12

Affine Transformations

Chaotic features of the World erase
And you will see its Beauty.
— Alexander A. Block (1880–1921)[1]

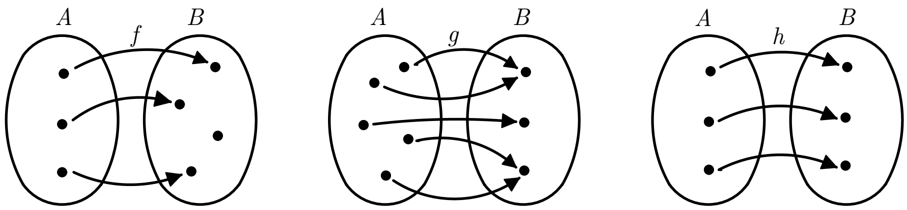
12.1 Introduction

Suppose we are struggling with a geometric problem concerning an arbitrary triangle or an arbitrary parallelogram. How often we would wish for the triangle to be an equilateral or $45^\circ - 90^\circ - 45^\circ$ triangle, or for the parallelogram to be a square! The solution is so easy in these cases. But we know that these would be just very particular instances of the problem. Solving them will make us feel better, but not much better. Well, the good news is that for *some* problems, solving just a particular instance turns out to be sufficient to claim that the problem is solved in complete generality! In this chapter we learn how to recognize some of these problems, and we justify such an approach.

We start by reviewing some familiar concepts. Let A and B be sets. A **function** or **mapping** f from A to B, denoted $f : A \to B$, is a set of ordered pairs (a, b), where $a \in A$ and $b \in B$, with the following property: for every $a \in A$ there exists a unique $b \in B$ such that $(a, b) \in f$. The fact that $(a, b) \in f$ is usually denoted by $f(a) = b$, and we say that f maps a to b. Another way to denote that f maps a to b is $f : a \mapsto b$; if it is clear which function is being discussed, we will often just write $a \mapsto b$. We also say that b is the **image** of a (in f), and that a is a **preimage** of b (in f). The set A is called the **domain** of f and the set B is the **codomain** of f. The set $f(A) = \{f(a) : a \in A\}$ is a subset of B, called the **range** of f.

A function $f : A \to B$ is **surjective** (or **onto**) if $f(A) = B$; that is, f is surjective if every element of B is the image of at least one element of A. A function $f : A \to B$ is **injective** (or **one-to-one**) if each element in the range of f is the image of *exactly* one element of A; that is, f is injective if $f(x) = f(y)$ implies $x = y$. A function $f : A \to B$ is **bijective** if it is both surjective and injective.

[1] Translated from the Russian by Vera Zubareva.



(i) f is injective (ii) g is surjective (iii) h is bijective

FIGURE 12.1.

If $f : A \to B$ and $g : B \to C$ are functions, then the **composition** of f and g, denoted $g \circ f$, is a function from A to C such that $(g \circ f)(a) = g(f(a))$ for any $a \in A$. The proof of Theorem 12.1 is left to the reader and can be found in many texts.

Theorem 12.1. *A composition of two bijections is a bijection.*

If $f : A \to B$, then $f^{-1} : B \to A$ is the **inverse** of f if $(f^{-1} \circ f)(a) = a$ for any $a \in A$ and $(f \circ f^{-1})(b) = b$ for any $b \in B$. A function f has an inverse if and only if f is a bijection.

Let $\mathbb{E}^2$ denote the Euclidean plane. Introducing a coordinate system[2] OXY on $\mathbb{E}^2$, we can identify every point P with the ordered pair of its coordinates (x_P, y_P); alternatively, P can be identified with its position vector, $\overrightarrow{OP} = \langle x_P, y_P \rangle$. The collection of all such vectors form a vector space,[3] namely $\mathbb{R}^2$. If $\vec{x}$ represents the vector with initial point at the origin and terminal point at (x_P, y_P), then $\overrightarrow{OP}$, $\langle x_P, y_P \rangle$, and $\mathbf{x}$ can also be used to denote $\vec{x}$.

A **transformation** of a set is a bijection of the set to itself. It is easy to see that any transformation $f : \mathbb{E}^2 \to \mathbb{E}^2$ corresponds to a bijection $\tilde{f} : \mathbb{R}^2 \to \mathbb{R}^2$, in that $\tilde{f}(\langle x_P, y_P \rangle) = \langle x_{P'}, y_{P'} \rangle$ whenever $f(P) = P'$. Since f and $\tilde{f}$ uniquely define one another within a fixed coordinate system, we will also refer to $\tilde{f}$ as a transformation of the plane, and we will write f to denote either a mapping of $\mathbb{E}^2$ to $\mathbb{E}^2$ or a mapping of $\mathbb{R}^2$ to $\mathbb{R}^2$. It will be clear from the context which of the two mappings f represents.

Just as any point P in OXY corresponds to a unique vector $\overrightarrow{OP}$, each figure φ in $\mathbb{E}^2$ uniquely corresponds to a set of vectors $\overrightarrow{OP}$ of $\mathbb{R}^2$, where $P \in \varphi$. We say that this set of vectors is a **figure in $\mathbb{R}^2$**, and we denote it again by φ. The set $f(\varphi)$ is defined as $\{f(P) : P \in \varphi \subseteq \mathbb{E}^2\}$, or $\{f(\overrightarrow{OP}) : \overrightarrow{OP} \in \varphi \subseteq \mathbb{R}^2\}$. It is not hard to make the relationship between point spaces and vector spaces more precise, but we will not do it here.[4] In fact, we freely interchange the representations of point and vector, (x, y) and $\langle x, y \rangle$, when they are domain elements of a function f.

Transformations of the plane and their application to solving geometry problems form the focus of this chapter. The transformations we study will be of two types, illustrated by the following examples:

$$f(\langle x, y \rangle) = \langle 2x - 3y, x + y \rangle \quad \text{and} \quad g(\langle x, y \rangle) = \langle 2x - 3y + 1, x + y - 4 \rangle.$$

---

[2] Recall that OXY denotes a coordinate system (not necessarily Cartesian) with axes $\overleftrightarrow{OX}$ and $\overleftrightarrow{OY}$.

[3] Students who have studied some linear algebra may recall that a vector space is a collection of objects on which an "addition" operation may be performed in such a way that nice properties like commutativity and the existence of additive inverses hold, but a precise definition of vector space is not necessary in order to continue reading.

[4] See, for example, [**34**], [**50**], or [**65**] for rigorous expositions.

At this point it is not obvious that f and g are bijections, but this will be verified later in the chapter. To get a more concrete sense of what f and g do, consider how they "transform" the vectors $\langle 0, 0\rangle$, $\langle 0, 1\rangle$, $\langle 1, 0\rangle$, and $\langle 1, 1\rangle$.

$\vec{x}$	$f(\vec{x})$	$g(\vec{x})$
$\langle 0, 0\rangle$	$\langle 0, 0\rangle$	$\langle 1, -4\rangle$
$\langle 0, 1\rangle$	$\langle -3, 1\rangle$	$\langle -2, -3\rangle$
$\langle 1, 0\rangle$	$\langle 2, 1\rangle$	$\langle 3, -3\rangle$
$\langle 1, 1\rangle$	$\langle -1, 2\rangle$	$\langle 0, -2\rangle$

Notice that the origin, $\vec{0}$, is fixed under f, while $g(\langle 0, 0\rangle) = \langle 1, -4\rangle$. Notice also that $f(\langle 0, 1\rangle + \langle 1, 0\rangle) = f(\langle 0, 1\rangle) + f(\langle 1, 0\rangle)$; again, this is not true of g. These properties of f are indicative of the **linearity** of that mapping. A function $T : \mathbb{R}^2 \to \mathbb{R}^2$ is called **linear** if $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$ for any vectors $\vec{x}$ and $\vec{y}$, and $T(k\vec{x}) = kT(\vec{x})$ for any vector $\vec{x}$ and scalar k. The reader can verify that these properties hold for f but not for g.

As will be shown later in this chapter, both f and g map a line segment to a line segment. Therefore, knowing where f and g map the points corresponding to the vectors $\langle 0, 0\rangle$, $\langle 0, 1\rangle$, $\langle 1, 1\rangle$, and $\langle 1, 0\rangle$ is sufficient for determining the image of the unit square, S, having vertices at these four points. Figure 12.2 shows S together with $f(S)$ and $g(S)$. Notice that both $f(S)$ and $g(S)$ are parallelograms; Theorem 12.7 will prove that this is not a coincidence.

12.2 Matrices

Transformations of $\mathbb{E}^2$ or $\mathbb{R}^2$ are often studied via another type of mathematical object, the matrix. Though the benefits of using the language of matrices are not striking when we study $\mathbb{E}^2$, matrices

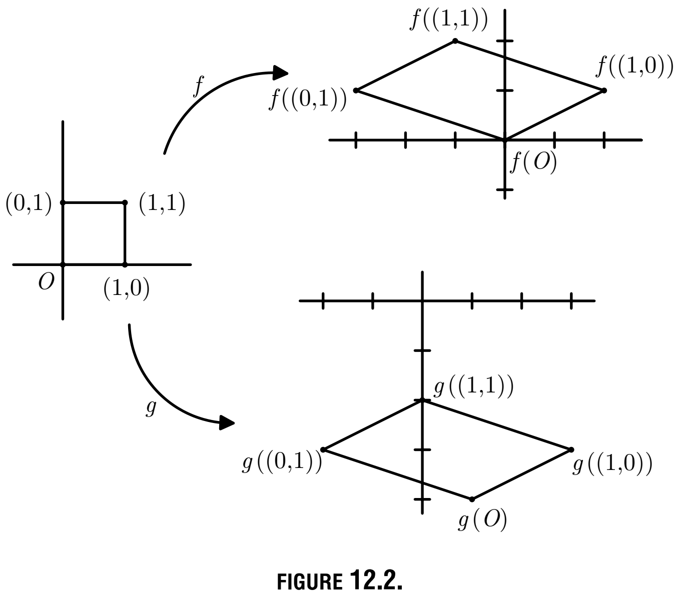


FIGURE 12.2.

turn out to be very convenient when generalizing geometric notions of the plane to spaces of higher dimensions.[5]

An $m \times n$ **matrix A** is a rectangular array of real numbers,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

The entry in the i^{th} row and the j^{th} column is denoted a_{ij}, and we often write $\mathbf{A} = [a_{ij}]$. Two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{A}' = [a'_{ij}]$ are called **equal** if they have an equal number of rows, an equal number of columns, and $a_{ij} = a'_{ij}$ for all i and j. When the matrix is $n \times n$, so that there are an equal number of rows and columns, the matrix is called a **square** matrix. Notice that a vector $\vec{v} = \langle v_1, v_2 \rangle$ can be thought of as the 1×2 matrix $[v_1 \ v_2]$, called a "row vector." It can also be thought of as a "column vector" by writing $\vec{v}$ as the 2×1 matrix $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$.

If $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ are both $m \times n$ matrices, then the **sum** $\mathbf{A} + \mathbf{B}$ is the $m \times n$ matrix $\mathbf{C} = [c_{ij}]$ in which $c_{ij} = a_{ij} + b_{ij}$. If $\mathbf{A} = [a_{ij}]$ is an $m \times n$ matrix and $c \in \mathbb{R}$, then the **scalar multiple** of $\mathbf{A}$ by c is the $m \times n$ matrix $c\mathbf{A} = [ca_{ij}]$. (That is, $c\mathbf{A}$ is obtained by multiplying each entry of $\mathbf{A}$ by c.)

The **product AB** of two matrices is defined when $\mathbf{A} = [a_{ij}]$ is an $m \times n$ matrix and $\mathbf{B} = [b_{ij}]$ is an $n \times p$ matrix. Then $\mathbf{AB} = [c_{ij}]$, where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. For example, if $\mathbf{A}$ is a 2×2 matrix, and $\mathbf{B}$ is a 2×1 matrix, then

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} \\ a_{21}b_{11} + a_{22}b_{21} \end{bmatrix}.$$

We say that here we multiply $\mathbf{A}$ by a (column) vector. Notice that $\mathbf{BA}$ is not defined in this case.

If $\mathbf{A}$ and $\mathbf{B}$ are both 2×2 matrices,

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

Although $\mathbf{BA}$ is defined in this case, *in general* $\mathbf{BA}$ is not equal to $\mathbf{AB}$. So matrix multiplication is not commutative. These two instances of matrix multiplication (when $\mathbf{A}$ is a 2×2 matrix and $\mathbf{B}$ is a 2×1 or a 2×2 matrix) are the only ones we will need in this book. In what follows, no matter whether $\vec{x}$ is a 1×2 vector or 2×1 vector, when it is used in the expression $\mathbf{A}\vec{x}$, it is always understood as a column vector, i.e., as a 2×1 matrix.

Theorem 12.2 summarizes some of the most useful properties of matrix operations. Its proof can easily be produced by the reader (part (4) is the most difficult) or may be found in a standard linear algebra text.

[5] Here, when we say "language," we mean the objects, their notation, operations on the objects, and properties of those operations – similar to the "languages" of trigonometry, algebra, logic, and calculus.

Theorem 12.2.

(1) *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $m \times n$ *matrices, then* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

(2) *If* $\mathbf{A}$, $\mathbf{B}$, *and* $\mathbf{C}$ *are* $m \times n$ *matrices, then* $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$.

(3) *Given an* $m \times n$ *matrix* $\mathbf{A}$, *there exists a unique* $m \times n$ *matrix* $\mathbf{B}$ *such that* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ *is the* **zero matrix** *(that is, the matrix with 0 in every entry).*

(4) *If* $\mathbf{A}$ *is an* $m \times n$ *matrix,* $\mathbf{B}$ *is an* $n \times p$ *matrix, and* $\mathbf{C}$ *is a* $p \times q$ *matrix, then* $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.

(5) *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $m \times n$ *matrices,* $\mathbf{C}$ *is an* $n \times p$ *matrix, and* $\mathbf{D}$ *is a* $q \times m$ *matrix, then* $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ *and* $\mathbf{D}(\mathbf{A} + \mathbf{B}) = \mathbf{DA} + \mathbf{DB}$.

(6) *If* $r, s \in \mathbb{R}$, $\mathbf{A}$ *is an* $m \times n$ *matrix, and* $\mathbf{B}$ *is an* $n \times p$ *matrix, then*

(a) $r(s\mathbf{A}) = (rs)\mathbf{A} = s(r\mathbf{A})$, *and*
(b) $\mathbf{A}(r\mathbf{B}) = r(\mathbf{AB})$.

(7) *If* $r, s \in \mathbb{R}$, *and* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $m \times n$ *matrices, then*

(a) $(r + s)\mathbf{A} = r\mathbf{A} + s\mathbf{A}$, *and*
(b) $r(\mathbf{A} + \mathbf{B}) = r\mathbf{A} + r\mathbf{B}$.

Using the notation of matrices, we can represent the functions

$$f(\langle x, y\rangle) = \langle 2x - 3y, x + y\rangle \quad \text{and} \quad g(\langle x, y\rangle) = \langle 2x - 3y + 1, x + y - 4\rangle$$

using matrix multiplication as follows. First, let $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, and let

$$\mathbf{A} = \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}.$$

Then

$$f(\vec{x}) = \mathbf{A}\vec{x} = \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

One way to think about the matrix $\mathbf{A}$ corresponding to the transformation f is that the columns of $\mathbf{A}$ specify the images of the vectors $\vec{i} = \langle 1, 0\rangle$ and $\vec{j} = \langle 0, 1\rangle$. Using matrix multiplication, we see that $\mathbf{A}\vec{i} = \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and $\mathbf{A}\vec{j} = \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$, as illustrated in Figure 12.3.

If we let $\vec{b} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$, then the same 2×2 matrix $\mathbf{A}$ gives

$$\begin{aligned} g(\vec{x}) = \mathbf{A}\vec{x} + \vec{b} &= \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 1 \\ -4 \end{bmatrix} \\ &= \begin{bmatrix} 2x - 3y \\ x + y \end{bmatrix} + \begin{bmatrix} 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 2x - 3y + 1 \\ x + y - 4 \end{bmatrix}. \end{aligned}$$

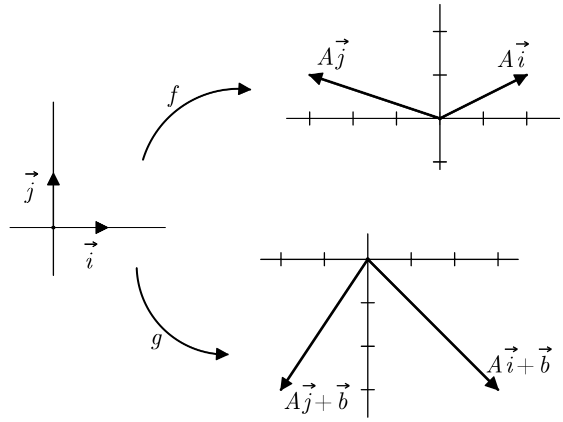


FIGURE 12.3.

Now, $\mathbf{A}\vec{i}+\vec{b}=\begin{bmatrix}2 & -3\\1 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}+\begin{bmatrix}1\\-4\end{bmatrix}=\begin{bmatrix}3\\-3\end{bmatrix}$, and
$\mathbf{A}\vec{j}+\vec{b}=\begin{bmatrix}2 & -3\\1 & 1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}+\begin{bmatrix}1\\-4\end{bmatrix}=\begin{bmatrix}-2\\-3\end{bmatrix}$, again illustrated in Figure 12.3.

Notice that using column form for vectors allows us to write the elements of the domain of f and g on the right side of the matrix representing the function, just as the variable is on the right when using the notation $f(x)$. If we compose two functions, f and g, where $f(\vec{x})=\mathbf{A}\vec{x}$ and $g(\vec{x})=\mathbf{B}\vec{x}$, then $(g\circ f)(\vec{x})=g(f(\vec{x}))=\mathbf{B}(\mathbf{A}\vec{x})=(\mathbf{BA})\vec{x}$. Hence the matrix that corresponds to the composition $g\circ f$ is **BA**.[6]

The 2×2 **identity matrix**, $\mathbf{I}_2=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$, has special significance. It is easy to check that $\mathbf{I}_2$ is the only matrix with the property that if **A** is any 2×2 matrix, $\mathbf{AI}_2=\mathbf{I}_2\mathbf{A}=\mathbf{A}$, and $\mathbf{I}_2\vec{x}=\vec{x}$ for each $\vec{x}$ in $\mathbb{R}^2$. Clearly $\mathbf{I}_2$ is a matrix analog of the number 1.[7]

Furthermore, for *some*[8] square matrices **A**, there exists a matrix **B** such that $\mathbf{AB}=\mathbf{BA}=\mathbf{I}_2$. It is easy to show that if **B** exists, then it is unique. Such a matrix **A** is called **invertible** or **nonsingular**, and the corresponding matrix **B** (more often denoted $\mathbf{A}^{-1}$) is called the **inverse** of **A**. For example, the matrix $\mathbf{A}=\begin{bmatrix}2 & -3\\1 & 1\end{bmatrix}$ is invertible, with $\mathbf{A}^{-1}=\begin{bmatrix}1/5 & 3/5\\-1/5 & 2/5\end{bmatrix}$, because

$$\begin{bmatrix}2 & -3\\1 & 1\end{bmatrix}\begin{bmatrix}1/5 & 3/5\\-1/5 & 2/5\end{bmatrix}=\begin{bmatrix}1/5 & 3/5\\-1/5 & 2/5\end{bmatrix}\begin{bmatrix}2 & -3\\1 & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}.$$

As $\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}_2$, the matrix $\mathbf{A}^{-1}$ is also invertible and **A** is its inverse.

[6] The order of matrices in this multiplication matches the order of the corresponding functions in the notation $g\circ f$, but the order in which the two functions are composed does not match the order in which they are written. For this reason, some authors prefer to replace the notation $f(\vec{x})$ with $(\vec{x})f$. Then $\vec{x}$ can be thought of as a row vector, and we write $(\vec{x})f=\vec{x}\mathbf{A}$. For f and g as in our case, this would make $\vec{x}\mathbf{AB}$ correspond to $(\vec{x}f)g$. While this notation may be less familiar, at least the orders match! One cannot have it all. . . .

[7] These statements can be made in greater generality. The $n\times n$ identity matrix, $I_n=[c_{ij}]$, is the matrix having $c_{ij}=1$ if $i=j$ and $c_{ij}=0$ otherwise. Then, if **A** is any $m\times n$ matrix, $\mathbf{AI}_n=\mathbf{I}_m\mathbf{A}=\mathbf{A}$.

[8] Actually, for *most* of them, but we will not discuss the meaning of "most" at this point.

Let $\mathbf{A}$ be an invertible matrix, let $\vec{b}$ be a vector, and let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined via $\vec{x} \mapsto \mathbf{A}\vec{x} + \vec{b}$. For any vector $\vec{y}$, the following are all equivalent.

$$\begin{aligned}
f(\vec{x}) &= \vec{y} \\
\mathbf{A}\vec{x} + \vec{b} &= \vec{y} \\
\mathbf{A}\vec{x} &= \vec{y} - \vec{b} \\
\mathbf{A}^{-1}(\mathbf{A}\vec{x}) &= \mathbf{A}^{-1}(\vec{y} - \vec{b}) \\
(\mathbf{A}^{-1}\mathbf{A})\vec{x} &= \mathbf{A}^{-1}\vec{y} - \mathbf{A}^{-1}\vec{b} \\
\mathbf{I}_2\vec{x} &= \mathbf{A}^{-1}\vec{y} - \mathbf{A}^{-1}\vec{b} \\
\vec{x} &= \mathbf{A}^{-1}(\vec{y} - \vec{b})
\end{aligned}$$

We conclude that f^{-1} exists and can be given by $f^{-1}(\vec{x}) = \mathbf{A}^{-1}(\vec{x} - \vec{b})$. (One can also easily check that for every vector $\vec{x}$, $(f^{-1} \circ f)(\vec{x}) = \vec{x}$ and $(f \circ f^{-1})(\vec{x}) = \vec{x}$.) Therefore, both f and f^{-1} are bijections on $\mathbb{R}^2$, also called **transformations of the plane**.

A transformation f of the plane of the form $f(\vec{x}) = \mathbf{A}\vec{x} + \vec{b}$ where $\mathbf{A}$ is an invertible matrix is called an **affine transformation** of the plane. Since $\mathbf{A}^{-1}$ is invertible if and only if $\mathbf{A}$ is, we have just proven the following.

Theorem 12.3. *An affine transformation of the plane has an inverse that is also an affine transformation of the plane.*

Obviously, it will be useful to know whether a given matrix has an inverse. Fortunately, there is a nice computational tool available for this. The **determinant** of a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the number $ad - bc$, denoted $\det \mathbf{A}$. The primary significance of the determinant follows from Theorem 12.4.

Theorem 12.4. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* 2×2 *matrices. Then*

(1) $\mathbf{A}$ *is invertible if and only if* $\det \mathbf{A} \neq 0$.

(2) *If* $\det \mathbf{A} \neq 0$*, then* $\mathbf{A}^{-1} = \dfrac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

(3) $\det(\mathbf{AB}) = (\det \mathbf{A})(\det \mathbf{B})$.

Proof. (3) Suppose that $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$. Then

$$\mathbf{AB} = \begin{bmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{bmatrix}.$$

Consequently,

$$\begin{aligned}
\det(\mathbf{AB}) &= (aa' + bc')(cb' + dd') - (ab' + bd')(ca' + dc') \\
&= aa'dd' + bb'cc' - ab'dc' - ba'cd' = (ad - bc)(a'd' - c'b') = (\det \mathbf{A})(\det \mathbf{B}).
\end{aligned}$$

(2) We demonstrate that $\mathbf{A}^{-1} = \dfrac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ by matrix multiplication:

$$\mathbf{A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad - bc & -ab + ba \\ cd - dc & -cb + da \end{bmatrix} = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = (ad - bc)\mathbf{I}_2.$$

By part (6) of Theorem 12.2, $\mathbf{A} \cdot \left(\frac{1}{\det \mathbf{A}}\right) \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \mathbf{I}_2$. It can similarly be demonstrated that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_2$.

(1) Part (2) above shows that if $\det \mathbf{A} \neq 0$, then $\mathbf{A}$ has an inverse.

Suppose that $\det \mathbf{A} = 0$. If $\mathbf{A}^{-1}$ exists, then $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_2$, and by Part (3) of this theorem, $(\det \mathbf{A})(\det \mathbf{A}^{-1}) = \det \mathbf{I}_2$. Since $\det \mathbf{I}_2 = 1 \cdot 1 - 0 \cdot 0 = 1$, this gives $0 \cdot \det \mathbf{A}^{-1} = 1$, a contradiction. □

Corollary 12.5. *A composition of affine transformations is an affine transformation.*

Proof. Let $f(\vec{x}) = \mathbf{A}\vec{x} + \vec{a}$ and $g(\vec{x}) = \mathbf{B}\vec{x} + \vec{b}$ be affine transformations. Then $(g \circ f)(\vec{x}) = g(f(\vec{x})) = \mathbf{B}(\mathbf{A}\vec{x} + \vec{a}) + \vec{b} = (\mathbf{BA})\vec{x} + (\mathbf{B}\vec{a} + \vec{b})$. Since $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices, $\mathbf{BA}$ is invertible. This can be seen in several ways.

Note that

$$(\mathbf{A}^{-1}\mathbf{B}^{-1})(\mathbf{BA}) = \mathbf{A}^{-1}(\mathbf{B}^{-1}\mathbf{B})\mathbf{A} = \mathbf{A}^{-1}(\mathbf{I}_2)\mathbf{A} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_2,$$

and similarly, $(\mathbf{BA})(\mathbf{A}^{-1}\mathbf{B}^{-1}) = \mathbf{I}_2$. Thus,

$$(\mathbf{BA})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}.$$

Therefore $\mathbf{BA}$ is invertible, and we conclude that $g \circ f$ is an affine transformation.

Alternatively, by Theorem 12.4(1), since $\mathbf{A}$ and $\mathbf{B}$ are invertible, $\det \mathbf{A}$ and $\det \mathbf{B}$ are both nonzero. Hence, by Theorem 12.4(3), $\det(\mathbf{BA}) = (\det \mathbf{B})(\det \mathbf{A}) \neq 0$. Therefore, by Theorem 12.4(1), $\mathbf{BA}$ is invertible, and we again conclude that $g \circ f$ is an affine transformation. □

The following simple theorem, whose proof is left to the reader, relates the determinant to collinearity of vectors.

Theorem 12.6. *Let* $\mathbf{A}$ *be a* 2×2 *matrix. Then the following statements are equivalent.*

(1) $\det \mathbf{A} = 0$.

(2) *The row vectors of* $\mathbf{A}$ *are collinear.*

(3) *The column vectors of* $\mathbf{A}$ *are collinear.*

Homotheties, in which the vector $\vec{x}$ is mapped to the vector $k\vec{x}$ where $k \neq 0$ (see Section 3.2.7), provide examples of one type of affine transformation. Two other kinds of affine transformations are of particular interest: translations and rotations.

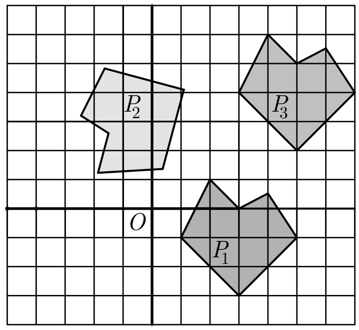


FIGURE 12.4.

A **translation** is an affine transformation of the form

$$f(\vec{x}) = \vec{x} + \vec{b} = \mathbf{I}_2\vec{x} + \vec{b}.$$

A translation can be pictured as "sliding" all points of the plane in the direction given by $\vec{b}$, by the distance $|\vec{b}|$.

A **rotation** is an affine transformation of the form

$$f(\vec{x}) = R_0^\theta(\vec{x}),$$

where $R_0^\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Under a rotation, the vector $\overrightarrow{OP}$ is mapped to the vector $\overrightarrow{OP'}$, where O is the origin, $m(\angle POP') = \theta$, and $|\overrightarrow{OP}| = |\overrightarrow{OP'}|$. This transformation can be pictured by imagining sticking a pin at the origin to fix that point, and then rotating the entire plane counterclockwise by the angle θ.

In Figure 12.4, the original figure, P_1, is mapped to P_2 via rotation by an angle of $120°$, and mapped to P_3 via translation by the vector $\langle 2, 5\rangle$. The effect of translations and rotations on conic sections will be explored in Section 12.5.

12.3 Properties

Some things never change.

— *Various*[9]

One of the essential aspects of affine transformations is that certain geometric properties are preserved, or *invariant*, under any affine transformation. If a geometric figure φ possesses a property that is invariant under affine transformations, then the image, $f(\varphi)$, under any affine transformation f will also have that property. Theorem 12.7 establishes the invariance of key properties under affine transformations. Note that the proof regularly uses the linearity of the function $\vec{x} \mapsto A\vec{x}$, i.e., the facts that $\mathbf{A}(t\vec{u}) = t(\mathbf{A}\vec{u})$ and $\mathbf{A}(\vec{u} + \vec{w}) = \mathbf{A}\vec{u} + \mathbf{A}\vec{w}$, where $\mathbf{A}$ is a 2×2 matrix and t is a scalar. Remembering that vectors can be thought of as 2×1 matrices, these facts follow from parts (6)(b) and (5), respectively, of Theorem 12.2.

[9] In the context of this section the phrase was used in the title of [**32**].

Theorem 12.7. *Let $f(\vec{x}) = \mathbf{A}\vec{x} + \vec{b}$ be an affine transformation. Then f*

(1) *maps a line to a line,*
(2) *maps a line segment to a line segment,*
(3) *preserves the property of parallelism among lines and line segments,*
(4) *maps an n-gon to an n-gon,*
(5) *maps a parallelogram to a parallelogram,*
(6) *preserves the ratio of lengths of two parallel segments, and*
(7) *preserves the ratio of areas of two figures.*

Proof.

(1) Let l be a line, and let $l : \vec{p} + t\vec{u}$, $t \in \mathbb{R}$, be an equation of l in vector form (as specified in Problem 11.7). Then, for every $t \in \mathbb{R}$,

$$f(\vec{p} + t\vec{u}) = \mathbf{A}(\vec{p} + t\vec{u}) + \vec{b} = (\mathbf{A}\vec{p} + \vec{b}) + t(\mathbf{A}\vec{u}) = \vec{p}_1 + t\vec{u}_1,$$

where $\vec{p}_1 = \mathbf{A}\vec{p} + \vec{b}$ and $\vec{u}_1 = \mathbf{A}\vec{u}$. Hence $f(l) = l_1$, where $l_1 : \vec{p}_1 + t\vec{u}_1, t \in \mathbb{R}$, is again a line.

(2) The proof is the same as that for (1), with t restricted to $[0, 1]$.

(3) Suppose that $l : \vec{p} + t\vec{u}$ and $m : \vec{q} + t\vec{v}$, $t \in \mathbb{R}$, are parallel lines. Then $\vec{v} = k\vec{u}$ for some $k \in \mathbb{R}$. Therefore,

$$\begin{aligned} f(\vec{p} + t\vec{u}) &= \mathbf{A}(\vec{p} + t\vec{u}) + \vec{b} = (\mathbf{A}\vec{p} + \vec{b}) + t(\mathbf{A}\vec{u}) = \vec{p}_1 + t\vec{u}_1 \quad \text{and} \\ f(\vec{q} + t\vec{v}) &= f(\vec{q} + t(k\vec{u})) = \mathbf{A}(\vec{q} + t(k\vec{u})) + \vec{b} \\ &= (\mathbf{A}\vec{q} + \vec{b}) + t(\mathbf{A}k\vec{u}) = \vec{q}_1 + t(k\vec{u}_1). \end{aligned}$$

That is, l and m are mapped to lines l_1 and m_1 that are parallel.

It is clear that for two line segments or a line and a line segment the proof is absolutely analogous.

(4) We prove this by strong induction on n. For the base case, when $n = 3$, consider a triangle T. Then T and its interior can be represented in vector form as $T : \vec{u} + s\vec{v} + t\vec{w}$, where $s, t \in [0, 1]$, $s + t \leq 1$, and the vectors $\vec{v}$ and $\vec{w}$ are not collinear. Then

$$\begin{aligned} f(T) = f(\vec{u} + s\vec{v} + t\vec{w}) &= \mathbf{A}(\vec{u} + s\vec{v} + t\vec{w}) + \vec{b} \\ &= (\mathbf{A}\vec{u} + \vec{b}) + s(\mathbf{A}\vec{v}) + t(\mathbf{A}\vec{w}) \\ &= \vec{u}_1 + s\vec{v}_1 + t\vec{w}_1, \end{aligned}$$

where $s, t \in [0, 1], s + t \leq 1$. By (3), $\vec{v}_1 = \mathbf{A}\vec{v}$ and $\vec{w}_1 = \mathbf{A}\vec{w}$ are not parallel. Thus, T is mapped to a triangle T_1, which completes the proof of the base case.

Now suppose that f maps each n-gon to an n-gon for all n, $3 \leq n \leq k$, and let $\mathcal{P}$ be a polygon with $k + 1$ sides. In the solution to Problem 3.2.30, we saw that every polygon with at least 4 sides has a diagonal contained completely in its interior. Let $\overline{AB}$ be such a diagonal in $\mathcal{P}$. This diagonal divides $\mathcal{P}$ into two polygons, $\mathcal{P}_1$ and $\mathcal{P}_2$, containing t and $k + 3 - t$ sides, respectively, for some t, $3 \leq t \leq k$. By the inductive hypothesis, $f(\mathcal{P}_1)$ and $f(\mathcal{P}_2)$ will be t-sided and $(k + 3 - t)$-sided polygons, respectively. Since each of these polygons will have the segment from $f(A)$ to $f(B)$ as a diagonal, the union of $\mathcal{P}_1$ and $\mathcal{P}_2$ will form a polygon with $k + 1$ sides, which concludes the proof.

(5) The proof that a parallelogram is mapped to a parallelogram is analogous to the proof that triangles get mapped to triangles in (4), by simply dropping the condition that $s + t \leq 1$.

(6) Consider parallel line segments, S_1 and S_2, given in vector form as $S_i : \vec{p}_i + t\vec{u}_i,\ t \in [0, 1]$. Because they are parallel, $\vec{u}_2 = k\vec{u}_1$ for some $k \in \mathbb{R}$. As $|\vec{u}_i|$ is the length of S_i, the ratio of lengths of S_2 and S_1 is $|k|$. From parts (1) and (2), S_i is mapped into a segment of length $|\mathbf{A}\vec{u}_i|$. Since $\mathbf{A}\vec{u}_2 = \mathbf{A}(k\vec{u}_1) = k(\mathbf{A}\vec{u}_1)$, $|\mathbf{A}\vec{u}_2| = |k||\mathbf{A}\vec{u}_1|$, which shows that the ratio of lengths of $f(S_2)$ and $f(S_1)$ is also $|k|$.

(7) We postpone discussion of the proof of this property until the end of this section.

□

Theorems 12.7 and 12.8 (to be proven below) are the vehicles by which we will be able to accomplish the goals promised at the beginning of the chapter – proving a geometric fact in complete generality simply by proving that it is true for a specific case.

Theorem 12.8. *(Fundamental Theorem of Affine Transformations) Given two ordered sets of three non-collinear points each, there exists a unique affine transformation f mapping one set onto the other.*

Proof. We first show that the special (ordered) triple of vectors,

$$\left\{\vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \vec{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\},$$

can be mapped by an appropriate affine transformation to an arbitrary (ordered) triple of vectors,

$$\left\{\vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \ \vec{q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, \ \vec{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}\right\},$$

which corresponds to three non-collinear points. Let

$$\mathbf{A} = \begin{bmatrix} q_1 - p_1 & r_1 - p_1 \\ q_2 - p_2 & r_2 - p_2 \end{bmatrix} \quad \text{and} \quad \vec{b} = \vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}.$$

One can immediately verify that

$$\mathbf{A}\vec{0} + \vec{b} = \vec{p}, \quad \mathbf{A}\vec{i} + \vec{b} = \vec{q}, \quad \text{and} \quad \mathbf{A}\vec{j} + \vec{b} = \vec{r}.$$

Note that the columns of $\mathbf{A}$ correspond to the vectors $\vec{q} - \vec{p}$ and $\vec{r} - \vec{p}$. Since the points (p_1, p_2), (q_1, q_2), and (r_1, r_2) are non-collinear, the vectors $\vec{q} - \vec{p}$ and $\vec{r} - \vec{p}$ are non-parallel vectors. Hence, by Theorem 12.6, the determinant of $\mathbf{A}$ is nonzero. Thus, by Theorem 12.4, $\mathbf{A}$ is invertible, and $f(\vec{x}) = \mathbf{A}\vec{x} + \vec{b}$ is an affine transformation by definition.

Let $\{\vec{p}, \vec{q}, \vec{r}\}$ and $\{\vec{p'}, \vec{q'}, \vec{r'}\}$ be two ordered triples of position vectors representing two arbitrary triples of non-collinear points. Using the result we have just proven, there exist affine transformations f and g mapping the special triple $\{\vec{0}, \vec{i}, \vec{j}\}$ to $\{\vec{p}, \vec{q}, \vec{r}\}$ and to $\{\vec{p'}, \vec{q'}, \vec{r'}\}$, respectively. Then $g \circ f^{-1}$ is an affine transformation that maps $\{\vec{p}, \vec{q}, \vec{r}\}$ to $\{\vec{p'}, \vec{q'}, \vec{r'}\}$. The uniqueness of this transformation is left to Problem 12.1. □

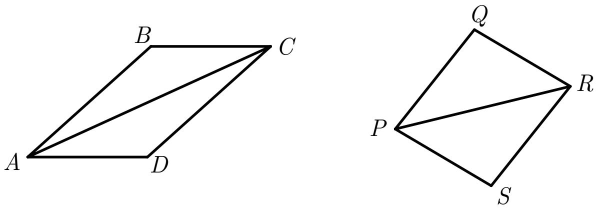


FIGURE 12.5.

Corollary 12.9.

(1) *Given any two triangles, there exists an affine transformation mapping one to the other.*
(2) *Given any two parallelograms, there exists an affine transformation mapping one to the other.*

Proof.

(1) By Theorem 12.8, the three vertices of one triangle can be mapped to the three vertices of any other triangle. Then use Theorem 12.7.
(2) Consider parallelograms $ABCD$ and $PQRS$, with diagonals $\overline{AC}$ and $\overline{PR}$, as shown in Figure 12.5.

By (1), there is an affine transformation, f, mapping $\triangle ABC$ to $\triangle PQR$, with $f(A) = P$, $f(B) = Q$, and $f(C) = R$. Furthermore, by Theorem 12.7(3), the images of lines AD and CD, namely $\overleftrightarrow{PS}$ and $\overleftrightarrow{RS}$, must be parallel to lines QR and QP, respectively. So, $f(D) = S$. □

Since, by Corollary 12.9, any triangle can be mapped to any other triangle, we say that all triangles are **affine equivalent**; likewise for all parallelograms. We conclude that, in particular, any triangle can be mapped by an affine transformation to an equilateral triangle or to a $45^\circ - 90^\circ - 45^\circ$ triangle, and every parallelogram can be mapped to a square.[10]

We now are prepared to discuss the general idea of a proof of property (7) of Theorem 12.7. First, impose upon the plane a grid of congruent squares. (See Figure 12.6(i).) The first four properties of Theorem 12.7 imply that an affine transformation f will map this grid of squares into a grid of parallelograms, and property (6) implies that these parallelograms are all congruent to each other. (See Figure 12.6(ii).)

Let φ_1 and φ_2 be two figures in the plane, with images $f(\varphi_1)$ and $f(\varphi_2)$, respectively, under the map. If the grid of squares is sufficiently fine, then the ratio of the number of squares in the interior of φ_1 to the number of squares in the interior of φ_2 will differ by arbitrarily little from the ratio Area(φ_1)/Area(φ_2). (Indeed, Area(φ_1)/Area(φ_2) is often defined as the limit of the ratio of the number of squares in φ_1 to the number of squares in φ_2 as the side of the square in the grid decreases indefinitely.[11]) Similarly, the ratio of the number of parallelograms in the interior of $f(\varphi_1)$

[10] Affine equivalent figures differ in shape, but not too much. This probably prompted Euler to introduce the term *'affinatas'* to identify transformations of the type $x' = x/m$, $y' = y/n$ in his *Introductio in analysin infinitorum* in 1748. The meanings of the word "affinity" include: a resemblance, or an inherent similarity between things.

[11] A proof of the existence of this limit requires rigorous calculus concepts, which are not assumed for this book.

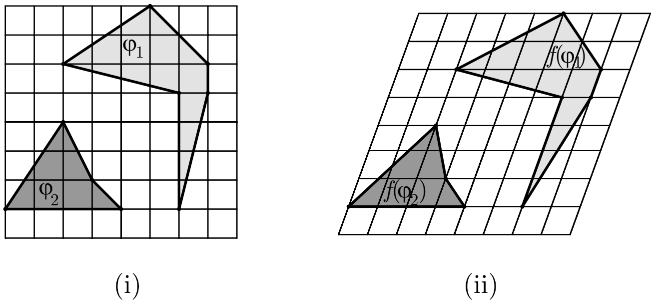


FIGURE 12.6.

to the number of parallelograms in the interior of $f(\varphi_2)$ will differ by arbitrarily little from the ratio Area $f(\varphi_1)$/Area $f(\varphi_2)$.

An equivalent way of stating property (7) of Theorem 12.7 is this: for every affine transformation f, there exists a positive real number k such that the area of every figure is altered by a factor of k, i.e., Area($f(\varphi)$) $= k\cdot$ Area(φ). In order to find k, we may concentrate on the change of area of the unit square defined by vectors $\vec{i}$ and $\vec{j}$. As previously noted, if $\mathbf{A}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the 2×2 matrix corresponding to an affine transformation f, the first column of $\mathbf{A}$ is $\vec{v} = f(\vec{i})$ and the second column is $\vec{w} = f(\vec{j})$. Under f, the unit square with sides given by $\vec{i}$ and $\vec{j}$ is mapped to a parallelogram with sides defined by $\vec{v} = \langle a, c\rangle$ and $\vec{w} = \langle b, d\rangle$. The area of the parallelogram can be found by subtracting the areas of two pairs of congruent triangles from the area of a rectangle. This is pictured in Figure 12.7 for the case when $a > b > 0$, and $d > c > 0$.

Therefore, the area of the parallelogram is

$$(a+b)(c+d) - 2\left(\frac{1}{2}(a+b)c\right) - 2\left(\frac{1}{2}b(c+d)\right) = ad - bc = \det\mathbf{A}.$$

By similar arguments one can show that essentially the same result holds if we remove the conditions imposed on a, b, c, and d. More precisely, the unit square defined by $\vec{i}$ and $\vec{j}$ is always mapped to a parallelogram having area equal to $|\det(\mathbf{A})|$. From this we conclude that the area of any figure is altered by a factor equalling the absolute value of the determinant of $\mathbf{A}$ under the transformation f.

Restating some parts of Theorem 12.7 in terms of invariants, we can say that certain properties of a figure, such as being a line, a segment, or a triangle, are invariant under affine transformations, as are ratios of lengths of parallel segments and ratios of areas of figures. The list can be continued. For example, the property of a segment being a median in a triangle, the property of a set of lines being concurrent, the property of a point being the centroid of a triangle, and the property of a quadrilateral being a trapezoid are all invariant under affine transformations.

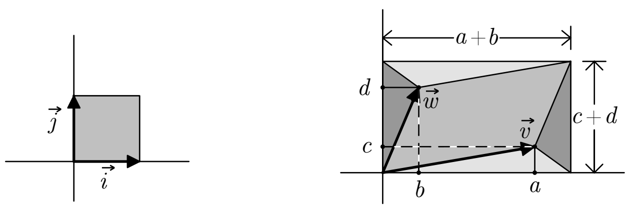


FIGURE 12.7.

On the other hand, there are many properties that are *not* invariant under affine transformations: the ratio of lengths of non-parallel segments, the property of lines being perpendicular, the property of a triangle being isosceles, the property of a quadrilateral being a rhombus, the property of a ray being the bisector of an angle, the property of a figure being a circle, the property of a point being the center of the in-circle of a triangle, etc.

12.4 Applications

A mathematician is a person who can find analogies between theorems; a better mathematician is one who can see analogies between proofs and the best mathematician can notice analogies between theories. One can imagine that the ultimate mathematician is one who can see analogies between analogies.

— Stefan Banach (1892–1945)

We begin with a theorem that we have seen before, but with a new proof that illustrates well the ideas of this chapter.

Theorem 12.10. *The three medians of a triangle are concurrent.*

Proof. Given a triangle ABC, by Corollary 12.9 there is an affine transformation, f, mapping $\triangle ABC$ to an equilateral triangle, $\triangle DEF$. By Theorem 12.7(2), f maps each side of $\triangle ABC$ to a side of $\triangle DEF$; we may assume that $\overline{AB}$ maps to $\overline{DE}$. Let C' be the midpoint of $\overline{AB}$, so that $AC' : C'B = 1 : 1$. By property (6) of Theorem 12.7, $f(C') = F'$ is the midpoint of $\overline{DE}$. Consequently, f maps the medians of $\triangle ABC$ to the medians of $\triangle DEF$.

Proving that the medians of $\triangle DEF$ are concurrent is easier than the general case, due to the many "symmetries" of an equilateral triangle. For example, in an equilateral triangle, the medians are also the perpendicular bisectors and the angle bisectors. These properties can be used to show that the three segments are concurrent, which will prove that the property holds for $\triangle ABC$ as well, and thus for all triangles. We leave the details to the reader. □

Note that we can also conclude that the point of concurrency of the medians (the centroid) divides each median in a ratio 2:1, starting from the vertex of the triangle. Triangles DGF' and FGD', as shown in Figure 12.8, are congruent $30° - 60° - 90°$ triangles. By properties of $30° - 60° - 90°$ triangles, $F'G : GD = 1 : 2$. By equating the lengths of congruent sides of the two triangles, $GD = GF$, so $F'G : GF = 1 : 2$. Because ratios of parallel segments are preserved under affine transformations, this ratio must also hold in an arbitrary triangle.

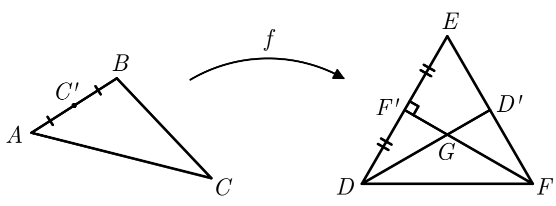


FIGURE 12.8.

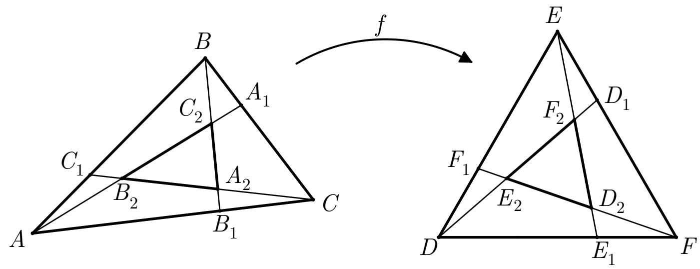


FIGURE **12.9.**

> **Theorem 12.11.** *Let f be an affine transformation and let $\mathcal{P}$ be a polygon. Then f maps the centroid of $\mathcal{P}$ to the centroid of $f(\mathcal{P})$.*

Proof. The discussion prior to the statement of the theorem establishes the result in the case where P is a triangle. The proof for the general case is left to Problem 12.5. □

Our proof of Theorem 12.10 used the method of affine transformations to re-prove a fact we have previously established. We know from earlier chapters that the three angle bisectors of a triangle and the three altitudes of a triangle are also concurrent. However, the method employed above does not work to prove the concurrence of these latter trios; when a triangle is mapped via an affine transformation onto an equilateral triangle, the property of a segment being an angle bisector or an altitude is not necessarily preserved. The mapping of medians to medians is a consequence of the invariance of ratios of parallel line segments, a property that is not relevant to angle bisectors or altitudes.

Example 76. Let A_1, B_1, and C_1 be points on the sides $\overline{BC}$, $\overline{CA}$, and $\overline{AB}$, respectively, of $\triangle ABC$, such that

$$\frac{BA_1}{A_1C} = \frac{CB_1}{B_1A} = \frac{AC_1}{C_1B} = \frac{1}{2}.$$

Let A_2, B_2, and C_2 be the points of intersections of the segments BB_1 and CC_1, CC_1 and AA_1, and AA_1 and BB_1, respectively. (See Figure 76). Prove that

$$\frac{\text{Area}(\triangle A_2B_2C_2)}{\text{Area}(\triangle ABC)} = \frac{1}{7}.$$

Solution: As in the previous example, we use an affine transformation, f, that maps $\triangle ABC$ to an equilateral triangle, $\triangle DEF$. The points $D_1 = f(A_1)$, $E_1 = f(B_1)$, and $F_1 = f(C_1)$ will divide the sides of $\triangle DEF$ in the same 1:2 ratio. Therefore, $\overline{DF_1}$, $\overline{ED_1}$, and $\overline{FE_1}$ will all have the same length. Let us assume that this length is 1.

Let D_2, E_2, and F_2 be the points of intersections of the segments EE_1 and FF_1, FF_1 and DD_1, and DD_1 and EE_1, respectively. Rotating $\triangle DEF$ clockwise by 120° around its center, we see that $D_1 \mapsto E_1 \mapsto F_1 \mapsto D_1$. This implies that $\overline{DD_1} \mapsto \overline{EE_1} \mapsto \overline{FF_1} \mapsto \overline{DD_1}$, and therefore $D_2 \mapsto E_2 \mapsto F_2 \mapsto D_2$. This proves that $\triangle D_2E_2F_2$ is equilateral.

Using the Cosine theorem for $\triangle DF_1F$, we get

$$FF_1 = \sqrt{1^2 + 3^2 - 2 \cdot 1 \cdot 3 \cdot \cos(\pi/3)} = \sqrt{7}.$$

Now, $\triangle DE_2F_1 \sim \triangle DED_1$, since they have two pairs of congruent angles. Thus,

$$\frac{E_2F_1}{F_1D} = \frac{ED_1}{D_1D} \implies \frac{E_2F_1}{1} = \frac{1}{\sqrt{7}} \text{ and } \frac{DE_2}{DF_1} = \frac{DE}{DD_1} \implies \frac{DE_2}{1} = \frac{3}{\sqrt{7}}.$$

Noting that $FD_2 = DE_2$, we see that $D_2E_2 = \sqrt{7} - 1/\sqrt{7} - 3/\sqrt{7} = 3/\sqrt{7}$. This implies that $D_2E_2/DE = 1/\sqrt{7}$, and therefore

$$\frac{\text{Area}(\triangle D_2E_2F_2)}{\text{Area}(\triangle DEF)} = \left(\frac{D_2E_2}{DE}\right)^2 = \left(\frac{1}{\sqrt{7}}\right)^2 = \frac{1}{7}.$$

Since the ratio of areas is invariant under affine transformations, the result follows. □

The reader may recall that Example 76 was previously presented as Problem 5.22. A comparison of the solutions should reveal that the above solution is more straightforward than the one presented previously. In Problem 12.11, we consider a generalization of this example.

Example 77. Is there a non-regular pentagon with the property that each diagonal is parallel to one of its sides?

Solution: First, we note that it is easy to show that a regular pentagon, $\mathcal{P}_5$, has this property. (In Figure 12.10, $ABCDE$ is such a pentagon.) We leave this task to the reader.

Theorem 12.7(4) establishes that any affine transformation will map a pentagon to a pentagon. We wish to find an affine transformation f such that the image of $\mathcal{P}_5$ under f is not a regular pentagon. There are many such affine transformations. Consider, for example, an affine transformation under which three consecutive vertices of $\mathcal{P}_5$ are mapped to the vertices of an equilateral triangle. Then, the image of $\mathcal{P}_5$ under f is not regular, since one of the angles of $f(\mathcal{P}_5)$ has measure 60°. (In Figure 12.10, the regular pentagon $ABCDE$ is mapped to the pentagon $A'B'C'D'E'$ in which $\angle A'B'C'$ has measure 60°.) However, by Theorem 12.7(3), the property of parallelism among line segments is preserved under an affine transformation, so the image of $\mathcal{P}_5$ will be a non-regular pentagon having the desired property. □

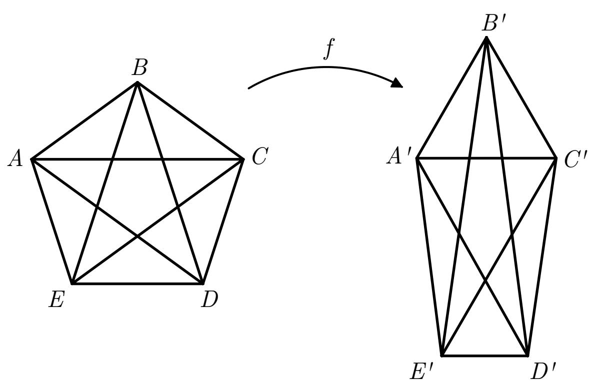


FIGURE 12.10.

12.5 Affine Transformations of Conic Sections

We have established that the image of an n-gon under an affine transformation is an n-gon, and the image of a parallelogram is a parallelogram. We next consider the effect of an affine transformation on the conic sections – ellipses, hyperbolas, and parabolas. Recall from Chapter 9 that any conic section can be represented by a second-degree equation having the general form

$$Ax^2 + Bxy + Cy^2 + Fx + Gy + H = 0,$$

where A, B, C, F, G, and H are real numbers. By Theorem 9.5, the equation represents an ellipse if $B^2 - 4AC < 0$, a parabola if $B^2 - 4AC = 0$, and a hyperbola if $B^2 - 4AC > 0$.

Theorem 12.12. *Let $f(\vec{x}) = \mathbf{A}\vec{x} + \vec{b}$, where $\mathbf{A}$ is an invertible 2×2 matrix, be an affine transformation. Then f maps an ellipse to an ellipse, a parabola to a parabola, and a hyperbola to a hyperbola.*

Proof. Suppose that the equation $Ax^2 + Bxy + Cy^2 + Fx + Gy + H = 0$ represents a non-degenerate conic, $\mathcal{F}$. If (x, y) is any point satisfying the equation, then the vector corresponding to this point, $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, is mapped to $f(\vec{x}) = \vec{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{A}\vec{x} + \vec{b}$. The inverse transformation, f^{-1}, is $\vec{x} \mapsto \mathbf{A}^{-1}\vec{x}' - \mathbf{A}^{-1}\vec{b}$. Therefore,

$$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} t \\ u \end{bmatrix},$$

for real numbers a, b, c, d, t, and u. With these values, $x = ax' + by' + t$ and $y = cx' + dy' + u$. Substituting these expressions into the equation $Ax^2 + Bxy + Cy^2 + Fx + Gy + H = 0$ that represents $\mathcal{F}$ results in a second-degree equation in x' and y'. Thus, $\mathcal{F}$ is mapped to another conic, $\mathcal{F}'$.

Note that $\mathcal{F}'$ cannot be a degenerate conic. A degenerate conic can only be a pair of lines, a single line, a point, or the empty set. By Theorem 12.7, if $\mathcal{F}'$ were a degenerate conic, the image of $\mathcal{F}'$ under f^{-1} would again be a pair of lines, a single line, a point, or the empty set. This contradicts our assumption that $\mathcal{F}$ is non-degenerate.

Replacing x and y in the equation $Ax^2 + Bxy + Cy^2 + Fx + Gy + H = 0$ with $x = ax' + by' + t$ and $y = cx' + dy' + u$ yields a second-degree equation corresponding to $\mathcal{F}'$:

$$\begin{aligned} &A(ax' + by' + t)^2 + B(ax' + by' + t)(cx' + dy' + u) + C(cx' + dy' + u)^2 \\ &\quad + F(ax' + by' + t) + G(cx' + dy' + u) + H = 0. \end{aligned}$$

When reduced, the discriminant of this equation is found to be

$$(ad - bc)^2(B^2 - 4AC),$$

where $B^2 - 4AC$ is the discriminant of the original conic, $\mathcal{F}$. As we've noted, the sign of the discriminant characterizes a non-degenerate conic. Since $(ad - bc)^2 > 0$, the sign of the discriminant is unchanged under affine transformation, and thus, the type of the conic is also unchanged. □

With Theorem 12.12, we have established that an affine transformation will send a conic to a conic of the same type. As with triangles and parallelograms, it turns out that we can actually do better than that: any ellipse can be mapped to any other ellipse under an affine transformation, and likewise for parabolas and hyperbolas.

Suppose that $\mathcal{E}$ is an ellipse with center at (h, k) and with major and minor axes of lengths $2a$ and $2b$. As discussed in the proof of Theorem 9.5, an ellipse is mapped to an ellipse under a translation or rotation. Under translation by $\vec{b} = \begin{bmatrix} -h \\ -k \end{bmatrix}$, the ellipse is mapped to a congruent ellipse with center at the origin. A rotation can be applied to the plane in order to align the major and minor axes of the ellipse with the x- and y-axes, respectively. The original ellipse, $\mathcal{E}$, has now been mapped to another ellipse, $\mathcal{E}'$, via the two specified affine transformations; $\mathcal{E}'$ can be represented by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Now apply a third affine transformation, $f(\vec{x}) = \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x/a \\ y/b \end{bmatrix}$.

Under this transformation, $\mathcal{E}'$ is mapped to an ellipse represented by the equation $x^2 + y^2 = 1$; that is, $\mathcal{E}'$ is mapped to the unit circle, $\mathcal{C}(O, 1)$. This proves the following theorem.

Theorem 12.13. *Given any ellipse, $\mathcal{E}$, there exists an affine transformation mapping $\mathcal{E}$ to the unit circle.*

From Theorem 12.13 follows Corollary 12.14, which establishes that all ellipses are affine equivalent.

Corollary 12.14. *Given any two ellipses, $\mathcal{E}_1$ and $\mathcal{E}_2$, there exists an affine transformation mapping $\mathcal{E}_1$ to $\mathcal{E}_2$.*

Proof. Consider ellipses $\mathcal{E}_1$ and $\mathcal{E}_2$. By Theorem 12.13, there are affine transformations f and g mapping $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively, to $\mathcal{C}(O, 1)$. By the definition of inverse mappings, g^{-1} is an affine transformation mapping $\mathcal{C}(O, 1)$ to $\mathcal{E}_2$. By the definition of composition of mappings, $g^{-1} \circ f$ is an affine transformation mapping $\mathcal{E}_1$ to $\mathcal{E}_2$. □

Similar techniques can be applied to show that all hyperbolas are affine equivalent and that all parabolas are affine equivalent. See Problems S12.6 and 12.6.

Example 78. Given an ellipse, $\mathcal{E}$, consider a set of parallel chords of $\mathcal{E}$. Prove that the midpoints of these chords form a diameter of the ellipse and the tangent lines to $\mathcal{E}$ at the endpoints of the diameter are parallel to the chords.

Solution: Let $\mathcal{E}$ be an ellipse with a set of parallel chords, $c_1, c_2, \ldots, c_n$, as shown in Figure 12.11. By Theorem 12.13, there is an affine transformation mapping $\mathcal{E}$ to the unit circle, $\mathcal{C}$. Under this mapping, the chords of $\mathcal{E}$ are mapped to a set of parallel chords of $\mathcal{C}$. Furthermore, the midpoints of the chords of $\mathcal{E}$ are mapped to the midpoints of chords of $\mathcal{C}$.

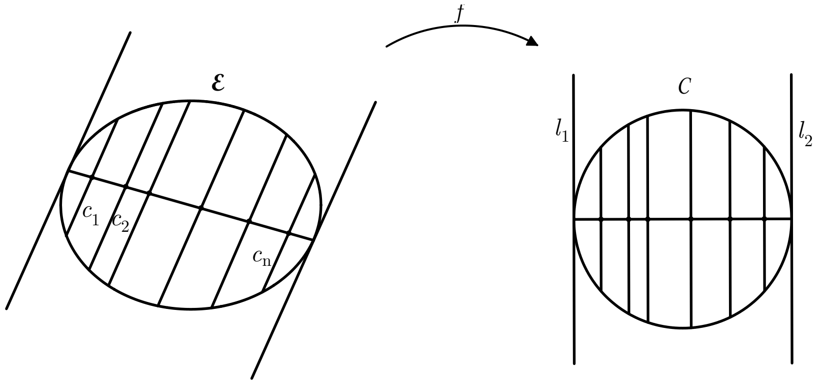


FIGURE 12.11.

By Theorem 4.6(1), the midpoint of a chord of a circle lies on a diameter perpendicular to the chord. Corollary 4.8 implies that the tangent lines, l_1 and l_2, to $\mathcal{C}$ at the endpoints of the diameter are perpendicular to it, and hence parallel to the set of chords of $\mathcal{C}$. This proves the theorem for $\mathcal{C}$. By Theorem 12.7, the properties of a point bisecting a segment, segments being parallel, collinearity, and tangency are all invariant under an affine transformation, so the statement holds for $\mathcal{E}$ as well. □

We invite the reader to compare this solution to that of Example 46 and Problem 9.12. In the case of the ellipse, which solution do you like more?

12.6 Problems

It's not that I'm so smart, it's just that I stay with problems longer.
— Albert Einstein (1879–1955)

12.1 Prove the uniqueness of the map in Theorem 12.8.

12.2 Given two trapezoids, is there always an affine transformation mapping one to the other?

12.3 Prove that the line joining the point of intersection of the extensions of the nonparallel sides of a trapezoid to the point of intersection of its diagonals bisects each base of the trapezoid.

12.4 Prove that all chords of an ellipse that cut off a region of constant area are tangent to a concentric similar (and similarly oriented) ellipse.

12.5 Complete the details of the proof of Theorem 12.11. (We note that the **centroid** of a polygon is the centroid of the set of its vertices.)

12.6 Given any parabola, $\mathcal{P}$, prove that there exists an affine transformation mapping $\mathcal{P}$ to the parabola given by the equation $y = x^2$.

12.7 Let A_1, B_1, and C_1 be points on the sides $\overline{BC}$, $\overline{CA}$, and $\overline{AB}$, respectively, of $\triangle ABC$, having the property that $BA_1/A_1C = CB_1/B_1A = AC_1/C_1B$. Prove that the centroids of $\triangle ABC$, $\triangle A_1B_1C_1$, and the triangle formed by lines AA_1, BB_1, and CC_1 coincide.

12.8 Let l be a line passing through the vertex M of parallelogram $MNPQ$ and intersecting lines NP, PQ, and NQ in points R, S, and T, respectively. Prove that $1/MR + 1/MS = 1/MT$.

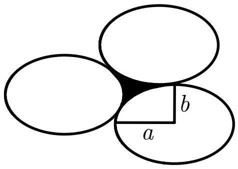


FIGURE 12.12.

12.9 Prove that an ellipse with semi-axes of lengths a and b has area πab.

12.10 Suppose that an ellipse touches the sides AB, BC, CD, and DA of a parallelogram $ABCD$ at the points P, Q, R, and S, respectively. Prove that the lengths CQ, QB, BP, and CR satisfy $\frac{CQ}{QB} = \frac{CR}{BP}$. (This problem and its solution are from [**7**].)

12.11 Let A_1, B_1, and C_1 be points on the sides $\overline{BC}$, $\overline{CA}$, and $\overline{AB}$, respectively, of $\triangle ABC$, having the property that $BA_1/A_1C = \alpha$, $CB_1/B_1A = \beta$, and $AC_1/C_1B = \gamma$. Let $\triangle DEF$ be the triangle bounded by $\overline{AA_1}$, $\overline{BB_1}$, and $\overline{CC_1}$. Find $\dfrac{\text{Area}(\triangle DEF)}{\text{Area}(\triangle ABC)}$.

12.12 Consider three ellipses that are congruent, similarly oriented (that is, all major axes are parallel), and which touch externally in pairs. (See Figure 12.12.)

Prove that the area of the curvilinear triangle bounded by them (the shaded area in Figure 12.12) is independent of their position. Then, find the area of the curvilinear triangle if the length of each major axis is a and the length of each minor axis is b.

12.13 Prove that a necessary and sufficient condition for a triangle inscribed in an ellipse to have maximum area is that the centroid of the triangle coincides with the center of the ellipse. Generalize the problem for an inscribed n-gon with $n \geq 3$.

12.7 Supplemental Problems

$S12.1$ Suppose f is an affine transformation of $\mathbb{E}^2$ such that $f((2,3)) = (3,-1)$, $f((2,1)) = (1,2)$ and $f((1,0)) = (0,1)$.

(a) Find $f((-2,5))$.
(b) Let Φ be a figure having area 5 square units. What is the area of $f(\Phi)$?

$S12.2$ Each vertex of a triangle is joined to two points of the opposite side that divide the side into three congruent segments. Consider the hexagon formed by these six segments. Prove that the three diagonals joining opposite vertices of the hexagon are concurrent.

$S12.3$ Let $ABCD$ be a trapezoid with $\overline{BC} \,||\, \overline{AD}$. Let the line through B parallel to the side CD intersect the diagonal AC at point P, and the line through C parallel to the side AB intersect the diagonal BD at point Q. Prove that $\overline{PQ}$ is parallel to the bases of the trapezoid.

$S12.4$ Is it always possible to use an affine transformation of a plane to map an altitude of a triangle to a bisector of the image of the triangle (not necessarily at the corresponding vertex)?

$S12.5$ Let $\mathcal{E}$ be an ellipse with center C. If f is any affine transformation, prove that $f(C)$ is the center of the ellipse $f(\mathcal{E})$.

$S12.6$ Given any hyperbola, $\mathcal{H}$, prove that there exists an affine transformation mapping $\mathcal{H}$ to the hyperbola given by $xy = 1$.

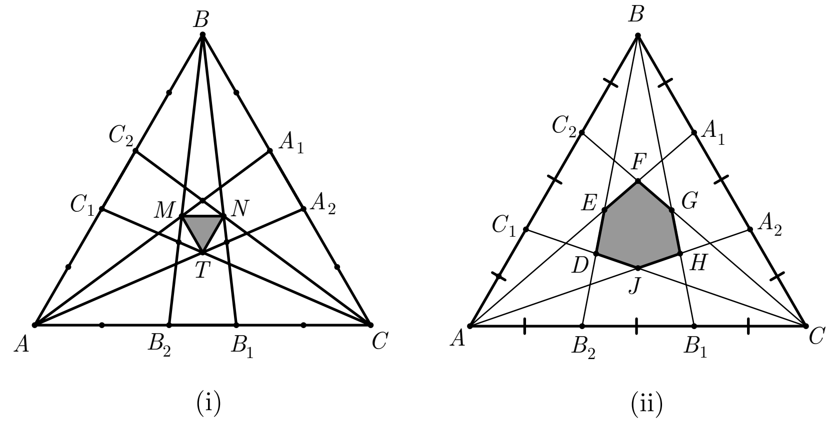


FIGURE 12.13.

*S*12.7 Let A_1, B_1, C_1, and D_1 be points on the sides CD, DA, AB, and BC, respectively, of a parallelogram $ABCD$ such that

$$\frac{CA_1}{CD} = \frac{DB_1}{DA} = \frac{AC_1}{AB} = \frac{BD_1}{BC} = \frac{1}{3}.$$

Show that the area of the quadrilateral formed by lines AA_1, BB_1, CC_1, and DD_1 is one thirteenth of the area of $ABCD$.

*S*12.8 Let n be a positive integer and consider an equilateral triangle ABC with unit side lengths. Let $\overline{A_1A_2}$, $\overline{B_1B_2}$, and $\overline{C_1C_2}$ be segments of length $1/(2n+1)$ lying on and centered at the midpoints of sides BC, AC, and AB, as shown in Figure 12.13(i).

(a) Let M be the intersection of segments BB_2 and AA_1, let N be the intersection of segments BB_1 and CC_2, and let T be the intersection of segments AA_2 and CC_1. Find MN and use it to find the area of $\triangle MNT$.

(b) Let P be the intersection of $\overline{AA_1}$ and $\overline{CC_2}$. Find the area of $\triangle MPN$.

(c) Suppose that each vertex of $\triangle ABC$ is joined to two points of the opposite side that divide the side into three congruent segments. Find the area of the hexagon formed by these six segments. (In Figure 12.13(ii), the hexagon is $DEFGHJ$.)

(d) Part (c) can be generalized.[12] For a positive odd integer m, divide each side of a triangle into m congruent segments and connect the endpoints of the middle segment on each side to the vertex opposite that side. These six segments bound a hexagonal region in the interior of the triangle. Determine the area of this hexagon as a fraction of the area of the original triangle. See Figure 12.14 for an illustration in the case where $m = 5$.

*S*12.9 How many ellipses can pass through four given points with no three of them being collinear? What if instead of the ellipses we consider parabolas?

*S*12.10 Given three non-collinear points in the plane. Find the locus of all points of all parabolas passing through them.

---

*S*12.11 Prove that any five points in a plane such that no three of them are collinear must lie on a unique conic that is either an ellipse, a hyperbola, or a parabola.

[12] The result given in part (c) is known as Marion's Theorem. The generalization given in part (d) was found by Ryan Morgan in 1994, when he was a tenth grader.

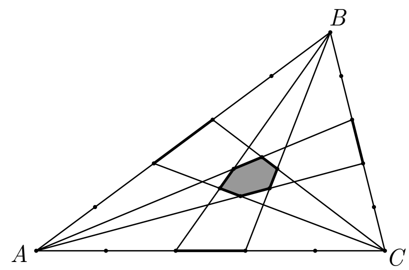


FIGURE 12.14.

$S12.12$ Suppose an affine transformation maps a circle to itself. Prove that the transformation is either a rotation or a symmetry with respect to a line.

$S12.13$ Prove that a necessary and sufficient condition for a triangle circumscribed around an ellipse to have minimum area is that the centroid of the triangle coincides with the center of the ellipse. Can you generalize the problem for an inscribed n-gon with $n \geq 3$?