begin{vmatrix} 1 & x & y & x^2 & y^2 & x\, y \\ 1 & x_1 & y_1 & {x_1}^2 & {y_1}^2 & x_1\, y_1 \\ 1 & x_2 & y_2 & {x_2}^2 & {y_2}^2 & x_2\, y_2 \\ 1 & x_3 & y_3 & {x_3}^2 & {y_3}^2 & x_3\, y_3 \\ 1 & x_4 & y_4 & {x_4}^2 & {y_4}^2 & x_4\, y_4 \\ 1 & x_5 & y_5 & {x_5}^2 & {y_5}^2 & x_5\, y_5 \end{vmatrix} = 0$ passes through all points (x_i, y_i). **[Interpolating conic]**

Circles also have a simple characterisation in Cartesian coordinates.

4. Find the equation of the circle passing through the non-collinear points (x_1, y_1), (x_2, y_2) and (x_3, y_3). **[Circumcircle equation]**

The determinant formula is not constrained to Cartesian coordinates; it can be used to find interpolating curves in any coordinate system, such as projective homogeneous coordinates, areal coordinates, complex numbers and even polar coordinates. As we cover the other coordinate systems in greater depth later in the book, it is worth messing around with polar coordinates here.



- The point with *polar coordinates* $P = \langle r,\ \theta\rangle$ in the Euclidean plane is defined such that $O\,P$ has length r and makes an angle of θ with the positive x-axis. In Cartesian coordinates, $P = (r\cos\theta,\ r\sin\theta)$. **[Definition of polar coordinates]**

Although the value of r is uniquely defined, θ is not; adding or subtracting multiples of $2\,\pi$ will describe the same point. This is a consequence of the periodicity of the elementary trigonometric functions.

5. Let $Q = \langle r_1,\ \theta_1\rangle$ be a point on the polar plane. Show that the equation of the circle with centre Q and radius a is given by $r^2 + {r_1}^2 - 2\,r\,r_1\cos(\theta - \theta_1) = a^2$. **[Polar equation of a circle]**

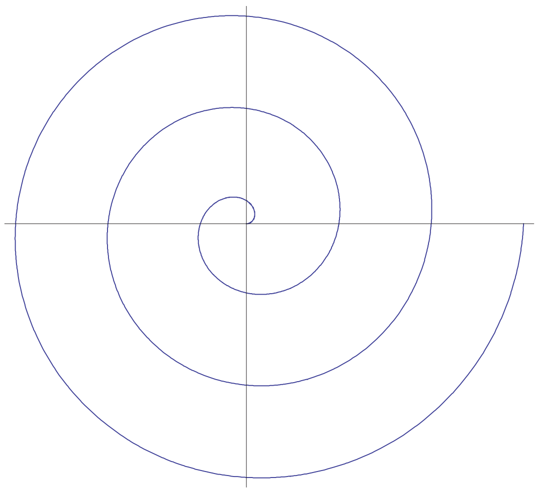
6. Hence show that a circle has general equation $A\,r^2 + B\,r\cos\theta + C\,r\sin\theta + D = 0$. **[General polar equation of a circle]**

It now becomes more obvious why this should work: the general equation for a circle in Cartesian coordinates is $A\left(x^2 + y^2\right) + B\,x + C\,y + D = 0$, and we have $x^2 + y^2 = r^2$, $x = r\cos\theta$ and $y = r\sin\theta$.

7. Find the equation, in polar coordinates, of the circle passing through the non-collinear points $\langle r_1,\ \theta_1\rangle$, $\langle r_2,\ \theta_2\rangle$ and $\langle r_3,\ \theta_3\rangle$. **[Circumcircle equation for polar coordinates]**

If three of the points are collinear, the term in r^2 vanishes and we are left with the equation of a line.

A curve which is particularly amenable to expressing in polar coordinates is the *Archimedean spiral*. If the spiral is centred on the origin, then it has polar equation $r = \frac{h}{2\,\pi}\,(\theta - \phi)$. h is the separation between successive turns of the spiral, and ϕ is the angle at which is emerges from the origin.

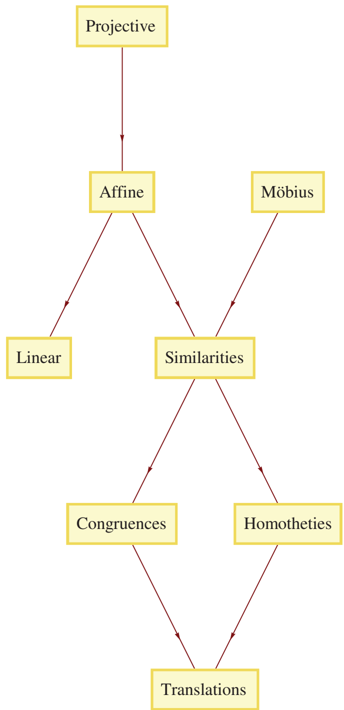


8. Find the equation for an Archimedean spiral of centre O passing through $\langle r_1, \theta_1 \rangle$ and $\langle r_2, \theta_2 \rangle$. **[Interpolating spiral]**

Adding multiples of 2π to either of the angles can alter the number of turns on the spiral and its direction. There is not a unique interpolating spiral with centre O passing through two given points; there are countably infinitely many.

Geometric transformations

So far, we have considered linear transformations. If we compose an arbitrary linear transformation with an arbitrary translation, then we obtain an *affine* transformation. Affine transformations have all the geometric properties of linear transformations, but do not necessarily preserve the origin. They are a special case of *projective* transformations, which are covered in a later chapter.



Affine transformations are projective transformations which preserve the line at infinity. Linear transformations also preserve the origin, whereas similarities preserve (or reverse) the circular points at infinity (thus mapping circles to circles). Congruences are similarities with a determinant of ± 1, whereas homotheties are similarities which preserve the direction of all lines (thus all points on the line at infinity). Translations (and reflections in a point) lie in the intersection of congruences and homotheties.

Do not worry if these terms are unfamiliar to you; they are explained properly in later chapters.

Scalar product

Let $\underline{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ and $\underline{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$ be three vectors in $\mathbb{R}^3$.

- The dot product (or inner product, or scalar product) $\underline{a}\cdot\underline{b} = a_1\, b_1 + a_2\, b_2 + a_3\, b_3 = |\underline{a}|\ |\underline{b}| \cos\theta$, where θ is the angle between the vectors a and b. **[Definition of dot product]**

The dot product is commutative and distributive, so $\underline{a}\cdot\underline{b} = \underline{b}\cdot\underline{a}$ and $\underline{a}\cdot(\underline{b}+\underline{c}) = \underline{a}\cdot\underline{b} + \underline{a}\cdot\underline{c}$.

9. Prove that, for every triangle $A\,B\,C$, we have $a^2 = b^2 + c^2 - 2\,b\,c\cos A$. **[Law of cosines]**

The dot product generalises to vectors in $\mathbb{R}^n$. This allows us to interchange between trigonometric, geometric and algebraic inequalities.

- The following three statements are all equivalent:
  - $\cos\theta \le 1$, with equality if and only if $\theta = \pi\, n$ for some integer n;
  - $\underline{a}\cdot\underline{b} \le |\underline{a}|\ |\underline{b}|$, with equality if and only if the vectors have the same direction;
  - $a_1\, b_1 + a_2\, b_2 + \ldots + a_n\, b_n \le \sqrt{a_1{}^2 + a_2{}^2 + \ldots + a_n{}^2}\ \sqrt{b_1{}^2 + b_2{}^2 + \ldots + b_n{}^2}$, with equality if and only if $a_i = \lambda\, b_i$ for some scalar $\lambda \in [0, \infty]$. **[Cauchy-Schwarz inequality]**

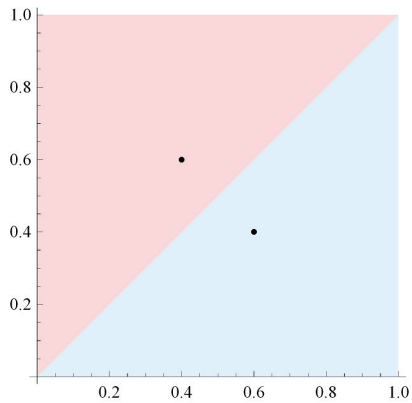
We can generalise the idea of a vector to a more abstract object, and thus extend the Cauchy-Schwarz inequality even further. See *Introduction to Inequalities* (Bradley) for an example of this.

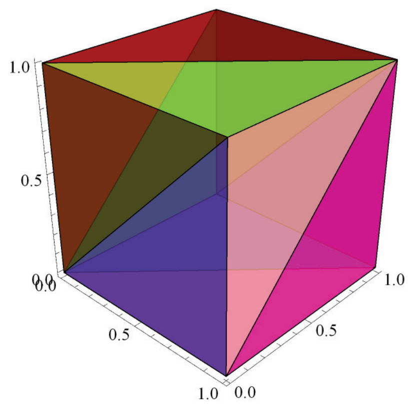
Another application of the dot product in inequalities is a proof of the *rearrangement inequality*. That states that if we have two non-negative sequences of equal length and multiply corresponding terms, the product is greatest when the sequences are sorted in the same order.

- Suppose that $a_1 \ge a_2 \ge \ldots \ge a_n \ge 0$ and $b_1 \ge b_2 \ge \ldots \ge b_n \ge 0$ are two decreasing sequences of non-negative integers. Then $\sum_{i=1}^{n} a_i\, b_i \ge \sum_{i=1}^{n} a_i\, b_{\sigma(i)}$ for any permutation σ. **[Rearrangement inequality]**

Proof:

We can prove this by considering the vectors $\underline{a}$ and $\underline{b}$ in the space $\mathbb{R}^n$. Observe that all $n!$ vectors in $\{\underline{b}_\sigma\}$ (the set of vectors obtained by permuting the elements of $\underline{b}$) are of equal length, so lie on a sphere with centre $\underline{0}$. The dot product $\sum_{i=1}^{n} a_i\, b_{\sigma(i)}$ of the vectors $\underline{a}$ and $\underline{b}_\sigma$ is greatest when the angle between them is smallest, which occurs when $\underline{a}$ and $\underline{b}_\sigma$ are closest (as all vectors in $\{\underline{b}_\sigma\}$ are of equal length). So, this has been converted into the equivalent problem of proving that $\underline{b}$ is the closest vector to $\underline{a}$ in $\{\underline{b}_\sigma\}$. We consider the *Voronoi diagram* of $\mathbb{R}^n$, which is simply a division of space depending on which $\underline{b}_\sigma$ is closest.





The diagrams above illustrate the cases when $n = 2$ or $n = 3$. The Voronoi diagram is created by the set of planes of the form $x_i = x_j$, which each partition space into the regions $x_i > x_j$ and $x_i < x_j$. This means that the regions of the Voronoi diagram are determined by the ordering of the elements; in the case where $n = 3$, we have six tetrahedral regions, namely $x_1 > x_2 > x_3$ and the five other permutations. As the elements of $\underline{a}$ and $\underline{b}$ are ordered in the same way, they must inhabit the same region. Hence, $\underline{b}$ is the closest vector in $\{\underline{b}_\sigma\}$ to $\underline{a}$, and we are finished.

Vector and triple products

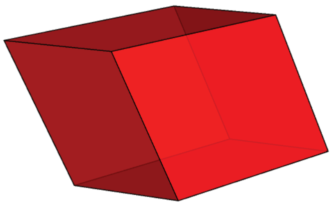
So far, we are able to 'multiply' two vectors in $\mathbb{R}^n$, resulting in a scalar. We can also define a vector (cross) product, which is specific to $\mathbb{R}^3$. (There is also a 7-dimensional version based on the octonion algebra, but that is outside the scope of the book.)

- The cross product (or vector product, or exterior product) $\underline{a} \times \underline{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} = \det \begin{pmatrix} \underline{i} & a_1 & b_1 \\ \underline{j} & a_2 & b_2 \\ \underline{k} & a_3 & b_3 \end{pmatrix}$, where $\underline{i}$, $\underline{j}$, $\underline{k}$ are the unit vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, respectively. **[Definition of cross product]**

The cross product is **anti-**commutative and distributive, so $\underline{a} \times \underline{b} = -\underline{b} \times \underline{a}$ and $\underline{a} \times (\underline{b} + \underline{c}) = \underline{a} \times \underline{b} + \underline{a} \times \underline{c}$. The vector $\underline{a} \times \underline{b}$ is perpendicular to both $\underline{a}$ and $\underline{b}$, and its magnitude is equal to the area of the parallelogram with vertices $\{0, \underline{a}, \underline{b}, \underline{a} + \underline{b}\}$.

Finally, we define the scalar triple product, which is the volume of the parallelepiped with vertices $\{0, \underline{a}, \underline{b}, \underline{c}, \underline{a} + \underline{b}, \underline{b} + \underline{c}, \underline{c} + \underline{a}, \underline{a} + \underline{b} + \underline{c}\}$.

- $\underline{a} \cdot (\underline{b} \times \underline{c}) = (\underline{a} \times \underline{b}) \cdot \underline{c} = \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$. **[Scalar triple product]**



Sir William Rowan Hamilton once had an epiphany whilst crossing a bridge, and carved the formula $i^2 = j^2 = k^2 = i\,j\,k = -1$ into one of the stones. This defines an extension to the complex numbers, which has four orthogonal units $(1, i, j, k)$ as opposed to two. A *Hamiltonian quaternion* is a number of the form $w + x\,i + y\,j + z\,k$, where $w, x, y, z \in \mathbb{R}$. Using a slight abuse of notation, this can be written as $w + \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. A scalar added to a vector?! We can multiply two quaternions $p = a + \underline{b}$ and $q = c + \underline{d}$ together to give the quaternion $p\,q = (a\,c - \underline{b} \cdot \underline{d}) + (a\,\underline{d} + c\,\underline{b} + \underline{b} \times \underline{d})$. Multiplication of quaternions is associative and distributive, but not commutative; $p\,q \neq q\,p$ in general. This is inherited from the non-commutativity of the cross product.

The quaternion $w + x\,i + y\,j + z\,k$ has a norm of $\sqrt{w^2 + x^2 + y^2 + z^2}$. As with complex numbers, $|p|\ |q| = |p\,q|$ for any $p, q \in \mathbb{H}$, where $\mathbb{H}$ is the set of all quaternions.

Solutions

1. Using matrix multiplication, we get $A = \begin{pmatrix} a \\ d \\ g \end{pmatrix}$, $B = \begin{pmatrix} b \\ e \\ h \end{pmatrix}$ and $C = \begin{pmatrix} c \\ f \\ i \end{pmatrix}$.

2. $\det\begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix} = x^3 + y^3 + z^3 - 3\,x\,y\,z$, as the NW-SE diagonals are x^3, y^3, z^3 and the NE-SW diagonals are each $x\,y\,z$.

3. We deduct the first column from the other two, obtaining $\det\begin{pmatrix} 1 & 0 & 0 \\ x & y-x & z-x \\ x^3 & y^3-x^3 & z^3-x^3 \end{pmatrix}$. Applying the recursion formula reduces this to $\det\begin{pmatrix} y-x & z-x \\ y^3-x^3 & z^3-x^3 \end{pmatrix}$. We then divide the first column by $y-x$ and multiply the entire determinant by $y-x$, obtaining $(y-x)\det\begin{pmatrix} 1 & z-x \\ y^2+x^2+x\,y & z^3-x^3 \end{pmatrix}$. Applying a similar factorisation to the second column results in $(y-x)(z-x)\det\begin{pmatrix} 1 & 1 \\ x^2+y^2+x\,y & x^2+z^2+x\,z \end{pmatrix}$. Leibniz's formula can now be used to expand the determinant, giving $(y-x)(z-x)\left(x\,y-x\,z+y^2-z^2\right)$. The quadratic factorises to $(y-z)(x+y+z)$, so the entire determinant is equal to $(y-x)(z-x)(y-z)(x+y+z)$.

4. $\det\begin{pmatrix} 1 & x & y & x^2+y^2 \\ 1 & x_1 & y_1 & x_1{}^2+y_1{}^2 \\ 1 & x_2 & y_2 & x_2{}^2+y_2{}^2 \\ 1 & x_3 & y_3 & x_3{}^2+y_3{}^2 \end{pmatrix} = 0$ will suffice, as the general equation for a circle is $A + B\,x + C\,y + D\left(x^2+y^2\right) = 0$.

5. Let $P = \langle r, \theta\rangle$ be a point on the circle, so $P\,Q = a$. By using the cosine rule, we have $a^2 = r^2 + r_1{}^2 - 2\,r\,r_1\cos(\theta-\theta_1)$.

6. Using the compound angle formula, we get $r^2 - 2\,r\,r_1\cos\theta_1\cos\theta - 2\,r\,r_1\sin\theta_1\sin\theta + r_1{}^2 - a^2 = 0$. By altering θ_1 and r_1, we can change the coefficients of $r\sin\theta$ and $r\cos\theta$ to anything. Similarly, altering a enables us to change the constant term. Multiplying out by a constant scaling factor enables the coefficient of r^2 to be changed. Hence, the general equation is simply $A\,r^2 + B\,r\cos\theta + C\,r\sin\theta + D = 0$.

7. $\det\begin{pmatrix} 1 & r^2 & r\sin\theta & r\cos\theta \\ 1 & r_1{}^2 & r_1\sin\theta & r_1\cos\theta \\ 1 & r_2{}^2 & r_2\sin\theta & r_2\cos\theta \\ 1 & r_3{}^2 & r_3\sin\theta & r_3\cos\theta \end{pmatrix} = 0$.

8. The general spiral has equation $A + B\,r + C\,\theta = 0$, so an interpolating spiral is $\det\begin{pmatrix} 1 & r & \theta \\ 1 & r_1 & \theta_1 \\ 1 & r_2 & \theta_2 \end{pmatrix} = 0$.

9. Consider the triangle $O\,A\,B$. $\left(|\underline{a}-\underline{b}|^2\right) = (\underline{a}-\underline{b})\cdot(\underline{a}-\underline{b}) = \left(|\underline{a}|^2\right) + \left(|\underline{b}|^2\right) - 2\,\underline{a}\cdot\underline{b}$. The last term equates to $-2\,|\underline{a}|\;|\underline{b}|\cos\theta$. This is the cosine rule, as required.