1

Pappos's Theorem: Nine Proofs and Three Variations

Bees, then, know just this fact which is of service to themselves, that the hexagon is greater than the square and the triangle and will hold more honey for the same expenditure of material used in constructing the different figures. We, however, claiming as we do a greater share in wisdom than bees, will investigate a problem of still wider extent, namely, that, of all equilateral and equiangular plane figures having an equal perimeter, that which has the greater number of angles is always greater, and the greatest plane figure of all those which have a perimeter equal to that of the polygons is the circle.

Pappos of Alexandria, ca. 340 CE

Everything in the world is strange and marvelous to well-open eyes.

José Ortega y Gasset

We will begin our journey through *projective geometry* in a slightly uncommon way. We will have a very close look at one particular geometric theorem—namely *The hexagon theorem of Pappos*. Pappos of Alexandria lived around 290–350 CE and was one of the last great Greek geometers of antiquity. He was the author of several books (some of them are unfortunately lost) that covered large parts of the mathematics known at that time. Among other topics, his work addressed questions in mechanics, dealt with the volume/circumference properties of circles, and even gave a solution to the angle trisection problem (with the additional help of a conic). The reader may take this first chapter as a kind of overture to the remainder of the book in which several topics that are important later on are introduced. Without any harm one can also skip this chapter on first reading and come back to it later.

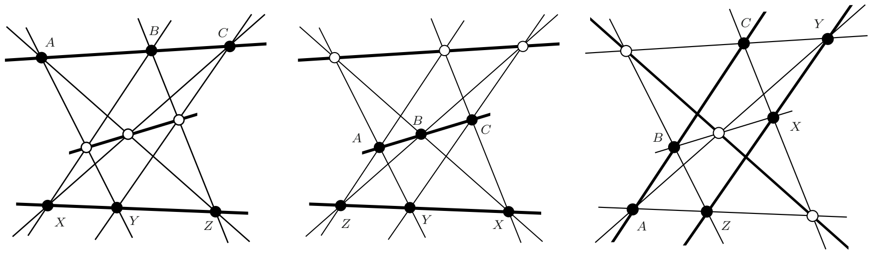


Fig. 1.1 Three versions of Pappos's theorem.

1.1 Pappos's Theorem and Projective Geometry

The theorem that we will investigate here is known as *Pappos's hexagon theorem* and usually attributed to Pappos of Alexandria (though it is not clear whether he was the first mathematician who knew about this theorem). We will later see that this theorem is special in several respects. Perhaps the most important property is that in a certain sense Pappos's theorem is the *smallest* theorem expressible in elementary terms only. The only objects involved in the statement of Pappos's theorem are *points* and *lines*, and the only relation needed in the formulation of the theorem is incidence. Properly stated, the theorem consists only of nine points and nine lines, and there is no such theorem with fewer items. Another remarkable fact is that the incidence configuration underlying Pappos's theorem has beautiful symmetry properties. Some of them are obvious, some of them slightly hidden.

Theorem 1.1 (Pappos's hexagon theorem). *Let A, B, C be three points on a straight line and let X, Y, Z be three points on another line. If the lines $\overline{AY}$, $\overline{BZ}$, $\overline{CX}$ intersect the lines $\overline{BX}$, $\overline{CY}$, $\overline{AZ}$, respectively, then the three points of intersection are collinear.*

Here *intersecting* means that two lines have exactly one point in common. The nine points of Pappos's theorem are the two triples of points on the initial two lines and the three points of intersection, which finally turn out to be collinear. The nine lines are the two initial lines, the six zigzag lines between the points, and finally the line on which the three intersection points lie. Figure 1.1 shows several instances of Pappos's theorem. The six black points correspond to the initial points, whereas the three white points are the intersections that turn out to be collinear. Observe that in our examples the positions of the nine points and lines (taken as a set) are identical. However, the role of the initial two triples of points is played by different points in each example. The first example shows the picture most often drawn in textbooks, with the final conclusion line between the two initial lines. The second picture shows that the roles of these three lines can be freely interchanged. The

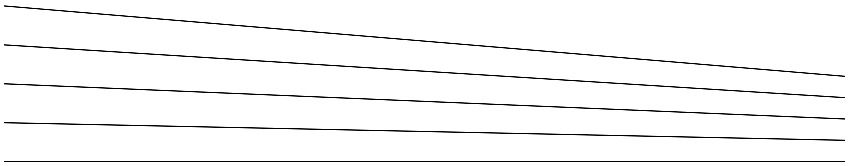


Fig. 1.2 An almost parallel bundle of lines that meet at a point far on the right.

last picture shows that also one of the inner lines can play the role of the conclusion line (by symmetry of the construction this line can be an arbitrary inner line). In fact, the automorphism group of the combinatorial structure behind Pappos's theorem admits that any pair of lines that do not have a point of the configuration in common can be taken as initial lines for the theorem.

The exact formulation of the theorem already has some subtleties, which we want to mention here. The theorem as stated above requires that the pairs of lines $(\overline{AY}, \overline{BX})$, $(\overline{BZ}, \overline{CY})$, and $(\overline{CX}, \overline{AZ})$ actually intersect, so that we can speak of the collinearity of the *intersection points*. Stated as in Theorem 1.1, Pappos's theorem is perfectly valid in Euclidean geometry. However, if we interpret it in Euclidean geometry it does not exhaust its full generality. There are essentially two different ways in which it can happen that two lines a and b may not intersect in Euclidean geometry. Either they are identical (then they have infinitely many points in common) or they are parallel (then they have no point in common). Now, *projective geometry* is an extension of Euclidean geometry in which points are added that are infinitely far away. By this we can properly speak of the intersection of parallel lines (the intersection point lies at infinity) and we get an interpretation of Pappos's theorem in which all instances of parallelism are covered as well.

The essence of real projective geometry may be summarized in the following two sentences: *Bundles of parallel lines meet at an infinite point. All infinite points are incident to a line at infinity.* Thus (real) projective geometry is an extension of Euclidean geometry by certain elements at infinity. In the next two chapters we will elaborate in depth on this extension of Euclidean geometry. In this chapter we will be content with a kind of pre-formal understanding of it.

Imagine a horizontal line a and a line b that is almost parallel to it. Both lines meet (since they are not parallel), but the point of intersection will be relatively far out. If the line b has a small negative slope, the intersection point will be far to the right of the picture. If the slope of b is small but positive, the intersection point will be far to the left. What happens if we move line b continuously from the situation with small negative slope via zero slope to the situation with small positive slope? The point of intersection will first

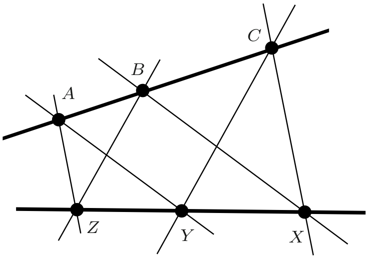


Fig. 1.3 Euclidean version of Pappos's theorem.

move farther and farther to the right (in fact, it can be arbitrarily far away). In the situation with zero, slope both lines are parallel and the intersection point vanishes. After this, the point comes back from a very far position on the left side. Projective geometry now eliminates the special case of parallel lines by postulating an additional point at infinity on the parallels. Figure 1.2 shows a bundle of lines that meet in a point very far out on the right. If this point is moved to infinity, then the lines will eventually become parallel.

It is important to notice that in the concept of projective geometry one assumes the existence of *many* different points at infinity: one for each bundle of parallel lines. All these points together form the line at infinity ℓ_∞. By introducing these additional elements, special cases get eliminated from geometry. As a matter of fact, these extensions imply that in the projective plane any two distinct points will have a unique line connecting them and any two distinct lines will have a unique point of intersection (it just may be at infinity). Furthermore, from an intrinsic viewpoint of the projective plane the infinite elements are indistinguishable from the finite elements. They have exactly the same incidence properties. (For more details see the next chapter.)

1.2 Euclidean Versions of Pappos's Theorem

By passing to a projective framework we get two kinds of benefit. First of all, we extend the scope in which Theorem 1.1 (in exactly the same formulation) is valid. Any point or any line may as well be located at an infinite position—the theorem remains true (we will prove this later). On the other hand, we may get interesting Euclidean specializations of Pappos's theorem by sending elements to infinity. One of them is given by the theorem below:

Theorem 1.2 (A Euclidean version of Pappos's theorem). *Consider two straight lines a and b in Euclidean geometry. Let A, B, C be three points on a and let X, Y, Z be three points on b. Then the following holds: If $\overline{AY} \parallel \overline{BX}$ and $\overline{BZ} \parallel \overline{CY}$ then automatically $\overline{AZ} \parallel \overline{CX}$.*

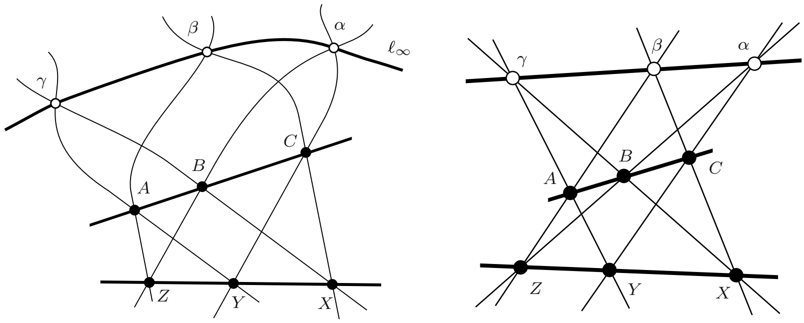


Fig. 1.4 Euclidean version of Pappos's theorem with points at infinity and line at infinity added (left). The straight version (right).

For a drawing of this theorem see Figure 1.3. Figure 1.4 illustrates how the parallelism of lines is translated to the projective setup. If $\overline{AY} \parallel \overline{BX}$ then these two lines intersect (projectively) at a point γ at infinity. Similarly we get an infinite intersection α for $\overline{BZ} \parallel \overline{CY}$. Pappos's theorem (in its projective version) states that γ and α and the intersection β of $\overline{AZ}$ with $\overline{CX}$ are collinear. Since γ and α span the line ℓ_∞ at infinity, $\overline{AZ}$ and $\overline{CX}$ must be parallel as well. In other words, the conclusion line (i.e. the line that encodes the final conclusion of the theorem) has been sent to infinity. The drawing on the right shows a straightened version of the situation with the conclusion line at a finite location. Observe the similarity of the combinatorics. Once we have introduced the concept of *projective transformation*, we will see that by a suitable transformation we can send any instance of Pappos's theorem to the above situation. Thus our Euclidean version is essentially equivalent to the full Pappos's theorem and not just a special case of it.

We will start our collection of proofs with two proofs of Theorem 1.2. It should be remarked in advance that most of our proofs will be algebraic and rely on translations of geometric facts to algebraic identities. There is a general problem with algebraic proofs: *one should never divide by zero!* This seemingly obvious fact leads to many difficulties and misunderstandings when geometric theorems are concerned. Very often, proofs work perfectly in generic situations in which no points or lines coincide or additional collinearities occur, but in certain degenerate cases they may break down. In fact, many algebraic proofs given in geometry textbooks suffer from this (d)effect and a whole branch of current ongoing research deals with the proper treatment of nondegeneracy conditions. The very statement of Theorem 1.1 carries nondegeneracy conditions in stating that the three crucial pairs of lines should actually intersect.

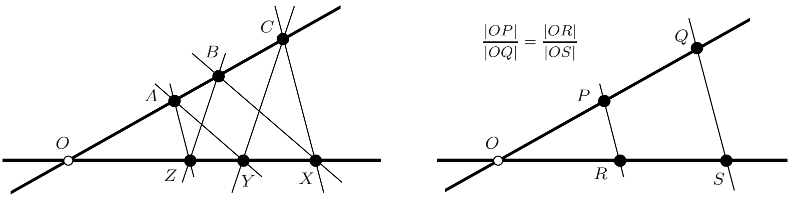


Fig. 1.5 Euclidean version of Pappos's theorem (left). Relation of parallels and segment ratios (right).

In our investigations we will bypass these degeneracy problems by assuming a few (rather strong) generic nondegeneracy properties. All nine points of the configuration should be distinct and all nine lines of the configuration should be distinct. If for a certain proof additional nondegeneracy assumptions are necessary, we will state them in the context of the proof.

Our first proof is extremely simple but (in its naive version) also of limited scope. It will be based on ratios of segment lengths. We present the proof in a version that works only under the following two additional assumptions: *The two initial lines must intersect in a point O. The triples of points on these lines should not be separated by O.* By introducing oriented lengths the proof can be easily extended to get rid of the second assumption. But we will not do this here.

Proof one: segment ratios. By $|PQ|$ we denote the distance between two points P and Q. Our first proof relies on the following fact, which is well known from school lessons on elementary geometry (compare Figure 1.5, right). Let a and b be two lines intersecting at O and let P and Q be two points on a not separated by O. Similarly, let R and S be two points on b not separated by O. Then $\overline{PR}$ and $\overline{QS}$ are parallel if and only if

$$\frac{|OP|}{|OQ|} = \frac{|OR|}{|OS|}.$$

Using this fact and the hypotheses of the theorem, the parallelism of $\overline{AY}$ and $\overline{BX}$ implies that

$$\frac{|OA|}{|OB|} = \frac{|OY|}{|OX|}.$$

Similarly, the parallelism of $\overline{BZ}$ and $\overline{CY}$ implies that

$$\frac{|OB|}{|OC|} = \frac{|OZ|}{|OY|}.$$

Since none of the six points are allowed to coincide with O, none of the denominators in the above expression are zero. Multiplying the two left sides of the equations and the two right sides of the equations and canceling the terms $|OB|$ and $|OY|$, we obtain

$$\frac{|OA|}{|OC|} = \frac{|OZ|}{|OX|}.$$

This in turn is equivalent to the fact that $\overline{AZ}$ and $\overline{CX}$ are parallel. □

At first sight the above proof seems to be very simple and elegant: Multiply two equations, cancel out terms, and get the result. Unfortunately, it has several drawbacks. One of the main problems is that we translated parallelism into ratios of lengths of segments. This translation works correctly only if the decisive points are not separated by the intersection of the lines. One can circumvent this problem by considering *oriented* line segments. The sign of the ratios used in our proof will be negative if the points are separated by O, and positive otherwise. However, to make this formally correct one should provide a case-by-case analysis that proves that the signs really have the desired behavior. A closer look shows that the proof is problematic, since we introduced the auxiliary point O and we made the proof dependent on its existence. The complete proof breaks down if the lines a and b are parallel and point O does not exist at all. In fact, the Euclidean version of Pappos's theorem does not at all depend on these special position requirements. The following proof uses only the six points of Theorem 1.2. However, we will need three slightly less trivial facts concerning polynomials and oriented areas of triangles and quadrangles.

Fact 1: Oriented triangle area.
For three points A, B, C with coordinates (a_x, a_y), (b_x, b_y), and (c_x, c_y) we can express the oriented area of the triangle $\Delta(A, B, C)$ by a polynomial in the coordinates. To be more specific, the desired polynomial is

$$\frac{1}{2}\det\begin{pmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ 1 & 1 & 1 \end{pmatrix} = \frac{1}{2}(a_x b_y + b_x c_y + c_x a_y - a_x c_y - b_x a_y - c_x b_y).$$

In fact, the specific shape of this polynomial is not important for our next proof. What is more important is the meaning of *oriented*: If the sequence of points (A, B, C) is in counterclockwise order, then the area will be calculated with positive sign. If they are in clockwise order, we will get a negative sign. If the three points are collinear, then the triangle vanishes and the area will be zero. We will denote the triangle area by $\mathbf{area}(A, B, C)$.

Fact 2: Oriented quadrangle area.
The oriented area of a quadrangle $\square(A, B, C, D)$ can be defined as

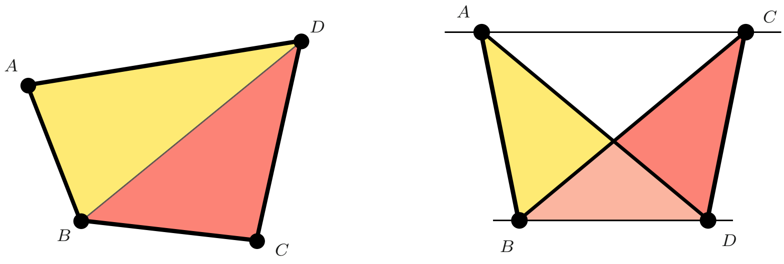


Fig. 1.6 Area of a quadrangle. The convex case (left) and a self-intersecting zero-area case (right).

$$\mathbf{area}(A,B,C,D) = \mathbf{area}(A,B,D) + \mathbf{area}(B,C,D).$$

This function is again a polynomial in the coordinates of the points. If the boundary of this triangle (the polygonal chain from A to B to C to D and back to A) is free of self-intersections, then the usual area is calculated (with sign depending of the orientation). However, if the polygon has self-intersections, then one of the triangles in the sum contributes a positive value and the other a negative value. The area of a self-intersecting quadrangle (A,B,C,D) is zero if and only if the two triangles involved in the sum have equal areas with opposite signs. Since both triangles share the edge (B,D), the zero case implies that A and C have the same altitude over this edge. In other words, the line through A and C is parallel to the line through B and D. Altogether we obtain

$$\overline{AC} \parallel \overline{BD} \text{ if and only if } \mathbf{area}(A,B,C,D) = 0.$$

Fact 3: Zero polynomials.
If a polynomial in several variables is zero in a full-dimensional region of the space of parameters, then it must be the zero polynomial. In other words, if we have a polynomial that evaluates to zero at a certain point and also for all small perturbations away from that point, then it must be the zero polynomial.

Now we have collected everything to formulate a proof of Pappos's theorem by area arguments. The following proof was given as a motivating example by D. Fearnly Sander in an article on the conceptual power of areas for theorem-proving [39].

Proof two: area method. Consider six points A,B,C,X,Y,Z in the Euclidean plane located at positions that roughly resemble the situation in Figure 1.7 on the left. This figure can be considered as being composed of two triangles $\Delta(A,C,B)$, $\Delta(X,Y,Z)$, and two quadrangles $\square(B,Y,X,A)$

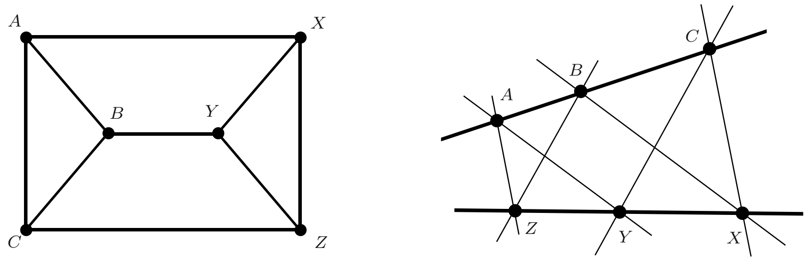


Fig. 1.7 Pappos proof by the area method.

and $\square(C, Z, Y, B)$. The sum of the oriented areas (with counterclockwise vertex labels) of these tiles equals the area of the surrounding quadrangle $\square(C, Z, X, A)$. Thus we have

$$\begin{aligned} &+ \mathbf{area}(A, C, B) \\ &+ \mathbf{area}(X, Y, Z) \\ &+ \mathbf{area}(B, Y, X, A) \\ &+ \mathbf{area}(C, Z, Y, B) \\ &- \mathbf{area}(C, Z, X, A) = 0. \end{aligned}$$

The expression on the left is obviously a polynomial, and it does not depend on the exact position of the points (since for our argument only the fact that all involved polygons are labeled counterclockwise and the fact that the inner tiles decompose the outer quadrangle were relevant). Hence by Fact 3 this formula must hold for arbitrary positions of the six points—even in degenerate cases. Now let the six points correspond to the points in Pappos's theorem. The hypotheses of Theorem 1.2 state that (A, B, C) and (X, Y, Z) are two collinear triples of points. Furthermore, we have $\overline{AY} \parallel \overline{XB}$ and $\overline{BZ} \parallel \overline{YC}$. In terms of areas, this means that

$$\mathbf{area}(A, C, B) = \mathbf{area}(X, Y, Z) = \mathbf{area}(B, Y, X, A) = \mathbf{area}(C, Z, Y, B) = 0.$$

This implies immediately that we also have

$$\mathbf{area}(C, Z, X, A) = 0,$$

since otherwise the above area-sum formula would be violated. Hence we have $\overline{AZ} \parallel \overline{XC}$ and the theorem is proved. $\square$

This proof is conceptually far less trivial than our first one, but as a benefit we get several things for free. In essence, the proof says that if four of the areas in the formula above vanish, then the last one has to vanish as well. In this form the theorem holds without any restrictions. It covers even the case of coinciding points.

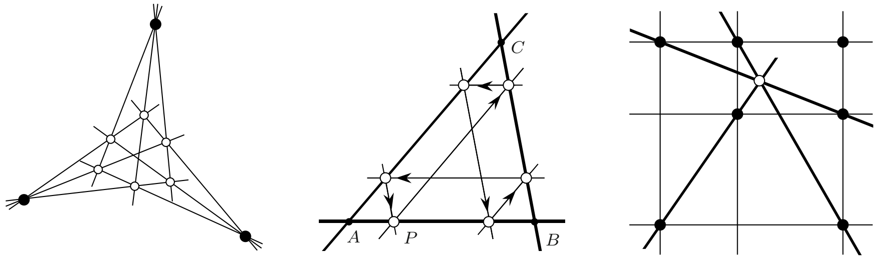


Fig. 1.8 Three versions of Pappos's Theorem.

As a second benefit we may observe that this proof is very useful for generalizations. We may consider the drawing in Figure 1.7 as the projection of a three-dimensional prism over a triangle. The five faces of the prism (two triangles and three quadrangles) correspond to the five areas involved in the proof. We can play a similar game with every three-dimensional polyhedron that has only triangles and quadrangles in its boundary. This gives an infinite collection of incidence theorems for which Pappos's theorem is the smallest example. The reader is invited to explore this field on his/her own. For instance, what is the corresponding theorem if we consider a cube as the underlying combinatorial structure?

Before we start to investigate proofs of Pappos's theorem based on concepts of projective geometry we will present some other interesting instances of Pappos's theorem. They are drawn in Figure 1.8. Lines that seem to be parallel in the drawings are really assumed to be parallel. The first picture shows a nice instance that reveals the order-three symmetry that is inherent to Pappos's theorem. The other two pictures show Euclidean specializations in which some of the points are sent to infinity. So the Euclidean instance in the second drawing could be formulated as follows.

Theorem 1.3 (Another Euclidean version of Pappos's theorem). *Start with a triangle A, B, C. Draw a point P on the line $\overline{AB}$. From there draw a parallel to $\overline{AC}$ and form the intersection with $\overline{BC}$. From this intersection draw a parallel to $\overline{AB}$ and form the intersection with $\overline{AC}$ and continue this procedure as indicated in the picture. After six steps you will reach point P again.*

The patient reader is invited to find out how the drawings in Figure 1.8 correspond to the labeling in our original version of the theorem.

1.3 Projective Proofs of Pappos's Theorem

In this section we want to present proofs in which (in contrast to the last section) we make no particular use of parallelism. All proofs in this section will rely on the collinearity properties of points only. In this respect these proofs are projective in nature, since incidence and collinearity are genuine projective concepts, while parallels are not.

The main algebraic tool used in this section is *homogeneous coordinates*, which will be introduced in much detail in later chapters. In contrast to the usual (x, y)-coordinates in the plane, homogeneous coordinates present points in the plane by three coordinates (x, y, z). Coordinate vectors that differ only by a nonzero scalar multiple are considered to be equivalent. The zero vector $(0, 0, 0)$ is excluded from consideration. Thus the nonzero points in a one-dimensional subspace of $\mathbb{R}^3$ represent the same point. A usual Euclidean plane H can be embedded in a homogeneous framework in the following way. Embed H as an affine subspace of $\mathbb{R}^3$ that does not contain the origin. Each point p of H corresponds to the one-dimensional subspace V_p spanned by p and may be represented by any nonzero vector of V_p. Conversely, each homogeneous vector (x, y, z) spans a subspace $V_{(x,y,z)}$. In general, this subspace intersects the embedded plane H at some point p. This is the point that corresponds to (x, y, z). It may happen that $V_{(x,y,z)}$ does not intersect H (this happens whenever the subspace is parallel to H). Then there is no Euclidean point associated to (x, y, z). In this case this homogeneous coordinate vector represents an infinite point (see Chapter 3 for details). Thus the finite and the infinite points can be represented by homogeneous coordinates in a completely generalized manner.

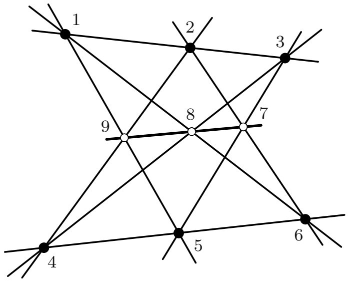
Collinearity of points in H translates to the fact that the three points in $\mathbb{R}^3$ lie in a single plane (the plane spanned by the corresponding line and the origin of $\mathbb{R}^3$). Thus if $A = (x_1, y_1, z_1)$, $B = (x_2, y_2, z_2)$, and $C = (x_3, y_3, z_3)$ are homogeneous coordinates of points, then one can test collinearity by checking the condition

$$\det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} = 0.$$

This condition works for finite as well as for infinite points. The following proof is based on this observation.

Proof three: determinant cancellations. For matters of better readability we have exchanged the labels of the points by simple digits from 1 to 9 (see Figure 1.9). For the proof we need the additional nondegeneracy condition that the triple of points $(1, 4, 7)$ is *not collinear*. The generic nondegeneracy conditions (no identical points and no identical lines) should still be valid.

Assume that $(1, 4, 7)$ is not collinear. After a suitable affine transformation (which does not affect the incidence relations of points and lines) we may assume without loss of generality that $(1, 4, 7)$ forms an equilateral triangle.



$$\begin{array}{r|ccc|}
1 & 1 & 0 & 0 \\
2 & a & b & c \\
3 & d & e & f \\
4 & 0 & 1 & 0 \\
5 & g & h & i \\
6 & j & k & l \\
7 & 0 & 0 & 1 \\
8 & m & n & o \\
9 & p & q & r
\end{array}
\qquad
\begin{array}{l}
[1,2,3]=0 \Longrightarrow \quad ce{=}bf \\
[1,5,9]=0 \Longrightarrow \quad iq{=}hr \\
[1,6,8]=0 \Longrightarrow \quad ko{=}ln \\
[2,4,9]=0 \Longrightarrow \quad ar{=}cp \\
[2,6,7]=0 \Longrightarrow \quad bj{=}ak \\
[3,4,8]=0 \Longrightarrow \quad fm{=}do \\
[3,5,7]=0 \Longrightarrow \quad dh{=}eg \\
[4,5,6]=0 \Longrightarrow \quad gl{=}ij \\
\hline
[7,8,9]=0 \Longleftarrow \quad mq{=}np
\end{array}$$

Fig. 1.9 Determinant cancellation for Pappos's theorem.

Now we embed the plane in which our configuration resides into three-space in such a way that the points 1, 4, and 7 are at the three-dimensional unit vectors $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$.

Since the configuration is now embedded in $\mathbb{R}^3$, each point is represented by three-dimensional (homogeneous) coordinates. Three points P, Q, R in our picture are collinear if and only if the determinant of the 3×3 matrix formed by their coordinates is zero. We abbreviate this determinant by $[PQR]$. The matrix in Figure 1.9 represents the coordinates of the configuration.

The letters in the matrix represent the coordinates of the remaining points. The generic nondegeneracy assumptions imply that none of the letters can be 0. This can be seen as follows. The triple of points $(3,4,7)$ cannot be collinear, since otherwise two of the configuration lines would coincide. However, the determinant formed by these points equals exactly a. Thus we get

$$0 \neq \det \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = a.$$

A similar argument works for each of the other variables.

With our special choice of coordinates, each of the eight collinearities of the hypotheses can be expressed as the vanishing of a certain 2×2 subdeterminant of the coordinate matrix. If we write down all these equations (compare Figure 1.9), multiply all left sides, and multiply all right sides, we are left with another equation $mq = np$, which translates back to the collinearity of $(7,8,9)$. By our nondegeneracy assumptions, all variables involved in the proof will be nonzero; therefore the cancellation process is feasible. $\square$

A proof that is essentially based on this structure first appeared in [14]. This proof carries remarkable symmetric structures concerning the cancellation patterns among the determinants. Structurally, it reduces to the facts that all collinearities correspond to 2×2 determinants and that each letter occurs on the left as well as on the right. The first fact is highly dependent on the choice of our basis, since only the zeros in the unit vectors are allowed to express each of the collinearities as a 2×2 determinant.

One can circumvent this problem by an even more abstract approach. Instead of dealing with concrete coordinates of points, we may deal with general properties of determinants. A fundamental role in this context is played by the *Grassmann-Plücker relations*. These relations state that for arbitrary five points A, B, C, D, E in the projective plane the following relation holds among the determinants of the homogeneous coordinates:

$$[ABC][ADE] - [ABD][ACE] + [ABE][ACD] = 0.$$

This remarkable identity is of fundamental importance for projective geometry, and we will dedicate a large part of Chapter 6 to it. For now we take the identity as an algebraic fact. On it we base our next proof.

Proof four: Grassmann-Plücker relations. We again assume that $(1, 4, 7)$ is not collinear. We consider the fact that $(1, 2, 3)$ is collinear in our theorem. Taking this Grassmann-Plücker relation

$$[147][123] - [142][173] + [143][172] = 0$$

together with the fact that $[123] = 0$, we obtain

$$[142][173] = [143][172].$$

For each of the eight collinearities of the hypotheses we can get one such equation:

$$\begin{array}{lcl}
[147][123] - [142][173] + [143][172] = 0 & \Longrightarrow & [142][173] = [143][172] \\
[147][159] - [145][179] + [149][175] = 0 & \Longrightarrow & [145][179] = [149][175] \\
[147][186] - [148][176] + [146][178] = 0 & \Longrightarrow & [148][176] = [146][178] \\
[471][456] - [475][416] + [476][415] = 0 & \Longrightarrow & [475][416] = [476][415] \\
[471][483] - [478][413] + [473][418] = 0 & \Longrightarrow & [478][413] = [473][418] \\
[471][429] - [472][419] + [479][412] = 0 & \Longrightarrow & [472][419] = [479][412] \\
[714][726] - [712][746] + [716][742] = 0 & \Longrightarrow & [712][746] = [716][742] \\
[714][753] - [715][743] + [713][745] = 0 & \Longrightarrow & [715][743] = [713][745]
\end{array}$$

Multiplying again all left sides and all right sides of the equations (and taking care of the signs of the determinants) and canceling out terms that occur on both sides, we end up with the equation

$$[718][749] = [719][748].$$

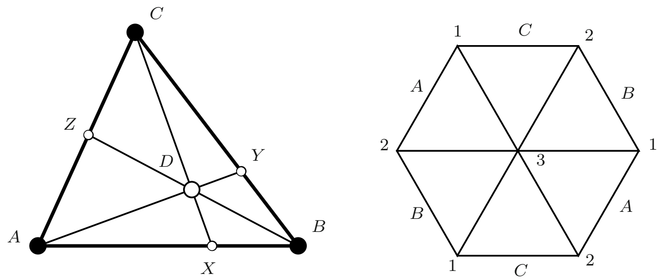


Fig. 1.10 Ceva's theorem $\frac{|AX|}{|XB|} \cdot \frac{|BY|}{|YC|} \cdot \frac{|CZ|}{|ZA|} = 1$ (left). The pasting scheme for the proof (right).

(The cancellation is feasible since all involved determinants will be nonzero by our nondegeneracy conditions.) By the Grassmann-Plücker relation

$$[714][789] - [718][749] + [719][748] = 0,$$

this implies that $[714][789] = 0$. Since [147] was assumed to be nonzero, this implies that $[789] = 0$, which in turn is equivalent to the collinearity of $(7, 8, 9)$. □

This proof is very similar to the previous one. However, working directly on the level of determinants makes the special choice of the basis no longer necessary. There are amazingly many theorems in projective geometry that can be proved by this generic determinant calculus, and one can even base methods for automatic theorem-proving on them. (For details on this subject see [15, 30, 109].)

Our next proof reveals a topological structure that underlies Pappos's theorem. The proof can be thought of as gluing together several triangular shapes to form a closed oriented surface. The fact that the surface is closed (has no boundary) corresponds to the conclusion of the theorem.

For this proof to work out we need a kind of basic building block: The *theorem of Ceva* (see for instance [28]). Ceva's theorem states that if in a triangle the sides are cut by three concurrent lines that pass through the corresponding opposite vertices, then the product of the three (oriented) length ratios along each side equals 1.

In fact, this theorem is almost trivial if one views the length ratios as ratios of certain triangle areas. For this observe that if the line (A, B) is cut by the line (C, D) at a point X, then we have

$$\frac{|AX|}{|XB|} = -\frac{\mathbf{area}(C, X, A)}{\mathbf{area}(C, X, B)} = -\frac{\mathbf{area}(C, D, A)}{\mathbf{area}(C, D, B)}, \qquad (*)$$

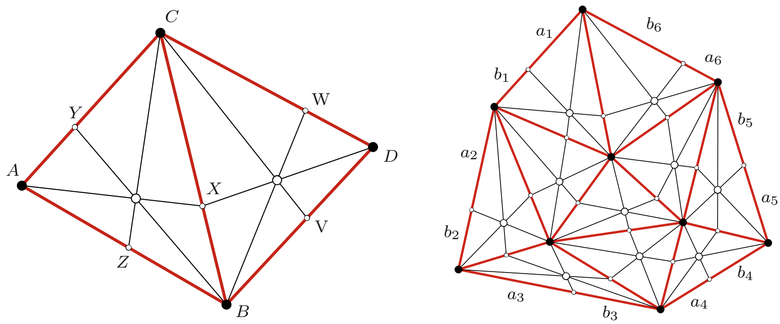


Fig. 1.11 Pasting copies of Ceva's theorem.

where $\mathbf{area}(A, B, C)$ denotes the oriented area of the triangle (A, B, C). In order to prove Ceva's theorem, we consider the obvious identity

$$\frac{\mathbf{area}(CDA)}{\mathbf{area}(CDB)} \cdot \frac{\mathbf{area}(ADB)}{\mathbf{area}(ADC)} \cdot \frac{\mathbf{area}(BDC)}{\mathbf{area}(BDA)} = -1$$

(note that the oriented triangle area is an alternating function and that each triangle in the denominator occurs as well in the numerator). Applying the above identity $(*)$, we immediately get Ceva's theorem. The converse of Ceva's theorem holds as well: If the product of the three ratios equals 1, then the three lines in the interior will meet.

Now consider the situation in which two Ceva triangles are glued together along an edge in a way such that they share the point on this edge. Multiplying the two Ceva expressions, we see that the ratio on the inner edge cancels, and we are left only with terms that live on the boundary of the figure (see Figure 1.11 (left)). We obtain

$$\frac{|AZ|}{|ZB|} \cdot \frac{|CY|}{|YA|} \cdot \frac{|BV|}{|VD|} \cdot \frac{|DW|}{|WC|} = 1.$$

We can extend this process to an arbitrary collection of triangles that are glued edge to edge. An edge can be used either by only one triangle (then it is a boundary edge) or by exactly two triangles. The whole collection of patched triangles should be orientable (thus we obtain an orientable triangulated 2-manifold with boundary). All triangles of the collection should be equipped with Ceva configurations that have the additional property that points on interior edges are shared by the Ceva configurations of two adjacent triangles. We consider the product of all corresponding Ceva expressions. After cancellation of the ratios that correspond to inner edges we are left with an expression that contains only oriented length ratios from the boundary. For instance, in the situation of Figure 1.11 (right) we get

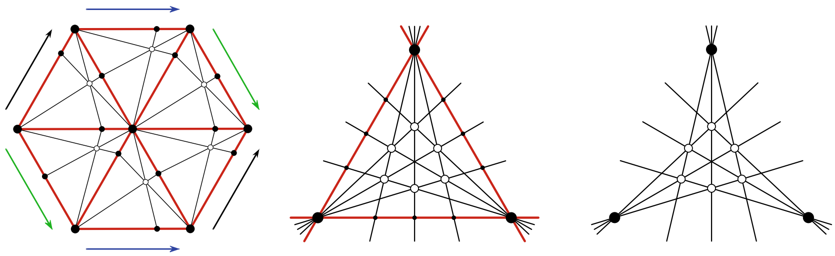


Fig. 1.12 Creating Pappos's theorem from six copies of Ceva's theorem.

$$\frac{a_1}{b_1} \cdot \frac{a_2}{b_2} \cdot \frac{a_3}{b_3} \cdot \frac{a_4}{b_4} \cdot \frac{a_5}{b_5} \cdot \frac{a_6}{b_6} = 1.$$

The inner part of the structure cancels completely and does not contribute to the product on the boundary. Now, if we have a collection of triangles that has nothing more than a triangular boundary (i.e., a 2-manifold with a single triangular hole), then the Ceva condition on the whole is automatically satisfied, and we can paste in a final triangle that carries a Ceva configuration. In other words, if we have an orientable triangulated 2-manifold *without boundary* and we have a Ceva configuration on all triangles but one (such that the edge points are shared), then a Ceva configuration can automatically be put on the final triangle. This is an incidence theorem. We now will show that using the right manifold, Pappos's theorem can be put in exactly this form.

Proof five: pasting copies of Ceva's theorem. Consider six triangles that are arranged as in Figure 1.10 on the right. Furthermore, identify opposite edges of the hexagon as indicated in the drawing. This can be done by placing the six triangles one over the other (think of the hexagon as made of paper and fold it appropriately) and gluing together corresponding opposite edges. Now place a Ceva configuration on each of the edges in a way such that whenever two triangles meet at an edge, the corresponding two points on this edge are identified. Our considerations above show that if the edge points are located such that five of the triangles carry proper Ceva configurations, then the last Ceva configuration is satisfied automatically. The figure in the middle shows the situation after all the triangle edges have been identified. Observe that the points on the edges of the outer triangle as well as the edges themselves do not contribute to the incidence theorem. What is left after these elements are deleted is exactly a drawing of Pappos's theorem. □

A proof very similar to this was given by H.S.M. Coxeter and S.L. Greitzer [28]. Their proof was based on Menelaus configurations instead of Ceva configurations but is essentially similar. In [110] one can find an elaborate treatment of the question of which geometric theorems can be proved by similar manifold arguments.

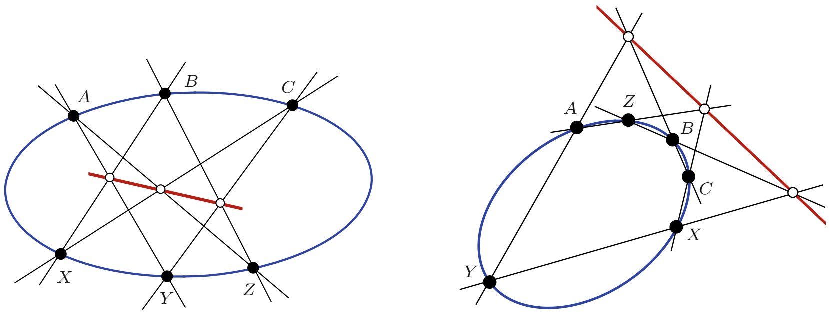


Fig. 1.13 Two instances of Pascal's theorem.

1.4 Conics

This section deals with generalizations and variations of Pappos's theorem. In particular, we will study what happens if we consider pairs of lines as degenerate cases of a degree-two curve (an ellipse, hyperbola, or parabola) in the plane. Degree-two curves are often also called *conics*, and they correspond to solutions of (homogeneous) quadratic equations in homogeneous coordinates. More specifically, a conic in the plane is characterized by six homogeneous parameters a, b, c, d, e, f and consists of the set of all points with homogeneous coordinates (x, y, z) that satisfy the equation

$$a \cdot x^2 + b \cdot y^2 + c \cdot xy + d \cdot xz + e \cdot yz + f \cdot z^2 = 0.$$

Let (x, y, z) be a solution of this equation. Since the total degree in x, y, z of each summand is the same (namely two), every scalar multiple $\lambda \cdot (x, y, z)$ is also a solution of this equation. Thus we may think of each solution as a point in the real projective plane. The totality of these points forms the conic. The geometric form of the conic depends on the special values of the parameters. Projectively, there is no difference between ellipse, hyperbola, and parabola. These three cases simply reflect different ways in which the line at infinity ℓ_∞ intersects the conic. If there is no intersection, the conic is an ellipse; if there are two intersections, the conic is a hyperbola (it has two infinite points, which correspond to the two asymptotes); if there is just one intersection, the conic is a parabola (which turns out to be a limit case between the two other possibilities).

There is one interesting special case that is also important from a projective point of view: the conic may degenerate into two lines (which may even coincide). This happens whenever the the term $ax^2+by^2+cxy+dxz+eyz+fz^2$ factorizes into two linear components:

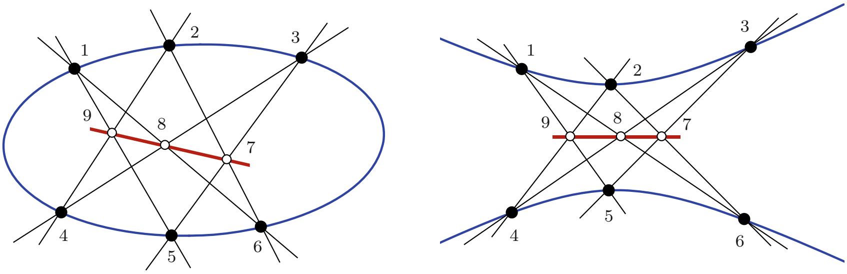


Fig. 1.14 Deformations of Pascal's theorem and labeling for the proof.

$$ax^2 + by^2 + cxy + dxz + eyz + fz^2 = (\alpha_1 x + \beta_1 y + \gamma_1 z) \cdot (\alpha_2 x + \beta_2 y + \gamma_2 z).$$

In this case the conic consists of two lines, each one described by the linear equation in one of the factors.

In general, five points in the projective plane determine a unique conic passing through each of them. Thus it is a truly projective condition whether six points lie on a common conic or not. In Chapter 10 we will prove that six points A, B, C, X, Y, Z are on a common conic if and only if the following condition among the determinants of the homogeneous coordinates holds:

$$[ABC][AYZ][XBZ][XYC] = [XYZ][XBC][AYC][ABZ].$$

We will use this nice characterization to prove the following well-known variation (or better generalization) of Pappos's theorem:

Theorem 1.4 (Variation 1: Pascal's Theorem). *Let A, B, C, X, Y, Z be six points on a conic. If the lines $\overline{AY}$, $\overline{BZ}$, $\overline{CX}$ intersect the lines $\overline{BX}$, $\overline{CY}$, $\overline{AZ}$ respectively, then the three points of intersection are collinear.*

Two instances of the theorem can be found in Figure 1.13. Pascal's theorem is named after the famous Blaise Pascal and was discovered by (the 16-year-old) Pascal in 1640. This is about 1300 years after the discovery of Pappos's theorem. Nevertheless, it is obviously a generalization of Pappos's theorem. If the conic in Pascal's theorem degenerates to consist of two lines, then we immediately obtain Pappos's theorem. We will prove Theorem 1.4 by a determinant cancellation argument similar to the one used in our fourth proof. Figure 1.14 shows two instances of Pascal's theorem one with an ellipse and one with a hyperbola. If we smoothly deform the first into the second, we will pass through the degenerate situation that resembles Pappos's theorem.

Proof six: Pascal's theorem. Again we assume for nondegeneracy reasons that no points and no lines of the theorem coincide. For the labeling in the proof we refer to Figure 1.14. Consider the following determinant equations:

$$\begin{array}{lcr}
\textbf{conic:} & \Rightarrow & [125]\,[136]\,[246]\,[345] = +\,[126]\,[135]\,[245]\,[346] \\
[159] = 0 & \Longrightarrow & [157]\,[259] = -\,[125]\,[597] \\
[168] = 0 & \Longrightarrow & [126]\,[368] = +\,[136]\,[268] \\
[249] = 0 & \Longrightarrow & [245]\,[297] = -\,[247]\,[259] \\
[267] = 0 & \Longrightarrow & [247]\,[268] = -\,[246]\,[287] \\
[348] = 0 & \Longrightarrow & [346]\,[358] = +\,[345]\,[368] \\
[357] = 0 & \Longrightarrow & [135]\,[587] = -\,[157]\,[358] \\
\hline
[987] = 0 & \Longleftarrow & [287]\,[597] = +\,[297]\,[587]
\end{array}$$

The first line encodes that the points $1, \ldots, 6$ lie on a conic. The next six lines are consequences of Grassmann-Plücker relations and the six collinearity hypotheses of our theorem. If we multiply all expressions on the left and all expressions on the right and cancel determinants that occur on both sides, we end up with the last expression, which (under the nondegeneracy assumption that $[157] \neq 0$) implies the desired collinearity of $(7, 8, 9)$. □

Similar to Pappos's theorem, there is a variety of reformulations and specializations. A nice reformulation is the following: *If a hexagon is inscribed in a conic in the projective plane, then the opposite sides of the hexagon meet in three collinear points.* Or if one prefers a Euclidean variant of this in which the conclusion line is sent to infinity, one could state the following: *If a hexagon is inscribed in a conic and two pairs of opposite edges are parallel, then so is the third pair.* There is another nice way to derive even more incidence theorems from Pascal's theorem. Assume that the conic has a fixed position. If two of the points in Pascal's theorem that are joined by a line continuously approach each other until they meet, their joining line will in the limit case become a tangent to the conic at the position where the two points are located. Thus we obtain as limit cases situations in which also tangents are involved (observe that tangents are proper concepts of projective geometry).

Instances of degenerate versions are given in Figure 1.15. The leftmost picture shows a smallest degenerate situation. The label 15 symbolizes that points 1 and 5 are identified. The labeling is consistent with the labeling in Figure 1.14. The join of 1 and 5 becomes the tangent at the point 15. One can also read the construction in the reverse direction. If a conic $\mathcal{C}$ and a point 15 on it are given, then one can construct the tangent at 15 by choosing four arbitrary points $2, 3, 4, 6$ on $\mathcal{C}$ and constructing the joins and intersections as given by the picture to arrive finally at point 9 another point on the tangent. This fact was also known to Pascal, and is one of the main applications of his theorem. The second picture shows in essence the same situation as the first one. However, here the point 15 has been sent to infinity and the corresponding tangent is located at the line at infinity. By this the conic becomes a parabola, and the two other lines through 15 become parallel to the symmetry axis of the parabola. Now the theorem reads as follows: *Start*

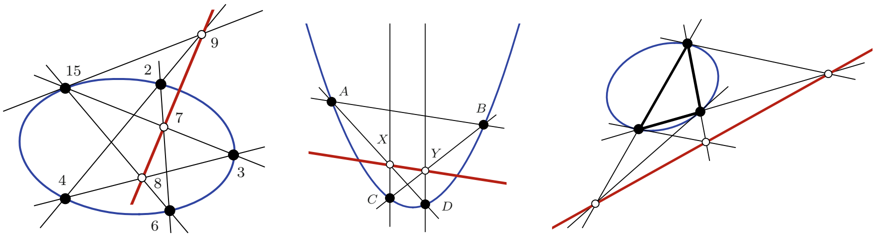


Fig. 1.15 Degenerate versions of Pascal's theorem.

with four points A, B, C, D on a parabola. Draw two lines through C and D parallel to the symmetry axis of the parabola. Intersect them with $\overline{AD}$ and $\overline{BC}$, respectively. Then the join of the two intersections is parallel to the join of A and B. The right figure shows an even more degenerate situation: *Inscribe a triangle into a conic. Form the tangents at the vertices. Intersect them with the opposite sides of the triangle. The three intersections are collinear.*

1.5 More Conics

We can think of Pascal's theorem being derived from Pappos's theorem by considering two lines that do not have a configuration point in common as a (degenerate) conic. Pascal's theorem says that the theorem stays valid even if the conic is not degenerate. The same process can be applied two more times to obtain a theorem with three conics and three lines. For this consider the left part of Figure 1.16. The blue conic arises from merging the upper and the lower lines of the drawing. The red and the green conics arise from merging two other lines. Amazingly, the new configuration still forms a theorem. If all incidences except for the blue line are satisfied as indicated in the picture, then the three white points are automatically collinear (we will prove this in a minute). First we observe that there are two combinatorially different ways of merging three pairs of lines in Pappos's theorem to three conics. The second possibility is shown in Figure 1.16 on the right. Also in this case we get a theorem. To see that they are combinatorially different, observe that in one picture the three lines meet in a point; in the other one they don't. Both theorems are an instance of an even more general fact that is a consequence of *Bézout's theorem* from algebraic geometry (see [19, 40]). An *algebraic curve* of degree d is the zeroset of a homogeneous polynomial of degree d. Thus conics are algebraic curves of degree 2. Bézout's theorem can be stated in the following way: *If an algebraic curve of degree n and an algebraic curve of degree m intersect, then either the number of intersections is finite and less*

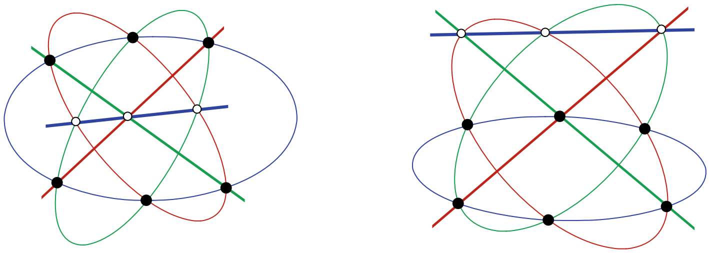


Fig. 1.16 Generalizations of Pascal's Theorem

than or equal to $n \cdot m$, or the curves intersect in infinitely many points and share a component. Now we can prove the following very strong statement:

Theorem 1.5 (Variation 2: Cayley-Bacharach-Chasles theorem). *Let A and B be two curves of degree three intersecting in nine proper points. If six of these points are on a conic, the remaining three points are collinear.*

Proof seven: algebraic curves. Let A and B be the curves and let $p_A(x, y, z)$ and $p_B(x, y, z)$ be the corresponding homogeneous polynomials of degree three. Bézout's theorem implies that if the two curves A and B have only finitely many points in common, then they can have at most nine points of intersection. Call them $1, \ldots, 9$. And assume that $1, \ldots, 6$ are on a conic C with polynomial p_C. We will prove that $7, 8, 9$ are collinear. Consider a linear combination $p_\mu = p_A + \mu \cdot p_B$ of the two polynomials for some real parameter μ. The polynomial p_μ has the following properties. First it is again a degree-three polynomial. Second, it passes through all nine points $1, \ldots, 9$ (each of these points is a zero of both p_A and p_B, so it is also a zero of any linear combination of them). Now consider an additional point q on the conic C distinct from $1, \ldots, 6$. There is a μ such that p_μ also passes through q (to find μ we just have to solve a linear equation $p_A(q) + \mu p_B(q) = 0$). Consider p_μ with this specific value μ. The curve p_μ passes through $1, \ldots, 6$ and through q. Thus it shares these *seven* points with the conic C. Bézout's theorem implies that p_μ must have C as one component. Thus we have $p_\mu = p_C \cdot L$ with a linear equation L (otherwise p_μ cannot have degree three). This implies that the points $7, 8, 9$ are all contained in the line described by the linear equation L. □

The situation of the theorem is sketched in Figure 1.17. This theorem was independently discovered by several people. Most probably Chasles was the first to discover this theorem, in a slightly more general version in 1885. As so often in mathematics, the theorem is usually attributed to others, in this case namely to Cayley and to Bacharach, who published similar results

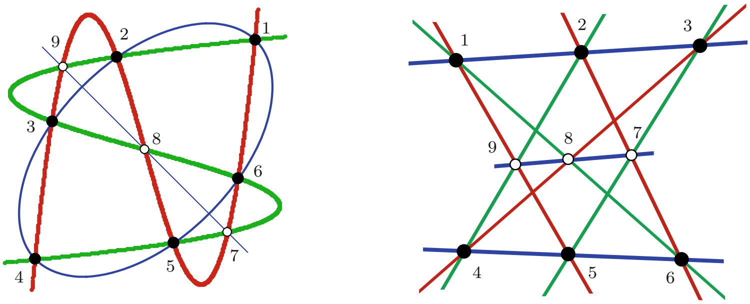


Fig. 1.17 If two cubics intersect in nine points six of which are on a conic, then the remaining three points are collinear.

later than Chasles (for a historic account see [38, 63]). The theorems shown in Figure 1.16 are immediate specializations of this theorem. There the two curves of degree three decompose into the product of a quadratic curve (the conic) and a linear curve (the line). So the two red components of the picture form one curve of degree three, and the two green components form the other one. The rest is a literal application of the above theorem. One can even go one step further and consider Pappos's original theorem as a direct consequence of Theorem 1.5. For this one simply has to consider three of the lines as one cubic and another three as the other cubic. The color coding in Figure 1.17 makes the decomposition clear.

1.6 Complex Numbers and Circles

We are almost at the end of our journey around Pappos's theorem. In this section we want to take the considerations of the last chapter still a little further and draw a surprising connection to the geometry of circles in the plane. Circles are an intrinsically Euclidean concept. Thus if we do so we have again to talk about the exact position of our line at infinity ℓ_∞. As already mentioned in Section 1.3, homogeneous coordinates can be considered as embedding the Euclidean plane into $\mathbb{R}^3$ at some affine hyperplane. This time (and this will be done quite often later in the book) we will choose the affine hyperplane $\{(x, y, z) \mid z = 1\}$ for this embedding. Thus a point with Euclidean coordinates (x, y) can be represented by homogeneous coordinates $(x, y, 1)$ or any nonzero scalar multiple of this vector. The infinite points are those with coordinates $(x, y, 0)$.

We now want to study *circles* under this special embedding. A circle is a special conic. Thus we want to find out which quadratic equations will correspond to circles. A circle is usually given by its center (c_x, c_y) and a radius r. In Euclidean geometry the circle equation can be written as

$$(x - c_x)^2 + (y - c_y)^2 - r^2 = 0.$$

Expanding this term and interpreting it in homogeneous coordinates with $z = 1$, we can rewrite it as

$$\begin{aligned}(x - c_x \cdot z)^2 + (y - c_y \cdot z)^2 - r^2 \cdot z^2 &= \\ x^2 - 2c_x xz + c_x^2 z^2 + y^2 - 2c_y yz + c_y^2 z^2 - r^2 z^2 &= \\ x^2 + y^2 - 2c_x xz - 2c_y yz + (c_x^2 + c_y^2 - r^2)z^2 &= 0.\end{aligned}$$

The last line gives the interpretation of the circle in parameters of a general conic. The circle is a special conic for which the coefficients of x^2 and y^2 are equal and the coefficient of xy vanishes.

There is a surprising (and very deep) connection between circles and complex numbers. Let us investigate what happens when we intersect a circle with the line at infinity. In other words, we search for solutions of the above equation with $z = 0$. Clearly the solution must be complex, since no circle has real intersections with the line at infinity (this property is possessed only by hyperbolas and parabolas). In the circular case for $z = 0$ the equation degenerates to

$$x^2 + y^2 = 0.$$

Up to scalar multiples we get the two solutions

$$\mathtt{I} = (1, i, 0) \qquad \text{and} \qquad \mathtt{J} = (1, -i, 0).$$

These solutions are complex points at the line at infinity. Moreover (and this is an important fact!), they do not depend on the specific choice of the specific circle. Thus we can say *All circles pass through* I *and* J *and any conic passing through these points is a circle.*

This fact is perhaps the most important connection of Euclidean and projective geometry. It allows us to express relations about circles as incidence relations of conics that involve the points I and J. In a very strong sense we could say that *every Euclidean incidence theorem can be expressed as a projective theorem in which two points play the special roles of* I *and* J. In a sense, Chapters 16 to 26 of this book are dedicated to the elaboration of this fact. Here we will make a small application of it in the context of Pappos's theorem. Consider again the two generalizations given in Figure 1.16. These two pictures are reproduced again in the first row of Figure 1.18. In these pictures the points in which three conics meet are marked by white dots. In the same way as we assumed in Section 1.2 that a certain line is located at the line at infinity we will now assume that in each picture two of these points are located at the points I and J. All other points should stay at real positions. A conic that passes through I and J is a circle. Thus the conics in our theorem become circles (this is similar to the effect that two lines become

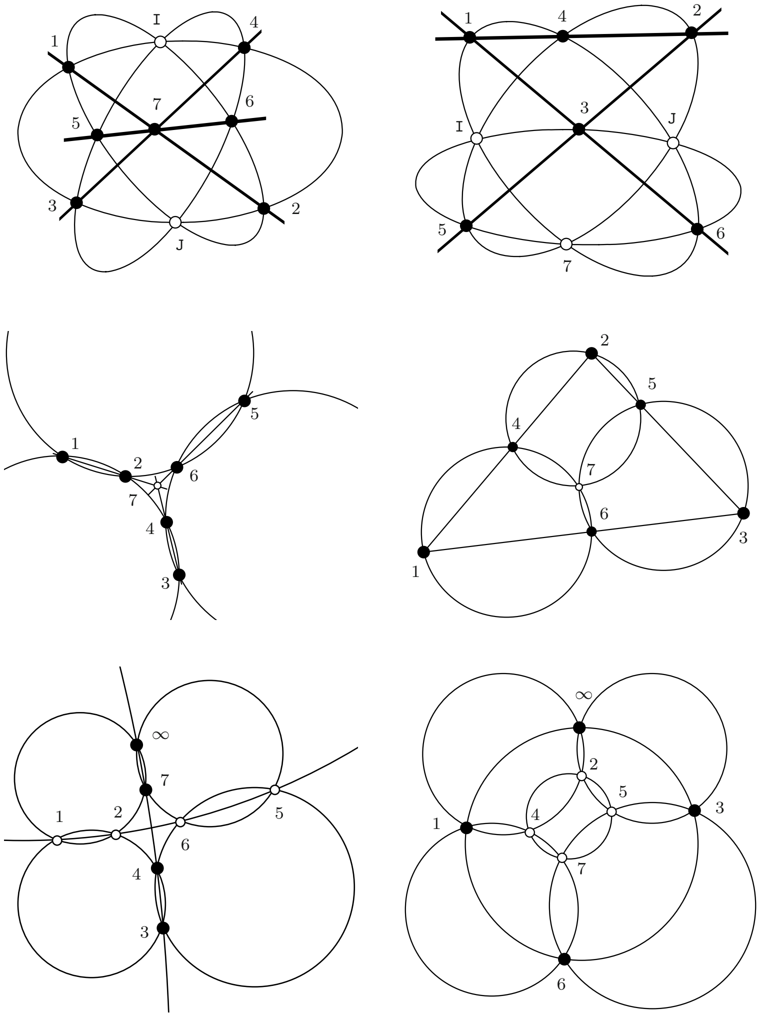


Fig. 1.18 Metamorphoses of theorems.

parallel if their point of intersection is located at an infinite position). So the two theorems can be interpreted as Euclidean theorems about seven points, three lines, and three circles. The corresponding pictures are shown in Figure 1.18 in the second row. For instance, the first of these two theorems can be stated as follows: *Given three circles that intersect mutually in two points, the three lines spanned by the intersections of each pair of circles meet in a point.* The meeting point corresponds to point 7 in the original theorem.

We can even go one step further. We can interpret straight lines as circles with infinite radius. There is a particular way of extending Euclidean

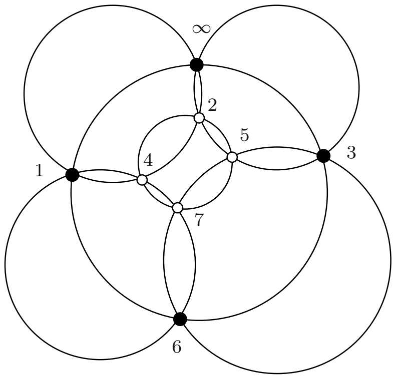


Fig. 1.19 Miquel's theorem.

geometry that reflects this way of thinking. For this we introduce *one* point ∞ at infinity and assume that straight lines are circles that in particular contain this point. (A word of caution: one should not confuse this extension of Euclidean geometry by *one* point with the projective plane we introduced earlier. In the projective plane a line at infinity was introduced. The extension by only one point used here has something to do with the *projective complex line* and is called the *one-point compactification of the Euclidean plane* and will be investigated later, in Chapter 17.

In this setup we no longer have to distinguish between lines and circles. Lines are just circles of infinite radius. In this interpretation our two theorems could be stated as theorems on six circles and eight points (we interpret the infinite point ∞ just as an ordinary point). The last row of Figure 1.18 gives a drawing of the situation in which ∞ is located at a finite position. For instance, the second theorem (which is a well-known fact from circle geometry) could be stated as follows.

Theorem 1.6 (Variation 3: Miquel's theorem). *Consider four points A, B, C, D on a circle. Draw four more circles C_1, C_2, C_3, C_4 that pass through the pairs of points (A, B), (B, C), (C, D), and (D, A), respectively. Now consider the other intersections of C_i and C_{i+1} for $i = 1, \ldots, 4$ (indices modulo 4). These four intersections are again cocircular.*

We will give an elementary proof of this theorem by calculations of angle sums. The basic fact that we will need for this proof is illustrated in Figure 1.20. If we consider a secant AB of a circle and if we look at this secant from two other different points C and D of the circle (which are on the same side of AB), we will see the secant in the same angle. If the points C and D are at opposite sides of the secant we will have complementary angles. Observe that the angles in Figure 1.20 are assumed to be oriented angles. Thus the complementary angle has to be counted with negative sign. If one takes care of the orientation of the angles one could say that the difference of the two angles at C and D will in both cases be a multiple of π. Conversely, four points A, B, C, D lie on a common circle if the difference of the angles (under which AB is seen) at C and D is a multiple of π. Thus we get a characterization of four points on a circle in terms of angles.

In principle, Miquel's theorem can now easily be proven by considering angle sums among the six involved circles. However, we here will prefer a more algebraic approach that expresses the angle relations in terms of complex numbers. For this assume that all eight points in the picture are finite and consider the picture of Miquel's theorem embedded in the complex number plane $\mathbb{C}$. We consider A, B, C, D from Figure 1.20 as complex numbers. Then, for instance, $A - C$ forms a complex number that points in the direction from C to A. Forming the quotient $\frac{A-C}{B-C}$, we get a complex number whose argument (the angle with respect to the real axis) is exactly the angle at point C. Similarly, $\frac{A-D}{B-D}$ gives a complex number that describes the angle at point D. We can compare these two angles by forming again the quotient of these two numbers: $\frac{A-C}{B-C} / \frac{A-D}{B-D}$. This number will be real if and only if the two angles differ by a multiple of π.

Taking everything together, we get the following characterization of four points being cocircular (possibly with infinite radius): *Four points A, B, C, D in the complex plane are cocircular if and only if*

$$\frac{(A-C)(B-D)}{(B-C)(A-D)}$$

is a real number.

The above expression is called a *cross-ratio*, and we will later on see that cross-ratios play a fundamental and omnipresent role in projective geometry (see Chapters 4 and 5). With the help of cross-ratios we can easily state a proof of Miquel's theorem.

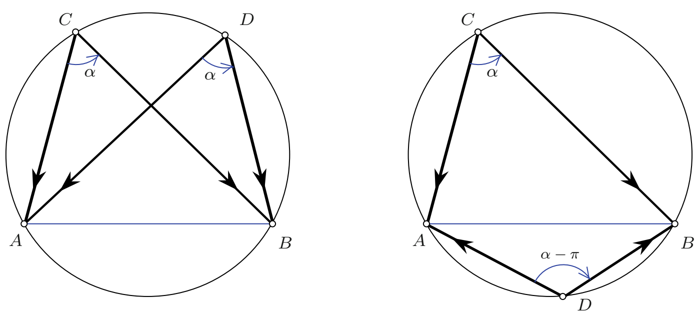


Fig. 1.20 Angles in a circle.

Proof eight: cross-ratio cancellations. Assume that the quadruples of points (A, B, C, D), (A, B, E, F), (B, C, F, G), (C, D, G, H), (D, A, H, E) are cocircular. From this we obtain that the following cross-ratios are all real:

$$\tfrac{(A-B)(C-D)}{(C-B)(A-D)},\quad \tfrac{(F-B)(A-E)}{(A-B)(F-E)},\quad \tfrac{(C-B)(F-G)}{(F-B)(C-G)},\quad \tfrac{(H-D)(C-G)}{(C-D)(H-G)},\quad \tfrac{(A-D)(H-E)}{(H-D)(A-E)}.$$

Multiplying all these numbers and canceling terms that occur in the numerator as well as in the denominator, we are left with the expression

$$\frac{(F-G)(H-E)}{(H-G)(F-E)}.$$

Since this expression is the product of real numbers, it must itself be real. By our above observations this expresses exactly the cocircularity of (E, F, G, H), which is the conclusion of our theorem. □

1.7 Finally...

We will end this section with an almost trivial proof of Pappos's theorem in its full generality by simply expanding an algebraic term. Still we need a little preparation for this. Again consider the original points of Pappos's theorem expressed in homogeneous coordinates. Thus we assume that the drawing plane H is again embedded in $\mathbb{R}^3$ at a position that does not contain the origin of $\mathbb{R}^3$. As before, each point p is represented by a three-dimensional vector (x, y, z). This time we will take *all* points of $\mathbb{R}^3 - \{(0,0,0)\}$ into account. For this we identify the vector (x, y, z) with all of its nonzero scalar multiples $(\lambda x, \lambda y, \lambda z)$, $\lambda \neq 0$. By this $\mathbb{R}^3 - \{(0,0,0)\}$ is divided into equivalence classes. Each equivalence class represents a point of the projective plane. A point of the drawing plane H can be represented by its actual (x, y, z) position or by any nonzero scalar multiple of it. Conversely, for a point (x, y, z) of $\mathbb{R}^3 - \{(0,0,0)\}$ we consider the line $l_{(x,y,z)}$ through it and the origin. The point in H that is represented by (x, y, z) is the intersection of $l_{(x,y,z)}$ and H. If this intersection does not exist, (x, y, z) represents an infinite point.

In this setup a straight line g in H may be considered a two-dimensional linear space spanned by the elements of g and the origin of $\mathbb{R}^3$. Such a line may by represented by a linear equation

$$\{(x, y, z) \in \mathbb{R}^3 - \{(0,0,0)\} \mid ax + by + cz = 0\}$$

given by parameters $(a, b, c) \in \mathbb{R}^3 - \{(0,0,0)\}$. Thus points as well as lines are represented by nonzero vectors in $\mathbb{R}^3$. A line g is incident to a point p if and only if the standard scalar product $\langle p, g\rangle$ is zero.

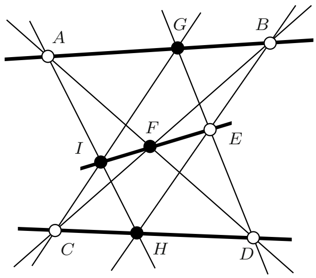


Fig. 1.21 A construction sequence for Pappos's theorem.

This observation gives us the key to a very elegant method of calculating the line that connects two points p and q. We simply need a vector g with the property $\langle p, g\rangle = \langle q, g\rangle = 0$. Such a vector can simply be calculated by the cross product $p \times q$. Similarly, the intersection of two lines g and h asks for a vector p with the property $\langle p, g\rangle = \langle p, h\rangle = 0$. Thus the intersection can be calculated by $g \times h$. So we can apply the cross product to calculate intersections and joins in projective geometry. (We will learn much more of this in Chapter 3.)

What happens if we try to form the join of two identical points p and q? If p and q represent the same point, they must be scalar multiples of each other: $q = \lambda p$. Performing the cross product, we obtain $p \times q = p \times \lambda p = \lambda(p \times p) = (0, 0, 0)$. Obtaining a zero vector as result is an indication of performing a degenerate operation. A similar effect results when we try to intersect two identical lines.

How can we test collinearity of three points p, q, r? The points are collinear if and only if the representing vectors are linearly dependent in $\mathbb{R}^3$. Thus we can test collinearity by the condition $\det(p, q, r) = 0$.

Now we can express Pappos's theorem as a sequence of nested cross products and a determinant. Expanding the final term and observing that it is zero will prove the theorem.

Proof nine: brute force. We give a construction sequence for the Pappos's configuration. We start with five free points A, B, C, D, E (compare Figure 1.21). The coordinates for the remaining four points in the construction can be calculated by

$$\begin{aligned} F &= (A \times D) \times (B \times C), \\ G &= (A \times B) \times (D \times E), \\ H &= (C \times D) \times (B \times E), \\ I &= (A \times H) \times (C \times G). \end{aligned}$$

Testing the final collinearity boils down to testing whether $\det(E, F, I) = 0$. The following session of the computer algebra program *Mathematica* shows an evaluation of these expressions. All output except for the final result has been suppressed. The final “0” proves Pappos’s theorem.

```
In[49]:= cross[{a_, b_, c_}, {x_, y_, z_}] := {b*z - c*y, -a*z + c*x, a*y - b*x}

In[50]:= a = {a1, a2, a3};
         b = {b1, b2, b3};
         c = {c1, c2, c3};
         d = {d1, d2, d3};
         e = {e1, e2, e3};
         f = cross[cross[a, d], cross[b, c]];
         g = cross[cross[a, b], cross[d, e]];
         h = cross[cross[c, d], cross[b, e]];
         i = cross[cross[a, h], cross[c, g]];
         Det[{e, f, i}]

Out[50]= 0
```

What does this evaluation indeed prove? It shows that when we perform the construction sequence independently of the initial choice of the coordinates of A, B, C, D, E, the final determinant will be zero. This may happen for two different reasons. Either during the construction sequence we run into a degenerate situation (such as the intersection of identical lines) that introduce a zero vector as an intermediate result. Or all operations were valid (this will be the case for almost every instance) and the final points E, F, I are indeed collinear. □

A word of caution: The last proof is very general and seems to be straightforward. Still, the help of a computer is essential here. Performing the calculations by hand would require one to perform all cross products and to evaluate the final determinant. The final term has altogether 15456 summands of degree 15. They can be canceled in pairs, which gives the final result.