Edward J. Specht (1915–2011)
Harold T. Jones (1925–1995)
Keith G. Calkins
Donald H. Rhoads

Supplement to Euclidean Geometry and its Subgeometries

February 2016

Preface

This Supplement is a companion to the book by Edward J. Specht, Harold T. Jones, Keith G. Calkins, and Donald H. Rhoads entitled *Euclidean Geometry and its Subgeometries* published in 2015 by Birkhäuser. We shall refer to this work as *Specht*.

Chapter 1 of this Supplement is an expansion of *Specht* Ch.18 (Section 18.2) in *Specht*. It duplicates some of the material found there, as well as parts of *Specht* Ch.1, Section 1.5. Chapter 2 defines and develops complex numbers on the coordinate plane.

Chapter 3 develops the notion of the length of an arc; Chapter 4 uses arc length to define the circular functions sin and cos in a treatment originated by our first author, Edward Specht. Chapter 5 builds on the previous two chapters to define angle measure.

Chapter 6 is a leisurely exploration of properties of polygons on a Pasch plane, eventuating in a proof of the Jordan Curve Theorem for the polygonal case. It does not try to achieve this proof in the most economical way.

The final chapters are essentially fragments which were left over from the main development, but which might have some interest for their own sakes.

Chapter 7 is a proof of "Property PE," which says that given a line $\mathcal{L}$ on a Pasch plane and a point Q not on $\mathcal{L}$, there exists a line $\mathcal{M}$ containing Q that is parallel to $\mathcal{L}$. This was proved in Chapter 8 of *Specht* as part of neutral geometry. Here we prove it for a Pasch plane (without assuming the existence of a reflection set) on which the line $\mathcal{L}$ has been ordered according to *Specht* Ch.6 and Axiom LUB holds.

Chapter 8 shows that in *Specht* Ch.8 Definition NEUT.2, property R.6 (existence of midpoints) is a consequence of properties R.1 through R.5, provided Axiom PW holds.

Citations and references. In this Supplement, we will often refer to theorems, definitions, and remarks, both from this Supplement, and from *Specht*. Our preferred (brief) style of reference will be simply by label, acronym, and number, as, for instance, "Theorem ISM.5" or "Definition VEC.12." When first referencing items of a particular acronym from *Specht*, we will include additional labeling, as, for instance, "*Specht* Ch.12 Theorem ISM.5." Repeated uses of the same acronym will usually revert to the shorter style.

Condensed Table of Contents for Specht, *Euclidean Geometry and its Subgeometries*

References to acronyms other than those listed above will be to items in this Supplement.

Contents

List of Figures

Chapter 1
The Plane as a Vector Space (VEC)

Dependencies: *This chapter is dependent on* Euclidean Geometry and its Subgeometries, *by Specht, Jones, Calkins, and Rhoads, published by Birkhäuser, 2015*
Acronym: *VEC*
Terms defined: *addition, scalar product on the plane; coordinatization (right or left-handed), axes, origin, clockwise and counterclockwise; first and second coordinates on a plane; vector or linear space, vector space isomorphism, coordinatization map; linearly independent, span, basis, dimension; ordered triples, n-tuples; vertical, horizontal, slope, norm, inner (dot) product, orthogonal; linear mapping, sum and scalar product of linear mappings, matrix, determinant; affine mapping*

The first part of this chapter is an expansion of Chapter 18 of *Euclidean Geometry and its Subgeometries* by Specht, Jones, Calkins, and Rhoads, hereafter referred to as *Specht*. This duplicates some of the material found in *Specht* Chapter 18, as well as parts of Chapter 1, Section 1.5. Later in the present chapter we provide some results on linear and affine mappings which are relevant to the main development in *Specht*.

In Section 18.3 of *Specht* Ch.18 we assigned a real number to each point on an arbitrary line in a Euclidean/LUB plane. This process might be characterized as *coordinatizing* the line. In Section 18.4 of the same chapter we briefly outlined the process by which the Euclidean/LUB plane itself may be coordinatized, assigning to each point on it a pair (a, b) of numbers.[1] Here we develop this process in greater detail.

[1] It is possible to coordinatize Euclidean space, assigning to each point a triple (a, b, c) of real numbers, but we do not pursue this.

Here, references to items labeled VEC will be to the current chapter; all other references are to *Specht*. In particular, this chapter contains numerous references to Chapter 18 of that work, which uses acronyms REAL and RR.

We refer the reader to the note **Citations and references** at the end of the Preface to this Supplement and to the abbreviated Table of Contents (with acronyms) for *Specht*.

1.1 Operations on the plane

Throughout this chapter, $\mathbb{P}$ will denote a Euclidean/LUB plane as defined in *Specht* Ch.18 Definition REAL.2.

Definition VEC.1 (A) For each $A \in \mathbb{P} \setminus \{O\}$, define τ_A to be the translation of $\mathbb{P}$ such that $\tau_A(O) = A$. *Specht* Ch.12 Theorem ISM.5 says that such a translation exists and is unique.

(B) Define $\tau_O = \imath$, the identity.

(C) For any A and B in $\mathbb{P}$, define

$$A + B = (\tau_B \circ \tau_A)(O) = \tau_B(\tau_A(O)) = \tau_B(A).$$

The operation $+$ is called **addition** and $A + B$ is the **sum** of A and B.

Remark VEC.2 (A) The operation $+$ from Definition VEC.1 applied to points on a line $\mathbb{L}$ through O is identical to the operation $\oplus$ from *Specht* Ch.14 Definition OF.1(A) and (C).

(B) Since we have made the identification of a line on the plane with the real numbers, we abandon the use of the symbol $\oplus$ and henceforth will use simply $+$. However, in cases where we wish to emphasize that we are adding two points in a single line through O, we may revert temporarily back to $\oplus$. Also, if we should have occasion to multiply points on such a line we may continue to use $\odot$—at this point there is no definition of the product of arbitrary points on the plane.

Theorem VEC.3 *The Euclidean/LUB plane $\mathbb{P}$ is an Abelian group under the operation $+$.*

Proof. Let $\mathbb{T} = \{\alpha \mid \alpha \text{ is a translation of } \mathbb{P} \text{ or } \alpha = \imath\}$, then by Theorem ISM.8(A) $\mathbb{T}$ is an Abelian group under composition of mappings. Routine calculations based on Definition VEC.1 confirm that $\mathbb{P}$ is an Abelian group under the operation $+$; we leave these to the reader as Exercise VEC.1. $\square$

Theorem VEC.4 *If O, A, and B are noncollinear, then $A+B$ is the fourth corner of the parallelogram whose other corners are O, A, and B.*

Proof. This is an immediate consequence of Exercise ISM.2. $\square$

Remark VEC.5 (A) Note that τ_A is the translation that not only maps O to A but also maps B to $A+B$. Also, $\tau_{A-B}(B) = (A-B)+B = A$ so τ_{A-B} maps B to A.

(B) If A and B are any two points then $\tau_{-B}(B) = O$ and $\tau_{-B}(A) = A-B$. By *Specht* Ch.8 Theorem NEUT.15 (since τ_{-B} is an isometry) $\tau_{-B}(\overline{AB}) = \overline{(A-B)O}$ and hence $\overline{AB} \cong \overline{O(A-B)}$.

(C) The line $\mathcal{L} = \overleftrightarrow{OA}$ can be built into an ordered field using the machinery of *Specht* Chapter 14; by Theorem OF.10(A)(1) of that chapter, for each $A \in \mathbb{P}$, $-A = \mathcal{R}_O(A)$. Hence for any A, $\mathcal{R}_O(\overline{OA}) = \overline{O\mathcal{R}_O(A)} = \overline{O(-A)}$ and $\overline{OA} \cong \overline{O(-A)}$. (*cf* Theorem OF.10(A)(4).)

(D) Combining parts (B) and (C), we have $\overline{AB} \cong \overline{O(A-B)} \cong \overline{O(B-A)}$.

Definition VEC.6 For every point $A \in \mathbb{P}$, and every real number x, define xA as in Definition REAL.19 (and summarized in Theorem REAL.20), where the line $\overleftrightarrow{OA}$ has been built into an ordered field. xA is called the **scalar product** of x and A, and the number x is called a **scalar**.

Remark VEC.6.1 (A) By Theorem REAL.37 we know that for each real number x there exists a dilation δ_x with fixed point O such that for all $A \in \mathbb{P} \setminus \{O\}$, $xA = \delta_x(A)$.

(B) Definition VEC.6 depends explicitly on the fact that the line $\overleftrightarrow{OA}$ has been built into a field. The multiplicative field properties of this line are essential to the development of the properties of scalar product such as those stated in Theorems REAL.23 and Corollary REAL.35.1.

Theorem VEC.7

(A) *For every $A \in \mathbb{P} \setminus \{O\}$, $\overleftrightarrow{OA} = \{xA \in \mathbb{P} \mid x \in \mathbb{R}\}$. That is, every line through the origin is the set of all scalar multiples of any non-O point in that line.*

Moreover, if A and B are any points in $\mathbb{P}$ and x and y are any real numbers,

(B) $x(yA) = (xy)A$, *(scalar multiplication is associative)*

(C) $x(A+B) = xA + xB$, *(scalar multiplication is distributive with respect to addition of points)*

(D) $(x+y)A = xA + yA$, *(scalar multiplication is distributive with respect to addition of scalars)*

(E) $1A = A$, *and*

(F) $xA = O$ *iff* $x = 0$ *or* $A = O$ *(or both).*

Proof. (A) $\overleftrightarrow{OA}$ is a line through the origin O and therefore is a fixed line for the dilation δ_x, whose existence is noted in Remark VEC.6.1, and hence if $x \neq 0$, $xA \in \overleftrightarrow{OA}$. By Theorem REAL.35(A), for every $A' \in \overleftrightarrow{OA} \setminus \{O\}$, there exists a real number $t \neq 0$ such that $tU_1 = A'$ and a real number $s \neq 0$ such that $sU_1 = A$, so that $A' = \frac{t}{s}A$.

(B) This is *Specht* Ch.18 Theorem REAL.23.

(C) This is Theorem REAL.32.

(D) This is Theorem REAL.31.

(E) $1A = A$ is immediate from *Specht* Ch.17 Definition QX.1(C).

(F) If $x = 0$ or $A = O$ (or both), $xA = O$ by Definition REAL.19(A)(1). If $xA = O$, then by Theorem REAL.25 $xA = x(U \odot A) = xU \odot A = O$ and by *Specht* Ch.14 Theorem OF.10(H), $xU = O$ or $A = O$. If $xU = O$ then $x = 0$ by Corollary REAL.34(B).

In parts (E) and (F) we used the $\odot$ symbol because we were referring back to the product operation used in OF and REAL on the line $\mathbb{L}$. $\square$

1.2 Vector spaces, $\mathbb{R}^2$ and isomorphisms

Theorem VEC.8 *Let $\mathbb{P}$ be a Euclidean/LUB plane, and let O be its origin. Let $\mathbb{L}_1$ and $\mathbb{L}_2$ be lines in $\mathbb{P}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$. Using the machinery of Chapters 14 and 18 of Specht, build each of the lines $\mathbb{L}_1$ and $\mathbb{L}_2$ into an ordered field which is isomorphic to $\mathbb{R}$, the set of all real numbers, and let U_1 and U_2, respectively, be their units, so that both U_1 and U_2 correspond to the real number 1 under their respective isomorphisms.*

(A) *For every $A \in \mathbb{P}$, there exist unique real numbers a and b such that $A = aU_1 + bU_2$.*

(B) *$aU_1 + bU_2 = O$ iff $a = b = O$.*

(C) *If $A \notin \mathbb{L}_1 \cup \mathbb{L}_2$, so that both $a \neq 0$ and $b \neq 0$, $\overline{O(aU_1)} \cong \overline{(bU_2)A}$ and $\overline{O(bU_2)} \cong \overline{(aU_1)A}$.*

Proof. If A is any point on $\mathbb{P}$, by Axiom PS there exists a unique line $\mathcal{M}_1$ containing the point A such that either $\mathcal{M}_1 = \mathbb{L}_1$ (in case $A \in \mathbb{L}_1$) or $\mathcal{M}_1 \parallel \mathbb{L}_1$; and there exists a unique line $\mathcal{M}_2$ such that either $\mathcal{M}_2 = \mathbb{L}_2$ (in case $A \in \mathbb{L}_2$) or $\mathcal{M}_2 \parallel \mathbb{L}_2$.

By Exercise I.1, $\mathcal{M}_1$ intersects $\mathbb{L}_2$ in exactly one point, which we shall call A_2, and $\mathcal{M}_2$ intersects $\mathbb{L}_1$ in exactly one point which we call A_1. By Theorem REAL.35, there exists a unique real number a such that $A_1 = aU_1$ and a unique real number b such that $A_2 = bU_2$. Since A uniquely determines $\mathcal{M}_1$ and $\mathcal{M}_2$, and these lines uniquely determine the points A_1 and A_2, which in turn uniquely determine a and b, a and b are uniquely determined by A.

Moreover, $A \in \mathbb{L}_1$ iff $A_2 = O$ iff $b = 0$, in which case
$$A = A_1 + O = A_1 + A_2 = aU_1 + bU_2;$$
$A \in \mathbb{L}_2$ iff $A_1 = O$ iff $a = 0$, in which case
$$A = O + A_2 = A_1 + A_2 = aU_1 + bU_2;$$
and $A = O$ iff $A \in \mathbb{L}_1 \cap \mathbb{L}_2$ iff $a = b = 0$, and again in this case
$$A = O + O = aU_1 + bU_2.$$

If $A \in \mathbb{P} \setminus (\mathbb{L}_1 \cup \mathbb{L}_2)$, by Theorem VEC.4, $aU_1 + bU_2$ is the fourth corner of the parallelogram of which O, aU_1, bU_2 are the other three corners. Since $\mathcal{M}_1$ contains the point A_1 and $\mathcal{M}_2$ contains the point A_2 and are parallel to (or equal to) $\mathbb{L}_1$ and $\mathbb{L}_2$, respectively, they are the same, respectively, as the sides $\overleftrightarrow{(aU_1)(aU_1 + bU_2)}$ and $\overleftrightarrow{(aU_2)(aU_1 + bU_2)}$ of this parallelogram. Since both $\mathcal{M}_1$ and $\mathcal{M}_1$ contain A, $A = aU_1 + bU_2$. This completes the proof of parts (A) and (B).

(C) The quadrilateral $\square O(aU_1)A(bU_2)$ is a parallelogram because $\mathcal{M}_1 \parallel \mathbb{L}_1$ and $\mathcal{M}_2 \parallel \mathbb{L}_2$. The result follows from *Specht* Ch.11 Theorem EUC.12(A). $\square$

Definition VEC.9 (A) In Theorem VEC.8, if $\mathbb{L}_1 \perp \mathbb{L}_2$, the two units U_1 and U_2, together with their lines $\mathbb{L}_1$ and $\mathbb{L}_2$ will be referred to as a **coordinatization** of $\mathbb{P}$. A coordinatization based on lines $\mathbb{L}_1$ and $\mathbb{L}_2$ and their units U_1 and U_2 will be referred to as the **coordinatization** (U_1, U_2). (*cf* Definition VEC.14(B).)

(B) $\mathbb{L}_1$ and $\mathbb{L}_2$ are the **axes** of this coordinatization, and O is its **origin**. For a visualization, see Figure 1.1.

(C) Interpreting $\mathbb{P}$ as a physical plane, such as a sheet of paper or a chalkboard, if a person's right hand is placed with the palm toward the surface so that the index finger points in the direction of $\overrightarrow{OU_1}$ and the thumb points in the direction $\overrightarrow{OU_2}$, then the coordinatization of $\mathbb{P}$ is **right-handed**.

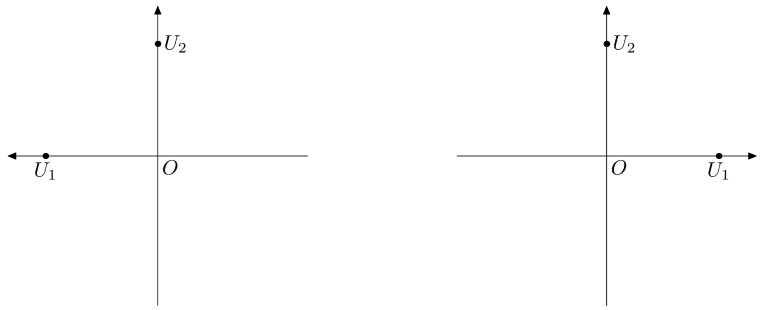


Fig. 1.1 Figures for Definition VEC.9: left-handed (left) and right-handed (right).

Whereas, if a person's left hand is placed palm toward the surface so that the index finger points in the direction of $\overrightarrow{OU_1}$ and the thumb points in the direction $\overrightarrow{OU_2}$, then the coordinatization of $\mathcal{P}$ is **left-handed**.

The rotation $\mathcal{P}$ such that $\rho(\overrightarrow{OU_1}) = \overrightarrow{OU_2}$ is **clockwise** for left-handed coordinatization and is **counterclockwise** for right-handed coordinatization.

Definition VEC.10 Let $\mathbb{R}^2$ denote the set

$$\mathbb{R} \times \mathbb{R} = \{(a, b) \mid a \text{ and } b \text{ are both members of } \mathbb{R}\},$$

that is, $\mathbb{R}^2$ is the *Cartesian product* of $\mathbb{R}$ and $\mathbb{R}$ (*cf Specht* Ch.1 Section 1.3). For any (a, b) and (c, d) in $\mathbb{R}^2$, and any real number x, define

(A) $(a, b) + (c, d) = (a + c, b + d)$ and

(B) $x(a, b) = (xa, xb)$.

(C) For any point $(a, b) \in \mathbb{R}^2$, we will refer to a as the **first coordinate** of (a, b), and to b as the **second coordinate**.

Theorem VEC.11 *For any (a, b), (c, d), and (e, f) in $\mathbb{R}^2$, and any real numbers x and y,*

(A) (1) $(a, b) + (c, d) = (c, d) + (a, b)$ *(addition is commutative).*

(2) $(a, b) + ((c, d) + (e, f)) = ((a, b) + (c, d)) + (e, f)$ *(addition is associative).*

(3) $(a, b) + (0, 0) = (a, b)$ ($(0, 0)$ *is the additive identity.*)

(4) $(-a, -b) + (a, b) = (0, 0)$ ($(-a, -b)$ *is the additive inverse of* (a, b).)

(B) (1) $x(y(a, b) = (xy)(a, b)$ *(scalar multiplication is associative).*

(2) $1(a, b) = (a, b)$.

(C) (1) $x((a, b) + (c, d)) = x(a, b) + x(c, d)$ *(scalar multiplication is distributive with respect to addition of points).*

(2) $(x+y)(a,b) = x(a,b)+y(a,b)$ *(scalar multiplication is distributive with respect to addition of scalars).*

(D) $x(a,b) = (0,0)$ *iff* $x = 0$ *or* $(a,b) = (0,0)$ *(or both).*

Proof. Using Definition VEC.10 and properties of real numbers,

(A) (1) $(a,b) + (c,d) = (a+c, b+d) = (c+a, d+b) = (c,d) + (a,b)$.

(2) $(a,b) + ((c,d) + (e,f)) = (a,b) + (c+e, d+f)$
$= ((a+(c+e), b+(d+e)) = ((a+c)+e, (b+d)+e)$
$= (a+c, b+d) + (e,f) = ((a,b) + (c,d)) + (e,f)$.

(3) $((a,b) + (0,0) = ((a+0, b+0) = (a,b)$.

(4) $(-a,-b) + (a,b) = (a-a, b-b) = (0,0)$.

(B) (1) $x(y(a,b) = x(ya, yb) = (xya, xyb) = (xy)(a,b)$.

(2) $1(a,b) = (1a, 1b) = (a,b)$.

(C) (1) $x((a,b) + (c,d)) = x((a+c, b+d)) = (x(a+c), x(b+d))$
$= (xa+xc, xb+xd) = (xa, xb) + (xc, xd)$
$= x(a,b) + x(c,d)$.

(2) $(x+y)(a,b) = (x+y)a, (x+y)b) = (xa+ya, xb+yb)$
$= (xa, xb) + (ya, yb) = x(a,b) + y(a,b)$.

(D) $x(a,b) = (0,0)$ iff $(xa, xb) = (0,0)$ iff $xa = 0$ and $xb = 0$ iff ($x = 0$ or $a = 0$) and ($x = 0$ or $b = 0$) iff $x = 0$ or ($a = 0$ and $b = 0$) iff $x = 0$ or $(a,b) = (0,0)$. □

Definition VEC.12 A **vector space**, or **linear space** over the field $\mathbb{R}$ of real numbers (called **scalars**) is a set $\mathbb{V}$ of elements called **vectors** satisfying the following conditions (A), (B), and (C):

(A) To every pair A and B of vectors in $\mathbb{V}$ there corresponds a vector $A+B$, called the **sum** of A and B, such that $\mathbb{V}$ forms an abelian group with respect to the operation $+$, that is,

(1) $A+B = B+A$ for all A and B in $\mathbb{V}$ (addition is **commutative**),

(2) $A+(B+C) = (A+B)+C$ for all A, B, and C in $\mathbb{V}$ (addition is **associative**),

(3) there exists in $\mathbb{V}$ a unique vector O (called the **origin** such that for every $A \in \mathbb{V}$, $A+O = O$ (O is the **additive identity**), and

(4) to every vector $A \in \mathbb{V}$ there corresponds a unique vector $-A$ such that $A+(-A) = O$ ($-A$ is the **additive inverse** of A.

(B) To every pair A and x, where $A \in \mathbb{V}$ and x is a real number, there corresponds a vector $xA \in \mathbb{V}$ called the **product**, or **scalar product** of x and A, such that

(1) $x(yA) = (xy)A$ for any real numbers x and y and every $A \in \mathbb{V}$ (multiplication by scalars is **associative**), and

(2) $1A = A$ for every vector $A \in \mathbb{V}$.

(C) (1) $x(A + B) = xA + xB$ for every real number x and all vectors A and B in $\mathbb{V}$ (scalar multiplication is **distributive** with respect to addition of vectors), and

(2) $(x + y)A = xA + yA$ for all real numbers x and y and every vectors $A \in \mathbb{V}$ (scalar multiplication is **distributive** with respect to addition of scalars).

(D) If $\mathbb{V}$ is a vector space, a subset $\mathbb{U} \subseteq \mathbb{V}$ is a **subspace** of $\mathbb{V}$ iff it is a vector space under the same operations as in $\mathbb{V}$. A subspace $\mathbb{U}$ of $\mathbb{V}$ is said to be a **proper** subspace if there exists at least one point $A \in \mathbb{V}$ such that $A \notin \mathbb{U}$.

Remark VEC.12.1 (A) By Theorem VEC.3 and Theorem VEC.7(A), (B), (C), and (D), $\mathbb{P}$ is an additive abelian group and a vector space. By Theorem VEC.11, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is a vector space.

(B) The reader should verify for herself that $\mathbb{R}$ is a vector space over itself—the vector space axioms are just a subset of the field axioms.

(C) The word *vector* in the term *vector space* does not imply their visualization as arrows on the plane in various locations (that is, "bound vectors"); if they are visualized as arrows, the initial point is always the origin O.

(D) The word *space* in the terms *vector* or *linear space* does not imply that it is ordinary "space" as in the incidence axioms. A vector space is a very general concept encompassing lines, planes, space, and spaces of higher dimension—see also Remark VEC.17.

(E) A subset $\mathbb{U}$ of $\mathbb{V}$ is a subspace iff for all A and B in $\mathbb{U}$ and every real number t, both $A+B \in \mathbb{U}$ and $tA \in \mathbb{U}$. This is because all the computational properties of $\mathbb{V}$ are "inherited" by $\mathbb{U}$. The set $\{O\}$ is a trivial subspace of $\mathbb{V}$; also, $\mathbb{V}$ is a subspace of itself.

Theorem VEC.13 *Every proper subspace (other than the trivial subspace O) of $\mathbb{V}$ of the plane $\mathbb{P}$ is a line through the origin O.*

Proof. Let $A \neq O$ be a point of $\mathbb{V}$. Then $\overleftrightarrow{OA} \subseteq \mathbb{V}$. If $\mathbb{U}$ is a proper subspace of $\mathbb{P}$ we show that it cannot contain any points other than those of $\overleftrightarrow{OA}$.

Suppose the contrary, that $B \neq O$ is also a point of $\mathbb{V}$ and $B \notin \overleftrightarrow{OA}$. Then also $\overleftrightarrow{OB} \subseteq \mathbb{V}$. If C is any point of $\mathbb{P}$, let $\mathcal{M}$ be the line through C which is parallel to $\overleftrightarrow{OA}$, as guaranteed by Axiom PS. Since $\overleftrightarrow{OB}$ intersects $\overleftrightarrow{OA}$ at

O it must also intersect $\mathcal{M}$ at some point D, by *Specht* Ch.2 Exercise IP.4. Then $C + (-D) \in \overleftrightarrow{OA}$. Hence $C - D + D = C \in \overleftrightarrow{OA} + \overleftrightarrow{OB}$ and therefore $\overleftrightarrow{OA} + \overleftrightarrow{OB} = \mathbb{P}$, and $\mathbb{V} = \mathbb{P}$, so it is not a proper subspace. Therefore the only proper subspaces of $\mathbb{P}$ are lines through the origin. $\square$

Definition VEC.14 (A) Let $\mathbb{P}$ be the Euclidean/LUB plane, O be its origin, and let $\mathbb{L}_1$ and $\mathbb{L}_2$ be perpendicular lines in $\mathbb{P}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$ which have been built into ordered fields with U_1 and U_2, respectively, as their units. Define λ to be a mapping from $\mathbb{P}$ to $\mathbb{R}^2$ as follows: for every $A = aU_1 + bU_2 \in \mathbb{P}$, define $\lambda(A) = \lambda(aU_1 + bU_2) = (a, b)$.

(B) The mapping λ defined in part (A) may be referred to as the **coordinatization map** belonging to the coordinatization (U_1, U_2) (*cf* Definition VEC.9(A)).

(C) A mapping Φ from a vector space $\mathbb{V}$ to a vector space $\mathbb{U}$ is a **vector space isomorphism**, or, if the context is well understood, simply an **isomorphism** iff

(1) Φ is a group isomorphism between the two spaces (as additive groups), and

(2) for every real number x and every $A \in \mathbb{V}$, $\Phi(xA) = x\Phi(V)$.

Again, as with isomorphisms of groups and fields, if two vector spaces are isomorphic, they cannot be distinguished algebraically and hence may be identified.

Remark VEC.15 It is easy to show that a bijection Φ of $\mathbb{V}$ to $\mathbb{U}$ is a vector space isomorphism iff for all real numbers x and y and all members A and B of $\mathbb{V}$, $\Phi(xA + yB) = x\Phi(A) + y\Phi(B)$. The proof of this is Exercise VEC.2.

Theorem VEC.16 *Let $\mathbb{P}$, $\mathbb{R}^2$, $\mathbb{L}_1$, $\mathbb{L}_2$ and λ be as in Definition VEC.14.*

(A) *λ is a bijection onto $\mathbb{R}^2$ and is a vector space isomorphism between $\mathbb{P}$ and $\mathbb{R}^2$.*

(B) *If $A \neq O$ is any point of $\mathbb{L}_1$, then $\mathbb{L}_1 = \{xA \mid x \in \mathbb{R}\}$ and*

$\lambda(\mathbb{L}_1) = \{(x, 0) \mid x \in \mathbb{R}\}$.

If $B \neq O$ is any point of $\mathbb{L}_2$, then $\mathbb{L}_2 = \{yB \mid y \in \mathbb{R}\}$ and

$\lambda(\mathbb{L}_2) = \{(0, y) \mid y \in \mathbb{R}\}$.

(C) *A line $\mathcal{L} \parallel \mathbb{L}_1$ iff for some point $C \in \mathbb{P} \setminus \mathbb{L}_1$, $\mathcal{L} = \mathbb{L}_1 + C$ iff for some real number $c \neq 0$, $\lambda(\mathcal{L}) = \{(x, c) \mid x \in \mathbb{R}\}$.*

A line $\mathcal{L} \parallel \mathbb{L}_2$ *iff for some point* $D \in \mathbb{P} \setminus \mathbb{L}_2$, $\mathcal{L} = \mathbb{L}_2 + D$ *iff for some real number* $d \neq 0$, $\lambda(\mathcal{L}) = \{(d, y) \mid y \in \mathbb{R}\}$.

Proof. (A)(I) λ is a one-to-one mapping, for if $\lambda(A) = \lambda(B)$, where $A = aU_1 + bU_2$ and $B = cU_1 + dU_2$, then $(a, b) = (c, d)$. λ is onto $\mathbb{R}^2$, for if (a, b) is any member of $\mathbb{R}^2$, $\lambda(aU_1 + bU_2) = (a, b)$. To see that it is an isomorphism, we must prove properties (1) and (2) of Definition VEC.14(C).

(1) λ is a group isomorphism between the two spaces (as additive groups), since for every $A = aU_1 + bU_2$ and $B = cU_1 + dU_2$ in $\mathbb{P}$,

$$\begin{aligned}\lambda(A + B) &= \lambda(aU_1 + bU_2 + cU_1 + dU_2) = \lambda(aU_1 + cU_1 + bU_2 + dU_2)\\ &= \lambda((a + c)U_1 + (b + d)U_2) = (a + c, b + d)\\ &= (a, b) + (c, d) = \lambda(A) + \lambda(B).\end{aligned}$$

(2) For every $A = aU_1 + bU_2$ and every real number x, using part (1)

$$\begin{aligned}\lambda(xA) &= \lambda(x(aU_1 + bU_2)) = \lambda(xaU_1 + xbU_2) = (xa, xb)\\ &= x(a, b) = x\lambda(aU_1 + bU_2) = x\lambda(A).\end{aligned}$$

The proofs of parts (B) and (C) are Exercise VEC.4. □

Remark VEC.17 (A) One of our objectives here is to use the fact of isomorphism between $\mathbb{R}^2$ and $\mathbb{P}$ to simplify the way in which we think about points in the plane. There is something quite clumsy about having constantly to refer to a point A in the plane as $aU_1 + bU_2$. For many purposes it's easier to think of (and easier to write!) such a point as a pair (a, b) of real numbers.

In the following, we make the identification between $aU_1 + bU_2$ and (a, b), treating them as if they were the same thing. We will be switching notations back and forth at will—doing so is legitimate because of the isomorphism λ and Theorem VEC.16. In particular, it is legitimate to write part (C) of Theorem VEC.16 just above as:

A line $\mathcal{L} \parallel \mathbb{L}_1$ *iff for some real number* $c \neq 0$, $\mathcal{L} = \{(x, c) \mid x \in \mathbb{R}\}$.

A line $\mathcal{L} \parallel \mathbb{L}_2$ *iff for some real number* $d \neq 0$, $\mathcal{L} = \{(d, y) \mid y \in \mathbb{R}\}$.

Citations to Theorem VEC.16(C) will, without further reference, be considered to include this version.

(B) In the language of vector space theory, part (B) of Theorem VEC.8 ($aU_1 + bU_2 = O$ iff $a = b = 0$) says that the two vectors U_1 and U_2 are **linearly independent**. Also, the fact that every point A in $\mathbb{P}$ can be expressed as $A = aU_1 + bU_2$ for some a and b says that the two vectors U_1 and U_2 **span** the space $\mathbb{P}$. In any vector space, a set of vectors which is both linearly independent and spans the space is called a **basis** for the space.

It can be shown that every vector space has a basis, and that any two bases for a given vector space have the same number of vectors, that is they have the same cardinal number (*cf Specht*, Chapter 1, Section 1.4). The number of elements in a basis of a space is called its **dimension**. Two vector spaces which are isomorphic must have the same dimension. The plane $\mathbb{P}$ and $\mathbb{R}^2$ both have dimension 2. That $\mathbb{R}^2$ has dimension 2 can be seen without resorting to the isomorphism, by verifying that the set $\{(1,0),(0,1)\}$ is a basis for $\mathbb{R}^2$.

(C) It is quite natural to extend Definition VEC.10 to vector spaces consisting of **ordered triples** of real numbers, and with some effort we might show that *space* (as defined in the axioms for incidence geometry) is isomorphic to this vector space.

We may extend these notions further to n**-tuples** of real numbers. Define the sum of any two n-tuples $(a_1, a_2, ..., a_n)$ and $(b_1, b_2, ..., b_n)$ in $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}$, as

$$(a_1, a_2, ..., a_n) + (b_1, b_2, ..., b_n) = (a_1 + b_1, a_2 + b_2, ..., a_n + b_n),$$

and for any real number x define the scalar product

$$x(a_1, a_2, ..., a_n) = (xa_1, xa_2, ..., xa_n).$$

Then it is easy to show (as we did in Theorem VEC.11 for $\mathbb{R}^2$) that $\mathbb{R}^n$ is a vector space, and that the set

$$\{(1,0,0,...,0),(0,1,0,...,0),(0,0,1,...,0),...,(0,0,0,....,1)\}$$

is a basis for $\mathbb{R}^n$. Thus $\mathbb{R}^n$ has dimension n.

(D) In more advanced vector space theory, these notions are extended to spaces of infinite dimension. It is quite easy to see that the set of all real-valued functions defined on the unit interval [0,1] (or for that matter, defined on any other fixed interval or on the whole real line) is a vector space, under pointwise addition of functions and scalar multiplication. That is, for any two functions f and g define $f + g$ to be the function whose value at each x in their common domain is $f(x) + g(x)$, and for any real number t define tf to be the function whose value at each x in its domain is $tf(x)$. In general, spaces of functions do not have finite dimension.

1.3 Lines and their slopes

Again, $\mathbb{P}$ will denote the Euclidean/LUB plane, O its origin, and $\mathbb{L}_1$ and $\mathbb{L}_2$ will be perpendicular lines in $\mathbb{P}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$, which have been built into ordered fields with U_1 and U_2, respectively, as their units, so that (U_1, U_2) is a coordinatization of $\mathbb{P}$. Every point $aU_1 + bU_2 \in \mathbb{P}$ is identified with the point $(a, b) \in \mathbb{R}^2$ using the isomorphism λ, as in Definition VEC.14.

Definition VEC.18 (A) A line $\mathcal{L}$ on $\mathbb{P}$ is **vertical** iff $\mathcal{L}\ \mathbb{PE}\ \mathbb{L}_2$ (meaning that either $\mathcal{L} \parallel \mathbb{L}_2$ or $\mathcal{L} = \mathbb{L}_2$, as defined in *Specht* Ch.3 Definition CAP.10).

(B) A line $\mathcal{L}$ on $\mathbb{P}$ is **horizontal** iff $\mathcal{L}\ \mathbb{PE}\ \mathbb{L}_1$.

Remark VEC.18.1 (A) Part (C) of Theorem VEC.16 says that a line $\mathcal{L} \parallel \mathbb{L}_1$ (and is *horizontal*) iff for some real number $c \neq 0$, $\mathcal{L} = \{(x, c) \mid x \in \mathbb{R}\}$. Also, $\mathcal{L} \parallel \mathbb{L}_2$ (and is *vertical*) iff for some real number $d \neq 0$, $\mathcal{L} = \{(d, y) \mid y \in \mathbb{R}\}$.

(B) At the risk of seeming overly pedantic, we note the reasoning it takes to verify that any horizontal line is perpendicular to any vertical line:

By Definition VEC.9, $\mathbb{L}_1$ and $\mathbb{L}_2$ are perpendicular; by *Specht* Ch.8 Theorem NEUT.32 each is a fixed line for the reflection over the other; by *Specht* Ch.11 Corollary EUC.3.1 every line parallel to a fixed line for a reflection is a fixed line for that reflection. It follows that lines parallel to $\mathbb{L}_1$ are fixed lines for the reflection over $\mathbb{L}_2$, and by Theorem NEUT.32 are perpendicular to $\mathbb{L}_1$.

By a similar argument all lines parallel to $\mathbb{L}_2$ are perpendicular to $\mathbb{L}_1$. By Theorem EUC.3, any line parallel to $\mathbb{L}_1$ is perpendicular to any line parallel to $\mathbb{L}_2$.

Theorem VEC.19 *Let $\mathcal{L}$ be a line in $\mathbb{P}$ which is not vertical. There exists a real number m such that for any two points $X = (x_1, x_2)$ and $Y = (y_1, y_2)$ of $\mathcal{L}$, $m = \frac{y_2 - x_2}{y_1 - x_1}$.*

Proof. (Case 1: $\mathcal{L}$ is horizontal.) For any two points $X = (x_1, x_2)$ and $Y = (y_1, y_2)$ of $\mathbb{P}$, $x_2 = y_2$ so that for all X and Y in $\mathbb{P}$, $\frac{y_2 - x_2}{y_1 - x_1} = 0$.

(Case 2: $\mathcal{L}$ is not horizontal and $O \in \mathcal{L}$.) By Theorem REAL.37, for every real number $t \neq 0$ there exists a dilation δ_t with fixed point O such that for every $X \neq O$ in $\mathbb{P}$, $tX = \delta_t(X)$. From *Specht* Ch.3 Theorem CAP.18 $\mathcal{L}$ is a fixed line for δ_t so for every t and every point $X \in \mathcal{L}$, $\delta_t(X) \in \mathcal{L}$.

Assume that $\mathcal{L}$ has been built into an ordered field with unit $U = (a, b)$. Then by Corollary REAL.35.1, for any two non-O points $X = (x_1, x_2)$ and $Y = (y_1, y_2)$ of $\mathcal{L}$ there exists a real number t such that $tX = Y$. Then

$$tX = t(x_1, x_2) = (tx_1, tx_2) = Y = (y_1, y_2)$$

and $tx_1 = y_1$ and $tx_2 = y_2$. It follows that $\frac{x_2}{x_1} = \frac{tx_2}{tx_1} = \frac{y_2}{y_1}$ for any two points (x_1, x_2) and (y_1, y_2) in $\mathcal{L}$. Also,

$$\frac{y_2-x_2}{y_1-x_1} = \frac{tx_2-x_2}{tx_1-x_1} = \frac{(t-1)x_2}{(t-1)x_1} = \frac{x_2}{x_1} = \frac{y_2}{y_1},$$

Now suppose that $Z = (z_1, z_2)$ and $W = (w_1, w_2)$ are any distinct non-O points of $\mathcal{L}$. By the same reasoning,

$$\frac{z_2-w_2}{z_1-w_1} = \frac{z_2}{z_1} = \frac{x_2}{x_1}.$$

Therefore $\frac{y_2-x_2}{y_1-x_1}$ is independent of our choice of points X and Y on the line $\mathcal{L}$, so long as neither $X = O$ or $Y = O$.

Finally, if $X = (x_1, x_2) = (0, 0) = O$, and $Y = (y_1, y_2)$ is any other point on $\mathcal{L}$, $\frac{y_2-x_2}{y_1-x_1} = \frac{x_2}{x_1}$. If we let $m = \frac{x_2}{x_1}$, then $m = \frac{y_2-x_2}{y_1-x_1}$ for any choice of X and Y on $\mathcal{L}$.

(Case 3: $\mathcal{L}$ is not horizontal and $O \notin \mathcal{L}$.) Let $A = (a, b)$ be a point of $\mathcal{L}$, which will be fixed for the rest of this argument. Let

$$\mathcal{M} = \mathcal{L} - A = \{X - A \mid X \in \mathcal{L}\} = \tau_{-A}(\mathcal{L}).$$

By Definition CAP.6, $\mathcal{M}$ is a line passing through O which is parallel to $\mathcal{L}$. A point $X = (x_1, x_2) \in \mathcal{L}$ iff $X - A = (x_1 - a, x_2 - b) \in \mathcal{M}$, and a point $Y = (y_1, y_2) \in \mathcal{L}$ iff $Y - A = (y_1 - a, y_2 - b) \in \mathcal{M}$. Then

$$\frac{(y_2-b)-(x_2-b)}{(y_1-a)-(x_1-a)} = \frac{y_2-b-x_2+b}{y_1-a-x_1+a} = \frac{y_2-x_2}{y_1-x_1};$$

this fraction, however, has been shown in Case 2 to be independent of the points X and Y, hence independent of the points $X - A$ and $Y - A$. This proves the theorem. □

Definition VEC.20 Let $\mathcal{L}$ be a nonvertical line on $\mathbb{P}$, and let $X = (x_1, x_2)$ and $Y = (y_1, y_2)$ be points on $\mathcal{L}$. The **slope** of $\mathcal{L}$ is $m = \frac{y_2-x_2}{y_1-x_1}$.

Theorem VEC.21 *Let s be any real number and Q be any point on $\mathbb{P}$. Then there exists a unique line $\mathcal{L}$ through Q with slope s.*

Proof. (I: Existence.) Let $\mathcal{L}$ be the line through $Q = (q_1, q_2)$ and $(q_1 + 1, q_2 + s)$. Then the slope of $\mathcal{L}$ is $\frac{q_2+s-q_2}{q_1+1-q_1} = s$.

(II: Uniqueness.) Let $\mathcal{M}$ be any nonvertical line through Q which has the same slope s as $\mathcal{L}$. Let the intersection of $\mathcal{M}$ and the vertical line through $(q_1 + 1, q_2)$ be $T = (q_1 + 1, q_2 + t)$, for some real number t. The slope of $\mathcal{M}$ is $\frac{q_2+t-q_2}{q_1+1-q_1} = \frac{t}{1} = t$. But the slope of $\mathcal{M}$ is s, so $s = t$ and $T = (q_1+1, q_2+s) \in \mathcal{L}$ so that $\mathcal{L}$ and $\mathcal{M}$ have two points in common, and are the same line. □

Theorem VEC.22 (A) *If $\mathcal{L}$ and $\mathcal{M}$ are non-vertical lines on $\mathbb{P}$, then $\mathcal{L}$ and $\mathcal{M}$ have the same slope iff $\mathcal{L}$ $\mathbb{PE}$ $\mathcal{M}$.*

(B) *If $\mathcal{L}$ with slope s and $\mathcal{M}$ with slope t are distinct non-vertical lines on $\mathbb{P}$, then $\mathcal{L} \perp \mathcal{M}$ iff $st = -1$.*

Proof. (A) If $\mathcal{L} \neq \mathcal{M}$, let $A = (a,b) \in \mathcal{L}$ and $B = (c,d) \in \mathcal{M}$. Then $\mathcal{L} + (B - A) = \tau_{B-A}(\mathcal{L})$ is a line that is parallel or equal to $\mathcal{L}$, by Definition CAP.6. Since $A \in \mathcal{L}$, $A + (B - A) = B$ so $\mathcal{L} + (B - A)$ and $\mathcal{M}$ intersect at the point B, and therefore they are the same line. If $X = (x_1, x_2)$ is any point distinct from A in $\mathcal{L}$, then the point $Y = X + (B - A) = (x_1 + c - a, x_2 + d - b)$ is a point of $\mathcal{M}$ distinct from B, and the slope of $\mathcal{M}$ is $\frac{(x_2+d-b)-d}{(x_1+c-a)-c} = \frac{x_2-b}{x_1-a}$ which is the slope of $\mathcal{L}$. Therefore if $\mathcal{L}$ and $\mathcal{M}$ are parallel, they have the same slope.

Conversely, if these two lines have the same slope, then by the previous argument, $\mathcal{L} + (B - A)$ has the same slope as $\mathcal{L}$ because they are parallel, and since $\mathcal{L} + (B - A)$ and $\mathcal{M}$ intersect at B, by the uniqueness part of Theorem VEC.21, $\mathcal{L} + (B - A) = \mathcal{M}$ and hence $\mathcal{L} \parallel \mathcal{M}$.

(B) The proof is Exercise VEC.13. □

Theorem VEC.23 *Let $Q = (q_1, q_2)$ be a point on $\mathbb{P}$, s a real number, and let $\mathcal{L}$ be the line through Q with slope s. Then*

$$\mathcal{L} = \{(x_1, x_2) \mid x_2 - q_2 - s(x_1 - q_1) = 0\}.$$

That is to say, a point $(x_1, x_2) \in \mathcal{L}$ iff $x_2 = q_2 + s(x_1 - q_1)$.

Proof. Let $X = (x_1, x_2)$, then by the uniqueness part of Theorem VEC.21, $X \in \mathcal{L} \setminus \{(q_1, q_2)\}$ iff $\frac{x_2-q_2}{x_1-q_1} = s$. Hence $X \in \mathcal{L}$ iff $x_2 - q_2 - s(x_1 - q_1) = 0$. □

Theorem VEC.24 *Let $U = (u_1, u_2)$ and $V = (v_1, v_2)$ be distinct points on $\mathbb{P}$. Then $\overleftrightarrow{UV} = \{(x_1, x_2) \mid (v_2 - u_2)(x_1 - u_1) - (v_1 - u_1)(x_2 - u_2) = 0\}$.*

Proof. (Case 1: $\overleftrightarrow{UV}$ is nonvertical.) Let (x_1, x_2) be any ordered pair of real numbers. The slope of $\overleftrightarrow{UV}$ is $\frac{v_2-u_2}{v_1-u_1}$, by Theorem VEC.19. Let U play the role of Q in Theorem VEC.23; then by that theorem,

$$X \in \overleftrightarrow{UV} \text{ iff } x_2 - u_2 - \left(\frac{v_2-u_2}{v_1-u_1}\right)(x_1 - u_1) = 0,$$

i.e., $(v_1 - u_1)(x_2 - u_2) - (v_2 - u_2)(x_1 - u_1) = 0$.

(Case 2: $\overleftrightarrow{UV}$ is vertical.) Let $X = (x_1, x_2)$ be any point of $\mathbb{P}$. Then $\overleftrightarrow{UV}$ is vertical iff $u_1 = v_1$, so that $(v_1 - u_1)(x_2 - u_2) - (v_2 - u_2)(x_1 - u_1) = 0$ becomes $(v_2 - u_2)(x_1 - u_1) = 0$. This is true iff $x_1 = u_1$, which is to say, $X = (x_1, x_2) \in \overleftrightarrow{UV}$. □

Theorem VEC.25 *Let U and V be distinct points on $\mathbb{P}$, $U = (u_1, u_2)$, and $V = (v_1, v_2)$. Then*

$$\begin{aligned}\overleftrightarrow{UV} &= \{X \mid X = t(V-U)+U \text{ and } t \in \mathbb{R}\} \\ &= \{(x_1, x_2) \mid (x_1, x_2) = t(v_1 - u_1) + u_1 \\ &\qquad\qquad \text{and } t(v_2 - u_2) + u_2 \text{ and } t \in \mathbb{R}\}.\end{aligned}$$

Proof. (Case 1: $\overleftrightarrow{UV}$ is nonvertical.) $\overleftrightarrow{UV} - U$ is a line $\mathcal{L}$ which passes through O. Assume $\mathcal{L}$ has been built into an ordered field with $O \neq V - U$ as its origin and unit U. Then $V - U \neq O$ is a point of $\mathcal{L}$, and a point $X = (x_1, x_2) \in \overleftrightarrow{UV}$ iff $X - U = (x_1 - u_1, x_2 - u_2) \in \mathcal{L}$.

By Corollary REAL.35.1 and the fact that the line $\mathcal{L}$ is fixed for δ_t, where t is any real number, we know that $X = (x_1, x_2) \in \overleftrightarrow{UV}$ iff there exists a real number t such that $X - U = t(V - U)$. That is to say,

$$(x_1 - u_1, x_2 - u_2) = t(v_1 - u_1, v_2 - u_2) = (t(v_1 - u_1), t(v_2 - u_2)),$$
$$\text{or } (x_1, x_2) = (t(v_1 - u_1) + u_1, t(v_2 - u_2) + u_2).$$

This proves Case 1.

(Case 2: $\overleftrightarrow{UV}$ is vertical.) Let (x_1, x_2) be any ordered pair of real numbers. Then $\overleftrightarrow{UV}$ is vertical iff $u_1 = v_1$. Hence in this case $X \in \overleftrightarrow{UV}$ iff $x_1 = u_1$. We know there exists a real number t such that $x_2 = t(v_2 - u_2) + u_2$. Then for this value of t in particular, $x_1 = t(0) + u_1 = t(v_1 - u_1) + u_1$. Conversely, suppose $x_2 = t(v_2 - u_2) + u_2$; then since $u_1 = v_1$, $x_1 = t(0) + u_1 = t(v_1 - u_1)t + u_1$. $\square$

Theorem VEC.25.1 *Let U and V be distinct points on $\mathbb{P}$, $U = (u_1, u_2)$, and $V = (v_1, v_2)$. The mapping Θ of $\mathbb{R}$ onto $\overleftrightarrow{UV}$ such that for each real number t, $\Theta(t) = t(V - U) + U$ preserves order and betweenness. That is, for any real numbers r, s, and t,*

$$r < s \textit{ iff } \Theta(r) = r(V-U)+U < \Theta(s) = s(V-U)+U \textit{ and}$$
$$r\text{–}s\text{–}t \textit{ iff } (r(V-U)+U)\text{–}(s(V-U)+U)\text{–}(t(V-U)+U).$$

Proof. Build $\overleftrightarrow{UV}$ into an ordered field with origin $\hat{O} = U$ and unit $\hat{U} = V - U$. Then by Theorem VEC.25 $\overleftrightarrow{UV} = \{X \mid X = t\hat{U} + \hat{O} \text{ and } t \in \mathbb{R}\}$. By Theorem REAL.35, the mapping $\Theta(t) = t(V - U) + U = t\hat{U} + \hat{O} = t\hat{U}$ is order preserving, that is, $t < s$ iff $t\hat{U} < s\hat{U}$ iff $t(V - U) + U < s(V - U) + U$.

So for any three points $X = r\hat{U}$, $Y = s\hat{U}$, and $Z = t\hat{U}$ on $\overleftrightarrow{UV}$, by *Specht* Ch.6 Theorem ORD.6

$$X\text{–}Y\text{–}Z \text{ iff } (X < Y < Z \text{ or } Z < Y < X) \text{ iff } (r < s < t \text{ or } t < s < r)$$

which is true iff r–s–t. $\square$

1.4 Norms and inner products

In this section, $\mathbb{P}$ will denote the Euclidean/LUB plane, O its origin, and $\mathbb{L}_1$ and $\mathbb{L}_2$ will be perpendicular lines in $\mathbb{P}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$, which have been built into ordered fields with U_1 and U_2, respectively, as their units, so that (U_1, U_2) is a coordinatization of $\mathbb{P}$, and $\overline{OU_1} \cong \overline{OU_2}$. Every point $aU_1 + bU_2 \in \mathbb{P}$ is identified with the point $(a, b) \in \mathbb{R}^2$ using the isomorphism λ, as in Definition VEC.14.

Theorem VEC.26 *Let $A = a_1U_1 + a_2U_2$ be any member of $\mathbb{P}$. Then there exists a unique real number $c > 0$ such that $[\overline{OcU_1}] = [\overline{OA}]$ and $cU_1 = \Phi[\overline{OA}]$, where Φ is the mapping defined in Specht Ch.9 Definition FSEG.14.*

Proof. By Theorem FSEG.13 there exists a unique point $X \in \overrightarrow{OU_1}$ such that $[\overline{OX}] = [\overline{OA}]$. By Theorem REAL.35(A) there exists a unique real number c such that $X = cU_1$, and since $X \in \overrightarrow{OU_1}$, $c > 0$. $\square$

Definition VEC.26.1 Let A be any member of $\mathbb{P}$. Then if $A \neq O$, define the **norm** of A (denoted $\|A\|$) as the positive real number c such that $\overline{OA} \cong \overline{OcU_1}$ (*i.e.* $cU_1 = \Phi[\overline{OA}]$) the existence of which is guaranteed by Theorem VEC.26; if $A = O$ define $\|A\| = 0$.

By Definition OF.16, $\|A\|$ is the *length* of the segment $\overline{OA}$, or the *distance* from O to A.

Theorem VEC.26.2 *For any two vectors A and B in $\mathbb{P} \setminus \{O\}$,*
$$\|A\| = \|B\| \text{ iff } \overline{OA} \cong \overline{OB}.$$

Proof. $\overline{OA} \cong \overline{OB}$ iff $[\overline{OA}] = [\overline{OB}]$ iff $\Phi[\overline{OA}] = \Phi[\overline{OB}]$. $\square$

Theorem VEC.26.3 *For any two vectors A and B in $\mathbb{P}$,*
$$\overline{AB} \cong \overline{O(B-A)}.$$

Proof. Since τ_A is a translation, the lines $\overleftrightarrow{O(B-A)}$ and $\tau_A(\overleftrightarrow{O(B-A)}) = \overleftrightarrow{AB}$ are parallel (*cf* Definition CAP.6). By Theorem VEC.4 $\square AO(B-A)B$ is a parallelogram. By Theorem EUC.12(A), $\overline{AB} \cong \overline{O(B-A)}$, since these segments are opposite edges of this parallelogram. $\square$

Theorem VEC.26.4 *For any distinct vectors A and B in $\mathbb{P}$, $\|B - A\|$ is the length of the segment $\overline{AB}$, as defined in Definition OF.16.*

Proof. By Theorem VEC.26.3, $[\overline{AB}] = [\overline{O(B-A)}]$, so that by Definition OF.16, the length of the segment $\overline{AB}$ is $\Phi[\overline{AB}] = \Phi[\overline{O(B-A)}] = \|B-A\|$. $\square$

Theorem VEC.26.5 (Third form of the Pythagorean Theorem) *For any three distinct non-collinear vectors A, B, and C in $\mathbb{P}$,*

$$\|A-B\|^2 = \|B-C\|^2 + \|A-C\|^2 \textit{ iff } \angle ACB \textit{ is a right angle.}$$

Proof. By Theorem VEC.26 and Theorem VEC.26.3, there exist positive real numbers a, b, and c such that $aU_1 = \Phi[\overline{BC}] = \Phi[\overline{O(B-C)}]$, $bU_1 = \Phi[\overline{AC}] = \Phi[\overline{O(A-C)}]$, and $cU_1 = \Phi[\overline{AB}] = \Phi[\overline{O(A-B)}]$. By Definition VEC 26.1, $a = \|B-C\|$, $b = \|A-C\|$, and $c = \|A-B\|$.

In the notation of *Specht* Ch.15 Theorem SIM.23.1 (Second form of the Pythagorean Theorem), let $\hat{A} = aU_1 = \Phi[\overline{O(B-C)}]$, $\hat{B} = bU_1 = \Phi[\overline{O(A-C)}]$, and $\hat{C} = cU_1 = \Phi[\overline{O(A-B)}]$. Then by that theorem, $\angle ACB$ is right iff $(cU_1)^2 = (aU_1)^2 + (bU_1)^2$, and by Theorem REAL.25, this is $c^2U_1 = a^2U_1 + b^2U_1$, or

$$c^2U_1 - a^2U_1 - b^2U_1 = (c^2 - a^2 - b^2)U_1 = O.$$

By Corollary REAL.34(B) this is true iff $c^2 - a^2 - b^2 = 0$ or $c^2 = a^2 + b^2$. That is, $\|A-B\|^2 = \|B-C\|^2 + \|A-C\|^2$. $\square$

Theorem VEC.27 (A) *For any vector $A = a_1U_1 + a_2U_2 \in \mathbb{P}$, $\|A\|^2 = {a_1}^2 + {a_2}^2$.*

(B) *For any distinct vectors $A = a_1U_1 + a_2U_2$ and $B = b_1U_1 + b_2U_2$ in $\mathbb{P}$, $\|A-B\|^2 = (a_1-b_1)^2 + (a_2-b_2)^2$.*

Proof. (A) The quadrilateral $\square O(a_1U_1)A(a_2U_2)$ is a parallelogram and

$$\overleftrightarrow{O(a_1U_1)} \perp \overleftrightarrow{O(a_2U_2)}$$

so by Theorem EUC.3 $\overleftrightarrow{O(a_1U_1)} \perp \overleftrightarrow{(a_1U_1)A}$. By Theorem EUC.12(A)

$$\text{(i) } \overline{(a_1U_1)A} \cong \overline{O(a_2U_2)}.$$

The three points O, a_1U_1, and A form a triangle, where $\angle O(a_1U_1)A$ is right. Since $\overline{OU_2} \cong \overline{OU_1}$, by *Specht* Ch.13 Theorem DLN.17 and Theorem REAL.37,

$$\text{(ii) } \overline{O(a_2U_2)} \cong \overline{O(a_2U_1)}.$$

By Theorem VEC.26.3,

$$\text{(iii) } \overline{(a_1U_1)A} \cong \overline{O(A - a_1U_1)}.$$

Putting congruences (i), (ii) and (iii) together,

$$\overline{O(A-a_1U_1)} \cong \overline{(a_1U_1)A} \cong \overline{O(a_2U_2)} \cong \overline{O(a_2U_1)},$$

so by Definition VEC.26.1 $\|A - a_1U_1\| = \|a_2U_1\| = |a_2|$. Also by definition, $\|a_1U_1 - O\| = \|a_1U_1\| = |a_1|$. By Theorem VEC.26.5,

$$\begin{aligned}\|A\|^2 = \|A - O\|^2 &= \|a_1U_1 - O\|^2 + \|A - a_1U_1\|^2 \\ &= |a_1|^2 + |a_2|^2 = {a_1}^2 + {a_2}^2.\end{aligned}$$

(B) $$\begin{aligned}A - B &= a_1U_1 + a_2U_2 - b_1U_1 - b_2U_2 \\ &= a_1U_1 - b_1U_1 + a_2U_2 - b_2U_2 = (a_1 - b_1)U_1 + (a_2 - b_2)U_2\end{aligned}$$

so by part (A) $\|A - B\|^2 = (a_1 - b_1)^2 + (a_2 - b_2)^2$. $\square$

Remark VEC.28 (A) Without proof one cannot assume that for a given real number t, the point tU_1 rotated onto $\overrightarrow{OU_2}$ would be the same point as tU_2. That is, it's not automatic that the scale for scalar multiplication on $\mathbb{L}_1$ is the same as for scalar multiplication on $\mathbb{L}_2$. If $\overline{OU_1} \cong \overline{OU_2}$, Theorem REAL.37 assures us that these two scales are indeed the same, that is, $\overline{OtU_1} \cong \overline{OtU_2}$. This was important in the proof just above, because the norm is defined in terms of a scalar multiple on $\mathbb{L}_1$, but we wanted to express the result in terms of a scalar multiple on $\mathbb{L}_2$. That is why we invoked congruence (ii) in this proof.

(B) We defined the norm of a vector as a point of $\mathbb{P}$, in a manner quite specific to the plane. Under the identification between $\mathbb{P}$ and $\mathbb{R}^2$ provided by Definition VEC.14 and Theorem VEC.16, for any point $(a, b) \in \mathbb{R}^2$ the result of Theorem VEC.27(A) above becomes

$$\|(a, b)\|^2 = a^2 + b^2 \text{ or } \|(a, b)\| = \sqrt{a^2 + b^2}.$$

For any two points (a, b) and (c, d) of $\mathbb{R}^2$, part (B) of Theorem VEC.27 becomes

$$\begin{gathered}\|(a, b) - (c, d)\|^2 = (a - c)^2 + (b - d)^2 \\ \text{or } \|(a, b) - (c, d)\| = \sqrt{(a - c)^2 + (b - d)^2}.\end{gathered}$$

(C) In $\mathbb{R}^n$, expressions similar to those in (A) are generally used for the definition of norm. That is, for any point $(a_1, a_2, ..., a_n) \in \mathbb{R}^n$, the norm $\|(a_1, a_2, ..., a_n)\| = \sqrt{{a_1}^2 + {a_2}^2 + ... + {a_n}^2}$.

Norms on vector spaces have a number of useful properties; we state the most fundamental of these in the following Theorem; in more general vector spaces these three properties are sometimes used as a definition of a norm.

Theorem VEC.29 *For all points $A = (a, b)$ and $B = (c, d)$ in $\mathbb{R}^2$, and every real number x,*

(1) $\|A\| > 0$ iff $(a, b) \neq (0, 0)$;

*(2) $\|A + B\| \leq \|A\| + \|B\|$; that is $\|(a + c, b + d)\| \leq \|(a, b)\| + \|(c, d)\|$ (the **triangle inequality**); and*

(3) $\|xA\| = |x|\|A\|$, *that is* $\|x(a,b)\| = |x|\|(a,b)\|$ *(the* ***homogeneity property****)*.

Proof. The proof is Exercise VEC.6. $\square$

Definition VEC.30 (A) The **inner product** (sometimes called the **dot product**) of two vectors (a, b) and (c, d) in $\mathbb{R}^2$ is the real number $ac + bd$, and is denoted here by the symbol $(a, b) \bullet (c, d)$.

(B) Two vectors (a, b) and (c, d) are **orthogonal** iff $(a, b) \bullet (c, d) = 0$.

Remark VEC.31 (A) Properties of the inner product are listed and proved in Exercise VEC.7.

(B) In general vector spaces (over the field of real numbers) the inner product is defined as a real-valued function which satisfies the properties listed in Exercise VEC.7. The present definition (and its natural extensions) are valid only on $\mathbb{R}^n$. In more general theory, the inner product of two vectors A and B is usually denoted (A, B) but for our very limited treatment we avoid this notation since it may be confused with our notation for vectors in $\mathbb{R}^2$.

Theorem VEC.32 *For every vector* $A = (a, b)$, $A \bullet A = \|A\|^2$.

Proof. $A \bullet A = (a, b) \bullet (a, b) = a^2 + b^2 = \|A\|^2$. $\square$

Theorem VEC.33 *Two nonzero vectors* $A = (a, b)$ *and* $B = (c, d)$ *are orthogonal iff* $\overleftrightarrow{(0,0)(a,b)} \perp \overleftrightarrow{(0,0)(c,d)}$.

Proof. Using the properties of inner product from Exercise VEC.4,

$$\begin{aligned} |A - B|^2 &= (A - B) \bullet (A - B) = A \bullet A - 2A \bullet B + B \bullet B \\ &= |A|^2 + |B|^2 - 2A \bullet B. \end{aligned}$$

By the Pythagorean Theorem and its converse (*cf Specht* Ch.15 Theorem SIM.23) $\overleftrightarrow{OA} \perp \overleftrightarrow{OB}$ iff $|A - B|^2 = |A|^2 + |B|^2$. Hence $A \bullet B = 0$ iff $A \perp B$. $\square$

The equality $|A - B|^2 = |A|^2 + |B|^2 - 2A \bullet B$ is a generalization of the Pythagorean Theorem since it holds for any triangle.

1.5 Linear mappings

In this section, $\mathbb{P}$ will denote the Euclidean/LUB plane, O its origin, and $\mathbb{L}_1$ and $\mathbb{L}_2$ will be perpendicular lines in $\mathbb{P}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$, which have been built into ordered fields with U_1 and U_2, respectively, as their units, so that (U_1, U_2) is a coordinatization of $\mathbb{P}$. Every point $aU_1 + bU_2 \in \mathbb{P}$ is identified with the point $(a, b) \in \mathbb{R}^2$ using the isomorphism λ, as in Definition VEC.14.

Definition VEC.34 Let $\mathbb{V}$ be a vector space over the field of real numbers.

(A) A **linear mapping** (or **linear transformation** or **linear operator**) α on $\mathbb{V}$ is a mapping of $\mathbb{V}$ into $\mathbb{V}$ which satisfies the following conditions:

(1) for all A and B in $\mathbb{V}$, $\alpha(A + B) = \alpha(A) + \alpha(B)$, and

(2) for every $A \in \mathbb{V}$ and every real number t, $\alpha(tA) = t\alpha(A)$.

(B) The mapping O is the mapping such that $O(A) = O$ for every $A \in \mathbb{V}$.

(C) The mapping $-\alpha$ is the mapping such that for every $A \in \mathbb{V}$,

$$(-\alpha)(A) = -(\alpha(A)).$$

(D) (1) The **sum** of two linear mappings α and β on $\mathbb{V}$ is the mapping $\alpha + \beta$ such that for every $A \in \mathbb{V}$, $(\alpha + \beta)(A) = \alpha(A) + \beta(A)$.

(2) The **product** or **scalar product** of a real number t and a linear mapping α on $\mathbb{V}$ is the mapping $t\alpha$ such that for every $A \in \mathbb{V}$, $(t\alpha)(A) = t(\alpha(A))$.

Theorem VEC.35 *Let $\mathbb{V}$ be a vector space over the field of real numbers.*

(A) *If α and β are linear mappings on $\mathbb{V}$, and t is any real number, then the mappings $\alpha + \beta$ and $t\alpha$ are linear mappings on $\mathbb{V}$.*

(B) *The mappings O and $\imath$ are linear mappings on $\mathbb{V}$.*

(C) *For every linear mapping α on $\mathbb{V}$, the mapping $-\alpha$ as in Definition VEC.34(C) is a linear mapping.*

(D) *If α is any linear mapping on $\mathbb{V}$, then*

(1) *$\alpha(O) = O$; and*

(2) *for any $A \in \mathbb{V}$, $\alpha(-A) = -(\alpha(A))$.*

(E) *The set of all linear mappings on the vector space $\mathbb{V}$, with the definitions of sum and scalar product given in Definition VEC.34(C) is itself a vector space over the real numbers.*

Proof. The proof of parts (A) through (D) is Exercise VEC.8. The proof of part (E) is Exercise VEC.9. $\square$

Theorem VEC.36 (A) *A linear mapping α on a vector space $\mathbb{V}$ is not one-to-one iff for some $A \neq O$ in $\mathbb{V}$, $\alpha(A) = O$.*

(B) *If α is a linear mapping on a vector space $\mathbb{V}$, then the image $\alpha(\mathbb{V})$ of $\mathbb{V}$ under α is a subspace of $\mathbb{V}$.*

Proof. (A) If for some $A \neq O$ in $\mathbb{V}$, $\alpha(A) = O$, by Theorem VEC.35(D)(1) $\alpha(O) = O$ so α is not one-to-one. Conversely, if α is not one-to-one, there exist distinct points A and B of $\mathbb{V}$ such that $\alpha(A) = \alpha(B)$ and hence $\alpha(A-B) = O$.

(B) Let A and B be points of $\alpha(\mathbb{V})$ and let t be a real number. Then there exist points C and D of $\mathbb{V}$ such that $\alpha(C) = A$ and $\alpha(D) = B$. It follows that $\alpha(C + D) = \alpha(C) + \alpha(D) = A + B$ so that $A + B \in \alpha(\mathbb{V})$. Also, $\alpha(tC) = t\alpha(C) = tA$ so that $tA \in \alpha(\mathbb{V})$. By Remark VEC.12.1(E), $\alpha(\mathbb{V})$ is a subspace of $\mathbb{V}$. $\square$

Definition/Remark VEC.37 (A) In the remainder of this section we will often write points $X = (x_1, x_2)$ of $\mathbb{R}^2$ in the form $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. For each $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ define $\alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$.

(B) It should be noted here that the mapping α is often defined in terms of matrices and matrix multiplication. A **matrix** is a rectangular array of numbers such as $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. This one is a "square 2 by 2 matrix."

Two matrices can be multiplied provided the number of columns in the left-hand matrix (multiplicand) is the same as the number of rows in the right hand matrix. Thus, $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$ can be multiplied, but $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$ can't.

Multiplication is carried out by the "row by column rule." by which the ijth entry of the product (that is, the entry in the ith row and the jth column) is the sum of the products of the entries in the ith row of the left-hand matrix with the entries in the jth column of the right hand matrix. Thus, the product

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$
$$= \begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) & (a_{11}b_{13} + a_{12}b_{23}) \\ (a_{21}b_{11} + a_{22}b_{21}) & (a_{21}b_{12} + a_{22}b_{22}) & (a_{21}b_{13} + a_{22}b_{23}) \end{bmatrix}.$$

We can use the same rule to express the value of the mapping α on $\mathbb{R}^2$. Recall that for each $X \in \mathbb{R}^2$ we wrote $X = (x_1, x_2)$ as a matrix with 2 rows and 1 column, that is, as $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and we defined $\alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$. The right-hand side here is precisely the product $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ using the row by column rule. In this usage, the matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is said to be the **matrix of the mapping** α. If we know the values of α at the points $(1, 0)$ and $(0, 1)$, we can easily find the matrix of α, since

$$\alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \text{ and } \alpha \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}.$$

The **determinant** of the matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is the quantity $a_{11}a_{22} - a_{12}a_{21}$, which is usually denoted by the symbol $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$. If $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is the matrix of the mapping α we will often say that $a_{11}a_{22} - a_{12}a_{21}$ is the determinant of α.

Theorem VEC.38 *The mapping α defined in Definition VEC.37 is a linear mapping on $\mathbb{R}^2$.*

Proof. Let $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ be any two points of $\mathbb{R}^2$, and let t be any real number. Then

$$\alpha(X + Y) = \alpha(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) = \begin{bmatrix} a_{11}(x_1 + y_1) + a_{12}(x_2 + y_2) \\ a_{21}(x_1 + y_1) + a_{22}(x_2 + y_2) \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}x_1 + a_{11}y_1 + a_{12}x_2 + a_{12}y_2 \\ a_{21}x_1 + a_{21}y_1 + a_{22}x_2 + a_{22}y_2 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} + \begin{bmatrix} a_{11}y_1 + a_{12}y_2 \\ a_{21}y_1 + a_{22}y_2 \end{bmatrix} = \alpha(X) + \alpha(Y).$$

Also $\alpha(tX) = \alpha(\begin{bmatrix} tx_1 \\ tx_2 \end{bmatrix}) = \begin{bmatrix} a_{11}(tx_1) + a_{12}(tx_2) \\ a_{21}(tx_1) + a_{22}(tx_2) \end{bmatrix} = \begin{bmatrix} t(a_{11}(x_1) + a_{12}(x_2)) \\ t(a_{21}(x_1) + a_{22}(x_2)) \end{bmatrix}$

$= t\alpha(X)$. □

Theorem VEC.39 *The mapping α defined in Definition VEC.37 is one-to-one iff its determinant $a_{11}a_{22} - a_{12}a_{21} \neq 0$.*

Proof. We show that there exist real numbers x_1 and x_2, not both zero, such that $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$ iff the determinant $a_{11}a_{22} - a_{12}a_{21} = 0$. That is, the mapping α is not one-to-one iff the determinant is 0.

(I) Suppose the determinant $a_{11}a_{22} - a_{12}a_{21} = 0$. It is quite easy to show that the mapping is not one-to-one if none of the entries a_{11}, a_{22}, a_{12} or a_{21} is zero. Sorting through the various cases where one or more of the entries is zero is not difficult, but tedious, and we leave this work to the reader as Exercise VEC.10.

If all entries are non-zero, then let $x_1 = 1$ and let $x_2 = -\frac{a_{11}}{a_{12}}$. Then $a_{21} = \frac{a_{11}a_{22}}{a_{12}}$, so that

$a_{11}x_1 + a_{12}x_2 = a_{11} + a_{12}(-\frac{a_{11}}{a_{12}}) = a_{11} - a_{11} = 0$

and $a_{21}x_1 + a_{22}x_2 = a_{21} + a_{22}(-\frac{a_{11}}{a_{12}}) = \frac{a_{11}a_{22}}{a_{12}} - a_{22}(\frac{a_{11}}{a_{12}}) = 0$,

so that $\alpha(x_1, x_2) = (0, 0)$. But $(x_1, x_2) \neq (0, 0)$, so by Theorem VEC.36, α is not one-to-one.

(II) Suppose $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$, where not both x_1 and x_2 are zero.

If $x_1 \neq 0$, $a_{11} + a_{12}\frac{x_2}{x_1} = 0$ and $a_{21} + a_{22}\frac{x_2}{x_1} = 0$, so that there is no loss of generality to assume that $x_1 = 1$. Then the assumption takes the form $a_{11} + a_{12}x_2 = 0$ and $a_{21} + a_{22}x_2 = 0$, so that $x_2 = -\frac{a_{11}}{a_{12}}$ and also $x_2 = -\frac{a_{21}}{a_{22}}$ hence $\frac{a_{11}}{a_{12}} = \frac{a_{21}}{a_{22}}$ and $a_{11}a_{22} - a_{12}a_{21} = 0$. If $x_1 = 0$ and $x_2 \neq 0$ then $0 = a_{11}x_1 + a_{12}x_2 = a_{12}x_2$ and $0 = a_{21}x_1 + a_{22}x_2 = a_{22}x_2$, hence $a_{12} = 0$, $a_{22} = 0$ and $a_{11}a_{22} - a_{12}a_{21} = 0$. □

Theorem VEC.40 *A linear mapping α defined on $\mathbb{P}$ is one-to-one iff it is onto. That is, it is a bijection iff it is either one-to-one or onto.*

Proof. (I) Suppose α is onto $\mathbb{P}$ and not one-to-one. There exists some point $U_1 \neq O$ of such that $\alpha(U_1) = O$. Let U_2 be any point such that $\overleftrightarrow{OU_1} \perp \overleftrightarrow{OU_2}$. Build $\overleftrightarrow{OU_1}$ and $\overleftrightarrow{OU_2}$ into ordered fields with U_1 and U_2 as their units. α is onto means that for every point $B \in \mathbb{P}$, there exists a point A such that $\alpha(A) = B$.

By Theorem VEC.8, for some real numbers s and t, $A = sU_1 + tU_2$. Then

$$B = \alpha(A) = \alpha(sU_1) + \alpha(tU_2) = s\alpha(U_1) + t\alpha(U_2) = t\alpha(U_2)$$

since $\alpha(U_1) = O$.
Therefore, every point of $\mathbb{P}$ is a member of $\overleftrightarrow{O\alpha(U_2)}$. But by Axiom I.5(B) we know there are points of $\mathbb{P}$ that are not on this line, contradicting the assumption that α is not one-to-one.

(II) Conversely, suppose that α is not onto $\mathbb{P}$. Then $\alpha(\mathbb{P})$ is a proper subspace of $\mathbb{P}$, and by Theorem VEC.13 it is a line through the origin.

Again using Theorem VEC.8, let U_1 and U_2 be any two non-O points of $\mathbb{P}$ such that U_1, U_2 and O are noncollinear and assume that the lines $\overleftrightarrow{OU_1}$ and $\overleftrightarrow{OU_2}$ have been built into ordered fields such that for every point A of $\mathbb{P}$ there are real numbers a and b such that $A = aU_1 + bU_2$. Then both $\alpha(U_1)$ and $\alpha(U_2)$ are members of the line $\alpha(\mathbb{P})$. If either $\alpha(U_1) = O$ or $\alpha(U_2) = O$, α is not one-to-one. If neither, then by Corollary REAL.35.1 there exists a real number t such that $\alpha(U_1) = t\alpha(U_2)$, and

$$O = \alpha(U_1) - t\alpha(U_2) = \alpha(U_1) - \alpha(tU_2) = \alpha(U_1 - tU_2).$$

Since the lines $\overleftrightarrow{OU_1}$ and $\overleftrightarrow{OU_2}$ have only the point O in common, $U_1 \neq tU_2$ and hence $U_1 - tU_2 \neq O$. Therefore α is not one-to-one. □

Theorem VEC.41 *The linear mapping α defined on $\mathbb{R}^2$ by Definition VEC.37 is a bijection iff its determinant $a_{11}a_{22} - a_{12}a_{21} \neq 0$.*

Proof. By Theorem VEC.39 α is one-to-one iff $a_{11}a_{22} - a_{12}a_{21} \neq 0$. By Theorem VEC.40 α is a bijection iff it is one-to-one.

Note that here we have used the identification between $\mathbb{P}$ and $\mathbb{R}^2$ that was made in Theorem VEC.16 and Remark VEC.17(A). □

Theorem VEC.42 (A) *The set of all bijective linear mappings on a vector space $\mathbb{V}$ forms a group under composition of mappings.*

(B) *The group defined in part (A) is not abelian.*

Proof. (A) By elementary function theory, composition of mappings, hence of bijective linear maps, is associative. There exists an identity $\imath$ for composition, and it is a linear mapping. The composition of two bijections is a bijection, by elementary function theory. The composition of two linear mappings is linear, as can be seen from the following calculations.

Let α and β be linear maps. Then for every A and B in $\mathbb{V}$, and every real number t, we have

$$\begin{aligned}\alpha \circ \beta(A+B) &= \alpha(\beta(A+B)) = \alpha(\beta(A)+\beta(B)) \\ &= \alpha(\beta(A)) + \alpha(\beta(B)) = \alpha \circ \beta(A) + \alpha \circ \beta(B)\end{aligned}$$

and

$$\alpha \circ \beta(tA) = \alpha(\beta(tA)) = \alpha(t\beta(A)) = t\alpha(\beta(A)) = t\alpha \circ \beta(A).$$

Again by elementary function theory, every bijection has an inverse which is a bijection. It remains only to prove that the inverse of a linear mapping is linear. Let α be a bijective linear map, and let β be its inverse. Then for every A and B in $\mathbb{V}$, since α is onto, there exist points C and D in $\mathbb{V}$ such that $\alpha(C) = A$ and $\alpha(D) = B$, and by linearity of α, $\alpha(C+D) = A+B$ so that $\beta(A+B) = C+D$. Then $\beta(A+B) = C+D = \beta(A)+\beta(B)$. Also, for any real number t, $\alpha(tC) = t\alpha(C) = tA$ so that $\beta(tA) = \beta(\alpha(tC)) = tC = t\beta(A)$. Hence β is linear.

(B) The proof is Exercise VEC.11. $\square$

Theorem VEC.43 *Let α be a bijective linear map on a vector space $\mathbb{V}$. Then α has a fixed point other than O iff the mapping $\alpha - \imath$ is not one-to-one.*

Proof. α has a fixed point other than O iff for some point $A \neq O$, $\alpha(A) = A$, that is, $\alpha(A) - A = O$, or the mapping $\alpha - \imath$ is not one-to-one. $\square$

1.6 Affine mappings and belineations

In this section, $\mathbb{P}$ will denote the Euclidean/LUB plane, O its origin, and $\mathbb{L}_1$ and $\mathbb{L}_2$ will be perpendicular lines in $\mathbb{P}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$, which have been built into ordered fields with U_1 and U_2, respectively, as their units, so that (U_1, U_2) is a coordinatization of $\mathbb{P}$. Every point $aU_1 + bU_2 \in \mathbb{P}$ is identified with the point $(a, b) \in \mathbb{R}^2$ using the isomorphism λ, as in Definition VEC.14.

Definition VEC.44 Let $\mathbb{V}$ be a vector space. A mapping β of $\mathbb{V}$ to $\mathbb{V}$ is an **affine** mapping iff there exists a linear mapping α on $\mathbb{V}$ and a point D of $\mathbb{V}$ such that for every $A \in \mathbb{V}$, $\beta(A) = \alpha(A) + D$. That is, $\beta = \tau_D \circ \alpha$, where τ_D is the translation such that $\tau(O) = D$.

When we wish to emphasize the relationship between α and β we will refer to β as an affine mapping **associated** with the linear mapping α, or to α as the linear mapping associated with the affine mapping β.

Remark VEC.45 An affine mapping β on $\mathbb{R}^2$ (identified with $\mathbb{P}$) takes the form

$$\alpha \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + d_1 \\ a_{21}x_1 + a_{22}x_2 + d_2 \end{bmatrix}, \text{ where } D = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}.$$

Theorem VEC.46 (A) *Let α be a linear mapping on a vector space $\mathbb{V}$ and let β be an affine mapping associated with α. Then α is a bijection iff β is a bijection.*

(B) *Let α be a linear mapping on $\mathbb{R}^2$ and β be an affine mapping associated with α. Then β is a bijection iff the determinant $a_{11}a_{22} - a_{12}a_{21}$ is non-zero.*

Proof. (A) For some $D \in \mathbb{V}$, $\beta(A) = \tau_D \circ \alpha$. Since translations are bijections (*cf* Theorem ISM.6 and Theorem NEUT.11), by elementary function theory, β is a bijection iff α is a bijection.

(B) Follows directly from Part (A) and Theorem VEC.41. $\square$

Theorem VEC.47 *If f_1, h_1, f_2, and h_2 are real numbers such that $(f_1, f_2) \neq (0, 0)$, then $\mathcal{L} = \{(f_1t + h_1, f_2t + h_2) \mid t \in \mathbb{R}\}$ is the line through $(f_1 + h_1, f_2 + h_2)$ and (h_1, h_2).*

Proof. This is an immediate consequence of Theorem VEC.25, if we let $v_i = f_i - h_i$ and $u_i = h_i$ for $i = 1, 2$. $\square$

Theorem VEC.48 *Let D be a point of $\mathbb{R}^2$ and let α be linear mapping on $\mathbb{R}^2$. Define β to be the mapping on $\mathbb{R}^2$ such that for every $A \in \mathbb{R}^2$, $\beta(A) = \alpha(A) + D$ (so that β is an affine mapping associated with α). If the determinant of α is nonzero, then*

(A) *β is a belineation, that is, a bijection preserving betweenness; and*

(B) *β is a collineation, mapping lines to lines.*

Proof. (A) By Theorem VEC.46 both α and β are bijections. Let A, B, and C be points of $\mathbb{R}^2$; using Definition LC.8 (from *Specht*, Chapter 21, section 21.5.2) and Theorem LC.12, A–B–C iff there exists a real number t such that $0 < t < 1$ and $B = A + t(C - A)$. Then

$$\begin{aligned} \beta(B) &= \alpha(B) + D = \alpha((A) + t(C - A)) + D = \alpha(A) + \alpha(t(C - A)) + D \\ &= \alpha(A) + t\alpha(C - A) + D = \alpha(A) + t\alpha(C) - t\alpha(A) + D \\ &= \alpha(A) + D + t\alpha(C) + tD - t\alpha(A) - tD \end{aligned}$$

$$= \alpha(A) + D + t[(\alpha(C) + D) - (\alpha(A) + D)]$$
$$= \beta(A) + t[\beta(C) - \beta(A)],$$

so that $\beta(A)$–$\beta(B)$–$\beta(C)$ and β is a belineation.

(B) By *Specht* Ch.21 Theorems LC.2 through LC.20, $\mathbb{R}^2$ is a Pasch plane, since all the axioms through Axiom PSA hold. Then by *Specht* Ch.7 Theorem COBE.2, β is a collineation. □

Theorem VEC.49 *Every belineation of* $\mathbb{R}^2$ *is an affine mapping of* $\mathbb{R}^2$.

Proof. Let γ be any belineation of $\mathbb{R}^2$. By *Specht* Ch.19 Theorem AA.10, γ is determined by its values on any three noncollinear points. Since $\begin{bmatrix}0\\0\end{bmatrix}$, $\begin{bmatrix}1\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\end{bmatrix}$ are noncollinear, let q_1, q_2, r_1, r_2, s_1, and s_2 be real numbers such that

$$\gamma\left(\begin{bmatrix}0\\0\end{bmatrix}\right) = \begin{bmatrix}q_1\\q_2\end{bmatrix}, \gamma\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}r_1\\r_2\end{bmatrix}, \gamma\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}s_1\\s_2\end{bmatrix},$$

and let β be the affine mapping such that

$$\beta\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}(r_1 - q_1)x_1 + (s_1 - q_1)x_2 + q_1\\(r_2 - q_2)x_1 + (s_2 - q_2)x_2 + q_2\end{bmatrix};$$

then

$$\beta\left(\begin{bmatrix}0\\0\end{bmatrix}\right) = \begin{bmatrix}q_1\\q_2\end{bmatrix}, \beta\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}r_1\\r_2\end{bmatrix}, \beta\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}s_1\\s_2\end{bmatrix}.$$

Hence $\gamma = \beta$. Here we have followed Martin [3].[2] □

Theorem VEC.50 (A) *A mapping* α *is a belineation of* $\mathbb{R}^2$ *with fixed point* $(0,0)$ *iff it is a bijective linear map of* $\mathbb{R}^2$.

(B) *The set of all belineations of* $\mathbb{R}^2$ *with fixed point* $(0,0)$ *is a group under composition of mappings.*

Proof. (A) By Theorem VEC.49, a belineation of $\mathbb{R}^2$ is a bijective affine mapping β whose value at each point A is $\beta(A) = \alpha(A)+D$ (α is its associated linear map). Since α is a linear map, $\alpha(0,0) = (0,0)$, so $\beta(0,0) = (0,0)$ iff $D = (0,0)$, which means that $\beta = \alpha$. Hence, every belineation of $\mathbb{R}^2$ with fixed point $(0,0)$ is a bijective linear map. Conversely, by Theorem VEC.48

[2] George E. Martin, *Transformation Geometry, An Introduction to Symmetry*, Springer, 1982 (Theorem 15.11).

every bijective linear map is a belineation, and every linear map has $(0,0)$ as a fixed point.

(B) By Theorem VEC.42 the set of all bijective linear maps forms a group under composition. □

Theorem VEC.51 *Let α be a belineation of $\mathbb{R}^2$ with fixed point $(0,0)$. Then by Theorem VEC.50, α is a bijective linear map. Hence by Definition VEC.37 and Theorem VEC.38, there exist real numbers a_{11}, a_{22}, a_{12} and a_{21} such that for every member $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ of $\mathbb{R}^2$, $\alpha\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$. Then α has no other fixed points iff $(a_{11}-1)(a_{22}-1) - a_{12}a_{21} \neq 0$.*

Proof. By Theorem VEC.43, α has a fixed point other than $(0,0)$ iff the mapping $\alpha - \imath$ is not one-to-one.

For every every member $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ of $\mathbb{R}^2$,

$$(\alpha - \imath)\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} (a_{11}-1)x_1 + a_{12}x_2 \\ a_{21}x_1 + (a_{22}-1)x_2 \end{bmatrix}.$$

By Theorem VEC.39, this mapping is one-to-one, hence has no fixed points other than $(0,0)$ iff $(a_{11}-1)(a_{22}-1) - a_{12}a_{21} \neq 0$. □

Theorem VEC.52 *A belineation φ of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ is an isometry iff for every two distinct members (x_1, x_2) and (y_1, y_2) of $\mathbb{R}^2$,*
$$\|\varphi(x_1,x_2) - \varphi(y_1,y_2)\| = \|(x_1,x_2) - (y_1,y_2)\| = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}.$$

Proof. (I) Let (x_1, x_2), (y_1, y_2), and (z_1, z_2) be three noncollinear members of $\mathbb{R}^2$. If the equality holds for every two distinct members of $\mathbb{R}^2$, then

$$\|\varphi(x_1,x_2) - \varphi(y_1,y_2)\| = \|(x_1,x_2) - (y_1,y_2)\|,$$
$$\|\varphi(x_1,x_2) - \varphi(z_1,z_2)\| = \|(x_1,x_2) - (z_1,z_2)\|, \text{ and}$$
$$\|\varphi(y_1,y_2) - \varphi(z_1,z_2)\| = \|(y_1,y_2) - (z_1,z_2)\|.$$

By Theorem VEC.26

$$\overline{\varphi(x_1,x_2)\varphi(y_1,y_2)} \cong \overline{(x_1,x_2)(y_1,y_2)},$$
$$\overline{\varphi(x_1,x_2)\varphi(z_1,z_2)} \cong \overline{(x_1,x_2)(z_1,z_2)}, \text{ and}$$
$$\overline{\varphi(y_1,y_2)\varphi(z_1,z_2)} \cong \overline{(y_1,y_2)(z_1,z_2)}.$$

By Theorem NEUT.62 there exists an isometry ψ of $\mathbb{R}^2$ such that

$$\psi(\triangle(x_1,x_2)(y_1,y_2)(z_1,z_2)) = \triangle\varphi(x_1,x_2)\varphi(y_1,y_2)\varphi(z_1,z_2).$$

Since the values of φ and ψ agree on the three noncollinear points (x_1, x_2), (y_1, y_2), and (z_1, z_2), by Theorem AA.10 $\varphi = \psi$.

(II) If φ is an isometry of $\mathbb{R}^2$, by Theorem NEUT.15,
$$\varphi(\overline{(x_1,x_2)(y_1,y_2)}) = \overline{\varphi(x_1,x_2)\varphi(y_1,y_2)}$$
and by Definition NEUT.6(B)
$$\overline{(x_1,x_2)(y_1,y_2)} \cong \overline{\varphi(x_1,x_2)\varphi(y_1,y_2)}.$$
By Theorem VEC.26.3, this is
$$\overline{O((x_1,x_2)-(y_1,y_2))} \cong \overline{O(\varphi(x_1,x_2)-\varphi(y_1,y_2))}.$$
and by Theorem VEC.26.2 this is equivalent to
$$\|(x_1,x_2)-(y_1,y_2)\| = \|\varphi(x_1,x_2)-\varphi(y_1,y_2)\|. \quad \square$$

Theorem VEC.53 *There exists a belineation of $\mathbb{R}^2$ which has no fixed point and is not an isometry.*

Proof. By Theorem VEC.48, an affine mapping whose associated linear map has non-zero determinant, is a collineation.

Let φ be the linear map of $\mathbb{R}^2$ such that for every member $(x_1,x_2) \in \mathbb{R}^2$, $\varphi(x_1,x_2) = (x_1, 2x_2)$. That is,
$$\varphi \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \cdot x_1 + 0 \cdot x_2 \\ 0 \cdot x_1 + 2 \cdot x_2 \end{bmatrix},$$
and the determinant of φ is $a_{11}a_{22} - a_{12}a_{21} = 1 \cdot 2 - 0 \cdot 0 = 2 \neq 0$, so that φ is a collineation. Let ψ be the affine map such that for every member (x_1,x_2) of $\mathbb{R}^2$, $\psi(x_1,x_2) = (x_1+1, x_2)$. Then
$$\psi \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \cdot x_1 + 0 \cdot x_2 + 1 \\ 0 \cdot x_1 + 1 \cdot x_2 + 0 \end{bmatrix}, \text{ where } D = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$
and the determinant of the linear mapping associated with ψ is
$$a_{11}a_{22} - a_{12}a_{21} = 1 \cdot 1 - 0 \cdot 0 = 1 \neq 0,$$
so that ψ is a collineation. Let $\theta = \psi \circ \varphi$. Then
$$\theta(x_1,x_2) = \psi(\varphi(x_1,x_2)) = \psi(x_1, 2x_2) = (x_1+1, 2x_2).$$

If θ had a fixed point (x_1,x_2), $x_1 = x_1 + 1$ and $1 = 0$, which is false, so θ has no fixed point. If θ were an isometry,
$$\begin{aligned}\|\theta(x_1,x_2) - \theta(y_1,y_2)\| &= \|(x_1+1, 2x_2) - (y_1+1, 2y_2)\| \\ &= \|(x_1,x_2) - (y_1,y_2)\|,\end{aligned}$$
so that $\sqrt{(x_1-y_1)^2 + 4(x_2-y_2)^2} = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$, and thus $4 = 1$ which is false. By Theorem VEC.52, θ is not an isometry. $\square$

1.7 Exercises for vector spaces

Exercise VEC.1* Complete the routine computations necessary to prove Theorem VEC.3, that is, show that $\mathbb{P}$ is an Abelian group under the operation $+$.

Exercise VEC.2* Prove Remark VEC.15: a bijection mapping Φ from $\mathbb{V}$ to $\mathbb{U}$ is a (vector space) isomorphism iff for all real numbers x and y and all members A and B of $\mathbb{V}$, $\Phi(xA + yB) = x\Phi(A) + y\Phi(B)$.

Exercise VEC.3* Assuming the hypotheses of Theorem VEC.8, let $A = aU_1 + bU_2$ and $B = cU_1 + dU_2$ be points of $\mathbb{P}$. Rewrite the assertions (B) through (F) of Theorem VEC.7 in terms of U_1 and U_2.

Exercise VEC.4* Prove Theorem VEC.16(B) and (C).

Exercise VEC.5* Let $\mathbb{P}$ be a Euclidean/LUB plane, and let $Q = (q_1, q_2)$ and $R = (r_1, r_2)$ be distinct points of $\mathbb{P}$. Then the point $M = (\frac{q_1+r_1}{2}, \frac{q_2+r_2}{2})$ is the midpoint of $\overline{QR}$.

Exercise VEC.6* Prove Theorem VEC.29: for any points $A = (a, b)$ and $B = (c, d)$ in $\mathbb{P} = \mathbb{R}^2$, and any real number x,

(1) $\|A\| = \|(a, b)\| \geq 0$;

(2) $\|A + B\| = \|(a + c, b + d)\| \leq \|(a, b)\| + \|(c, d)\| = \|A\| + \|B\|$; and

(3) $\|xA\| = \|x(a, b)\| = |x|\|(a, b)\| = |x|\|A\|$.

Exercise VEC.7* Let A, B, and C be vectors and t be a real number; then

(A) $A \bullet B = B \bullet A$;

(B) $t(A \bullet B) = (tA) \bullet B = A \bullet (tB)$;

(C) $A \bullet (B + C) = A \bullet B + A \bullet C$; and

(D) $A \bullet A = \|A\|^2 > 0$, and $A \bullet A = 0$ iff $A = O$.

Exercise VEC.8* Prove Theorem VEC.35(A) through (D):

(A) If α and β are linear mappings on $\mathbb{V}$, and t is any real number, then the mappings $\alpha + \beta$ and $t\alpha$ are linear mappings on $\mathbb{V}$.

(B) The mappings O and $\imath$ are linear mappings on $\mathbb{V}$.

(C) For every linear mapping α on $\mathbb{V}$, the mapping $-\alpha$ as in Definition VEC.34(C) is a linear mapping.

(D) If α is any linear mapping on $\mathbb{V}$, then

(1) $\alpha(O) = O$; and

(2) for any $A \in \mathbb{V}$, $\alpha(-A) = -(\alpha(A))$.

Exercise VEC.9* Prove Theorem VEC.35(E): the set of all linear mappings on the vector space $\mathbb{V}$, with the definitions of sum and scalar product given in Definition VEC.34(C) is itself a vector space over the real numbers.

Exercise VEC.10* Complete the proof of Theorem VEC.39: if the determinant $a_{11}a_{22} - a_{12}a_{21} = 0$ and one or more of the entries a_{11}, a_{22}, a_{12} or a_{21} is zero, show that the mapping is not one-to-one.

Exercise VEC.11* Prove part (B) of Theorem VEC.42, by giving a counterexample showing that the set of all bijective linear mappings on the vector space $\mathbb{R}^2$ is not abelian.

Exercise VEC.12* Let $\mathbb{P}$ be a Euclidean/LUB plane, and let $\mathcal{L}$ be a line with slope s on $\mathcal{P}$ which is neither vertical nor horizontal. Let Q be a point on $\mathcal{L}$ such that $Q = (q_1, q_2)$, and let $X \neq Q$ be a point on $\mathcal{L}$ such that $X = (x_1, x_2)$.

If $s > 0$, then $x_1 < q_1$ and $x_2 < q_2$, or $x_1 > q_1$ and $x_2 > q_2$.

If $s < 0$, then $x_1 < q_1$ and $x_2 > q_2$, or $x_1 > q_1$ and $x_2 < q_2$.

Exercise VEC.13* Prove Theorem VEC.22(B): iff $\mathcal{L}$ and $\mathcal{M}$ be distinct lines which are neither vertical nor horizontal and which have respective slopes s and t, then $\mathcal{L} \perp \mathcal{M}$ iff $st = -1$.

Exercise VEC.14* Let a, b, and c be real numbers such that a and b are not both zero. Then $\mathcal{L} = \{(x_1, x_2) \mid ax_1 + bx_2 + c = 0\}$ is characterized as follows:

(I) If $b = 0$, then $a \neq 0$ and $\mathcal{L}$ is the vertical line through $(\frac{-c}{a}, 0)$.

(II) If $b \neq 0$, then $\mathcal{L}$ is the line through $(0, \frac{-c}{b})$ with slope $\frac{-a}{b}$.

Exercise VEC.15* Let $\mathbb{P}$ be a Euclidean/LUB plane, and a, b, and c be real numbers where not both a and b are 0; let

$$\mathcal{L} = \{(x_1, x_2) \mid ax_1 + bx_2 + c = 0\}$$

and

$$\mathcal{M} = \{(x_1, x_2) \mid bx_1 - ax_2 + c = 0\}.$$

Then $\mathcal{L} \perp \mathcal{M}$.

Exercise VEC.16* The set of collineations of $\mathbb{R}^2$ with $(0, 0)$ as a sole fixed point, together with the identity mapping $\imath$, is not a group under composition of mappings.

Exercise VEC.17* There exist stretches S and T of $\mathbb{R}^2$ such that $T \circ S$ has only the fixed point $(0, 0)$.

Exercise VEC.18 Let k be a nonzero real number and let φ be a collineation of $\mathbb{R}^2$ such that for every member (x_1, x_2) of $\mathbb{R}^2$, $\varphi(x_1, x_2) = (kx_1, kx_2)$. Prove that

(1) φ is a dilation of $\mathbb{R}^2$ with fixed point $(0, 0)$ (*cf Specht* Ch.3 Theorem CAP.22).

(2) Using the equality $\varphi(x_1, x_2) = (kx_1, kx_2)$ prove that the set of dilations of $\mathbb{R}^2$ with fixed point $(0, 0)$, together with $\imath$ (the identity mapping of $\mathbb{R}^2$ is a group under composition of mappings.

Exercise VEC.19 Let k be a nonzero real numbers, φ be the collineation of $\mathbb{R}^2$ such that for every member (x_1, x_2) of $\mathbb{R}^2$ $\varphi(x_1, x_2) = (kx_1, x_2)$ and let ψ be the collineation of $\mathbb{R}^2$ such that for every member (x_1, x_2) of $\mathbb{R}^2$, $\psi(x_1, x_2) = (x_1, kx_2)$.

Prove: (1) φ is a stretch of $\mathbb{R}^2$ with axis $\mathbb{L}_1$.

(2) ψ is a shear of $\mathbb{R}^2$ with axis $\mathbb{L}_2$.

(3) $\varphi \circ \psi = \psi \circ \varphi$ is a dilation of $\mathbb{R}^2$ with fixed point $(0, 0)$.

(4) The set of stretches with axis $\mathbb{L}_1$ together with the identity $\imath$ is a group under composition of mappings. (Use the equality $\varphi(x_1, x_2) = (kx_1, k_2)$.

(5) The set of stretches with axis $\mathbb{L}_2$ together with the identity $\imath$ is a group under composition of mappings. (Use the equality $\psi(x_1, x_2) = (x_1, kx_2)$.

Exercise VEC.20 Let $\mathbb{V}$ be a set of collineations of $\mathbb{R}^2$ with the property that for every member φ of $\mathbb{V}$ there exist nonzero real numbers r and s such that for every member (x_1, x_2) of $\mathbb{R}^2$, $\varphi(x_1, x_2) = (rx_1, sx_2)$. Prove that $\mathbb{V} \cup \{\imath\}$ is an abelian group under composition of mappings.

1.8 Selected answers for vector spaces

Exercise VEC.1 Proof. Let A, B, and C be any points of $\mathbb{P}$, and let τ_A, τ_B, and τ_C be the translations in $\mathbb{T}$ such that $\tau_A(O) = A$, $\tau_B(O) = B$, and $\tau_C(O) = C$.

$A + B = (\tau_A \circ \tau_B)(O) \in \mathbb{P}$ since $\tau_A \circ \tau_B$ is a mapping of $\mathbb{P}$ to $\mathbb{P}$, so that $\mathbb{P}$ is closed under addition.

$A + (B + C) = (\tau_A \circ (\tau_B \circ \tau_C))(O) = ((\tau_A \circ \tau_B) \circ \tau_C)(O) = (A + B) + C$ so that addition is associative.

$A + B = (\tau_A \circ \tau_B)(O) = (\tau_B \circ \tau_A)(O) = B + A$, so addition is commutative.

$A + O = (\tau_A \circ \tau_O)(O) = (\tau_A \circ \imath)(O) = \tau_A(O) = A$ so O is the additive identity.

For any translation τ_A which maps O to A, there exists an inverse translation ${\tau_A}^{-1}$. If we define $-A = {\tau_A}^{-1}(O)$, $A + (-A) = (\tau_A({\tau_A}^{-1}(O)) = O$ so that $-A$ is the additive inverse of A. $\square$

Exercise VEC.2 Proof. If Φ is an isomorphism then for all real numbers x and y and all A and B in $\mathbb{V}$, $\Phi(xA+yB) = \Phi(xA)+\Phi(yB) = x\Phi(A)+y\Phi(B)$. Conversely, assume that for all real numbers x and y and all A and B in $\mathbb{V}$,

$$\Phi(xA + yB) = \Phi(xA) + \Phi(yB) = x\Phi(A) + y\Phi(B).$$

Let $x = y = 1$; then

$$\Phi(A + B) = \Phi(xA + yB) = \Phi(xA) + \Phi(yB) = \Phi(A) + \Phi(B).$$

Let $y = 0$; then

$$\Phi(xA) = \Phi(xA + 0 \cdot B) = x\Phi(A) + 0 \cdot \Phi(B) = x\Phi(A).$$

Therefore Φ is an isomorphism. $\square$

Exercise VEC.3 Proof.

(B) $x(y(aU_1 + bU_2)) = x(yA) = (xy)A = (xy)aU_1 + (xy)bU_2$,

(C) $$\begin{aligned} x(aU_1 + bU_2 + cU_1 + dU_2) &= x(A + B) = xA + xB \\ &= x(aU_1 + cU_1 + bU_2 + dU_2) \\ &= x(a + c)U_1 + x(b + d)U_2, \end{aligned}$$

(D) $(x+y)(aU_1+bU_2) = (x+y)A = xA+yA = x(aU_1+bU_2)+y(aU_1+bU_2)$,

(E) $1(aU_1 + bU_2) = 1A = aU_1 + bU_2$, and

(F) $x(aU_1 + bU_2) = xA = O = 0U_1 + 0U_2$ iff $x = 0$ or $A = O$, and by the proof of Theorem VEC.7, $A = O$ means that both $a = 0$ and $b = 0$. $\square$

Exercise VEC.4 Proof. (B) By Theorem VEC.7(A) a point $X \in \mathbb{L}_1$ iff for some real number x, $X = xA$, proving the first assertion. This is true in particular if $A = U_1$. By Theorem VEC.8(A) there exist real numbers a and b such that $X = aU_1 + bU_2$. Thus $X \in \mathbb{L}_1$ iff for some real number x

$$O = aU_1 + bU_2 - xU_1 = aU_1 - xU_1 + bU_2 = (a - x)U_1 + bU_2$$

which by Theorem VEC.8(B) is true iff $a = x$ and $b = 0$. Therefore $X = aU_1 + bU_2 \in \mathbb{L}_1$ iff $a = x$ and $b = 0$ iff $\lambda(X) = \lambda(aU_1 + bU_2) = (a, b) = (x, 0)$ for some real number x.

A similar proof shows that $X = aU_1 + bU_2 \in \mathbb{L}_2$ iff $a = 0$ and $b = y$ iff $\lambda(X) = \lambda(aU_1 + bU_2) = (a, b) = (0, y)$ for some real number y.

(C) Let $\mathcal{L}$ be a line in $\mathbb{P}$ which is parallel to $\mathbb{L}_1$, and let $C \in \mathcal{L}$, and τ_C be the unique translation such that $\tau_C(O) = C$. Then by Definition CAP.6 $\tau_C(\mathbb{L}_1)$ is a line which is either equal to or parallel to $\mathbb{L}_1$, and which contains the point C. By Axiom PS (from *Specht* Chapter 2) $\mathcal{L} = \tau_C(\mathbb{L}_1) = C + \mathbb{L}_1$, and since $C \in \mathcal{L}$ and $\mathcal{L} \parallel \mathbb{L}_1$, $C \notin \mathbb{L}_1$.

Conversely, if for some $C \in \mathbb{P}\backslash\mathbb{L}_1$, $\mathcal{L} = C+\mathbb{L}_1$, then since $\tau_C(\mathbb{L}_1) = C+\mathbb{L}_1$ and τ is a translation, either $C+\mathbb{L}_1 \parallel \mathbb{L}_1$ or $C+\mathbb{L}_1 = \mathbb{L}_1$. But $C = C+O \notin \mathbb{L}_1$ so the latter is ruled out, and $C + \mathbb{L}_1 \parallel \mathbb{L}_1$.

This shows that $\mathcal{L} \parallel \mathbb{L}_1$ iff there exists a $C \notin \mathbb{L}_1$ such that $\mathcal{L} = C + \mathbb{L}_1$.

By Theorem VEC.8(A) every point $C \in \mathbb{P}$ can be written as $C = eU_1+cU_2$ for some real numbers e and c. Suppose that for some $C \in \mathbb{P}\backslash\mathbb{L}_1$, $\mathcal{L} = C+\mathbb{L}_1$. By Part(B), $c \neq 0$, and also by Part (B) $X \in \mathbb{L}_1$ iff $X = xU_1$ for some real number x. Thus for every $X \in C + \mathbb{L}_1$, $X = eU_1 + cU_2 + xU_1 = yU_1 + cU_2$ for some real number $y = e + x$.

Conversely, suppose that there exists a real number $c \neq 0$ such that for every X, $X = xU_1 + cU_2$ for some real number x. Let $C = cU_2$, so that $C \notin \mathbb{L}_1$. Then since $xU_1 \in \mathbb{L}_1$, for every $X = cU_2 + xU_1$ where x is a real number, $X = cU_2 + xU_1 = C + xU_1 \in C + \mathbb{L}_1$.

This shows that for $C \notin \mathbb{L}_1$, $X \in C + \mathbb{L}_1$ iff there exists a real number $c \neq 0$ such that $X = cU_2 + xU_1$.

Therefore $\mathcal{L} \parallel \mathbb{L}_1$ iff there exists a point $C \in \mathcal{L}$ and a real number $c \neq 0$ such that $\mathcal{L} = C + \mathbb{L}_1 = \{X \mid X = cU_2 + xU_1 \text{ and } x \in \mathbb{R}\}$, and this is true iff $\lambda(\mathcal{L}) = \lambda(C + \mathbb{L}_1) = \{(x, c) \mid x \in \mathbb{R}\}$.

A similar proof shows that a line $\mathcal{L} \parallel \mathbb{L}_2$ iff for some point $D \in \mathbb{P} \setminus \mathbb{L}_2$, $\mathcal{L} = D + \mathbb{L}_2$ iff for some real number $d \neq 0$, $\lambda(\mathcal{L}) = \{(d, y) \mid y \in \mathbb{R}\}$. □

Exercise VEC.5 Proof. By Theorem VEC.25,

$$\begin{aligned}\overleftrightarrow{QR} &= \{X \mid X = t(R-Q) + Q \text{ and } t \in \mathbb{R}\}\\ &= \{(x_1, x_2) \mid (x_1, x_2) = t(r_1 - q_1) + q_1 \text{ and } t(r_2 - q_2) + q_2 \text{ and } t \in \mathbb{R}\}.\end{aligned}$$

If $t = 0$, $X = 0(R-Q) + Q = Q$ and if $t = 1$, $X = 1(R-Q) + Q = R$.
If $t = \frac{1}{2}$ then

$$X = \tfrac{1}{2}(R-Q) + Q = \tfrac{1}{2}R - \tfrac{1}{2}Q + Q = \tfrac{1}{2}R + \tfrac{1}{2}Q = \tfrac{R+Q}{2}$$

and this point belongs to $\overleftrightarrow{QR}$. But $\frac{R+Q}{2} = (\frac{q_1+r_1}{2}, \frac{q_2+r_2}{2})$ which is M. By Theorem VEC.25.1, since $0 < \frac{1}{2} < 1$, Q–M–R so $M \in \overleftrightarrow{QR}$. Now

$$\begin{aligned}\|Q - M\| &= \sqrt{(q_1 - (\tfrac{r_1+q_1}{2}))^2 + (q_2 - (\tfrac{r_2+q_2}{2}))^2}\\ &= \sqrt{(\tfrac{q_1}{2} - \tfrac{r_1}{2})^2 + (\tfrac{q_2}{2} - \tfrac{r_2}{2})^2},\end{aligned}$$

and

$$\begin{aligned}\|R - M\| &= \sqrt{(r_1 - (\tfrac{r_1+q_1}{2}))^2 + (r_2 - (\tfrac{r_2+q_2}{2}))^2}\\ &= \sqrt{(\tfrac{r_1}{2} - \tfrac{q_1}{2})^2 + (\tfrac{r_2}{2} - \tfrac{q_2}{2})^2}\end{aligned}$$

$$= \sqrt{\left(-\left(\tfrac{q_1}{2} - \tfrac{r_1}{2}\right)\right)^2 + \left(-\left(\tfrac{q_2}{2} - \tfrac{r_2}{2}\right)\right)^2} = \|Q - M\|.$$

Hence by Theorem VEC.26.2, $\overline{RM} \cong \overline{QM}$ and by Definition NEUT.6(C), M is the midpoint of $\overline{QR}$. $\square$

Exercise VEC.6 Proof.

(1) $\|(a,b)\| = \sqrt{a^2+b^2} \geq 0$, by properties of real numbers.

(2) For any real numbers a, b, c and d

$$0 \leq (bc - ad)^2 = b^2c^2 - 2abcd + a^2d^2$$

so $b^2c^2 + a^2d^2 \geq 2abcd$. Then

$$\begin{aligned} 0 \leq (ac+bd)^2 &= a^2c^2 + 2abcd + b^2d^2 \leq a^2c^2 + b^2c^2 + a^2d^2 + b^2d^2 \\ &= (a^2+b^2)(c^2+d^2) = \|A\|^2\|B\|^2 \end{aligned}$$

hence $ac + bd \leq \|A\|\|B\|$. Then

$$\begin{aligned} \|A+B\|^2 &= \|(a,b)+(c,d)\|^2 = \|(a+c,b+d)\|^2 \\ &= (a+c)^2 + (b+d)^2 = a^2 + c^2 + 2ac + b^2 + d^2 + 2bd \\ &= (a^2+b^2) + (c^2+d^2) + 2(ac+bd) \\ &\leq \|A\|^2 + \|B\|^2 + 2\|A\|^2\|B\|^2 = (\|A\| + \|B\|)^2. \end{aligned}$$

Hence $\|A+B\| \leq \|A\| + \|B\|$.

(3) $$\begin{aligned} \|xA\| = \|x(a,b)\|^2 &= \|(xa,xb)\|^2 = (xa)^2 + (xb)^2 \\ &= x^2a^2 + x^2b^2 = x^2(a^2+b^2) = |x|^2(a^2+b^2) = |x|^2\|A\|^2 \end{aligned}$$

so $\|xA\| = \|x(a,b)\| = |x|\|(a,b)\| = |x|\|A\|$. $\square$

Exercise VEC.7 Proof. Let $A = (a_1,a_2)$, $B = (b_1,b_2)$, $C = (c_1,c_2)$, where a_1, a_2, b_1, b_2, c_1, and c_2 are real numbers.

(1) $A \bullet B = a_1b_1 + a_2b_2 = b_1a_1 + b_2a_2 = B \bullet A$.

(2) $t(A \bullet B) = t(a_1b_1 + a_2b_2)$,

$$\begin{aligned} (tA) \bullet B &= (t(a_1,a_2)) \bullet (b_1,b_2) = (ta_1,ta_2) \bullet (b_1,b_2) \\ &= t(a_1 \bullet b_1) + t(a_2 \bullet b_2) = t(a_1b_1 + a_2b_2), \\ A \bullet (tB) &= (a_1,a_2) \bullet (t(b_1,b_2)) = (a_1,a_2) \bullet (tb_1,tb_2) \\ &= a_1(tb_1) + a_2(tb_2) = t(a_1b_1 + a_2b_2). \end{aligned}$$

(3) $$\begin{aligned} (a_1,a_2) \bullet ((b_1,b_2) + (c_1,c_2)) &= (a_1,a_2)(b_1+c_1,b_2+c_2) \\ &= (a_1(b_1+c_1), a_2(b_2+c_2)) = (a_1b_1 + a_1c_1, a_2b_2 + a_2c_2) \\ &= (a_1,a_2) \bullet (b_1,b_2) + (a_1,a_2) \bullet (c_1,c_2). \end{aligned}$$

(4) $A \bullet A = a_1^2 + a_2^2 = \|A\|^2$. $\square$

Exercise VEC.8 Proof. Prove Theorem VEC.35 parts (A) through (D): for all A and B in $\mathbb{V}$, and all real numbers t and s,

(A) $$\begin{aligned} (\alpha+\beta)(A+B) &= \alpha(A+B) + \beta(A+B) \\ &= \alpha(A) + \alpha(B) + \beta(A) + \beta(B) \end{aligned}$$

$$= (\alpha(A) + \beta(A)) + (\alpha(B) + \beta(B))$$
$$= (\alpha + \beta)(A) + (\alpha + \beta)(B);$$

and

$$\begin{aligned}(\alpha + \beta)(tA) &= \alpha(tA) + \beta(tA) = t\alpha(A) + t\beta(A)\\ &= t(\alpha(A) + \beta(A)) = t(\alpha + \beta)(A).\end{aligned}$$

Thus $\alpha + \beta$ is linear.

$$\begin{aligned}(t\alpha)(A + B) &= t(\alpha(A + B)) = t(\alpha(A) + \alpha(B))\\ &= t(\alpha(A)) + t(\alpha(B)) = (t\alpha)(A) + (t\alpha)(B)\end{aligned}$$

and by associativity of scalar product,

$$\begin{aligned}(t\alpha)(sA) &= t(\alpha(sA)) = t(s\alpha(A)) = (ts)\alpha(A)\\ &= (st)\alpha(A) = s(t\alpha(A) = s(t\alpha)(A).\end{aligned}$$

Thus $t\alpha$ is linear.

(B) $O(A + B) = O = O(A) + O(B)$; $O(tA) = O = t \cdot O = t \cdot O(A)$.

$\imath(A + B) = A + B = \imath(A) + \imath(B)$; $\imath(tA) = tA = t \cdot \imath(A)$.

(C) (1) $\alpha(A + B) + (-\alpha(A)) + (-\alpha(B))$

$$= \alpha(A) + \alpha(B) + (-\alpha(A)) + (-\alpha(B))$$
$$= \alpha(A) + (-\alpha(A)) + \alpha(B) + (-\alpha(B)) = O$$

which shows that $-\alpha(A + B) = (-\alpha(A)) + (-\alpha(B))$.

(2) $\alpha(tA) + t((-\alpha)(A)) = \alpha(tA) + t(-\alpha(A)) = t(\alpha(A)) + t(-\alpha(A))$

$$= t(\alpha(A) + (-\alpha(A))) = t(O) = O,$$

which shows that $(-\alpha)(tA) = -(\alpha(tA)) = t((-\alpha)(A))$.

(D) (1) By property (A)(2) of Definition VEC.34,

$$\alpha(O) = \alpha(O \cdot A) = O \cdot \alpha(A) = O.$$

(2) $\alpha(A) + \alpha(-A) = \alpha(A + (-A)) = \alpha(O) = O$,

so $\alpha(-A) = -\alpha(A)$. □

Exercise VEC.9 Proof. Prove Theorem VEC.35(E). We key the various parts of this proof to the properties listed in Definition VEC.12. Let α, β, and γ be linear mappings on $\mathbb{V}$, and let t and s be any real numbers.

(A) By Exercise VEC.8(A), $\alpha + \beta$ and $t\alpha$ are linear mappings. Let A and B be any members of $\mathbb{V}$.

(1) $(\alpha + \beta)(A) = \alpha(A) + \beta(A) = \beta(A) + \alpha(A) = (\beta + \alpha)(A)$

so $\alpha + \beta = \beta + \alpha$.

$$\begin{aligned}(2)\ (\alpha + (\beta + \gamma))(A) &= \alpha(A) + (\beta + \gamma)(A) = \alpha(A) + (\beta(A) + \gamma(A))\\ &= (\alpha(A) + \beta(A)) + \gamma(A) = (\alpha + \beta)(A) + \gamma(A)\\ &= ((\alpha + \beta) + \gamma)(A)\end{aligned}$$

so $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

(3) O as defined in Definition VEC.34(B) is a linear mapping by Exercise VEC.6(B). Then $(\alpha + O)(A) = \alpha(A) + O(A) = \alpha(A) + O = \alpha(A)$ so $\alpha + O = \alpha$. If β is a linear mapping such that $\alpha + \beta = \alpha$, then for every $A \in \mathbb{V}$, $\alpha(A) + \beta(A) = \alpha(A)$ so that $\beta(A) = O$ and thus $\beta = O$.

(4) $-\alpha$ as defined in Definition VEC.34(C) is a linear mapping by Exercise VEC.6(C). Then

$$(\alpha + (-\alpha))(A) = \alpha(A) + (-\alpha)(A) = \alpha(A) + (-\alpha(A)) = O = O(A)$$

so that $\alpha + (-\alpha) = O$. If there is a linear mapping β such that $\alpha + \beta = O$, then $O = (\alpha + \beta)(A) = \alpha(A) + \beta(A)$ so that $\beta(A) = -\alpha(A)$ for every A and hence $\beta = -\alpha$.

(B) (1) $(t(s\alpha))(A)) = t((s\alpha)(A)) = t(s(\alpha(A))) = (ts)\alpha(A) = (ts\alpha)(A)$, so that $t(s\alpha) = (ts)\alpha$.

(2) $(1\alpha)(A) = 1(\alpha(A)) = \alpha(A)$ so $1\alpha = \alpha$.

(C) (1) $$\begin{aligned}(t(\alpha + \beta))(A) &= t(\alpha(A) + \beta(A)) = t(\alpha(A)) + t(\beta(A)) \\ &= (t\alpha)(A) + (t\beta)(A) = (t\alpha) + (t\beta))(A)\end{aligned}$$

so that $t(\alpha + \beta) = t\alpha + t\beta$.

(2)$(t + s)\alpha(A) = t(\alpha(A)) + s(\alpha(A))) = (t\alpha)(A) + (s\alpha)(A)$

so that $(t + s)\alpha = t\alpha + s\alpha$. $\square$

Exercise VEC.10 Proof. In this argument we will refer to the numbers $a_{11}, a_{22}, a_{12}, a_{21}$ as the "entries."

(Case 1) If 3 entries are 0 but a_{22} or a_{12} is nonzero, let $x_1 = 1$ and $x_2 = 0$; and the result $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$ will follow.

(Case 2) If 3 entries are 0 but a_{11} or a_{21} is nonzero, let $x_1 = 0$ and $x_2 = 1$; then $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$.

(Case 3) If 2 entries are 0 but a_{11} and a_{12} are nonzero, let $x_1 = a_{12}$ and $x_2 = -a_{11}$; then $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$.

(Case 4) If 2 entries are 0 but a_{21} and a_{22} are nonzero, let $x_1 = a_{22}$ and $x_2 = -a_{21}$; then $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$.

(Case 5) If 2 entries are 0 but a_{11} and a_{21} are nonzero, let $x_1 = 0$ and $x_2 = 1$; then $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$.

(Case 6) If 2 entries are 0 but a_{12} and a_{22} are nonzero, let $x_1 = 1$ and $x_2 = 0$; then $a_{11}x_1 + a_{12}x_2 = 0$ and $a_{21}x_1 + a_{22}x_2 = 0$.

(Case 7) If 2 entries are 0 but a_{11} and a_{22} are nonzero, then either a_{12} or a_{21} is zero, and this reduces to Case 1 or 2.

(Case 8) If 2 entries are 0 but a_{12} and a_{21} are nonzero, then either a_{11} or a_{22} is zero, and this reduces to Case 1 or 2.

(Case 9) If a single entry is zero and all others are non-zero, then by $a_{11}a_{22} - a_{12}a_{21} = 0$ one other entry must be zero, and this reduces to one of the Cases 3 through 8. $\square$

Exercise VEC.11 Proof. Let α be the linear mapping on $\mathbb{R}^2$ such that $\alpha(1,0) = (0,1)$ and $\alpha(0,1) = (1,0)$; and let β be the linear mapping on $\mathbb{R}^2$ such that $\beta(1,0) = (1,1)$ and $\beta(0,1) = (1,-1)$.

In the form of the mapping given in Definition VEC.37, for α, $a_{11} = 0$, $a_{12} = 1$, $a_{21} = 1$ and $a_{22} = 0$, so that the determinant is $0 \cdot 0 - 1 \cdot 1 = -1 \neq 0$. For β (using bs in place of as), $b_{11} = 1$, $b_{12} = 1$, $b_{21} = 1$ and $b_{22} = -1$, so that the determinant is $1 \cdot (-1) - 1 \cdot 1 = -2 \neq 0$. Thus both α and β are bijections by Theorem VEC.41.

Then $\beta(\alpha(1,0)) = \beta(0,1)) = (1,-1)$, and $\beta(\alpha(0,1)) = \beta(1,0)) = (1,1)$, but $\alpha(\beta(1,0)) = \alpha(1,1)) = (1,1)$, and $\alpha(\beta(0,1)) = \alpha(1,-1) = (-1,1)$ so that $\alpha \circ \beta \neq \beta \circ \alpha$. $\square$

Exercise VEC.12 Proof. By Definition VEC.20 $s = \frac{x_2 - q_2}{x_1 - q_1}$. Hence if $s > 0$, then $x_1 - q_1$ and $x_2 - q_2$ are both positive or both negative. Hence $q_1 < x_1$ and $q_2 < x_2$ or $q_1 > x_1$ and $q_2 > x_2$. If $s < 0$, then $x_2 - q_2$ is negative and $x_1 - q_1$ is positive, or $x_2 - q_2$ is positive and $x_1 - q_1$ is negative so that either $x_1 < q_1$ and $x_2 > q_2$ or $x_1 > q_1$ and $x_2 < q_2$. $\square$

Exercise VEC.13 Proof. Let the slope of $\mathcal{L}$ be s and the slope of $\mathcal{M}$ be t.

(I) If $\mathcal{L} \perp \mathcal{M}$, let $Q = (q_1, q_2)$ be their point of intersection, and let $\mathcal{N}$ be the vertical line through $(q_1 + 1, q_2)$. Then the point of intersection of $\mathcal{L}$ and $\mathcal{N}$ is $S = (q_1 + 1, q_2 + s)$ and the point of intersection of $\mathcal{M}$ and $\mathcal{N}$ is $T = (q_1 + 1, q_2 + t)$.

(II) Conversely, if $st = -1$, then s and t are of opposite parity (*i.e.* one is positive and the other is negative), and $s \neq t$. By Theorem VEC.20 $\mathcal{L}$ and $\mathcal{M}$ are not parallel, so must intersect at some point $Q = (q_1, q_2)$. Let $\mathcal{N}$ be the vertical line through $(q_1 + 1, q_2)$. Neither $\mathcal{L}$ nor $\mathcal{M}$ is vertical so there must be a point of intersection of $\mathcal{N}$ with each of them. The point of intersection of $\mathcal{L}$ and $\mathcal{N}$ is $S = (q_1 + 1, q_2 + s)$ and the point of intersection of $\mathcal{M}$ and $\mathcal{N}$ is $T = (q_1 + 1, q_2 + t)$, as before.

(III) The third form of the Pythagorean Theorem (Theorem VEC.26.5) says that $\angle SQT$ is right iff $\|S - T\|^2 = \|S - Q\|^2 + \|T - Q\|^2$.

By Theorem VEC.27,

$\|S - T\| = \sqrt{(q_1 + 1 - q_1 - 1)^2 + (q_2 + s - q_2 - t)^2} = \sqrt{(s - t)^2} = |s - t|,$

$\|S-Q\| = \sqrt{(q_1+1-q_1)^2+(q_2+s-q_2)^2} = \sqrt{1^2+s^2} = \sqrt{1+s^2}$, and
$\|T-Q\| = \sqrt{(q_1+1-q_1)^2+(q_2+t-q_2)^2} = \sqrt{1^2+t^2} = \sqrt{1+t^2} = 1$.

Hence $\|S-T\|^2 = \|S-Q\|^2 + \|T-Q\|^2$ iff $|s-t|^2 = 1+s^2+(1+t^2)$ iff $s-t^2 = s^2-2st+t^2 = 1+s^2+1+t^2$ iff $s^2-2st+t^2 = 1+s^2+1+t^2$ iff $-2st = 2$ iff $st = -1$. □

Exercise VEC.14 Proof. For reference, the equation of the line is $ax_1 + bx_2 + c = 0$.

(I) If $b = 0$, then $\mathcal{L} = \{(x_1,x_2) \mid x_1 = \frac{-c}{a}\}$. It thus contains the point $(\frac{-c}{a}, 0)$. By Theorem VEC.16 $\mathcal{L}$ is parallel to $\mathbb{L}_2$ and by Definition VEC.18 it is a vertical line.

(II) If $b \neq 0$, then the point $(0, \frac{-c}{b})$ is on $\mathcal{L}$, since $a \cdot 0 + b(\frac{-c}{b}) + c = 0$. $(1, \frac{-a-c}{b})$ is a point on the line since $a \cdot 1 + b(\frac{-a-c}{b}) + c = 0$. Thus by Definition VEC.20 the slope of $\mathcal{L}$ is

$$\frac{\frac{-a-c}{b} - \frac{-c}{b}}{1-0} = -\frac{a}{b},$$

so that $\mathcal{L}$ is the line through $(0, \frac{-c}{b})$ with slope $-\frac{a}{b}$. □

Exercise VEC.15 Proof. For reference, $\mathcal{L} = \{(x_1,x_2) \mid ax_1 + bx_2 + c = 0\}$ and $\mathcal{M} = \{(x_1,x_2) \mid bx_1 - ax_2 + c = 0\}$.

(Case 1: $a = 0$.) Then $\mathcal{L} = \{(x_1,x_2) \mid x_2 = \frac{-c}{b}\}$ and $\mathcal{M} = \{(x_1,x_2) \mid x_1 = \frac{-c}{b}\}$. By Theorem VEC.16(C), $\mathcal{L} \parallel \mathbb{L}_1$ and $\mathcal{M} \parallel \mathbb{L}_2$. By Definition VEC.18, $\mathcal{L}$ is horizontal and $\mathcal{M}$ is vertical, and by Remark VEC.18.1(B) $\mathcal{L} \perp \mathcal{M}$.

(Case 2: $b = 0$.) In this case $\mathcal{L} = \{(x_1,x_2) \mid x_1 = \frac{-c}{a}\}$ and $\mathcal{M} = \{(x_1,x_2) \mid x_2 = \frac{c}{a}\}$. By Theorem VEC.16(C), $\mathcal{L} \parallel \mathbb{L}_2$ and $\mathcal{M} \parallel \mathbb{L}_1$. By Definition VEC.18, $\mathcal{L}$ is vertical and $\mathcal{M}$ is horizontal, and by Remark VEC.18.1(B) $\mathcal{L} \perp \mathcal{M}$.

(Case 3: $a \neq 0$ and $b \neq 0$.) By Exercise VEC.14(II), the slope of $\mathcal{L}$ is $-\frac{a}{b}$. Replacing, in the argument of Exercise VEC.14(II), a with b and b with $-a$, we see that the slope of $\mathcal{M}$ is $\frac{b}{a}$. Thus, the product of the two slopes is -1, and by Exercise VEC.13, $\mathcal{L} \perp \mathcal{M}$. □

Exercise VEC.16 Proof. Let S and T be the collineations of $\mathbb{R} \times \mathbb{R}$ such that for all members of (x,y) of $\mathbb{R} \times \mathbb{R}$, $S(x,y) = \begin{bmatrix} 2x \\ \frac{y}{2} \end{bmatrix}$ and $T(x,y) = \begin{bmatrix} \frac{x}{2} \\ \frac{y}{2} \end{bmatrix}$. Since $\begin{vmatrix} 2-1 & 0 \\ 0 & \frac{1}{2}-1 \end{vmatrix} \neq 0$ and $\begin{vmatrix} \frac{1}{2}-1 & 0 \\ 0 & \frac{1}{2}-1 \end{vmatrix} \neq 0$, by Theorem VEC.51 S and

T each have only the fixed point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, whereas $(T \circ S)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ \frac{y}{4} \end{bmatrix}$, every member $\begin{bmatrix} x \\ 0 \end{bmatrix}$ of $\left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} \mid x \in \mathbb{R} \right\}$ is a fixed point of $T \circ S$. Moreover, $T \circ S \neq \imath$, so that the product $T \circ S$ is neither the identity nor a collineation with only $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ as a fixed point. Hence the set is not closed under composition, and is not a group. $\square$

Exercise VEC.17 Proof. Let r and t be real numbers different from 0 and S and T be the collineations of $\mathbb{R} \times \mathbb{R}$ such that for all $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R} \times \mathbb{R}$,

$$S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + ry \\ y \end{bmatrix} \text{ and } T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y + tx \end{bmatrix}.$$

Then $(T \circ S)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + ry \\ tx + (1 + tr)y \end{bmatrix}$. Note that every member of the horizontal axis $\{(x, 0) \mid x \in \mathbb{R}\}$ is a fixed point of S and every member of the vertical axis $\{(0, y) \mid y \in \mathbb{R}\}$ is a fixed point of T. However, since $\begin{vmatrix} 0 & r \\ t & 1 + tr \end{vmatrix} \neq 0$, by Theorem VEC.51, $T \circ S$ has only the fixed point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. $\square$

Chapter 2
The Field of Complex Numbers (CX)

Dependencies: *Euclidean Geometry and its Subgeometries (Specht); Chapter 1 of this supplement*
Acronym: *CX*
New terms defined: *product of points on the plane, complex number, purely imaginary, real (complex numbers); real and imaginary parts, modulus, absolute value, complex conjugate (of a complex number)*

In this chapter, $\mathbb{C}$ will be the plane $\mathbb{P}$ which, in the previous chapter (acronym VEC) was made into a vector space.

The only product defined in a general vector space is the product of a real number and a vector, that is, the *scalar* product. Products of vectors are not defined.[1] However, the plane (the archetypical two-dimensional vector space) is quite special among vector spaces, because a coherent and useful notion of product *can* be constructed on it, and this product makes it into a field—the field of complex numbers.

Multiplication operations have been defined on n-tuples of real numbers in the cases where $n = 4$ (*quaternions*) and $n = 8$ (*octonians*), in addition to ordinary vector space addition on these spaces. But these systems are not fields, as in them multiplication is not commutative. Attempts have been made to define a multiplication operation on n-tuples of real numbers for $n \geq 3$, so that the field properties hold. There are algebraic theorems which show that such attempts cannot succeed. (*cf* John L. Kelley, *Algebra: A Modern Introduction*, D. Van Nostrand, 1965 [2]).

[1] Except in the vector space $\mathbb{R}$ where the vector space and the field are the same.

This chapter is dependent on all of *Specht*, and includes references such as "Theorem ROT.15" which refer to items in that work. The note **Citations and references** at the end of the Preface to this Supplement explains the conventions we use for such citations, and an abbreviated Table of Contents (with acronyms) for *Specht* is included for the reader's guidance.

References in this chapter to items labeled VEC or CX are to this Supplement; all other references are to *Specht*.

2.1 Definitions and theorems for complex numbers

We now define the product of two points in $\mathbb{C}$ and show that with this definition, together with vector space addition, $\mathbb{C}$ is a field. We will use the symbol "$\cdot$" for this new product. With one notable exception, if A and B are two points of the plane $\mathbb{C}$ which are collinear with the origin O, their product $A \odot B$ as members of the line $\overleftrightarrow{OA}$ will not be the same as their product according to our new definition. Indeed, the two products are wildly different—the product $A \cdot B$ will in general not even belong to $\overleftrightarrow{OA}$. The exception is the case where $\overleftrightarrow{OA} = \mathbb{L}_1$, where the two products agree. The new product "$\cdot$" will be an extension to the whole plane of the product $\odot$ on $\mathbb{L}_1$. If we have occasion to discuss the ordered field multiplication of points on a single line (other than $\mathbb{L}_1$), we may use the notation $\odot$.

Definition CX.1 In this definition and the rest of the chapter, $\mathbb{C}$ will denote the Euclidean/LUB plane with origin O, $\mathbb{L}_1$ and $\mathbb{L}_2$ will be perpendicular lines in $\mathbb{C}$ such that $\mathbb{L}_1 \cap \mathbb{L}_2 = \{O\}$, which have been built into ordered fields (using the machinery of *Specht* Chapter 14) with U_1 and U_2, respectively, as their units, where $\overline{OU_1} \cong \overline{OU_2}$.

Addition of points in $\mathbb{C}$ is defined as in Definition VEC.1 and scalar product as in Definition VEC.6, so that $\mathbb{C}$ is a vector space under the operation $+$ and scalar product. The norm $\|A\|$ of a point A is as in Definition VEC.26.1, and every point $A = aU_1 + bU_2 \in \mathbb{C}$ is identified with the point $(a, b) \in \mathbb{R}^2$ using the vector space isomorphism λ, as in Definition VEC.14.

(A) Define the mapping ρ_A as follows:

(i) If $A \in \mathbb{C} \setminus \overleftrightarrow{OU_1}$ then ρ_A is the unique rotation of $\mathbb{C}$ about O such that $\rho_A(\overrightarrow{OU_1}) = \overrightarrow{OA}$. The existence and uniqueness of this rotation is guaranteed by *Specht* Ch.10 Theorem ROT.15.

(ii) If $A \in \overrightarrow{OU_1}$, then $\rho_A = \imath$ (the identity).

(iii) If $A \in \overrightarrow{OU_1'}$, where U_1'–O–U_1, then $\rho_A = \mathcal{R}_O$, where $\mathcal{R}_O$ is the point reflection about O (*cf* Definition ROT.1(B), *etc.*).

(B) Define the mapping δ_A as follows:

(i) If $A \in \mathbb{C} \setminus \overrightarrow{OU_1}$ then δ_A is the unique dilation of $\mathbb{C}$ with fixed point O such that $\delta_A(U_1) = {\rho_A}^{-1}(A)$.

This is equivalent to $\delta_A(\rho_A(U_1)) = A$, for by *Specht* Ch.13 Theorem DLN.7(E),

$$\delta_A(\rho_A(U_1)) = \rho_A(\delta_A(U_1)) = \rho_A({\rho_A}^{-1}(A)) = A.$$

A dilation cannot have two fixed points, so if ${\rho_A}^{-1}(A) = U_1$ (*i.e.* $A = \rho_A(U_1)$), define $\delta_A = \imath$.

(ii) If $A \in \overrightarrow{OU_1}$, then δ_A is the unique dilation of $\mathbb{C}$ with fixed point O such that $\delta_A(U_1) = A$. (A dilation cannot have two fixed points, so if $A = U_1$, define $\delta_A = \imath$.)

The existence and uniqueness of these dilations is guaranteed by Theorem DLN.7.

(C) Define the **product** $A \cdot B$ of A and B as follows:

(i) If $A = O$ or $B = O$ (or $A = O = B$), then $A \cdot B = O$.

(ii) If A and B are both members of $\mathbb{C} \setminus \{O\}$, then $A \cdot B = \delta_A(\rho_A(B))$.

The operation $\cdot$ on $\mathbb{C}$ is called **multiplication**.

(D) The points of the set $\mathbb{C}$, which has now been equipped with the operations $+$ and $\cdot$, are called **complex numbers**.

Remark CX.2 (A) Suppose $A \in \overrightarrow{OU_1'}$, where U_1'–O–U_1. By Definition CX.1(A)(iii), $\rho_A = \mathcal{R}_O$ and $\delta_A(U_1) = {\rho_A}^{-1}(A) = \mathcal{R}_O(A)$, because $\mathcal{R}_O^{-1} = \mathcal{R}_O$ (*cf* Corollary ROT.6).

(B) Applying Definition CX.1(C) to the line $\mathbb{L}_1 = \overleftrightarrow{OU_1}$ confirms that this definition agrees with $\odot$ from *Specht* Ch.14 Definition OF.1. For if $A \in \overrightarrow{OU_1}$ then $\rho_A = \imath$ and $\delta_A(\rho_A(U_1)) = \delta_A(\imath(U_1)) = A$ so that $\delta_A \circ \rho_A$ is the dilation δ_A of Definition OF.1(B). If $A \in \overrightarrow{OU_1'}$ then $\rho_A = \mathcal{R}_O$ and by part (A) of this remark, $\delta_A(U_1) = \mathcal{R}_O(A)$ so that

$$(\delta_A \circ \rho_A)(U_1) = \delta_A(\rho_A(U_1)) = \delta_A(\mathcal{R}_O(A)) = \mathcal{R}_O(\mathcal{R}_O(A)) = A$$

which agrees with the dilation δ_A as defined in Definition OF.1(B).

Therefore the present definition of $\cdot$ is an extension to the whole plane of the notion of product in Definition OF.1(D).

(C) Upon reflection, it will be seen that Definition CX.1 is the formalization in this context of the usual definition for products of complex numbers. Applying the rotation ρ_A to the point B is the same as "adding the angles of A and B"; applying the dilation δ_A "multiplies the modulii" of the points A and B, where the modulus of a complex number A is the length of the segment $\overline{OA}$, or the distance from O to A, as defined in Definition OF.16.

That is to say, applying δ_A to $\rho_A(B)$ "stretches" or "shrinks" $\rho_A(B)$ by the same ratio as δ_A "stretches" or "shrinks" U_1 to get A.

The case where $A \in \overrightarrow{OU_1'}$ where U_1'–O–U_1 can be a bit tricky to visualize. The rotation $\rho_A = \mathcal{R}_O$ maps the point B to $-B$, and then the dilation δ_A stretches that point in the same way that it stretches U_1 to $-A \in \overrightarrow{OU_1}$.

Theorem CX.3 *For any points A and B in $\mathbb{C} \setminus \{O\}$*
(A) $(\delta_A \circ \rho_A)(U_1) = A$,
(B) $A \cdot B = \delta_A(\rho_A(B)) = (\delta_A \circ \rho_A \circ \delta_B \circ \rho_B)(U_1)$, *and*
(C) $\delta_{A \cdot B} = \delta_A \circ \delta_B$ *and* $\rho_{A \cdot B} = \rho_A \circ \rho_B$.

Proof. (A) (I) If $A \in \overrightarrow{OU_1}$, then $\rho_A = \imath$, $\rho_A(U_1) = U_1$, and

$$\delta_A(\rho_A(U_1)) = \delta_A(U_1) = A.$$

(II) If $A \in \mathbb{C} \setminus \overrightarrow{OU_1}$, then by *Specht* Ch.13 Theorem DLN7(E) $\delta_A(\rho_A(U_1)) = \rho_A(\delta_A(U_1))$ and by Definition CX.1(B)(i) this is $\rho_A(\rho_A^{-1}(A)) = A$. Note that if $A \in \overrightarrow{OU_1'}$, where U_1'–O–U_1, this last calculation becomes $\mathcal{R}_O(\mathcal{R}_O(A)) = A$.

(B) By Definition CX.1(C) and part (A) applied to B, we have

$$A \cdot B = \delta_A(\rho_A(B)) = \delta_A(\rho_A(\delta_B(\rho_B(U_1)))) = (\delta_A \circ \rho_A \circ \delta_B \circ \rho_B)(U_1).$$

(C) In this part of the proof, when we say a mapping is a rotation* we mean that it is either a rotation about O or the identity $\imath$; when we say a mapping is a dilation* we mean that it is either a dilation with fixed point O or the identity $\imath$.

By Theorem DLN.20, the union of the set of all rotations about O and the set of all dilations with fixed point O, together with their compositions and the identity map $\imath$ comprises an abelian group under composition. Thus by Theorem DLN.20 and part (B) above, for any A and B in $\mathbb{C} \setminus \{O\}$,

$$A \cdot B = (\delta_A \circ \rho_A \circ \delta_B \circ \rho_B)(U_1) = (\delta \circ \rho)(U_1)$$

where ρ is a rotation* and δ is a dilation*.

By *Specht* Ch.3 Theorem CAP.18, $\overleftrightarrow{O(A \cdot B)}$ is a fixed line for δ^{-1}, so either $\rho(U_1) \in \overrightarrow{O(A \cdot B)}$ or $\rho(U_1) \in \overrightarrow{O(-A \cdot B)}$. If the latter is true, define δ' so that for every point $X \in \mathbb{C}$, $\delta'(X) = -\delta(X)$, and ρ' so that for every point $X \in \mathbb{C}$, $\rho'(X) = -\rho(X)$; then $\rho'(U_1) \in \overrightarrow{O(A \cdot B)}$. Then

$$\begin{aligned}(\delta' \circ \rho')(U_1) &= \delta'(-\rho(U_1)) = -\delta'(\rho(U_1)) \\ &= -(-\delta(\rho(U_1))) = \delta(\rho(U_1)) = A \cdot B.\end{aligned}$$

Hence in either case, there exists a dilation* δ and a rotation* ρ such that $(\delta \circ \rho)(U_1) = A \cdot B$, $\rho(U_1) \in \overrightarrow{O(A \cdot B)}$, and $\delta \circ \rho = (\delta_A \circ \delta_B) \circ (\rho_A \circ \rho_B)$. Both δ and $\delta_A \circ \delta_B$ are dilations* and both ρ and $\rho_A \circ \rho_B$ are rotations*. Both $\rho(U_1)$ and $(\rho_A \circ \rho_B)(U_1)$ are in $\overrightarrow{O(A \cdot B)}$ and $\overleftrightarrow{O\rho(U_1)} \cong \overleftrightarrow{OU_1} \cong \overleftrightarrow{O(\rho_A \circ \rho_B(U_1))}$ so by Property R.4 of *Specht* Ch.8 Definition NEUT.2, $\rho(U_1) = (\rho_A \circ \rho_B)(U_1)$ and thus $\rho = \rho_A \circ \rho_B$. ρ is therefore the rotation* that maps $\overrightarrow{OU_1}$ to $\overrightarrow{O(A \cdot B)}$, that is, $\rho = \rho_A \circ \rho_B = \rho_{A \cdot B}$.

Then δ is the dilation* that maps $\rho(U_1)$ to $A \cdot B$ and hence is the dilation* that maps U_1 to $\rho^{-1}(A \cdot B)$; by Theorem CAP.24 there is only one such dilation* and hence $\delta = \delta_A \circ \delta_B = \delta_{A \cdot B}$. $\square$

Theorem CX.4 *$\mathbb{C} \setminus \{O\}$ is an Abelian group under the operation "$\cdot$", where U_1 is the identity and for every $A \in \mathbb{C} \setminus \{O\}$ the inverse of A is the point $\delta_A^{-1}(\rho_A^{-1}(U_1))$.*

Proof. Let $\mathbb{G}$ be the union of the set of all rotations $\mathbb{C}$ about O, the set of all dilations of $\mathbb{C}$ with fixed point O, the set of all compositions of mappings in these two sets, and $\{\imath\}$.

By Theorem DLN.20 $\mathbb{G}$ is an Abelian group under the composition of mappings. The proof that $\mathbb{C} \setminus \{O\}$ is an Abelian group under the operation $\cdot$ consists of a series of calculations using Definition CX.1 and Theorem CX.3, and is left to the reader as Exercise CX.1. $\square$

Theorem CX.5 *Let φ be a collineation of the Euclidean/LUB plane $\mathbb{C}$ such that $\varphi(0) = 0$ and let S and T be members of $\mathbb{C}$. Then $\varphi(S + T) = \varphi(S) + \varphi(T)$.*

Proof. (Case 1: $S = O$.) By Definition CX.1 each side of the above equality is equal to $\varphi(T)$.

(Case 2: $S \neq O$.) τ_S is the translation of $\mathbb{C}$ such that $\tau_S(0) = S$. Then by Theorem CAP.13 $\varphi \circ \tau_S \circ \varphi^{-1}$ is a translation of $\mathbb{C}$. Since $(\varphi \circ \tau_S \circ \varphi^{-1})(O) = \varphi(S)$, $\varphi \circ \tau_S \circ \varphi^{-1} = \tau_{\varphi(S)}$ and so $\varphi \circ \tau_S = \tau_{\varphi(S)} \circ \varphi$ and thus $(\varphi \circ \tau_S)(T) = (\tau_{\varphi(S)} \circ \varphi)(T)$, i.e., $\varphi(\tau_S(T)) = \tau_{\varphi(S)}(\varphi(T))$, or $\varphi(S+T) = \varphi(S) + \varphi(T)$. $\square$

Theorem CX.6 (Distributive Property) *Let A, B, and C be members of $\mathbb{C}$. Then $A \cdot (B + C) = (A \cdot B) + (A \cdot C)$.*

Proof. Define $\varphi = \delta_A \circ \rho_A$ as defined in Definition CX.1(A) and (B). Then φ is a collineation with $\varphi(O) = O$, since both δ_A and ρ_A are collineations with fixed point O.

(Case 1: $A = O$.) By Definitions VEC.1 and CX.1 $A \cdot (B+C) = \varphi_A(B+C)$. By Theorem CX.5

$$\varphi_A(B+C) = \varphi_A(B) + \varphi_A(C) = A \cdot B + A \cdot C.$$

(Case 2: $A \neq O$.) Then $A \cdot (B+C) = \varphi_A(B+C)$. By Theorem CX.5

$$\varphi_A(B+C) = \varphi_A(B) + \varphi_A(C) = A \cdot B + A \cdot C. \quad \square$$

Theorem CX.7 *The Euclidean/LUB plane $\mathbb{C}$ under the operations of addition and multiplication is a field.*

Proof. This is a synthesis of Theorems VEC.3, CX.4, and CX.6. $\square$

Note that $\mathbb{C}$ is not an ordered field.

Theorem CX.8 *Let $\mathbb{C}$ be the complex plane, and (as in our overall assumptions) let $\mathbb{L}_1$ be a line in $\mathbb{C}$ with origin O which has been built into an ordered field with unit U_1, the multiplicative identity for $\mathbb{C}$. Let t be any real number, and let δ_t be as in Definition CX.13.2, that is, for all points $A \in \mathbb{L}_1$, $\delta_t(A) = tA$.*

(A) *For all points $A \in \mathbb{C}$, $\delta_t(A) = tA$ (not just for points $A \in \mathbb{L}_1$, as in Definition CX.13.2).*

(B) *If U is a point of $\mathbb{C} \setminus \{\mathbb{L}_1\}$ such that $\overline{OU_1} \cong \overline{OU}$, then for all real numbers $t \neq 0$, $\overline{OtU_1} \cong \overline{OtU}$.*

(C) *Let A be any point in $\mathbb{C} \setminus \{O\}$, and suppose the line $\mathcal{L} = \overleftrightarrow{OA}$ is built into an ordered field with origin O and unit U. Then $\delta_t(A) = tA = tU_1 \cdot A$.*

(D) *In particular, for any $A \in \mathbb{L}_2$, $\delta_t(A) = tA = tU_1 \cdot A$.*

Proof. (A) This is *Specht* Ch.18 Theorem REAL.37.

(B) Let ρ be the rotation of $\mathbb{C}$ such that $\rho(U_1) \in \overrightarrow{OU}$, so that ρ maps $\overleftrightarrow{OU_1}$ to $\overleftrightarrow{OU}$ and ρ^{-1} maps $\overleftrightarrow{OU}$ to $\overleftrightarrow{OU_1}$.

Since ρ is an isometry, $\overline{OU} \cong \overline{OU_1} \cong \overline{O\rho(U_1)}$ and by Property R.4 of *Specht* Ch.8 Definition NEUT.2, $\rho(U_1) = U$, that is, $U_1 = \rho^{-1}(U)$. By Theorem DLN.17 and Theorem NEUT.15,

$$\overline{O(\delta_t(U))} = \delta_t(\overline{OU}) \cong \delta_t(\overline{OU_1}) = \overline{O(\delta_t(U_1))}.$$

(C) It isn't possible to directly use Theorem REAL.25 to show that $tU_1 \cdot A = t(U_1 \cdot A) = tA$, because that theorem applies only to the product of points in a line $\mathbb{L}$ through O, whilst U_1 and A are points in different lines through the origin.

Let ρ be as defined in the proof of part (B). Then let $A' = \rho^{-1}(A)$, so that $\rho(A') = A$.

By Definition CX.1, the complex number product $tU_1 \cdot A = \delta_{tU_1}(\rho_{tU_1}(A))$ by Definition CX.1(C)

(Case 1: $t > 0$.) $tU_1 \in \overrightarrow{OU_1}$ so that by Definition CX.1(A) $\rho_{tU_1} = \imath$ and hence

$$
\begin{aligned}
tU_1 \cdot A &= \delta_{tU_1}(\rho_{tU_1}(A)) = \delta_{tU_1}(A) = \delta_{tU_1}(\rho(A')) \\
&= \rho(\delta_{tU_1}(A')) && \text{commutativity (Theorem DLN.7(E))} \\
&= \rho(tU_1 \odot A') && \text{by Definition OF.1(D) (both } A' \text{ and } tU_1 \in \mathbb{L}_1) \\
&= \rho(tU_1 \cdot A') && \cdot \text{ and } \odot \text{ are the same on } \mathbb{L}_1 \\
&= \rho(t(U_1 \cdot A')) && \text{by Theorem REAL.25 (both } A' \text{ and } tU_1 \in \mathbb{L}_1) \\
&= \rho(tA') && \text{by identity of } U_1 \\
&= \rho(\delta_t(A')) && \delta_t \text{ as in Definition REAL.38} \\
&= \delta_t(\rho(A')) && \text{commutativity (Theorem DLN.7(E))} \\
&= \delta_t(A) && \text{by definition} \\
&= tA && \text{Theorem REAL.37 applies } \delta_t \text{ to all points.}
\end{aligned}
$$

(Case 2: $t < 0$.) $-t = (-1)t$, so by Case 1, $((-1)t)U_1 \cdot A = ((-1)t)A$. The left-hand side of this equality is $((-1)(tU_1) \cdot A = (-1)(tU_1 \cdot A)$ by Theorems REAL.23 and REAL.25. The right-hand side is $(-1)(tA)$ by Theorem REAL.25, so we have $(-1)(tU_1 \cdot A) = (-1)(tA)$; multiplying both sides by -1 and using Theorem REAL.25 and arithmetic, we have $1(tU_1 \cdot A) = 1(tA)$ which is $tU_1 \cdot A = tA$ by Theorem VEC.7(E).

(D) This is part (A) where $\mathcal{L} = \mathbb{L}_{\in}$. $\square$

Remark CX.9 Theorem CX.8 shows that the comparative scales for scalar multiplication in $\mathbb{L}_1$ and another line through the origin are the same if the segments from O to their respective units are congruent. In particular this is true for $\mathbb{L}_2$. Moreover, scalar multiplication tA agrees with complex number multiplication $tU_1 \cdot A$.

Theorem CX.9 *If A and B are distinct members of $\mathbb{C} \setminus \overleftrightarrow{OU_1}$, then $\triangle OU_1B$ is similar to $\triangle OA(B \cdot A)$.*

Proof. By Definition CX.1 and the fact that a rotation is an isometry, $\angle U_1OB \cong \angle AO(B \cdot A)$ (*cf* Theorem NEUT.15). By Theorem DLN.14, dilations preserve angles so that $\angle OU_1B \cong \angle OA(B \cdot A)$. By Definition

SIM.7, $[\overline{OU_1}][\overline{OA}] = [\overline{OA}]$ so that $[\overline{OA}] = [\overline{OA}] \oplus [\overline{OU_1}]$; also $[\overline{OA}][\overline{OB}] = [\overline{O(A \cdot B)}]$ so that $[\overline{OA}] = [\overline{O(A \cdot B)}] \oplus [\overline{OB}]$. Therefore the ratios of corresponding edges of $\triangle OU_1B \sim \triangle OA(B \cdot A)$ are the same. By Theorem SIM.18 $\triangle OU_1B \sim \triangle OA(B \cdot A)$. $\square$

2.2 Computation with complex numbers

In this section, $\mathbb{C}$ is the complex plane, which is equipped with origin O and operations $+$ and $\cdot$. $\mathbb{L}_1$ and $\mathbb{L}_2$ are perpendicular lines on $\mathcal{P}$ intersecting at O, which are built into ordered fields with units U_1 and U_2 for $\mathbb{L}_1$ and $\mathbb{L}_2$, respectively. We assume that U_2 is chosen so that $\overline{OU_2} \cong \overline{OU_1}$. By Theorem CX.4 U_1 is the multiplicative identity for $\mathbb{C}$. By Theorem VEC.8 every complex number Z can be written as $Z = xU_1 + yU_2$ where x and y are uniquely determined real numbers. $+$ will denote addition, and $\cdot$ will denote multiplication of two complex numbers.

Definition CX.10 (A) If $Z = xU_1 + yU_2$ is any complex number, x is said to be the **real part** of Z, and y is the **imaginary part** of Z.

(B) A complex number Z is said to be **real** iff $Z \in \mathbb{L}_1$ and hence $Z = xU_1$; it is **(purely) imaginary** iff $Z \in \mathbb{L}_2$ and hence $Z = yU_2$.

Theorem CX.11

(A) *If x and y are any real numbers, then $xU_1 + yU_2 = O$ iff $x = y = 0$.*

(B) *For all real numbers x and y, $xU_1 \cdot yU_1 = (xy)U_1$.*

(C) *For all real numbers x and y, $xU_2 \cdot yU_2 = (xy)(U_2 \cdot U_2)$.*

(D) *For all real numbers x and y, $xU_1 \cdot yU_2 = yU_2 \cdot xU_1 = (xy)U_2$.*

(E) *$U_2 \cdot U_2 = -U_1$.*

(F) *$0U_1 = 0U_2 = O$.*

Proof. (A) This is Theorem VEC.8(B).

(B) Since $\cdot$ is the same as $\odot$ on $\mathbb{L}_1$, this follows from two applications of Theorem REAL.23 and one of Theorem REAL.25 as follows:

$$\begin{aligned} xU_1 \cdot yU_1 &= xU_1 \odot yU_1 = x(U_1 \odot yU_1) = x(yU_1 \odot U_1) \\ &= x(y(U_1 \odot U_1)) = (xy)(U_1 \odot U_1) = (xy)U_1. \end{aligned}$$

(C) By Theorem CX.8, $xU_2 = xU_1 \cdot U_2$ and $yU_2 = yU_1 \cdot U_2$. Then by part (C) and commutativity,

$$\begin{aligned} xU_2 \cdot yU_2 &= xU_1 \cdot U_2 \cdot yU_1 \cdot U_2 = xU_1 \cdot yU_1 \cdot U_2 \cdot U_2 \\ &= (xy)U_1 \cdot U_2 \cdot U_2 = (xy)(U_2 \cdot U_2), \end{aligned}$$

using Theorem CX.8 once again.

(D) By Theorem CX.8, $yU_2 = yU_1 \cdot U_2$. Using part (B),

$$xU_1 \cdot yU_2 = xU_1 \cdot yU_1 \cdot U_2 = (xy)U_1 \cdot U_2 = (xy)U_2.$$

(E) By Definition CX.1, since ρ_{U_2} is the rotation that maps U_1 to U_2, $\delta_{U_2} = \imath$. By Theorem ROT.15(A), $\rho_{U_2} = \mathcal{R}_{\mathcal{M}} \circ \mathcal{R}_{\mathbb{L}_1}$, where $\mathcal{M}$ is the line of symmetry of $\angle U_1OU_2$ (and also of its vertical angle $\angle(-U_1)O(-U_2)$). Then $\mathcal{R}_{\mathbb{L}_1}(U_2) \in \overrightarrow{O(-U_2)}$ and $\mathcal{R}_{\mathcal{M}}(-U_2) \in \overrightarrow{O(-U_1)}$. Hence

$$U_2 \cdot U_2 = \rho_{U_2}(U_2) = \mathcal{R}_{\mathcal{M}}(\mathcal{R}_{\mathbb{L}_1}(U_2)) \in \overrightarrow{O(-U_1)}.$$

The mapping ρ_{U_2}, and hence also the mapping $\rho_{U_2} \circ \rho_{U_2}$ is an isometry so $\overline{O(U_2 \cdot U_2)} = \overline{O(\rho_{U_2}(\rho_{U_2}(U_1)))} \cong \overline{OU_1}$. Also $\mathcal{R}_O(U_1) = -U_1$, so that $\overline{OU_1} \cong \overline{O(-U_1)}$ and $\overline{O(U_2 \cdot U_2)} \cong \overline{O(-U_1)}$. By Property R.4 of Definition NEUT.2, $U_2 \cdot U_2 = -U_1$.

(F) Immediate from Definition REAL.19(A)(1). $\square$

Remark/Definition CX.12 (A) With the rules in Theorem CX.11 just above, together with the rules for addition and the distributive law which were developed earlier in the chapter, we have everything we need to do complex number arithmetic. At this point we abandon the use of capital letters A, B, X, and the like for complex numbers; henceforth we will use lower case letters. We will usually reserve the later letters of the alphabet (z and w in particular) for non-real complex numbers and use the earlier letters for real numbers.

(B) We abandon O as our designation of the origin, in favor of the number zero (0).

(C) We will use 1 for U_1, the unit in $\mathbb{L}_1$, which is the multiplicative identity for $\mathbb{C}$. This gives us the freedom to write $xU_1 \cdot yU_1$ as simply xy.

(D) We will write i in place of U_2. Be sure not to confuse the symbol i with the symbol $\imath$ for the identity.

Thus, instead of writing a complex number as $Z = xU_1 + yU_2$, we will write $z = x + yi$. It is also quite legitimate to write a complex number $Z = xU_1 + yU_2$ as the ordered pair (x, y).

Theorem CX.13 (Restatement of Theorem CX.12) *Using the notation just introduced, the statement that every complex number Z can be written as $Z = xU_1 + yU_2$ where x and y are uniquely determined real num-*

bers becomes "every complex number z can be written as $z = x + yi$ where x and y are uniquely determined real numbers".

(A) *If x and y are any real numbers, then $x + yi = 0$ iff $x = y = 0$.*

(B) *For all real numbers x and y, $xy = xy$!!*[2]

(C) *For all real numbers x and y, $(xi)(yi) = (xy)i^2$.*

(D) *For all real numbers x and y, $x(yi) = (xy)i$.*

(E) $ii = i^2 = -1$.

(F) $0i = 0$.

Theorem CX.14 (Computation involving complex numbers.) *Let a, b, c, and d be real numbers, so that $a+bi$ and $c+di$ are complex numbers. Then*

(A) $(a + bi) + (c + di) = (a + c) + (b + d)i$.

(B) $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$.

(C) $(a + bi)(a - bi) = a^2 + b^2$.

(D) *If $c + di \neq 0$, then* $\frac{a+bi}{c+di} = \frac{ac+bd}{c^2+d^2} + \frac{-ad+bc}{c^2+d^2}i$.

Proof. We use the various rules of Theorem CX.13 as well as the field properties of $\mathbb{C}$.

(A) Follows immediately from commutativity of addition.

(B) $(a + bi)(c + di) = ac + bdi^2 + adi + bci = (ac - bd) + (ad + bc)i$.

(C) $(a + bi)(a - bi) = a^2 - b^2 i^2 = a^2 + b^2$.

(D) $\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{ac-bdi^2-adi+bci}{c^2+d^2} = \frac{ac+bd}{c^2+d^2} + \frac{-ad+bc}{c^2+d^2}i$. □

Definition CX.15 (A) The **modulus** or **absolute value** of a complex number $z = a+bi$ is its norm $\|z\|$, which is written in the context of complex numbers, as $|z|$.

(B) If $z = a + bi$ is any complex number, $\bar{z} = a - bi$ is its **complex conjugate**.

Remark CX.16 (A) Recalling Definition VEC.6, if $z = a+bi$ is a complex number and t is a real number, the *scalar product* $tz = ta + tbi$.

(B) By Theorem VEC.27(A), for any complex number $z = a + bi$, $|z| = \sqrt{a^2 + b^2}$.

(C) For any complex number $z = a + bi$,

$$z\bar{z} = (a + bi)(a - bi) = a^2 + abi - bai - b^2 i^2 = a^2 - b^2(-1) = a^+b^2 = |z|^2.$$

[2] Notation can obscure what is really going on, but in this case it's ok, since we have proved already that $xU_1 \cdot yU_1 = (xy)U_1$.

(D) $a + bi = 0 + 0i$ iff $|a + bi| = 0$. For if $|a + bi| = a^2 + b^2 = 0$ then since both $a^2 \geq 0$ and $b^2 \geq 0$, both must be zero. Here *Specht* Ch.14 Theorem OF.10(C) is used twice.

(E) Notice that we have two uses of the symbol $|z|$; if z is a real number, Definition OF.13(B) applies. In this case $|z| = z$ if $z > 0$ and $|z| = -z$ if $z < 0$. If z is a non-real complex number, Definition CX.15(A) applies. In either case, $|z|$ is the *length* of the segment $\overleftrightarrow{0z}$, that is to say, the *distance* from 0 to z. The two definitions agree if z is a real number.

Some authors strongly prefer the terms **magnitude** or **modulus** rather than "absolute value" for complex numbers since the common method of finding the absolute value (as defined in Definition OF.13) of a real number doesn't work for complex numbers. This avoids the confusion possible when the same symbol is used for two different definitions. However, the term absolute value has become entrenched.

Theorem CX.17 *Let a, b, c, and d be real numbers. Then*

$$|(a + bi)(c + di)| = |a + bi||c + di|.$$

That is, the absolute value of the product of two complex numbers is the product of their absolute values.

Proof.
$$\begin{aligned} |(a+bi)(c+di)| &= |(ac-bd)+(ad+bc)i| \\ &= \sqrt{(ac-bd)^2+(ad+bc)^2} \\ &= \sqrt{a^2c^2-2acbd+b^2d^2+a^2d^2+2adbc+b^2c^2} \\ &= \sqrt{a^2c^2+a^2d^2+b^2c^2+b^2d^2} \\ &= \sqrt{a^2(c^2+d^2)+b^2(c^2+d^2)} \\ &= \sqrt{(a^2+b^2)(c^2+d^2)} \\ &= \sqrt{a^2+b^2}\sqrt{c^2+d^2} = |a+bi||c+di|. \quad \square \end{aligned}$$

Corollary CX.18 Let $a, b, c,$ and d be real numbers, then $(a+bi)(c+di) = 0+0i$ iff at least one of the complex numbers $a+bi$ or $c+di$ is the zero complex number $0 + 0i$.

Proof. The "if" part is immediate by definition, so we prove the converse. If $|a + bi||c + di| = 0$, by Exercise CX.3 either $|a + bi| = 0$ or $|c + di| = 0$ (or both). It is valid to use Theorem OF.10 since we are using only properties of real numbers here. By Remark CX.16(D), either $a = b = 0$ or $c = d = 0$, or both. $\square$

2.3 Exercises for complex numbers

Exercise CX.1* Complete the computations necessary to prove Theorem CX.4.

Exercise CX.2* Let X be any nonzero complex number and let Y be any member of $\overleftrightarrow{OX}$, then there exists a unique real number t such that $Y = tX$ (*cf* Definition CX.7).

Exercise CX.3* Let A and B be complex numbers. $A \odot B = O$ iff $A = O$ or $B = O$ (or both A and B are equal to zero).

Exercise CX.4 Following the lead of the last sentence in Remark CX.14, rewrite the conclusions of Theorem CX.15, using the notation (x, y) in place of $x + yi$.

2.4 Selected answers for complex numbers

Exercise CX.1 Proof. Let A, B, and C be any members of $\mathbb{C} \setminus \{O\}$. By Definition CX.1, $A \cdot B \in \mathbb{C} \setminus \{O\}$, so this set is closed under the operation.

(I) Associativity.

$$\begin{aligned}
A \cdot (B \cdot C) &= (A \cdot (\delta_B \circ \rho_B \circ \delta_C \circ \rho_C)(U_1)) && \text{by Theorem CX.3(B)} \\
&= \delta_A \circ \rho_A((\delta_B \circ \rho_B \circ \delta_C \circ \rho_C)(U_1)) && \text{Definition CX.1(C)} \\
&= (\delta_A \circ \rho_A) \circ (\delta_B \circ \rho_B \circ \delta_C \circ \rho_C)(U_1) && \\
&= (\delta_A \circ \rho_A \circ \delta_B \circ \rho_B) \circ (\delta_C \circ \rho_C)(U_1) && \text{associativity of bijections} \\
&= (\delta_A \circ \delta_B \circ \rho_A \circ \rho_B) \circ (\delta_C \circ \rho_C)(U_1) && \text{Theorem DLN.7(E)} \\
&= (\delta_{A \cdot B} \circ \rho_{A \cdot B}) \circ (\delta_C \circ \rho_C)(U_1) && \text{Theorem CX.3(C)} \\
&= (\delta_{A \cdot B} \circ \rho_{A \cdot B})((\delta_C \circ \rho_C)(U_1)) && \\
&= (A \cdot B) \cdot C && \text{Definition CX.1(C)}
\end{aligned}$$

(II) Identity. By Definition CX.1, $A \cdot U_1 = \delta_A \circ \rho_A(U_1) = A$, so that U_1 is the multiplicative identity.

(III) Commutativity. By Definition CX.1, and Theorem DLN.22 (commutativity of $\mathbb{G}$)

$$\begin{aligned}
A \cdot B &= \delta_A \circ \rho_A(\delta_B \circ \rho_B(U_1)) = (\delta_A \circ \rho_A \circ \delta_B \circ \rho_B)(U_1) \\
&= (\delta_B \circ \rho_B \circ \delta_A \circ \rho_A \circ)(U_1) = \delta_B \circ \rho_B(\delta_A \circ \rho_A \circ (U_1) = B \cdot A.
\end{aligned}$$

(IV) Inverses. Let $A^{-1} = \delta_A^{-1}(\rho_A^{-1}(U_1))$. Then by Theorem DLN.20(A) (commutativity of $\mathbb{G}$)

$$\begin{aligned} A \cdot A^{-1} &= \delta_A \circ \rho_A((\delta_A^{-1} \circ \rho_A^{-1})(U_1)) \\ &= (\delta_A \circ \delta_A^{-1} \circ \rho_A \circ \rho_A^{-1})(U_1) = \imath \circ \imath(U_1) = U_1. \quad \square \end{aligned}$$

Exercise CX.2 Proof. This is Corollary REAL.35.1. $\square$

Exercise CX.3 Proof. The proof of Exercise OF.10(H) is valid here since that proof uses only field properties and does not involve any properties of the relation $<$. $\square$

Chapter 3
Arc Length (ARC)

Dependencies: *Euclidean Geometry and its Subgeometries (Specht)*
Acronym:*ARC*
New terms defined: *arc, closed arc, rectifiable arc, arc length, summation, function of bounded variation, total variation*

This chapter defines rectifiable arcs, arc length, and functions of bounded variation. The following Chapter 4 defines the circular functions sin and cos and develops their properties, starting from an antiderivative of the function $f(x) = \frac{2}{1+x^2}$. Chapter 5 defines angle measure in terms of the length of arc of a unit circle, and proves several interesting results using this definition.

To understand these chapters the reader needs some background in calculus and analysis: we assume familiarity with limits, $\epsilon - \delta$ limit proofs, continuity and uniform continuity of functions, the mean value theorem, as well as Riemann sums and integrals.

Here we will be working with $\mathbb{R}^2$, the Cartesian coordinate plane. In *Specht* Ch.21 Theorem LC.44 we summarized a proof that all the axioms in that work hold for $\mathbb{R}^2$. Therefore, in the present chapter and the two that follow, we are free to use all the theorems and definitions from *Specht*.

For an explanation of our conventions for citations of items in *Specht*, we refer the reader to the note **Citations and references** at the end of the Preface to this Supplement, and to the abbreviated Table of Contents (with acronyms) included there.

In this chapter references to items labeled ARC will be to this Supplement; all other references are to *Specht*.

On notation: If a and b are real numbers, we will denote $\overset{[\leftrightarrow]}{ab}$ by $[a,b]$, $\overset{]\leftrightarrow[}{ab}$ by $]a,b[$, $\overset{[\leftrightarrow[}{ab}$ by $[a,b[$, and $\overset{]\leftrightarrow]}{ab}$ by $]a,b]$. If f is a function defined on real numbers, we may denote the set $\{f(x) \mid x \in [a,b]\}$ by $f[a,b]$ instead of the more formally correct symbol $f([a,b])$. If m and n are integers and $m < n$, we use the symbol $[m;n]$ to denote the set $\{k \mid k \text{ is an integer and } m \leq k \leq n\}$.

3.1 Definitions and theorems for arc length

Definition ARC.1 (I) A subset $\mathcal{C}$ of $\mathbb{R}^2$ is an **arc** iff a and b are real numbers such that $a \leq b$, and there exists a mapping f of $\overset{[\leftrightarrow]}{ab} = [a,b]$ into $\mathbb{R}^2$ such that

(a) $\mathcal{C} = f[a,b]$,
(b) f is one-to-one,
(c) f is continuous on $[a,b]$, and
(d) f^{-1} is continuous on $f[a,b]$.

A mapping with these properties is a **homeomorphism.** If $a = b$, the arc is the **trivial** arc, consisting of a single point. It is understood that an arc does not intersect itself.

A subset $\mathcal{C}$ of $\mathbb{R}^2$ is a **closed arc** iff a and b are real numbers such that $a < b$, and there exists a mapping f of $\overset{[\leftrightarrow]}{ab} = [a,b]$ into $\mathbb{R}^2$ such that (a) $\mathcal{C} = f[a,b]$, (b) the restriction of f to $[a,b[$ is one-to-one, (c) f is continuous on $[a,b]$, and (d) f^{-1} is continuous on $f[a,b[$, and (e) $f(a) = f(b)$. A *closed* arc intersects itself only at its endpoints.

If $\mathcal{C}$ and f are as defined above, we will say that $\mathcal{C}$ is an **arc generated by the function** f or an **arc generated by the function** f **over** $[a,b]$, or a **closed arc generated by the function** f **over** $[a,b]$**.**

If $f[a,b]$ is a closed arc generated by the function f over $[a,b]$, and if $a < x < b$, then both $f[a,x]$ and $f[x,b]$ are arcs (not closed).

If we need a symbol for an arc without reference to the function which generates it, we will generally use the the symbol $\mathcal{C}$ rather than $\mathcal{A}$, since $\mathcal{C}$ suggests the word *curve*, the name popularly given to what we have defined as *arc*.

(II) A **partition** of $[a,b]$ is a finite subset $\mathcal{P}[a,b] = \{a = x_0, x_1, ..., x_n = b\}$ of $[a,b]$ such that

(A) $n \geq 1$, so that $\mathcal{P}[a,b]$ contains at least the two elements a and b, and
(B) $a = x_0 < x_1 < ... < x_n = b$, that is, for every $k \in [1;n]$, $t_{k-1} < t_k$.

We will find it convenient at times to summarize the definition of partition by writing

$$\mathcal{P}[a,b] = \{a = x_0 < x_1 < ... < x_n = b\}.$$

A partition $\mathcal{P}[a,b]$ containing only the two elements a and b will be referred to as the *trivial* partition. Within a given argument, if it is clear what interval is being partitioned, we may write $\mathcal{P}[a,b]$ simply as $\mathcal{P}$. Note that this use of the word "partition" is related to, but not the same as its use in *Specht* Chapter 1, Section 1.4.

(III) If $\mathcal{P}_1[a,b]$ and $\mathcal{P}_2[a,b]$ are partitions of $[a,b]$, then $\mathcal{P}_2[a,b]$ is a **refinement** of $\mathcal{P}_2[a,b]$ iff $\mathcal{P}_1[a,b]$ is a proper subset of $\mathcal{P}_2[a,b]$.

(IV) Let f be a function with values in $\mathbb{R}$ or $\mathbb{R}^2$, defined on the interval $[a,b]$, and let $\mathcal{P}$ be a partition of $[a,b]$. The **summation of f over the partition $\mathcal{P}$** is

$$\mathcal{S}_{\mathcal{P}}(f) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|.$$

If it is desired to emphasize the domain of f we may write the sum as

$$\mathcal{S}_{\mathcal{P}[a,b]}(f) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|.$$

For a visualization see Figure 3.1.

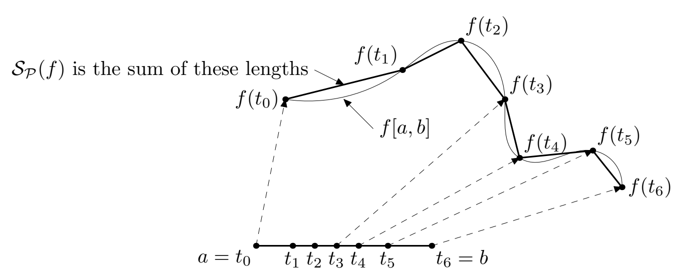


Fig. 3.1 Showing construction of the summation of f over the partition $\mathcal{P} = \{a = t_0 < t_1 < ... < t_6 = b\}$

(V) Let $\mathcal{P}[a,b] = \{a = t_0 < t_1 < ... < t_n = b\}$ be a partition of $[a,b]$. The **gauge** of $\mathcal{P}[a,b]$ is the number

$$\max\{t_k - t_{k-1} \mid k \in [1;n]\} = \max\{|t_k - t_{k-1}| \mid k \in [1;n]\}.$$

(VI) The arc (or closed arc) $f[a,b]$, as well as the mapping f, is said to be **rectifiable** iff there exists a positive number h such that for every partition $\mathcal{P}[a,b]$ of $[a,b]$, $\mathcal{S}_{\mathcal{P}}(f) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| < h$.

(VII) If the arc (or closed arc) $f[a,b]$ is rectifiable, then its **arc length**, or simply its **length**, is

$$\mathbb{L}(f[a,b]) = \text{lub}\{\mathcal{S}_{\mathcal{P}}(f) \mid \mathcal{P} \text{ is a partition of } [a,b]\}.$$

Remark ARC.2 (A) The arc length $\mathbb{L}(f[a,b])$ of a rectifiable arc $f[a,b]$ in $\mathbb{R}^2$ is always defined because the set of all sums $\mathcal{S}_{\mathcal{P}}(f)$, where $\mathcal{P}$ is a partition of $[a,b]$, is a set of real numbers which is bounded above, and such sets always have a least upper bound.

(B) If $a = b$, ($f[a,b]$ is the trivial arc) the only possible sum $\mathcal{S}_{\mathcal{P}}(f)$ is $|f(a) - f(b)| = 0$ so $\mathbb{L}(f[a,b]) = 0$. If $a \neq b$ then since f is one-to-one, $|f(a) - f(b)| \neq 0$ is a sum $\mathcal{S}_{\mathcal{P}}(f)$, and since the arc length is the least upper bound of all such sums, $\mathbb{L}(f[a,b]) > 0$.

(C) In Definition ARC.1(I), the definition of an arc (not closed) does not need to declare that the inverse of f is continuous; this is a consequence of a theorem from general topology which states that any one-to-one continuous function defined on a compact domain has a continuous inverse.

Theorem ARC.3 *Let $f[a,b]$ be an arc generated by the function f over $[a,b]$. If $\mathcal{P}_2$ and $\mathcal{P}_1$ are partitions of $[a,b]$, and $\mathcal{P}_2$ is a refinement of $\mathcal{P}_1$, then $\mathcal{S}_{\mathcal{P}_1}(f) \leq \mathcal{S}_{\mathcal{P}_2}(f)$.*

Proof. By Definition ARC.1(III), $\mathcal{P}_1 \subseteq \mathcal{P}_2$ and $\mathcal{P}_1 \neq \mathcal{P}_2$. Then there exists a natural number m and a sequence

$$\mathcal{P}_1 = \mathcal{Q}_1 \subseteq \mathcal{Q}_2 \subseteq ... \subseteq \mathcal{Q}_m = \mathcal{P}_2$$

of partitions of $[a,b]$ such that for every k with $1 < k \leq m$, there exists a real number t such that

$$\mathcal{Q}_k \setminus \mathcal{Q}_{k-1} = \{t\},$$

where t is a singleton, both x_{p-1} and x_p belong to $\mathcal{Q}_{k-1}$, and $x_{p-1} < t < x_p$. That is to say, each of the partitions $\mathcal{Q}_k$ contains one more point than its predecessor $\mathcal{Q}_{k-1}$.

Thus, it suffices to show that if $\mathcal{P}[a,b] = \{a = x_0 < x_1 < ... < x_n = b\}$ is a partition of $[a,b]$ and t is a real number such that $x_{k-1} < t < x_k$, the partition $\mathcal{Q}[a,b] = \mathcal{P}[a,b] \cup \{t\}$ satisfies

$$\mathcal{S}_{\mathcal{P}}(f) \leq \mathcal{S}_{\mathcal{Q}}(f).$$

This follows immediately from

$$|f(x_{k-1}) - f(x_k)| \leq |f(x_{k-1}) - f(t)| + |f(t) - f(x_k)|,$$

which is true by the triangle inequality for $\mathbb{R}^2$. $\square$

If $\mathcal{C}$ is a closed arc, for $\mathcal{S}_{\mathcal{P}}(f) \neq 0$ it is necessary for $\mathcal{P}$ to contain at least three points. Theorem ARC.3, however, is true in the case of a closed arc.

Theorem ARC.4 *Suppose that a and b are distinct real numbers, where $a < b$, and that $f[a,b]$ is an arc (or closed arc) generated by the function f over $[a,b]$.*

(A) *If $[d,e]$ is a proper subsegment of $[a,b]$, that is, $d < e$ and both d and e are members of $[a,b]$, but $[a,b] \neq [d,e]$, then if $f[a,b]$ is rectifiable, $f[d,e]$ is rectifiable.*

(B) *If c is a real number such that $a < c < b$,*

$$\mathbb{L}(f[a,b]) = \mathbb{L}(f[a,c]) + \mathbb{L}(f[c,b]).$$

Proof. (A) Let $\mathcal{P}[d,e] = \{d = x_0 < x_1 < ... < x_n = e\}$ be a partition of $[d,e]$. Then $\mathcal{Q}[a,b] = \mathcal{P}[d,e] \cup \{a,b\}$ is a partition of $[a,b]$;

(Case 1:) if $a \neq d$ and $b \neq e$ then

$$\mathcal{Q}[a,b] = \{a < d = x_0 < x_1 < ... < x_n = e < b\};$$

(Case 2:) if $a = d$ then $\mathcal{Q}[a,b] = \{a = d = x_0 < x_1 < ... < x_n = e < b\}$;

(Case 3:) if $b = e$ then $\mathcal{Q}[a,b] = \{a < d = x_0 < x_1 < ... < x_n = e = b\}$;

Since $f[a,b]$ is rectifiable, there exists a number $h > 0$ such that for every partition $\mathcal{R}[a,b]$ of $[a,b]$, $\mathcal{S}_{\mathcal{R}}(f) < h$; thus, in particular, $\mathcal{S}_{\mathcal{Q}}(f) < h$.
In Case 1, this means that (noting that $f(x_0) = f(d)$ and $f(x_n) = f(e)$)

$$\begin{aligned} h > \mathcal{S}_{\mathcal{Q}}(f) &= \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| + |f(x_0) - f(a)| + |f(b) - f(x_n)| \\ &\geq \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| = \mathcal{S}_{\mathcal{P}}(f), \end{aligned}$$

so that $\mathcal{S}_{\mathcal{P}}(f) < h$. Since $\mathcal{P}$ was initially chosen to be an arbitrary partition of $[d,e]$, this means that $f[d,e]$ is rectifiable.

Likewise, in Case 2,

$$h > \mathcal{S}_{\mathcal{Q}}(f) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| + |f(b) - f(x_n)| \geq \mathcal{S}_{\mathcal{P}}(f),$$

and in Case 3,

$$h > \mathcal{S}_{\mathcal{Q}}(f) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| + |f(x_0) - f(a)| \geq \mathcal{S}_{\mathcal{P}}(f),$$

so that in either case, $f[d,e]$ is rectifiable.

(B) (I) Let f_1 and f_2 be the restrictions of f to $[a,c]$ and $[c,b]$, respectively. By part (A) the arcs $f_1[a,c]$ and $f_2[c,b]$ are rectifiable, and

$$\mathbb{L}(f_1[a,c]) = \text{lub}\{\mathcal{S}_{\mathcal{Q}}(f_1) \mid \mathcal{Q} \text{ is a partition of } [a,c]\}, \text{ and}$$
$$\mathbb{L}(f_2[c,b]) = \text{lub}\{\mathcal{S}_{\mathcal{Q}}(f_2) \mid \mathcal{Q} \text{ is a partition of } [c,b]\}.$$

Let $\mathcal{P}$ be any partition of $[a, b]$; then $\mathcal{P} \cup \{c\}$ is a partition of $[a, b]$; define $\mathcal{P}_1 = (\mathcal{P} \cup \{c\}) \cap [a, c]$ and $\mathcal{P}_2 = (\mathcal{P} \cup \{c\}) \cap [c, b]$; then $\mathcal{P}_1$ is a partition of $[a, c]$ and $\mathcal{P}_1$ is a partition of $[c, b]$.

Then, for any partition $\mathcal{P}$ of $[a, b]$, and using Theorem ARC.3,

$$\mathcal{S}_{\mathcal{P}}(f) \leq \mathcal{S}_{(\mathcal{P} \cup \{c\})}(f) = \mathcal{S}_{\mathcal{P}_1}(f_1) + \mathcal{S}_{\mathcal{P}_2}(f_2) \leq \mathbb{L}(f[a, c]) + \mathbb{L}(f[c, b]).$$

so that

$$\begin{aligned} \mathbb{L}(f[a, b]) &= \operatorname{lub}\{\mathcal{S}_{\mathcal{P}}(f) \mid \mathcal{P} \text{ is a partition of } [a, b]\} \\ &\leq \mathbb{L}(f_1[a, c]) + \mathbb{L}(f_2[c, b]). \end{aligned}$$

(II) Conversely, let ϵ be any positive real number. By definition of $\mathbb{L}(f_1[a, c])$ and $\mathbb{L}(f_2[c, b])$ there exist partitions $\mathcal{P}_1$ of $[a, c]$ and $\mathcal{P}_2$ of $[c, b]$ such that

$$\mathcal{S}_{\mathcal{P}_1}(f_1) > \mathbb{L}(f_1[a, c]) - \tfrac{\epsilon}{2}, \text{ and } \mathcal{S}_{\mathcal{P}_2}(f_2) > \mathbb{L}(f_2[c, b]) - \tfrac{\epsilon}{2}.$$

Let $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$; $\mathcal{P}$ is a partition of $[a, b]$ and

$$\begin{aligned} \mathcal{S}_{\mathcal{P}}(f) = \mathcal{S}_{\mathcal{P}_1}(f_1) + \mathcal{S}_{\mathcal{P}_2}(f_2) &> \mathbb{L}(f_1[a, c]) - \tfrac{\epsilon}{2} + \mathbb{L}(f_2[c, b]) - \tfrac{\epsilon}{2} \\ &= \mathbb{L}(f_1[a, c]) + \mathbb{L}(f_2[c, b]) - \epsilon. \end{aligned}$$

Therefore if $\mathcal{Q}$ is any refinement of $\mathcal{P}$, using Theorem ARC.3,

$$\mathcal{S}_{\mathcal{Q}}(f) \geq \mathcal{S}_{\mathcal{P}}(f) > \mathbb{L}(f_1[a, c]) + \mathbb{L}(f_2[c, b]) - \epsilon.$$

Since we chose ϵ arbitrarily, this means that

$$\begin{aligned} \mathbb{L}(f[a, b]) &= \operatorname{lub}\{\mathcal{S}_{\mathcal{P}}(f) \mid \mathcal{P} \text{ is a partition of } [a, b]\} \\ &\geq \mathbb{L}(f_1[a, c]) + \mathbb{L}(f_2[c, b]). \end{aligned}$$

(III) That $\mathbb{L}(f[a, b]) = \mathbb{L}(f_1[a, c]) + \mathbb{L}(f_2[c, b])$ follows immediately from parts (I) and (II) above. $\square$

Definition ARC.5 Let a and b be distinct real numbers and let φ be a mapping of $[a, b] = \overleftrightarrow{ab}$ into $\mathbb{R}$. φ is of **bounded variation** on $[a, b]$ iff there exists a positive number h such that for every partition $\mathcal{P}[a, b]$ of $[a, b]$, $\mathcal{S}_{\mathcal{P}[a,b]}(\varphi) < h$. (Here, $\mathcal{S}_{\mathcal{P}[a,b]}(\varphi)$ is as defined in Definition ARC.1(IV).)

If φ is of bounded variation on $[a, b]$, then the **total variation** of φ on $[a, b]$ is

$$\mathbb{V}(\varphi_{[a,b]}) = \operatorname{lub}\{\mathcal{S}_{\mathcal{P}[a,b]}(\varphi) \mid \mathcal{P}[a, b] \text{ is a partition of } [a, b]\}.$$

Let t be any member of $]a, b] = \overleftrightarrow{ab}$. Then $\mathbb{V}(\varphi_{[a,t]})$ is the total variation of φ on $[a, t] = \overleftrightarrow{at}$.

Remark ARC.6 Our definition here of the total variation is the same as that in Definition ARC.1(VII) for arc length, except that now we are dealing with a real-valued function, rather than a function with values in $\mathbb{R}^2$.

The symbol $\mathbb{V}(\varphi[a, b])$ for the total variation of the function φ over the segment $[a, b]$ is completely analogous to the symbol $\mathbb{L}(f[a, b])$ for the arc length of an arc $f[a, b]$. The symbol $\mathcal{S}_{\mathcal{P}}(\varphi)$ means the same thing in both contexts.

Theorem ARC.7 *Suppose that a and b are distinct real numbers, where $a < b$, and that φ is a mapping of $[a, b]$ into $\mathbb{R}$. Let $[d, e]$ be a proper subsegment of $[a, b]$, that is, $d < e$ and both d and e are members of $[a, b]$, but $[a, b] \neq [d, e]$. Then*

(A) *if φ is of bounded variation on $[a, b]$, it is also of bounded variation on $[d, e]$; and*

(B) *if c is a real number such that $a < c < b$,*

$$\mathbb{V}(\varphi[a, b]) = \mathbb{V}(\varphi[a, c]) + \mathbb{V}(\varphi[c, b]).$$

We may express this last expression by saying that $\mathbb{V}(\varphi[a, t])$ is an additive *function of t.*

Proof. With the substitution where appropriate of $\mathbb{V}(\varphi[a, b])$ for $\mathbb{L}(f[a, b])$, the proof of Theorem ARC.4 is valid word-for-word for Theorem ARC.7. $\square$

Theorem ARC.8 *Let a and b be real numbers such that $a < b$ and let φ be a mapping of $[a, b] = \overleftrightarrow{ab}$ into $\mathbb{R}$ which is of bounded variation on $[a, b]$. If ψ is the mapping such that for every member x of $[a, b]$, $\psi(x) = \mathbb{V}(\varphi[a, x])$, then*

(A) *ψ is nondecreasing on $[a, b]$, and*

(B) *If φ is continuous on $[a, b]$, then ψ is continuous on $[a, b]$.*

Proof. (A) If s and t are members of $[a, b]$ such that $s < t$, then $\psi(t) = \mathbb{V}(\varphi[a, t])$ and $\psi(s) = \mathbb{V}(\varphi[a, s])$. By Theorem ARC.7(B),

$$\mathbb{V}(\varphi[a, t]) = \mathbb{V}(\varphi[a, s]) + \mathbb{V}(\varphi[s, t]),$$

so that, rearranging,

$$\psi(t) - \psi(s) = \mathbb{V}(\varphi[a, t]) - \mathbb{V}(\varphi[a, s]) = \mathbb{V}(\varphi[s, t]) \geq 0,$$

and thus $\psi(s) \leq \psi(t)$.

(B) We assume there exists a member x of $[a, b]$ such that ψ is not continuous at x and show that this assumption leads to a contradiction. Since ψ is nondecreasing the only discontinuity can be a jump discontinuity; either (Case 1) the limit from the right is greater than $\psi(x)$ or (Case 2) the limit from the left is less than $\psi(x)$.

(Case 1:) There exists a number $d > 0$ such that (limit from the right) $\lim_{t \to x} \psi(t) - \psi(x) = d$, where $t \in]x, b] = \overleftrightarrow{xb}$.

(I) We begin our proof by making the following observations labeled (a), (b), and (c):

From part (A), $\psi(t) = \mathbb{V}(\varphi[a,t])$ is an increasing function of t, and it is additive by Theorem ARC.7(B). Thus for all $t > x$,

$$\mathbb{V}(\varphi[a,t]) = \mathbb{V}(\varphi[a,x]) + \mathbb{V}(\varphi[x,t]).$$

so that

$$\mathbb{V}(\varphi[a,t]) - \mathbb{V}(\varphi[a,x]) = \mathbb{V}(\varphi[x,t]) \geq d.$$

Since $\lim_{t\to x} \mathbb{V}(\varphi[x,t]) = d$, there exists a number $t_0 > x$ such that for all t with $x < t \leq t_0$,

$$d \leq \mathbb{V}(\varphi[x,t]) < \tfrac{6d}{5}. \quad \text{(a)}$$

Also, since φ is continuous, the number t_0 may also be chosen so that for all t with $x < t \leq t_0$,

$$|\varphi(t) - \varphi(t_0)| < \tfrac{d}{5}. \quad \text{(b)}$$

For every partition $\mathcal{P} = \mathcal{P}[x,t]$ of $[x,t]$, where $x < t \leq t_0$, and applying (a)

$$\mathcal{S}_{\mathcal{P}[x,t]}(\varphi) < \mathbb{V}(\varphi[x,t]) < \tfrac{6d}{5}. \quad \text{(c)}$$

(II) By the definition of least upper bound, we can choose $\mathcal{P}_0 = \{x = x_1 < x_1 < x_2 < \ldots < x_n = t_0\}$ to be a partition of $[x,t_0]$ such that

$$\mathcal{S}_{\mathcal{P}_0}(\varphi) > \mathbb{V}(\varphi[x,t_0]) - \tfrac{d}{5}.$$

Then

$$\begin{aligned}\mathcal{S}_{\mathcal{P}_0}(\varphi) &= |\varphi(x_1) - \varphi(x_0)| + \sum_{k=2}^{n} |\varphi(x_k) - \varphi(x_{k-1})| \\ &< \tfrac{d}{5} + \sum_{k=2}^{n} |\varphi(x_k) - \varphi(x_{k-1})|,\end{aligned}$$

or, rearranging and using (I)(a) above,

$$\sum_{k=2}^{n} |\varphi(t_k) - \varphi(t_{k-1})| > \mathcal{S}_{\mathcal{P}_0}(\varphi) - \tfrac{d}{5} > \mathbb{V}(\varphi[x,t_0]) - \tfrac{d}{5} - \tfrac{d}{5} > d - \tfrac{2d}{5} = \tfrac{3d}{5}.$$

(III) Now let $\mathcal{P}_1 = \{x = y_0 < y_1 < y_2 < \ldots < y_m = x_1\}$ be a partition of $[x_0,x_1] = [x,x_1]$; since $\mathbb{V}(\varphi[x,x_1]) \geq d$, we can choose $\mathcal{P}_1$ so that

$$\mathcal{S}_{\mathcal{P}_1}(\varphi) > \mathbb{V}(\varphi[x,x_1]) - \tfrac{d}{5} \geq d - \tfrac{d}{5} = \tfrac{4d}{5}.$$

Let $\mathcal{Q} = \mathcal{P}_0 \cup \mathcal{P}_1$; then $\mathcal{Q}$ is a partition of $[x,t]$ and by (II),

$$\begin{aligned}\mathcal{S}_{\mathcal{Q}}(\varphi) &= \mathcal{S}_{\mathcal{P}_1}(\varphi) + \sum_{k=2}^{n} |\varphi(x_k) - \varphi(x_{k-1})| \\ &> \tfrac{4d}{5} + \tfrac{3d}{5} = \tfrac{7d}{5}\end{aligned}$$

But by (I)(c), $\mathcal{S}_{\mathcal{Q}}(\varphi) < \frac{6d}{5}$, so we have a contradiction.

(Case 2:) We follow the proof of Case 1. There exists a number $d > 0$ such that (limit from the left) $\psi(x) - \lim_{t\to x} \psi(t) = d$, where $t \in [a,x]$. We leave the rest of the proof to the reader as Exercise ARC.1. $\square$

Remark ARC.9 (A) A real-valued function defined on a set $\mathcal{E}$ of real numbers is said to be uniformly continuous iff for any real number $\epsilon > 0$, there exists a real number $\delta > 0$ such that for all x and y in $\mathcal{E}$, if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.

(B) A real-valued function defined on a set $\mathcal{E}$ of real numbers is said to be bounded iff there exists a real number $b > 0$, such that for every $x \in \mathcal{E}$, $|f(x)| < b$.

(C) It is well known from calculus that if a real-valued function f defined on a closed interval $[a, b]$ is continuous, it is uniformly continuous; from this it is easy to see that it is also bounded; for if ϵ and δ are as in (A), let n be any integer such that $n \frac{|b-a|}{\delta}$; then for any $x \in [a, b]$, $|x - a| < n(\delta)$ and there exists a subset $\{a = t_0 < t_1 < t_2 < ... < t_{n-1} < t_n = x\}$ such that for all $k \in [1; n]$, $|t_k - t_{k-1}| < \delta$, and hence for all $k \in [1; n]$, $|f(t_k) - f(t_{k-1})| < \epsilon$, so that

$$|f(a) - f(x)| \leq \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| < n\epsilon.$$

Then let $b = |f(a)| + n\epsilon$; for all $x \in [a, b]$, $|f(x)| \leq b$, so that f is bounded on $[a, b]$.

Theorem ARC.10 *Let f be a continuous function defined on the closed interval $[a, b]$ having values in $\mathbb{R}$ or $\mathbb{R}^2$. By Remark ARC.9, f is uniformly continuous on $[a, b]$.*

For each integer $j > 0$ let $\{\mathcal{P}_j\}$ be a partition of $[a, b]$ and let g_j be the gauge of the partition $\mathcal{P}_j$. Suppose that $\lim_{j \to \infty} g_j = 0$ and $\lim_{g_j \to 0} \mathcal{S}_{\mathcal{P}_j}(f)$ exists and is equal to some number L. Since for all j, $\mathcal{S}_{\mathcal{P}_j}(f) > 0$, $L > 0$.

(A) *Then for every partition $\mathcal{P}$ of $[a, b]$,*

$$\mathcal{S}_{\mathcal{P}}(f) \leq \lim_{g_j \to 0} \mathcal{S}_{\mathcal{P}_j}(f) = L.$$

(B) *The least upper bound*

$$\text{lub}\{\mathcal{S}_{\mathcal{P}}(f) \mid \mathcal{P} \text{ is a partition of } [a, b]\}.$$

exists, and is equal to L.

Proof. (A) Suppose the contrary is true, that for some partition $\mathcal{P} = \{a = x_0 < x_1 < x < 2 < ... < x_n = b\}$ of $[a, b]$,

$$\mathcal{S}_{\mathcal{P}}(f) > L.$$

Let $d = \mathcal{S}_{\mathcal{P}}(f) - L$, that is, $\mathcal{S}_{\mathcal{P}}(f) = L + d$ Since f is uniformly continuous, there exists a real number $\delta > 0$ such that for all x and y in $[a, b]$, if $|x - y| < \delta$, then $|f(x) - f(y)| < \frac{d}{3n}$. Let j_0 be so large that for all $j \geq j_0$,

(1) $g_j < \min\{x_k - x_{k-1} \mid k \in \{[1, n]\}$;

(2) $g_j < \delta$; and

(3) $|\mathcal{S}_{\mathcal{P}_j} - L| < \frac{d}{3}$, so that $\mathcal{S}_{\mathcal{P}_j} < L + \frac{d}{3}$.

Let $j \geq j_0$; for each $k \in [1; n]$, define $\mathcal{Y}_k = \mathcal{P}_j \cap]x_{k-1}, x_k[= \{y_0 < y_1 < y_2 < ... < y_{m_k}\}$ by (1) above, this set is nonempty, and $\mathcal{Y}_k$ is a partition of $[{y_0}^{(k)}, {y_{m_k}}^{(k)}]$. Then

$$\sum_{p=1}^{m_k} |f(y_p{}^{(k)}) - f(y_{p-1}{}^{(k)})| + |f(y_0{}^{(k)}) - f(x_{k-1})| + |f(x_k) - f(y_{m_k}{}^{(k)})|$$
$$= \mathcal{S}_{\mathcal{Y}_k}(f) + |f(y_0{}^{(k)}) - f(x_{k-1})| + |f(x_k) - f(y_{m_k}{}^{(k)})|$$
$$\geq |f(x_k) - f(x_{k-1})|; \quad (*)$$

this last inequality uses the fact that $\{x_{k-1}, y_0{}^{(k)}, y_1{}^{(k)}, y_2{}^{(k)}, ..., y_{m_k}{}^{(k)}, x_k\}$ is a refinement of the trivial partition $\{x_{k-1}, x_k\}$ of $[x_{k-1}, x_k]$. Here it is possible, but not necessary, for $x_{k-1} = y_0{}^{(k)}$ or $y_{m_k}{}^{(k)} = x_k$.

Now all the terms $|f(y_p{}^{(k)}) - f(y_{p-1}{}^{(k)})|$ (which belong to the summations $\sum_{p=1}^{m_k} |f(y_p{}^{(k)}) - f(y_{p-1}{}^{(k)})|$ over all partitions $\mathcal{Y}_k$) are also terms in the summation $\mathcal{S}_{\mathcal{P}_j}(f)$; therefore the sum of all these terms is less or equal to $\mathcal{S}_{\mathcal{P}_j}(f)$. It follows that

$$\mathcal{S}_{\mathcal{P}_j}(f) \geq \sum_{k=1}^{n} \mathcal{S}_{\mathcal{Y}_k}(f),$$

and using (*) above,

$$\mathcal{S}_{\mathcal{P}_j}(f) + \sum_{k=1}^{n} \left(|f(y_0{}^{(k)}) - f(x_{k-1})| + |f(x_k) - f(y_{m_k}{}^{(k)})|\right)$$
$$\geq \sum_{k=1}^{n} \mathcal{S}_{\mathcal{Y}_k}(f) + \sum_{k=1}^{n} \left(|f(y_0{}^{(k)}) - f(x_{k-1})| + |f(x_k) - f(y_{m_k}{}^{(k)})|\right)$$
$$\geq \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| = \mathcal{S}_{\mathcal{P}}(f).$$

Finally, since $|y_0{}^{(k)} - x_{k-1}| < \delta$ and $|x_k - y_{m_k}{}^{(k)}| < \delta$,

$$|f(y_0{}^{(k)}) - f(x_{k-1})| < \tfrac{d}{3n} \text{ and } |f(x_k) - f(y_{m_k}{}^{(k)})| < \tfrac{d}{3n},$$

so that

$$\sum_{k=1}^{n} \left(|f(y_0{}^{(k)}) - f(x_{k-1}{}^{(k)})| + |f(x_k) - f(y_{m_k}{}^{(k)})|\right) < \tfrac{2d}{3}$$

and

$$\mathcal{S}_{\mathcal{P}_j}(f) + \tfrac{2d}{3} > \mathcal{S}_{\mathcal{P}}(f), \text{ or } \mathcal{S}_{\mathcal{P}_j}(f) > \mathcal{S}_{\mathcal{P}}(f) - \tfrac{2d}{3}.$$

Recall from the beginning of the proof that $\mathcal{S}_{\mathcal{P}}(f) = L + d$, so that

$$\mathcal{S}_{\mathcal{P}_j}(f) > L + d - \tfrac{2d}{3} = L + \tfrac{d}{3}.$$

But from condition (3) above,

$$\mathcal{S}_{\mathcal{P}_j}(f) < L + \tfrac{d}{3},$$

a contradiction. It follows that for every partition $\mathcal{P}$ of $[a, b]$, $\mathcal{S}_{\mathcal{P}}(f) \leq L$.

(B) By part (A), L is an upper bound for

$$\{\mathcal{S}_{\mathcal{P}}(f) \mid \mathcal{P} \text{ is a partition of } [a, b]\}.$$

By definition of limit, for every $\epsilon > 0$, there exists a partition $\mathcal{P}_j$ such that $L - \mathcal{S}_{\mathcal{P}_j}(f) < \epsilon$. Therefore the least upper bound is greater or equal to L, hence equal to L. □

Theorem ARC.11 (Integral form for arc length) *Let a and b be real numbers such that $a < b$ and let α and β be continuous real valued functions* *defined on* $[a,b] = \overleftrightarrow{ab}$*, such that the function* $f(t) = (\alpha(t), \beta(t))$ *mapping* $[a,b]$ *to a subset* $\mathcal{C}$ *of* $\mathbb{R}^2$ *is continuous. Assume further that the derivatives* α' *and* β' *exist and are continuous on* $[a,b]$*, and that for every member* t *of* $[a,b]$*,* $\Big(\alpha'(t)\Big)^2 + \Big(\beta'(t)\Big)^2 > 0$*. Then*

(A) $f[a,b]$ *is a rectifiable arc (or closed arc), and*

(B) *for every member* x *of* $[a,b]$*,*

$$\mathbb{L}(f[a,x]) = \int_a^x \sqrt{\Big(\alpha'(u)\Big)^2 + \Big(\beta'(u)\Big)^2}\, du.$$

Proof. For each natural number j, let $\mathcal{P}_j[a,x]$ be a partition of $[a,x]$ such that for each j, $\mathcal{P}_{j+1}[a,x]$ is a refinement of $\mathcal{P}_j[a,x]$, and $\lim_{j\to\infty} g_j = 0$, where g_j is the gauge of $\mathcal{P}_j[a,x]$. For an arbitrary j, let $\mathcal{P}_j[a,x] = \{t_0 < t_1 < t_2 < ... < t_n\}$. Then $\mathcal{S}_{\mathcal{P}_j}(f)$ is

$$\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| = \sum_{k=1}^{n} \sqrt{\Big(\alpha(t_k) - \alpha(t_{k-1})\Big)^2 + \Big(\beta(t_k) - \beta(t_{k-1})\Big)^2}.$$

By the mean-value theorem for derivatives, for each $k \in [1;n]$ there exist numbers s_k and u_k of $]t_{k-1}, t_k[$ such that $\alpha(t_k) - \alpha(t_{k-1}) = (t_k - t_{k-1})\alpha'(s_k)$ and $\beta(t_k) - \beta(t_{k-1}) = (t_k - t_{k-1})\beta'(u_k)$. Then

$$\begin{aligned}\mathcal{S}_{\mathcal{P}_j}(f) &= \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| \\ &\le \sum_{k=1}^{n} \sqrt{(\alpha'(s_k))^2 + (\beta'(s_k))^2}(t_k - t_{k-1}) \\ &\quad + \sum_{k=1}^{n} \sqrt{|\beta'(u_k))^2 - (\beta'(s_k)^2|}(t_k - t_{k-1}) \quad (*).\end{aligned}$$

Since β' is continuous on $[a,b]$, it is bounded on $[a,b]$. Thus there exists a positive number h such that for every t belonging to $[a,b]$, $|\beta'(t)| \le \frac{h}{2(b-a)}$.

Since β' is uniformly continuous on $[a,b]$, for every positive number ϵ there exists a positive number δ such that for all numbers s and t belonging to $[a,b]$, if $|s-t| < \delta$, then $|\beta'(s) - \beta'(t)| < \frac{\epsilon^2}{2(b-a)h}$.

Thus if we choose the partition $\mathcal{P}_j[a,x] = \{t_0 < t_1 < t_2 < ... < t_n\}$ so that $g_j = \max\{(t_k - t_{k-1)} \mid k \in [1;n]\} < \delta$,

$$\begin{aligned}&\sum_{k=1}^{n} \sqrt{|(\beta'(u_k))^2 - (\beta'(s_k))|^2}(t_k - t_{k-1}) \\ &= \sum_{k=1}^{n} \sqrt{|\beta'(u_k) + \beta'(s_k)|}\sqrt{|\beta'(u_k) - \beta'(s_k)|}(t_k - t_{k-1}).\end{aligned}$$

For all k,

$$\begin{aligned}\sum_{k=1}^{n} \sqrt{|\beta'(u_k) + \beta'(s_k)|} &\le \sum_{k=1}^{n} \sqrt{|\beta'(u_k)| + |\beta'(s_k)|} \\ &\le \sqrt{|\beta'(u_k)|} + \sqrt{|\beta'(s_k)|} < 2\sqrt{|\beta'(t)|} \le 2\sqrt{\tfrac{h}{2(b-a)}}.\end{aligned}$$

Also, for all k,

$$\sqrt{|\beta'(u_k) - \beta'(s_k)|} < \sqrt{\tfrac{\epsilon^2}{2(b-a)h}}$$

so that

$$\begin{aligned}
&\sum_{k=1}^{n} \sqrt{|(\beta'(u_k))^2 - (\beta'(s_k))|^2}(t_k - t_{k-1}) \\
&\quad < 2\sqrt{\tfrac{h}{2(b-a)}} \cdot \sqrt{\tfrac{\epsilon^2}{2(b-a)h}} \sum_{k=1}^{n} (t_k - t_{k-1}) \\
&\quad\quad = \sqrt{\tfrac{\epsilon^2}{(b-a)}} \sum_{k=1}^{n} (t_k - t_{k-1}) = \tfrac{\epsilon}{(b-a)}(b-a) = \epsilon.
\end{aligned}$$

Therefore

$$\lim_{g_j \to 0} \sum_{k=1}^{n} \sqrt{|(\beta'(u_k))^2 - (\beta'(s_k))^2|}(t_k - t_{k-1}) = 0,$$

where taking the limit $\lim\limits_{g_j \to 0}$ means the same thing as taking the limit $\lim\limits_{j \to \infty}$ through the sequence $\mathcal{P}_j[a,x]$ of partitions of $[a,x]$.

By definition of the integral (from calculus), and by the argument just above, the limit of the right-hand side of (*) is

$$\int_a^x \sqrt{\Big(\alpha'(u)\Big)^2 + \Big(\beta'(u)\Big)^2}\, du + 0.$$

For each $j > 1$, $\mathcal{P}_{j+1}[a,x]$ is a refinement of $\mathcal{P}_j[a,x]$, so the sequence $\mathcal{S}_{\mathcal{P}_j}(f)$ (the left-hand side of (*)) is non-decreasing (with j), and has an upper bound since the right-hand side of (*) has a limit. Therefore the limit $\lim\limits_{g_j \to 0} \mathcal{S}_{\mathcal{P}_j}(f)$ exists; by Theorem ARC.10, this is

$$\mathbb{L}(f[a,x]) = \text{lub}\{\mathcal{S}_{\mathcal{P}}(f) \mid \mathcal{P} \text{ is a partition of } [a,x]\},$$

completing the proof. $\square$

Theorem ARC.12 (Arc length is a bicontinuous bijection) *Let a and b be real numbers such that $a < b$; let α and β be continuous real-valued functions defined on $[a,b]$, such that $f = (\alpha, \beta)$ is a one-to-one continuous function of $[a,b]$ into $\mathbb{R}^2$ (and thus a homeomorphism); assume also that the arc $f[a,b]$ is rectifiable.*

Let φ be the mapping defined on $[a,b]$ such that $\varphi(a) = 0$ and for every $t \in]a,b]$, $\varphi(t) = \mathbb{L}(f[a,b])$. Then

(A) *φ is increasing, one-to-one, and continuous on $[a,b]$;*

(B) *φ^{-1} exists, is increasing, and is continuous on $\overline{\varphi(a)\varphi(b)}$; and*

(C) *the mapping $\varphi \circ f^{-1}$ is a one-to-one mapping of $f([a,b]) = \mathcal{C}$ into $\mathbb{R}$; that is, if $\varphi(s) = \varphi(t)$ then $f(s) = f(t)$.*

Proof. (A) Let s and t be members of $[a,b]$ such that $s < t$.

(Case 1: $a = s < t \leq b$.) Let $\mathcal{P}[a,t]$ be a partition of $[a,t]$. Since f is one-to-one, every term of $\mathcal{S}_{\mathcal{P}[a,t]}(f)$ is positive (as opposed to non-negative). Therefore $\mathbb{L}(f[a,t])$ is positive and so $\varphi(a) = 0 < \varphi(t)$.

(Case 2: $a < s$.) Let $\mathcal{P}[s,t]$ be a partition of $[s,t]$. Since every term of $\mathcal{S}_{\mathcal{P}[s,t]}(f))$ is positive, $\mathbb{L}(f[s,t])$ is positive. By Theorem ARC.4, $\mathbb{L}(f[a,t]) = \mathbb{L}(f[a,s]) + \mathbb{L}(f[s,t])$, $\varphi(t) = \varphi(s) + \mathbb{L}(f[s,t])$, so $\varphi(s) < \varphi(t)$. Thus φ is increasing, and is one-to-one.

To show continuity of φ, let s and t be members of $[a,b]$ such that $s < t$, let n be a natural number, and let $\mathcal{P}[a,b] = \{a = x_0 < x_1 < x_2 < ... < x_n = b\}$ be a partition of $[a,b]$. Then

$$\begin{aligned}\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| &\leq \sum_{k=1}^{n} |\alpha(x_k) - \alpha(x_{k-1})| + \sum_{k=1}^{n} |\beta(x_k) - \beta(x_{k-1})| \\ &\leq \mathbb{V}(\alpha[s,t]) + \mathbb{V}(\beta[s,t]).\end{aligned}$$

Since $\mathcal{P}[a,b]$ is an arbitrary partition of $[a,b]$,

$$\mathbb{L}(f[a,t]) \leq \mathbb{V}(\alpha_{[s,t]}) + \mathbb{V}(\beta_{[s,t]}).$$

Let ϵ be any positive number. By Remark ARC.9(C) there exists a positive number δ such that for all members s and t of $[a,b]$ for which $0 \leq s < t \leq b$ and $t - s < \delta$, $\mathbb{L}(f[s,t]) < \epsilon$. Since $\mathbb{L}(f[s,t]) = \varphi(t) - \varphi(s)$, this means that φ is continuous on $[a,b]$.

(B) By Exercise ARC.3 φ^{-1} is increasing and continuous on $\overline{\varphi(a)\varphi(b)}$.

(C) Since f and φ are one-to-one, so are f^{-1} and $\varphi \circ f^{-1}$. $\square$

3.2 Exercises for arc length

Exercise ARC.1* Complete the proof of Theorem ARC.8(B), Case 2.

Exercise ARC.2* Let a and b be distinct real numbers, α and β be mappings of $[a,b]$ into $\mathbb{R}$ and $f = (\alpha, \beta)$, then:

(A) f is continuous on $[a,b]$ iff each of α or β is continuous on $[a,b]$.

(B) f is rectifiable iff each of α and β is of bounded variation on $[a,b]$.

Exercise ARC.3* Let a and b be real numbers such that $a < b$. If φ is a mapping of $[a,b]$ into $\mathbb{R}$ which is increasing and is continuous on $[a,b]$, then:

(I) φ^{-1} exists.

(II) φ^{-1} is increasing on $[\varphi(a), \varphi(b)]$.

(III) φ^{-1} is continuous on $[\varphi(a), \varphi(b)]$.

Exercise ARC.4* Let φ be a mapping of $\mathbb{R}$ into $\mathbb{R}$ which is increasing and continuous on $\mathbb{R}$, then:

(I) φ^{-1} exists.

(II) φ^{-1} is increasing on $\mathbb{R}$.

(III) φ^{-1} is continuous on $\mathbb{R}$.

3.3 Selected answers for arc length

Exercise ARC.1 Proof. We follow the proof of Case 1 of Theorem ARC.8(B). There exists a number $d > 0$ such that (limit from the left) $\psi(x) - \lim_{t \to x} \psi(t) = d$, where $t \in [a, x]$.

(I) $\psi(t) = \mathbb{V}(\varphi[a,t])$ is an increasing function of t, and is additive. For all $t < x$,

$$\mathbb{V}(\varphi[a,x]) = \mathbb{V}(\varphi[a,t]) + \mathbb{V}(\varphi[t,x]).$$

so that

$$\mathbb{V}(\varphi[a,x]) - \mathbb{V}(\varphi[a,t]) = \mathbb{V}(\varphi[t,x]) \geq d.$$

Since $\lim_{t \to x} \mathbb{V}(\varphi[x,t]) = d$, there exists a number $t_0 < x$ such that for all t with $x > t \geq t_0$,

$$d \leq \mathbb{V}(\varphi[t,x]) < \tfrac{6d}{5}. \quad \text{(a)}$$

Also, since φ is continuous, the number t_0 may also be chosen so that for all t with $x > t \geq t_0$,

$$|\varphi(x) - \varphi(t)| < \tfrac{d}{5}. \quad \text{(b)}$$

For every partition $\mathcal{P} = \mathcal{P}[t,x]$ of $[t,x]$, where $x > t \geq t_0$, and applying (a)

$$\mathcal{S}_{\mathcal{P}[t,x]}(\varphi) < \mathbb{V}(\varphi[t,x]) < \tfrac{6d}{5}. \quad \text{(c)}$$

(II) By the definition of least upper bound, we can choose $\mathcal{P}_0 = \{t_0 = x_1 < x_1 < x_2 < ... < x_n = x\}$ to be a partition of $[t_0, x]$ such that

$$\mathcal{S}_{\mathcal{P}_0}(\varphi) > \mathbb{V}(\varphi[0,x_n]) - \tfrac{d}{5}.$$

Then by (I)(b) above,

$$\begin{aligned}\mathcal{S}_{\mathcal{P}_0}(\varphi) &= \sum_{k=1}^{n-1} |\varphi(x_k) - \varphi(x_{k-1})| + |\varphi(x_n) - \varphi(x_{n-1})| \\ &< \sum_{k=1}^{n-1} |\varphi(x_k) - \varphi(x_{k-1})| + \tfrac{d}{5},\end{aligned}$$

or, rearranging and using (I)(a) above,

$$\sum_{k=1}^{n-1} |\varphi(x_k) - \varphi(x_{k-1})| > \mathcal{S}_{\mathcal{P}_0}(\varphi) - \tfrac{d}{5} > \mathbb{V}(\varphi[x,t_0]) - \tfrac{d}{5} - \tfrac{d}{5} > d - \tfrac{2d}{5} = \tfrac{3d}{5}.$$

(III) Now let $\mathcal{P}_1 = \{x_{n-1} = y_0 < y_1 < y_2 < ... < y_m = x\}$ be a partition of $[x_{n-1}, x_n] = [x_{n-1}, x]$; since $\mathbb{V}(\varphi[x_{n-1}, x]) \geq d$, we can choose $\mathcal{P}_1$ so that

$$\mathcal{S}_{\mathcal{P}_1}(\varphi) > \mathbb{V}(\varphi[x_{n-1},x]) - \tfrac{d}{5} \geq d - \tfrac{d}{5} = \tfrac{4d}{5}.$$

Let $\mathcal{Q} = \mathcal{P}_0 \cup \mathcal{P}_1$; then $\mathcal{Q}$ is a partition of $[t,x]$ and by (II),

$$\mathcal{S}_{\mathcal{Q}}(\varphi) = \sum_{k=1}^{n-1} |\varphi(x_k) - \varphi(x_{k-1})| + \mathcal{S}_{\mathcal{P}_1}(\varphi) > \tfrac{3d}{5} + \tfrac{4d}{5} = \tfrac{7d}{5}$$

But by (I)(c), $\mathcal{S}_{\mathcal{Q}}(\varphi) < \frac{6d}{5}$, so we have a contradiction. $\square$

Exercise ARC.2 Proof. (A) If α and β are each continuous on $[a, b]$, then for every member s of $[a, b]$ and for every positive number ϵ, there exists a positive number δ such that for every member t of $[a, b] \setminus \{s\}$, if $|t - s| < \delta$, then $|\alpha(t) - \alpha(s)| \leq \frac{\epsilon}{2}$ and $|\beta(t) - \beta(s)| < \frac{\epsilon}{2}$. Thus $|f(s) - f(t)| = \sqrt{\Big(\alpha(t) - \alpha(s)\Big)^2 + \Big(\beta(t) - \beta(s)\Big)^2} \leq |\alpha(t) - \alpha(s)| + |\beta(t) - \beta(s)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Hence if α and β are continuous on $[a, b]$, then f is continuous on $[a, b]$. Conversely, if f is continuous on $[a, b]$, then for every member s of $[a, b]$ and for every positive number ϵ, there exists a positive number ϵ, such that for every member t of $[a, b] \setminus \{s\}$, if $|t - s| < \delta$, then $f(t) - f(s)| < \epsilon$. Hence $|\alpha(t) - \alpha(s)| \leq \sqrt{\Big(\alpha(t) - \alpha(s)\Big)^2 + \Big(\beta(t) - \beta(s)\Big)^2} = |f(t) - f(s)| < \epsilon$ and $|\beta(t) - \beta(s)| \leq \sqrt{\Big(\alpha(t) - \alpha(s)\Big)^2 + \Big(\beta(t) - \beta(s)\Big)^2} = |f(t) - f(s)| < \epsilon$. Therefore α is continuous on $[a, b]$ and β is continuous on $[a, b]$.

(B) (I) If each of α or β is of bounded variation on $[a, b]$, then there exists an $h > 0$ such that for every partition $\mathcal{P}[a, b] = \{a = x_0 < x_1 < x_2 < ... < x_n = b\}$ of $[a, b]$, $\sum_{k=1}^{n} |\alpha(t_k) - \alpha(t_{k-1})| < \frac{h}{2}$ and $\sum_{k=1}^{n} |\beta(t_k) - \beta(t_{k-1})| < \frac{h}{2}$. Since for each member k of $[1; n]$, $|f(t_k) - f(t_{k-1})| = \sqrt{\Big(\alpha(t_k) - \alpha(t_{k-1})\Big)^2 + \Big(\beta(t_k) - \beta(t_{k-1})\Big)^2} \leq |\alpha(t_k) - \alpha(t_{k-1})| + |\beta(t_k) - \beta(t_{k-1})|$, so $\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| < h$. Thus f is rectifiable on $[a, b]$.

(II) Conversely, if f is rectifiable on $[a, b]$, then there exists an $h > 0$ such that for every partition $\mathcal{P}[s, t] = \{a = x_0 < x_1 < x_2 < ... < x_n = b\}$, $\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| < h$. Since for every member k of $[1; n]$, $|\alpha(t_k) - \alpha(t_{k-1})| \leq \sqrt{(\alpha(t_k) - \alpha(t_{k-1}))^2 + (\beta(t_k) - \beta(t_{k-1}))^2} = |f(t_k) - f(t_{k-1})| < h$ and $|\beta(t_k) - \beta(t_{k-1})| \leq |f(t_k) - f(t_{k-1})| < h$, so both α and β are of bounded variation on $[a, b]$. $\square$

Exercise ARC.3 Proof. (I) We prove that for every member u of $[a, b]$ there exists a unique number s of $[a, b]$ such that $\varphi(s) = u$. Assume that there exist distinct members s and t of $[a, b]$ such that $\varphi(s) = \varphi(t) = u$. We choose the notation so that $s < t$. Since φ is increasing on $[a, b]$, $\varphi(s) < \varphi(t)$, a contradiction.

(II) Let u and v be members of $\varphi[a, b]$ such that $u < v$. By part (I) there exist distinct and unique members s and t of $[a, b]$ such that $\varphi(s) = u$ and

$\varphi(t) = v$. If $t \leq s$, then $v \leq u$, contrary to the given fact that $u < v$. Hence $s < t$, that is $\varphi^{-1}(u) < \varphi^{-1}(v)$.

(III) Let u and v be members of $\varphi[a, b]$ such that $\varphi(a) \leq u < v \leq \varphi(b)$. By part (I) there exist unique members s and t of $[a, b]$ such that $\varphi(s) = u$ and $\varphi(t) = v$. We prove that $\varphi\,]s, t[=]u, v[=]\varphi(s), \varphi(t)[$.

(A) If q is any member of $]s, t[$, then $s < q < t$ so $\varphi(s) < \varphi(q) < \varphi(t)$. Thus $\varphi(q) \in]\varphi(s), \varphi(t)[=]u, v[$. Since q is any member of $]s, t[$, $\varphi\,]s, t[\subseteq]u, v[$.

Conversely, if w is any member of $]u, v[$, then by the continuity of φ and the intermediate value theorem there exists a unique member q of $]s, t[$ such that $\varphi(q) = w$. Since w is any member of $]u, v[$, $]u, v[\subseteq]\varphi(s), \varphi(t)[$, and therefore $]u, v[=]\varphi(s), \varphi(t)[$, and $\varphi^{-1}\,]\varphi(s), \varphi(t)[=]s, t[$.

(B) We now prove that φ^{-1} is continuous on $[\varphi(a), \varphi(b)]$.

(Case 1: φ^{-1} is continuous at $\varphi(a)$.) Let $\epsilon > 0$, and let $s \in]a, b[$ (so that $\varphi(s) \in]\varphi(a), \varphi(b)[$) be a number such that $|s - a| < \epsilon$. Let $\delta = |\varphi(a) - \varphi(s)|$; then for every $u \in]\varphi(a), \varphi(b)[$ such that $|\varphi(a) - u| < \delta$, $u \in]\varphi(a), \varphi(s)[$, so $\varphi^{-1}(u) \in]a, s[$, and hence $|a - \varphi^{-1}(u)| < \epsilon$. Therefore φ^{-1} is continuous at $\varphi(a)$.

(Case 2: φ^{-1} is continuous at $\varphi(b)$.) The proof is similar to that for Case 1 with obvious substitutions of b for a.

(Case 3: φ^{-1} is continuous at every point $u \in]\varphi(a), \varphi(b)[$.) Since $u \in]\varphi(a), \varphi(b)[$, $\varphi^{-1}(u) \in]a, b[$. Let $\epsilon > 0$. Let s and t be points of $]a, b[$ such that $a < s < \varphi^{-1}(u) < t < b$ and $|s - \varphi^{-1}(u)| < \epsilon$ and $|t - \varphi^{-1}(u)| < \epsilon$. Then $\varphi(a) < \varphi(s) < u < \varphi(t) < \varphi(b)$. Let $\delta = \min\{|\varphi(s) - u|, |\varphi(t) - u|\}$. Then if $|x - u| < \delta$, $\varphi(s) < x < \varphi(t)$; so $s < \varphi^{-1}(x) < t$, and hence $|\varphi^{-1}(x) - \varphi^{-1}(u)| < \epsilon$. Therefore φ^{-1} is continuous at u. $\square$

Exercise ARC.4 Proof. The proofs of parts (I) and (II) are word-for-word identical to those for Exercise ARC.3, with $\mathbb{R}$ substituted for $[a, b]$.

(III) Let u and v be members of $\mathbb{R}$ such that $u < v$. By part (I) there exist unique members s and t of $\mathbb{R}$ such that $\varphi(s) = u$ and $\varphi(t) = v$. The proof that that $\varphi\,]s, t[=]u, v[=]\varphi(s), \varphi(t)[$ is exactly as in part (III) of Exercise ARC.3.

The proof that φ^{-1} is continuous on $\mathbb{R}$ is almost exactly the same as in Case 3 of part (III)(B) of Exercise ARC.3, with $\mathbb{R}$ substituted for $]a, b[$. $\square$

Chapter 4

The Real Functions Cosine and Sine (CS)

Dependencies: *Euclidean Geometry and its Subgeometries (Specht); Chapter 3 of this supplement*
Acronym: *CS*
New terms defined: *sine and cosine functions, periodic function, unit circle, circumference, the function cis*

In this chapter we define the circular functions sin and cos using a function $q(x)$ which is the inverse of the function

$$g(x) = \int_0^x \frac{2}{1+x^2}\,dt.$$

This function $q(x)$ turns out to be the restriction of $\tan(\frac{x}{2})$ to the interval $]-\pi, \pi[$.[1]

This chapter depends on the previous chapter; as in that chapter, we assume familiarity with calculus and are free to use any theorems and definitions from *Specht*. References such as "Theorem ROT.15" cite items from *Specht*, and references to items with acronyms ARC and CS are to the present Supplement. Again, we refer the reader to the note **Citations and references** at the end of the Preface to this Supplement, and to the abbreviated Table of Contents for *Specht* included there.

On notation: although sin, cos, and tan are functions, where no ambiguity arises we will use traditional shorthand, writing $\sin x$ for $\sin(x)$, $\cos x$ for $\cos(x)$, and $\tan x$ for $\tan(x)$. We will also write $\sin^2 x$ for $(\sin x)^2 = (\sin(x))^2$ and, when q is a function, $q^2(x)$ for $(q(x))^2$. We will use the notation f' to denote the derivative of a function f.

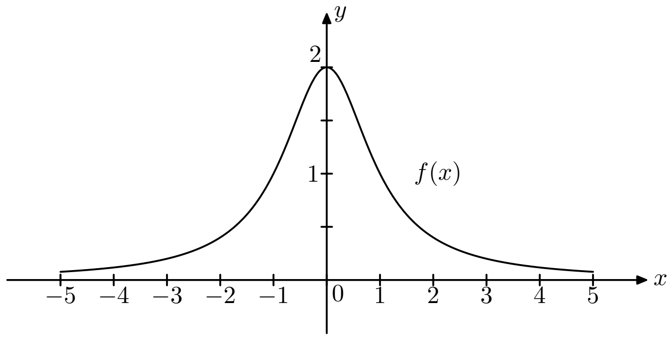
[1] This definition of the circular functions sin and cos appears to have originated with Edward Specht, the first author of *Euclidean Geometry and its Subgeometries.*

4.1 Basic properties of cosine and sine; periodicity

Definition CS.1 (A) For each real number x, define $f(x) = \dfrac{2}{1+x^2}$.

(B) For each real number x define $g(x) = \int_0^x f(t)\,dt$.

(C) Define $g(1) = k$.



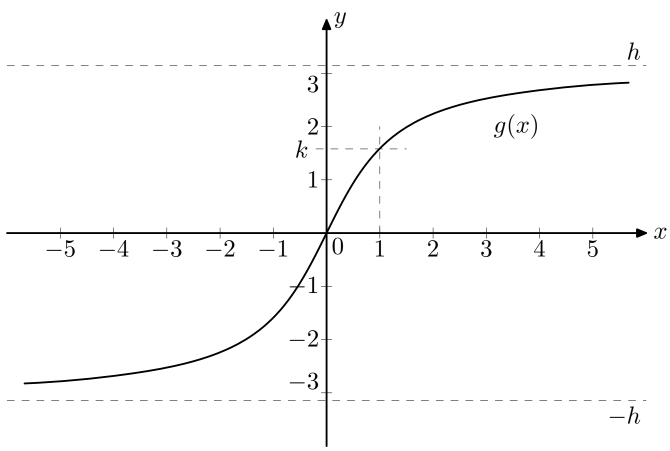


Fig. 4.1 Graphs of $f(x) = \frac{2}{1+x^2}$ (top) and $g(x) = \int_0^x f(t)\,dt$ (bottom) for Definition CS.1.

Theorem CS.2 (A) *The graph of g is symmetric with respect to the origin $O = (0, 0)$ and so is an odd function (meaning that for all x, $g(-x) = -g(x)$).*

(B) *g is an increasing and continuous function on $\mathbb{R}$.*

(C) *The limit $\lim_{x\to+\infty} g(x)$ exists, and is equal to a number $h \leq 4$. Also the limit $\lim_{x\to-\infty} g(x)$ exists and equals $-h$.*

(D) *$1 < g(1) = k < 2$ and also $g(1) = k < h$.*

Proof. See Figure 4.1 above. (A) is obvious from calculus.

(B) If $x_1 < x_2$, then

$$g(x_2) - g(x_1) = \int_0^{x_2} f(t)\,dt - \int_0^{x_1} f(t)\,dt = \int_{x_1}^{x_2} f(t)\,dt > 0.$$

Thus g is increasing. To see that it is continuous, note that for all x, $0 < f(x) \leq 2$; for any $\epsilon > 0$, define $\delta = \frac{\epsilon}{2}$; then for any $x \geq 0$ and $y \geq 0$, if $|y - x| < 2$,

$$|g(x) - g(y)| = |\int_0^{x} f(t)\,dt - \int_0^{y} f(t)\,dt| = |\int_{x}^{y} f(t)\,dt| < 2|x - y| = \epsilon.$$

(C) For $x \geq 0$ define $F(x) = \left\{\begin{matrix} 2 & \text{if } 0 \leq x \leq 1 \\ 2/x^2 & \text{if } x > 1 \end{matrix}\right\}$. Then for $0 \leq x \leq 1$, $f(x) < 2$; for $1 < x$, $f(x) < 1/x^2$. Therefore, for all $x \geq 0$, $f(x) \leq F(x)$, and

$$\begin{aligned} g(x) = \int_0^{x} f(t)\,dt &< \int_0^{x} F(t)\,dt = 2 + \int_1^{x} 2t^{-2}\,dt \\ &= 2 + 2(-1)(x^{-1} - 1^{-1}) = 2 - 2/x + 2 = 4 - 2/x < 4. \end{aligned}$$

Since $g(x)$ is an increasing function and bounded above by 4, the limit $h = \lim_{x\to+\infty} g(x)$ exists and is less than or equal to 4.

As we observed in part (A), the graph of g is symmetric with respect to $(0, 0)$, so $\lim_{x\to-\infty} g(x) = -h$.

(D) First note that $f(0) = 2$, $f(\frac{1}{2}) = \frac{8}{5}$, and $f(1) = 1$. Then if $0 \leq t \leq \frac{1}{2}$, $\frac{8}{5} \leq f(t) \leq 2$, and if $\frac{1}{2} \leq t \leq 1$, $1 \leq f(t) \leq \frac{8}{5}$. Then

$$\tfrac{8}{5}(\tfrac{1}{2}) = \tfrac{4}{5} \leq \int_0^{\frac{1}{2}} f(t)\,dt \leq 2(\tfrac{1}{2}) = 1 \text{ and } 1(\tfrac{1}{2}) = \tfrac{1}{2} \leq \int_{\frac{1}{2}}^{1} f(t)\,dt \leq \tfrac{8}{5}(\tfrac{1}{2}) = \tfrac{4}{5},$$

so that $\frac{4}{5} + \frac{1}{2} = \frac{13}{10} \leq \int_0^{1} f(t)\,dt = \int_0^{\frac{1}{2}} f(t)\,dt + \int_{\frac{1}{2}}^{1} f(t)\,dt \leq 1 + \frac{4}{5} = \frac{9}{5}$.

This shows that $1 < \frac{13}{10} \leq g(1) = k \leq \frac{9}{5} < 2$; we have already seen that g is an increasing function, so $1 \leq g(1) = k < h$. $\square$

Definition CS.3 Since g is an increasing and continuous function mapping $\mathbb{R}$ into $]-h,+h[\subseteq \mathbb{R}$, by Exercise ARC.4, g has an inverse function which is also continuous and increasing on $\mathbb{R}$. We denote the inverse function of g by q. The graph of q (shown below in Figure 4.2) is the reflection over the line $\mathcal{L} = \{(x_1, x_2) \mid x_1 = x_2\}$ of the graph of g.

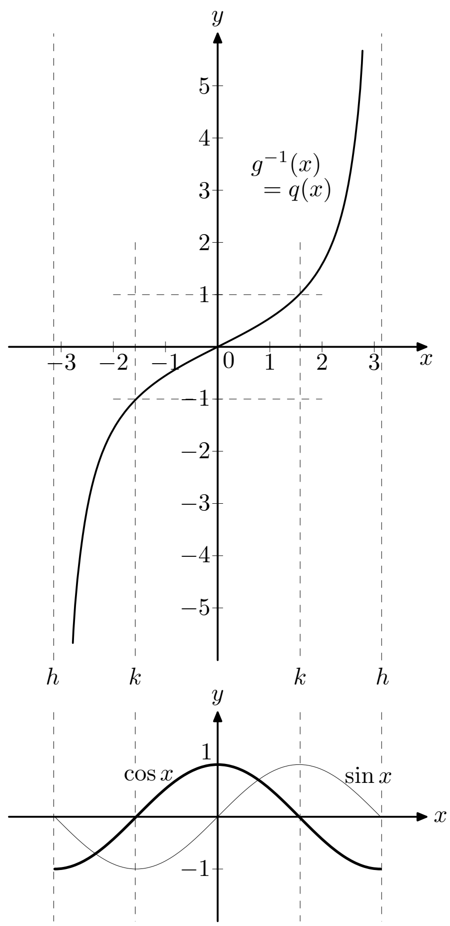


Fig. 4.2 The graphs of $q(x) = g^{-1}(x)$, $\sin x$, and $\cos x$ for Definition CS.3 and Heuristic Remark CS.4.

Heuristic Remark CS.4 At this stage we refer to the usual intuitive development of trigonometry, not to use as part of our development, but to give some guidance as to what definitions might be fruitful. From trigonometry, we have

$$\tan\tfrac{x}{2} = \frac{\sin\frac{x}{2}}{\cos\frac{x}{2}} = \frac{2\sin\frac{x}{2}\cos\frac{x}{2}}{2\cos^2\frac{x}{2}} = \frac{\sin x}{1+\cos x}.$$

This becomes

$$\frac{\sin x}{1+\cos x} = \frac{\sin^2 x}{\sin x(1+\cos x)} = \frac{1-\cos^2 x}{\sin x(1+\cos x)} = \frac{(1-\cos x)(1+\cos x)}{\sin x(1+\cos x)} = \frac{1-\cos x}{\sin x}.$$

Solving the two equations $\tan\frac{x}{2} = \frac{\sin x}{1+\cos x}$ and $\tan\frac{x}{2} = \frac{1-\cos x}{\sin x}$ for $\cos x$ and $\sin x$ in terms of $\tan\frac{x}{2}$ we get $\cos x = \frac{1-\tan^2\frac{x}{2}}{1+\tan^2\frac{x}{2}}$ and $\sin x = \frac{2\tan\frac{x}{2}}{1+\tan^2\frac{x}{2}}$. We use the last two equations to define sin and cos.

Definition/Remark CS.5

Define $\left\{\begin{aligned} \cos x &= \frac{1-q^2(x)}{1+q^2(x)} \\ \sin x &= \frac{2q(x)}{1+q^2(x)} \end{aligned}\right\}$ for $-h < x < h$. We would also like to define these functions at $x = -h$ and $x = h$, but the definition above won't work at these points because $q(-h)$ and $q(h)$ are undefined. So using the definitions of sin and cos on $]-h, h[$, we evaluate their limits at $-h$ and h and use them to complete the definitions.

$$\lim_{x\to h}\cos x = \lim_{x\to h}\frac{\frac{1}{q^2(x)}-1}{\frac{1}{q^2(x)}+1} = -1 \text{ so we define } \cos h = -1.$$

$$\lim_{x\to -h}\cos x = \lim_{x\to -h}\frac{\frac{1}{q^2(x)}-1}{\frac{1}{q^2(x)}+1} = -1 \text{ so we define } \cos(-h) = -1.$$

$$\lim_{x\to h}\sin x = \lim_{x\to h}\frac{\frac{2}{q(x)}}{\frac{1}{q^2(x)}+1} = 0 \text{ so we define } \sin h = 0.$$

$$\lim_{x\to -h}\sin x = \lim_{x\to -h}\frac{\frac{2}{q(x)}}{\frac{1}{q^2(x)}+1} = 0 \text{ so we define } \sin(-h) = 0.$$

This completes the definitions of sin and cos on $[-h, h]$. Their graphs are shown in the lower figure on the facing page.

Theorem CS.6 *For every* $x \in [-h, h]$

(A) sin *and* cos *are continuous at* x,

(B) $\cos^2 x + \sin^2 x = 1$,

(C) $|\cos x| \le 1$ *and* $|\sin(x)| \le 1$,

(D) $\cos 0 = 1$, $\sin 0 = 0$, $\cos k = \cos(-k) = 0$, $\sin k = 1$ *and* $\sin(-k) = -1$, *where* k *is the number whose existence is guaranteed by Theorem CS.2(D), such that* $k > 0$ *and* $q(k) = 1$, *and*

(E) $\cos x > 0$ *for* $-k < x < k$ *and* $\cos x < 0$ *for* $-h \le x < -k$ *and* $k < x \le h$*;* $\sin x > 0$ *for* $0 < x < h$ *and* $\sin x < 0$ *for* $-h < x < 0$.

(F) $\cos$ *is an increasing function on* $[-h, 0[$ *and a decreasing function on* $[0, h[$.

Proof. (A) sin and cos are continuous at h and $-h$ by Definition/Remark CS.5; they are continuous at all other points because q is a continuous function and the denominator $1 + q^2(x)$ is never zero.

By Definition/Remark CS.5 both (B) and (C) are true for $x = -h$ and $x = h$. So we may assume that $-h < x < h$.

(B) $\cos^2 x + \sin^2 x = \left(\frac{1-q^2(x)}{1+q^2(x)}\right)^2 + \left(\frac{2q(x)}{1+q^2(x)}\right)^2$
$= \frac{1-2q^2(x)+q^4+4q^2(x)}{(1+q^2(x))^2} = \frac{(1+q^2(x))^2}{(1+q^2(x))^2} = 1.$

(C) If $|\cos x| > 1$, then $\cos^2 x > 1$; if $|\sin x| > 1$, then $\sin^2 x > 1$; in either case $\cos^2 x + \sin^2 x$ would be greater than 1, contradicting part (1).

(D) The results all follow from the definitions of the functions sin and cos in Definition/Remark CS.5. $\cos 0 = 1$ and $\sin 0 = 0$ because $q(0) = 0$; $\cos k = \cos(-k) = 0$, $\sin k = 1$ and $\sin(-k) = -1$ because $q(k) = 1$ and $q(-k) = -1$.

(E) Recall that q is a one-to-one increasing function. In Definition/Remark CS.4, the numerator of the expression for $\cos x$ is $1 - q^2(x)$, and the denominator is always positive. For $-k < x < k$, $|q(x)| < 1$ so that $1 - q^2(x) > 0$. For $-h < x < -k$ $q(x) < -1$ so that $1 - q^2(x) < 0$; and for $k < x < h$, $q(x) > 1$ so that $1 - q^2(x) < 0$. Finally, $\cos(-h) = \cos(h) = -1$ by Definition/Remark CS.4.

The numerator of the expression for $\sin x$ is $2q(x)$ and the denominator is always positive; for $0 < x < h$, $q(x) > 0$; for $-h < x < 0$, $q(x) < 0$.

(F) As in (E), we use the fact that q is one-to-one and increasing. As x increases from $-h$ to $-k$ to 0, $\cos x = \frac{1-q^2(x)}{1+q^2(x)}$ increases from -1 to 0 to 1; as x increases from 0 to k to h, $\cos x = \frac{1-q^2(x)}{1+q^2(x)}$ decreases from 1 to 0 to -1. $\square$

Theorem CS.7 *If* $-h < x < h$*, the derivative* q' *of* q *exists at* x *and* $q'(x) = \frac{1+q^2(x)}{2}$.

Proof. For every real number x, $g(x) = \int\limits_0^x \frac{2}{1+t^2}\,dt$. By the Fundamental Theorem of Calculus, $g'(x) = \frac{2}{1+x^2}$. Since q is the inverse of g, and using the quotient rule for derivatives, $q'(x) = \frac{1}{g'(q(x))} = \frac{1+q^2(x)}{2}$. $\square$

Theorem CS.8 *If* $-h < x < h$, $\cos' x = -\sin x$ *and* $\sin' x = \cos x$.

Proof. (I) Since $\cos x = \frac{1-q^2(x)}{1+q^2(x))}$,

$$\cos' x = \frac{(1+q^2(x))(-2q(x)q'(x))-(1-q^2(x))2q(x)q'(x)}{(1+q^2(x))^2}$$
$$= \frac{[-2q(x)(1+q^2(x))-(1-q^2(x))2q(x)]q'(x)}{(1+q^2(x))^2}$$
$$= \frac{[-2q(x)-2q^3(x)-2q(x)+2q^3(x)]q'(x)}{(1+q^2(x))^2} = \left(\frac{-4q(x)}{(1+q^2(x))^2}\right)q'(x).$$

Using Theorem CS.7,

$$\cos' x = \left(\frac{-4q(x)}{(1+q^2(x))^2}\right)\left(\frac{1+q^2(x)}{2}\right) = \frac{-2q(x)}{1+q^2(x)} = -\sin x.$$

(II) Since $\sin x = \frac{2q(x)}{1+q^2(x)}$,

$$\sin' x = \frac{(1+q^2(x))2q'(x)-(2q(x))2q(x)q'(x)}{(1+q^2(x))^2} = \frac{(1+q^2(x)-2q^2(x))2q'(x)}{(1+q^2(x))^2}$$
$$= \left(\frac{1-q^2(x)}{(1+q^2(x))^2}\right)(2q'(x)) = \left(\frac{1-q^2(x)}{(1+q^2(x))^2}\right)\cdot 2\cdot\left(\frac{1+q^2(x)}{2}\right) = \frac{1-q^2(x)}{1+q^2(x)} = \cos x. \quad \square$$

Theorem CS.9 *In this theorem we confine our attention to the segment* $[-h, h]$ *because so far we have defined* cos *and* sin *only on that segment.*

(I) cos *has a maximum of* 1 *at* 0*, and a minimum of* -1 *at* $-h$ *and at* h*. These are relative and absolute maxima (minima).*

(II) sin *has a maximum of* 1 *at* k*, and a minimum of* -1 *at* $-k$*. These are relative and absolute maxima (minima).*

Proof. In this proof we will use the results of Theorem CS.6 and Theorem CS.8 without further reference.

(I) Since $\cos' x = -\sin x$ and $-\sin x$ is positive for $-k < x < 0$, is negative for $0 < x < h$, and is 0 at $x = 0$, 0 is a relative maximum of cos. We know that $\cos(0) = 1$, and since for all x, $\cos x \le 1$, this is an absolute maximum. Since $-\sin x$ is positive for $-h < x < 0$ and is negative for $0 < x < h$, cos has an absolute minimum of -1 at both h and $-h$.

(II) Exercise CS.1. $\square$

Definition CS.10 (A) Let p be a positive real number. A real valued function f is **periodic of period** p iff for every real number t, and every integer n, $f(t + np) = f(t)$.

(B) In Definition/Remark CS.5 we defined cos and sin on the interval $[-h, h]$. We now extend this definition by periodicity so that cos and sin are defined on $\mathbb{R}$.

For each $t \in \mathbb{R}$, there exists a unique integer n and a unique number $x \in]-h, h]$ such that $t = x + 2hn$. Define $\cos t = \cos x$ and $\sin t = \sin x$.

Theorem CS.11 *For every real number t,* $\cos(-t) = \cos t$ *and* $\sin(-t) = -\sin t$.

Proof. (Case 1: $-h < t < h$.) By Definition/Remark CS.5, $\cos(-t) = \frac{1-q^2(-t)}{1+q^2(-t)} = \frac{1-((-q(t))^2}{1+(-q(t))^2} = \frac{1-(q(t))^2}{1+(q(t))^2} = \cos t$ and $\sin(-t) = \frac{2q(-t)}{1+(q^2(-t))} = \frac{-2q(t)}{1+q^2(t)} = -\sin t$.

(Case 2: $|t| = h$.) By Definition/Remark CS.5, $\cos h = \cos(-h) = -1$. Moreover, $\sin(-h) = \sin h = 0$.

(Case 3:) If t is any real number, then by Definition CS.10 there exists an integer n such that $t = s + 2hn$, where $-h < s < h$. So, $\cos(-t) = \cos(-s - 2hn) = \cos(-s) = \cos s$ and $\cos t = \cos(s + 2hn) = \cos s$, so $\cos(-t) = \cos t$. Moreover, $\sin(-t) = \sin(= s - 2hn) = \sin(-s) = -\sin s = -\sin(s + 2hn) = -\sin t$. □

Theorem CS.12 *The functions* cos *and* sin *defined on* $\mathbb{R}$ *in Definition CS.10 are*

(A) *continuous, and*

(B) *periodic of period* $2h$.

Proof. (A) Since by definition, $\cos(-h) = \cos h = -1$, and $\sin(-h) = \sin h = 0$, and both functions are continuous on $[-h, h]$, they are continuous on $\mathbb{R}$.

(B) In this proof, let f be either cos or sin. Let t be any real number and n any integer, and let $x = t + 2hn$. There exist integers m_1 and m_2 such that $y_1 = t + 2hm_1 \in]-h, h]$, and $y_2 = x + 2hm_2 \in]-h, h]$; substituting the expression for x just above, we have $y_2 = x + 2hm_2 = t + 2hn + 2hm_2 = t + 2h(n + m_2) \in]-h, h]$.

The difference $y_2 - y_2 = 2h(n + m_2 - m_1)$ is an integral multiple of $2h$, and since both y_1 and y_2 belong to $]-h, h]$, $y_1 = y_2$.

$$y_1 = t + 2hm_1 = x + 2hm_2.$$

By Definition CS.10,

$$f(y_1) = f(t + 2hm_1) = f(t) \text{ and } f(y_1) = f(x + 2hm_2) = f(x),$$

so that $f(t) = f(x)$; thus f is periodic of period $2h$. □

4.2 Cosine, sine, and the unit circle

Definition CS.13 (A) Let r be any real number greater than 0; define $\mathcal{C}((0,0); r)$ to be the set $\{(x_1, x_2) \mid x_1^2 + x_2^2 = r^2\}$; this set is the **circle with**

center $(0,0)$ **and radius** r. We will have special interest in $\mathcal{C}((0,0);1) = \{(x_1,x_2) \mid x_1^2 + x_2^2 = 1\}$, which is called the **unit circle**.

(B) The **inside** of a circle $\mathcal{C}(O;r)$ is the set $\{X \mid \operatorname{dis}(X,O) < r\}$, denoted $\operatorname{ins}\mathcal{C}(O;r)$; the **enclosure** of $\mathcal{C}(O;r)$ is the set $\{X \mid \operatorname{dis}(X,O) \le r\}$, denoted $\operatorname{enc}\mathcal{C}(O;r)$. If $\mathcal{L}$ is a line and $O \in \mathcal{L}$, the set $\mathcal{L} \cap \operatorname{enc}\mathcal{C}(O;r)$ is **a diameter** of the circle; if $\mathcal{C}(O;r) \cap \mathcal{L} = \{X,Y\}$, then the number $\operatorname{dis}(X,Y)$ is called **the diameter** of $\mathcal{C}(O;r)$; clearly, the diameter of $\mathcal{C}(O;r)$ is equal to $2r$.

(C) The **circumference** of a circle is its arc length. The number π is defined to be the ratio of the circumference of a circle to its diameter. Since the diameter of the unit circle is 2, its circumference is 2π.

Theorem CS.14 (A) *The unit circle* $\mathcal{C}((0,0);1)$ *with center* $(0,0)$ *and radius 1 is the set*

$$\mathcal{E} = \{(\cos t, \sin t) \mid -h \le t \le h\}.$$

(B) *The unit circle is also equal to*

$$\begin{aligned}
&\{(\cos t, \sin t) \mid -h < t \le h\} = \{(\cos t, \sin t) \mid -h \le t < h\} \\
&\quad = \{(\cos t, \sin t) \mid 0 \le t \le 2h\} = \{(\cos t, \sin t) \mid 0 \le t < 2h\} \\
&\qquad = \{(\cos t, \sin t) \mid -2h < t \le 0\}.
\end{aligned}$$

Proof. (A) (I) By Definition/Remark CS.5 and Theorem CS.6(B), for all $t \in]-h,h]$, $\cos^2 t + \sin^2 t = 1$, so every member of $\mathcal{E}$ is a member of $\mathcal{C}((0,0);1)$.

(II) To prove that every member of $\mathcal{C}((0,0);1)$ is a member of $\mathcal{E} = \{(\cos t, \sin t) \mid -h \le t \le h\}$ we consider eight cases.

(Case 1: $x_1 = 1$ and $x_2 = 0$.) Let $t = 0$; then $x_1 = \cos 0 = 1$ and $x_2 = \sin 0 = 0$.

(Case 2: $x_1 = -1$ and $x_2 = 0$.) Let $t = h$ or $t = -h$; then $x_1 = \cos h = -1$ and $x_2 = \sin h = 0$.

(Case 3: $x_1 = 0$ and $x_2 = 1$.) Let $t = k$; then $x_1 = \cos k = 0$ and $x_2 = \sin k = 1$.

(Case 4: $x_1 = 0$ and $x_2 = -1$.) Let $t = -k$; then $x_1 = \cos(-k) = 0$ and $x_2 = \sin(-k) = -1$.

(Case 5: $0 < x_1 < 1$ and $0 < x_2 < 1$.) Since $\cos 0 = 1$ and $\cos k = 0$, by the Intermediate Value Theorem of calculus there exists a number $t \in]0,k[$ such that $\cos t = x_1$.

The function sin is continuous, $\sin 0 = 0$, $\sin k = 1$, and $0 < t < 1$, so that $0 < \sin t < 1$. By definition of the unit circle, ${x_1}^2 + {x_2}^2 = 1$; by Theorem CS.6(B) $\cos^2 t + \sin^2 t = 1$; since $\cos t = x_1$, $\cos^2 t + \sin^2 t = {x_1}^2 + \sin^2 t = 1$ so that $\sin^2 t = {x_2}^2$; hence $\sin t = x_2$.

(Case 6: $0 < x_1 < 1$ and $-1 < x_2 < 0$.) By the Intermediate Value Theorem there exists a number $t \in]-k, 0[$ such that $\cos t = x_1$. Since $\sin t \in]-1, 0[$ and $\cos^2 t + \sin^2 t = 1$, by reasoning similar to that in Case 5, $\sin t = x_2$.

(Case 7: $-1 < x_1 < 0$ and $0 < x_2 < 1$.) By the Intermediate Value Theorem there exists a number $t \in]k, h[$ such that $\cos t = x_1$. Since $\sin t \in]0, 1[$ and $\cos^2 t + \sin^2 t = 1$, by reasoning similar to that in Case 5, $\sin t = x_2$.

(Case 8: $-1 < x_1 < 0$ and $-1 < x_2 < 0$.) By the Intermediate Value Theorem there exists a number $t \in]-h, -k[$ such that $\cos t = x_1$. Since $\sin t \in]-1, 0[$ and $\cos^2 t + \sin^2 t = 1$, by reasoning similar to that in Case 5, $\sin t = x_2$.

(B) In part (A), the end point $(\cos h, \sin h) = (\cos(-h), \sin(-h))$ is included twice; thus either of the next two formulations is correct, since they merely omit this redundancy. The other formulations are true since cos and sin are periodic of period $2h$. □

Theorem CS.15 (A) *The arc length of the unit circle is* $2h$*, where* h *is the positive real number defined in Theorem CS.2(C).*

(B) *Let* $h = \lim_{x \to +\infty} g(x)$*, as defined in Theorem CS.2(C); then* $h = \pi$*.*

We will show that $k = \frac{\pi}{2}$ in Theorem CS.25.

Proof. (A) Since the unit circle is the set $\{(\cos t, \sin t) \mid 0 \leq t \leq 2h\}$, by Theorem ARC.11 its arc length is

$$\int_0^{2h} \sqrt{(\cos' t)^2 + (\sin' t)^2}\, dt = \int_0^{2h} \sqrt{(-\sin t)^2 + (\cos t)^2}\, dt = \int_0^{2h} 1\, dt = 2h.$$

(B) By part (A), the arc length of the unit circle is $2h$; by Definition CS.13(C), this is 2π. □

Definition CS.16 Define **cis** to be the function mapping $\mathbb{R}$ into $\mathbb{R}^2$, whose value at each t is $\operatorname{cis} t = (\cos t, \sin t)$.

Remark CS.17 (A) From Definition CS.16, $\operatorname{cis} 0 = (1, 0)$, $\operatorname{cis} k = (0, 1)$, and $\operatorname{cis} \pi = (-1, 0)$. The notation cis is intended to suggest the complex number $\cos t + i \sin t$. The reader should bear in mind that $\operatorname{cis} t$ refers to a point of the plane, specifically, a point of the unit circle, not to a number, as with $\cos t$ and $\sin t$.

(B) We will often use the symbol $\operatorname{cis}([a, b])$ to mean $\{\operatorname{cis}(x) \mid x \in [a, b]\}$, and $\operatorname{cis}([a, b[)$ to mean $\{\operatorname{cis}(x) \mid x \in [a, b[\}$. Again, as always, when we use the notation $[a, b]$ or $[a, b[$ to describe an interval of numbers, it is specifically understood that $a < b$.

(C) On a circle, each pair of distinct points $S = \operatorname{cis} s$ and $T = \operatorname{cis} t$ defines two arcs, and we need to specify which one we are talking about. We do this by specifying the interval upon which cis is defined. If $s < t$, we describe the arc traversed in the positive direction from s to t by $\operatorname{cis}[s, t]$.

The complementary arc defined by $\operatorname{cis} s$ and $\operatorname{cis} t$ (the one that "goes around the other way," containing the point $\operatorname{cis} 0$) is the set $\{\operatorname{cis} u \mid u \notin]s, t[\,\}$. If $s \neq 0$, this arc cannot be written as the image under cis of an interval which is a subset of $[0, 2\pi[$. It could be written as $\operatorname{cis}[t - 2\pi, s]$ or $\operatorname{cis}[t, s + 2\pi]$.

(D) In the development up through Theorem CS.9,the domain of definition of sin and cos was the interval $[-h, h] = [\pi, \pi]$. With Definition CS.10(B), we extended this definition to the whole real line; so now it does not really make any difference what interval of length 2π we use as a "primary" domain of definition. In Theorem CS.14(B) we made it "official" that any of the intervals $[-\pi, \pi]$, $]-\pi, \pi]$, $[-\pi, \pi[$, $[0, 2\pi]$, $[0, 2\pi[$, or $]-2\pi, 0]$ will do for this purpose, as cis maps any of these intervals onto the unit circle.

(E) Bearing in mind that where both end-points are included, the function cis is not one-to-one, the unit circle may be referred to in any of the following ways: $\operatorname{cis}[-\pi, \pi]$, $\operatorname{cis}\,]-\pi, \pi]$, $\operatorname{cis}[-\pi, \pi[$, $\operatorname{cis}[0, 2\pi]$, $\operatorname{cis}\,]0, 2\pi]$, $\operatorname{cis}[0, 2\pi[$; or, for that matter, as $\operatorname{cis}[a, a + 2\pi]$, $\operatorname{cis}\,]a, a + 2\pi]$, $\operatorname{cis}[a, a + 2\pi[$; or $\operatorname{cis}[a - 2\pi, a]$, $\operatorname{cis}\,]a - 2\pi, a]$ $\operatorname{cis}[a - 2\pi, a[$, where a can be any real number.

Theorem CS.18 *The mapping* cis *on the interval* $[0, 2\pi[$*, or on any interval* $[a, a + 2\pi[$ *or* $]a, a + 2\pi]$*, is continuous and one-to-one onto the unit circle; it is one-to-one on any interval* $[a, b]$ *where* $0 < b - a < 2\pi$*.*

Proof. Since both cos and sin are continuous, cis is continuous.

We show that cis is one-to-one on the interval $[-\pi, \pi[$. By Theorem CS.6(F), cos is an increasing function on $[-\pi, 0[$ and a decreasing function on $[0, \pi[$, hence is one-to-one on both intervals. Thus if $t > s$ and $\operatorname{cis} t = \operatorname{cis} s$, not both t and s can be in $[-\pi, 0[$ and not both can be in $[0, \pi[$. Therefore $t \in [0, \pi[$ and $s \in [-\pi, 0[$. Then by Theorem CS.6(E), $\sin t \geq 0$ and $\sin s < 0$, so that $\operatorname{cis} t \neq \operatorname{cis} s$, a contradiction.

If $x \in [-\pi, 0[$, then $\operatorname{cis}(x + 2\pi) = \operatorname{cis} x$ so the values taken by cis on $[0, 2\pi[$ are exactly those taken on $[-\pi, \pi[$. Hence if there are two points in $[0, 2\pi[$ where cis takes the same value, there must be two points in $[-\pi, \pi[$ where cis takes the same value, which we have shown to be impossible. Therefore cis is one-to-one on $[0, 2\pi[$, and by periodicity cis is one-to-one on any interval $[a, a+2\pi[$ or $]a, a+2\pi]$.

Finally, by Theorem CS.6(B), cis maps $[0, 2\pi[$ into the unit circle; and by Theorem CS.14 cis maps $[0, 2\pi[$ onto the unit circle. $\square$

Theorem CS.19 *Let φ be the mapping of $[0, 2\pi]$ into $\mathbb{R}$ such that for every member t of $[0, 2\pi]$, $\varphi(t) = \mathbb{L}(\operatorname{cis}[0, t])$, the length of the arc of the unit circle from $\operatorname{cis} 0 = (1, 0)$ to $\operatorname{cis} t$. Then $\varphi(0) = 0$, and*

(A) *for every number $t \in [0, 2\pi]$, $\varphi(t) = t$;*

(B) *if s and t are real numbers such that $0 \leq s < t \leq 2\pi$, then the arc length $\mathbb{L}(\operatorname{cis}[s, t])$ is $t - s$; and*

(C) *the arc length $\mathbb{L}(\operatorname{cis}[0, k])$ of the unit circle from $(1, 0) = \operatorname{cis} 0$ to $(0, 1) = \operatorname{cis} k$ is k and the arc length $\mathbb{L}(\operatorname{cis}[k, \pi])$ from $(0, 1) = \operatorname{cis} k$ to $(-1, 0) = \operatorname{cis} h$ is $\pi - k$.*

Proof. Note that for all numbers s, $(\cos' s)^2 + (\sin' s)^2 = (\sin s)^2 + (\cos s)^2 = 1$.

(A) By Theorem ARC.11, for every member t of $]0, 2\pi]$,

$$\varphi(t) = \int_0^t \left((\cos' s)^2 + (\sin' s)^2\right) ds = \int_0^t 1\, ds = t.$$

(B) Since $\operatorname{cis}[s, t]$ is the image under $\operatorname{cis} u$ of $[s, t]$, we may apply Theorem ARC.11, and the arc length $\mathbb{L}(\operatorname{cis}[s, t])$ is

$$\int_s^t \sqrt{\left(\cos' u\right)^2 + \left(\sin' u\right)^2}\, du = \int_s^t du = t - s.$$

(C) This follows immediately from part (B). $\square$

4.3 Sides of a line intersecting a circle

Remark CS.20 In the proofs of future theorems it will be important to be able to tell what points are on what sides of lines. In the next series of results, we will show explicitly that a line through two arbitrary points $A = \operatorname{cis} a$ and $B = \operatorname{cis} b$ on a unit circle (and this could easily be extended to any circle) divides it into two arcs as described in Chapter 3 Definition ARC.1, each of them being the image under the (continuous) mapping cis of an interval of real numbers.

These results will also show that, speaking informally, cis preserves "sense" on the unit circle; that is, if $a < b$, then rotating $A = \operatorname{cis} a$ to $B = \operatorname{cis} b$ is a rotation in the "positive direction"; or, moving t from a toward b moves $\operatorname{cis} t$ from A toward B. The problem with these statements is that the terms "sense," "positive direction," and "toward" are not well defined mathematical terms. In more formal language mathematical language, the idea can be expressed as follows: let $a < b < a + \pi$, so that the points $A = \operatorname{cis} a$ and $B = \operatorname{cis} b$ form an angle $\angle AOB$; then whenever $t \in]a, b[$, $\operatorname{cis} t$ belongs to the inside of this angle; more to the point, $\operatorname{cis} t$ lies on the $B = \operatorname{cis} b$-side of $\overleftrightarrow{OA}$.

Of course, all this is intuitively obvious, inasmuch as the mapping cis is continuous. But it does seem to require proof.

To start things off right, we should point out something else that many will consider quite obvious: a line can intersect a circle in at most two points, and a line contains a point on the inside of a circle iff it intersects the circle in two points. Moreover, if a line intersects a circle $\mathcal{C}(O; r)$ in two points A and B, the line of symmetry of $\angle AOB$ intersects the line $\overleftrightarrow{AB}$ at a point which is inside the circle. The proof of this is Exercise CS.13.

As an aid to keeping things straight, we strongly advise the reader to construct copious figures while reading the following proofs.

Theorem CS.21 *Let $A = \operatorname{cis} a$, $B = \operatorname{cis} b$, and $C = \operatorname{cis} c$ be points on the unit circle $\mathcal{C}(O; 1)$ such that $0 < b - a < 2\pi$. Then if $a < c < b$, every point $T = \operatorname{cis} t$ where $a < t < b$ belongs to the C-side of $\overleftrightarrow{AB}$.*

Proof. (A) Let $x = \operatorname{lub}\{t \mid \operatorname{cis}[c, t[\subseteq C\text{-side of } \overleftrightarrow{AB}\}$. If $\operatorname{cis} x = B$ then $\operatorname{cis}[c, b[\subseteq C$-side of $\overleftrightarrow{AB}$. If $\operatorname{cis} x \neq B$, $\operatorname{cis} x \notin \overleftrightarrow{AB}$ so it is either in the C-side or the side opposite the C-side of $\overleftrightarrow{AB}$. Thus there exists a number ϵ such that for all $Z \in \overleftrightarrow{AB}$, $|\operatorname{cis} x - Z| > \epsilon$. By continuity of cis, there exists $\delta > 0$ such that if $|x - w| < \delta$, then $|\operatorname{cis} x - \operatorname{cis} w| < \epsilon$.

(Case 1:) If $\operatorname{cis} x \in C$-side $\overleftrightarrow{AB}$, then because $x = \operatorname{lub}\{t \mid \operatorname{cis}[c, t[\subseteq C\text{-side of } \overleftrightarrow{AB}\}$ there exists a w such that $|w - x| < \delta$ and $\operatorname{cis} w$ is in the side of $\overleftrightarrow{AB}$ opposite the C-side.

(Case 2:) If $\operatorname{cis} x \notin C$-side $\overleftrightarrow{AB}$, then it must belong to the side opposite C, and there exists a w such that $|w - x| < \delta$ and $\operatorname{cis} w \in C$-side of $\overleftrightarrow{AB}$.

In either case $\operatorname{cis} x$ and $\operatorname{cis} w$ are on opposite sides of $\overleftrightarrow{AB}$, so by Theorem PSH.11 there exists a point $Z \in \overleftrightarrow{AB}$ such that $\operatorname{cis} x$–Z–$\operatorname{cis} w$, and $\epsilon < |\operatorname{cis} x - Z| < |\operatorname{cis} x - \operatorname{cis} w| < \epsilon$, a contradiction. Therefore all the points $\operatorname{cis} t$ where $t \in [c, b[$ belong to the C-side of $\overleftrightarrow{AB}$.

(B) Let $x = \operatorname{glb}\{t \mid \operatorname{cis}]t, c]\} \subseteq C$-side of $\overleftrightarrow{AB}$. If $\operatorname{cis} x = A$ then $\operatorname{cis}]x, c] \subseteq C$-side of $\overleftrightarrow{AB}$. If $\operatorname{cis} x \neq A$, $\operatorname{cis} x \notin \overleftrightarrow{AB}$ so it is either in the C-side or in the side opposite the C-side of $\overleftrightarrow{AB}$. Thus there exists a number ϵ such that for all $Z \in \overleftrightarrow{AB}$, $|\operatorname{cis} x - Z| > \epsilon$. By continuity of cis, there exists $\delta > 0$ such that if $|x - w| < \delta$, then $|\operatorname{cis} x - \operatorname{cis} w| < \epsilon$.

Substituting $x = \operatorname{glb}\{t \mid \operatorname{cis}\,]t, c]\}$ for $x = \operatorname{lub}\{t \mid \operatorname{cis}[c, t[\}$, the balance of the proof for (B) is almost word-for-word as in part (A). Therefore all the points $\operatorname{cis} t$ where $t \in\,]a, c]$ belong to the C-side of $\overleftrightarrow{AB}$, and it follows that every point of $\operatorname{cis}\,]a, b[$ is on the C-side. $\square$

Theorem CS.22 *Let* $A = \operatorname{cis} a$ *and* $B = \operatorname{cis} b$ *be any points on the unit circle* $\mathcal{C}(O; 1)$.

(A) *Either*

(I) *A–O–B and the notation may be chosen (and the points possibly renamed) so that* $b = a + \pi$, *or*

(II) *A–O–B is false and the notation may be chosen (and the points possibly renamed) so that* $b - a < \pi$.

(B) *In either case (I) or (II), the points A and B define two arcs on the unit circle,* $\operatorname{cis}[a, b]$ *and* $\operatorname{cis}[b, a + 2\pi]$*; and*

(1) *all the points of* $\operatorname{cis}\,]a, b[$ *are on the same side of* $\overleftrightarrow{AB}$*;*

(2) *all the points of* $\operatorname{cis}\,]b, a + 2\pi[$ *are on the same side of* $\overleftrightarrow{AB}$*; and*

(3) *every point of* $\operatorname{cis}\,]a, b[$ *is on the opposite side of* $\overleftrightarrow{AB}$ *from every point of* $\operatorname{cis}\,]b, a + 2\pi[$.

(C) *In case (II), if* $b - a < \pi$ *then* $\operatorname{cis}\,]a, b[\, \subseteq \operatorname{ins} \angle AOB$, *and* $\operatorname{cis}\,]b, a + 2\pi[\, \subseteq \operatorname{out} \angle AOB$*;*

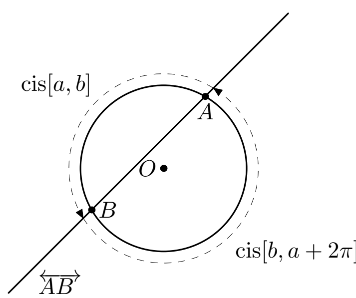


Fig. 4.3 A line divides a circle into two arcs.

Proof. See Figure 4.3. (Case I: *A–O–B*) We can choose a and b (possibly renaming the points) so that $b = a + \pi$ and therefore $a < b$. In this case, $\angle AOB$ is not defined. One of the arcs between A and B is $\operatorname{cis}[a, b]$. Also, $A = \operatorname{cis}(a + 2\pi)$ so that $b < a + 2\pi$, and one of the arcs between A and B is $\operatorname{cis}[b, a + 2\pi]$. Let $\mathcal{L}$ be the perpendicular bisector of $\overline{AB}$, C be the point of intersection of $\mathcal{L}$ with $\operatorname{cis}\,]a,b[$, and let C' be the point of intersection of $\mathcal{L}$ with $\operatorname{cis}\,]b,a+2\pi[$. These points are on opposite sides of the line $\overleftrightarrow{AB}$, and by Theorem CS.21, every point of $\operatorname{cis}\,]a,b[$ is on the C-side, and every point of $\operatorname{cis}\,]b,a+2\pi[$ is on the C'-side of $\overleftrightarrow{AB}$. This proves (B)(1), (2) and (3) for Case (I).

(Case II: A–O–B is false.) Initially choose a and b so that both $0 \leq a < 2\pi$, $0 \leq b < 2\pi$ and $a < b$. If $b - a < \pi$ we leave the notation "as is"; if $b - a > \pi$, then $a - b < -\pi$. Let $a' = b$ and $b' = a + 2\pi$, so that $\operatorname{cis} a'$ and $\operatorname{cis} b'$ are the original points A and B, which have now been renamed. Then $b' - a' = a + 2\pi - b = a - b + 2\pi < -\pi + 2\pi = \pi$. Either way we have found numbers a and b such that $\operatorname{cis} a$ and $\operatorname{cis} b$ are the given points A and B, $a < b$ and $b - a < \pi$. Moreover, $b < a + \pi < a + 2\pi$ and $\operatorname{cis} a = \operatorname{cis}(a + 2\pi)$.

This proves part (A) of the theorem, and shows that one of the arcs defined by A and B is $\operatorname{cis}[a,b]$, and the other is $\operatorname{cis}[b,a+2\pi]$. The first of these arcs subtends the angle $\angle AOB$.

Let A' and B' be points on the unit circle such that A'–O–A and B'–O–B, so that $A' = \operatorname{cis}(a+\pi)$ and $B' = \operatorname{cis}(b+\pi)$. Apply Theorem CS.21 to the sides of $\overleftrightarrow{AA'}$ and $\overleftrightarrow{BB'}$. If $a < t < b$, then $a < t < b < a + \pi$ so by Theorem CS.21, $T = \operatorname{cis} t$ is on the B-side of $\overleftrightarrow{AA'}$; also since $b - \pi < a < t < b$, T is on the A-side of $\overleftrightarrow{BB'}$; by Definition PSH.36 $T \in \operatorname{ins}\angle AOB$, so that $\operatorname{cis}\,]a,b[\, \subseteq \operatorname{ins}\angle AOB$, proving the first assertion of part(C).

If $b < t < a + 2\pi$ then either $b < t < b + \pi$ or $b + \pi \leq t < a + 2\pi$. Note that $b < a + \pi < b + \pi$ and the points $A' = \operatorname{cis}(a+\pi)$ and A are on opposite sides of $\overleftrightarrow{BB'}$; also $a + \pi < b + \pi < a + 2\pi$ and the points $B' = \operatorname{cis}(b+\pi)$ and B are on opposite sides of $\overleftrightarrow{AA'}$.

If $b < t < b+\pi$, by Theorem CS.21 $\operatorname{cis} t \in A' = \operatorname{cis}(a+\pi)$-side of $\overleftrightarrow{BB'}$ which is the side of $\overleftrightarrow{BB'}$ opposite A, hence $\operatorname{cis} t \in \operatorname{out}\angle AOB$. If $b+\pi \leq t < a+2\pi$, by Theorem CS.21 $\operatorname{cis} t \in B' = \operatorname{cis}(b+\pi)$-side of $\overleftrightarrow{AA'}$ which is the side of $\overleftrightarrow{AA'}$ opposite B, hence $\operatorname{cis} t \in \operatorname{out}\angle AOB$; thus $\operatorname{cis}\,]b,a+2\pi[\, \subseteq \operatorname{out}\angle AOB$, proving the second assertion of (C). By Theorem PSH.41, $\operatorname{ins}\angle AOB \cap \operatorname{out}\angle AOB = \emptyset$, so that $\operatorname{cis}\,]a,b[\, \cap \operatorname{cis}\,]b,a+2\pi[\, = \emptyset$.

Applying Theorem CS.21 to the sides of $\overleftrightarrow{AB}$, we find that all the points of $\operatorname{cis}\,]a,b[$ are on the same side of $\overleftrightarrow{AB}$, and all the points of $\operatorname{cis}\,]b,a+2\pi[$ are on the same side of $\overleftrightarrow{AB}$.

Every point of $\mathcal{C}(O;1) \setminus \{A,B\}$ is a member of either $\operatorname{cis}\,]a,b[$ or $\operatorname{cis}\,]b,a+2\pi[$, and therefore to one of the sides of $\overleftrightarrow{AB}$. Let $\overleftrightarrow{CC'}$ be the line of symmetry of $\angle AOB$, where both C and C' are points on the circle. By Exercise CS.13

$\overleftrightarrow{CC'}$ intersects $\overleftrightarrow{AB}$ at a point D such that $\operatorname{dis}(O,D) < 1$, so that C–D–C', and hence C and C' are on opposite sides of $\overleftrightarrow{AB}$.

Choose the notation so that $C \in \operatorname{cis}\,]a,b[$, so that $\operatorname{cis}\,]a,b[\subseteq C$-side of $\overleftrightarrow{AB}$. Since C' belongs to the side opposite C, $C' \notin \operatorname{cis}\,]a,b[$, and since $\operatorname{cis}\,]a,b[\cup \operatorname{cis}\,]b,a+2\pi[= \mathcal{C}(O;1) \setminus \{A,B\}$, $C' \in \operatorname{cis}\,]b,a+2\pi[$, and hence $\operatorname{cis}\,]b,a+2\pi[\subseteq C'$-side of $\overleftrightarrow{AB}$. Then every point of $\operatorname{cis}\,]a,b[$ is on the same side of $\overleftrightarrow{AB}$, and every point of $\operatorname{cis}\,]b,a+2\pi[$ is on the opposite side. This proves parts (B)(1),(2) and (3) for case (II). □

4.4 Isometry preserves arc length; $k = \frac{\pi}{2}$; summary

Lemma CS.23 (Preservation of arc length) *Let φ be an isometry of the plane such that $\varphi(O) = O'$; then φ preserves distance and maps the unit circle $\mathcal{C}(O;1)$ onto the unit circle $\mathcal{C}(O';1)$. Let s and t be real numbers such that $s < t$ and $t - s < 2\pi$, so that $\operatorname{cis}[s,t]$ is an arc on $\mathcal{C}(O;1)$ whose end points are $\operatorname{cis} s$ and $\operatorname{cis} t$.*

(A) *The set $\varphi(\operatorname{cis}[s,t])$ is an arc on the unit circle $\mathcal{C}(O';1)$ with end points $\varphi(\operatorname{cis} s)$ and $\varphi(\operatorname{cis} t)$, and φ maps $\operatorname{cis}[s,t]$ one-to-one onto $\varphi(\operatorname{cis}[s,t])$.*

(B) *The arc length $\mathbb{L}(\varphi(\operatorname{cis}[s,t]))$ of $\varphi(\operatorname{cis}[s,t])$ equals the arc length $\mathbb{L}(\operatorname{cis}[s,t])$ of $\operatorname{cis}[s,t]$.*

Proof. (A) Let $f = \varphi \circ \operatorname{cis}$. *Specht* Ch.21 Theorem LC.25.1 shows that isometries preserve distance. Since $\varphi(O) = O'$ this implies that the points $f(s) = \varphi(\operatorname{cis} s)$ belong to the unit circle $\mathcal{C}(O';1)$. It also implies that φ is continuous, and therefore $f = \varphi \circ \operatorname{cis}$ is continuous. By Theorem CS.18 cis is one-to-one on $[s,t]$ and since φ is an isometry, it is one-to-one, so that f is one-to-one on $[s,t]$, and its restriction to that interval has an inverse. Therefore $f = \varphi \circ \operatorname{cis}$ is a continuous one-to-one mapping of $[s,t]$ into the unit circle $\mathcal{C}(O';1)$.

Thus s and t define not only the arc $\operatorname{cis}[s,t]$, but also the image $\varphi(\operatorname{cis}[s,t]) = f[s,t]$, which is an arc of the unit circle $\mathcal{C}(O';1)$.

Moreover, $\operatorname{cis} u \in \operatorname{cis}[s,t]$ iff $u \in [s,t]$ iff $f(u) \in \varphi(\operatorname{cis}[s,t])$ so that the mapping φ maps $\operatorname{cis}[s,t]$ one-to-one onto $\varphi(\operatorname{cis}[s,t])$. This completes the proof of (A).

(B) Let $\mathcal{P} = \{s - t = t_0 < t_1 < t_2 < ... < t_n = 0\}$ be a partition of $[s,t]$. Since φ preserves distance, for each $j \in \{1,2,...,n\}$,

$$|\operatorname{cis} t_j - \operatorname{cis} t_{j-1}| = |\varphi(\operatorname{cis} t_j) - \varphi(\operatorname{cis} t_{j-1})| = |f(t_j) - f(t_{j-1})|.$$

The length of the arc $\varphi(\mathrm{cis}[s,t])$ is the least upper bound of all summations

$$\mathcal{S}_{\mathcal{P}}(f) = \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})|$$

which are exactly the same as the summations

$$\mathcal{S}_{\mathcal{P}}(\mathrm{cis}) = \sum_{j=1}^{n} |\,\mathrm{cis}(t_j) - \mathrm{cis}(t_{j-1})|$$

which define the length of the arc $\mathrm{cis}[s,t]$. Therefore its arc length is the same as that of its image under φ. $\square$

Remark CS.24 (A) In the next theorem we remedy a deficiency of our development so far. In Theorem CS.2(C) we established the number h (which has turned out to be π) as the limit $\lim_{x\to\infty} g(x)$. In part (D) of the same theorem we showed that the number $k = g(1)$ (that is, that number k for which $q(k) = 1$) is between 1 and 2 and is less than h. In Theorem CS.6(D) we showed that $\cos k = 0$ and in Theorem CS.9 we showed that $\sin k = 0$, so that $\mathrm{cis}\, k = (0,1)$. It's "obvious" from the picture that the arc length from $\mathrm{cis}\, 0$ to $\mathrm{cis}\, k$ is half that from $\mathrm{cis}\, 0$ to $\mathrm{cis}\, \pi$; but this hasn't been proved. We have not established the relation between k and h.

If it had been simple to calculate numerically the integral of $f(t)$ from 0 to 1, and from 0 to "infinity," we might have established this relation ere this, but since the standard method of calculation of these integrals involves the *arctan* function, which presupposes the definition of *tan* which in turn presupposes the definitions of *sin* and *cos*, and these functions are what we are trying to define, such an argument would be circular. In parts (C) and (D) of the next theorem, we prove that indeed, $k = h/2 = \pi/2$.

Theorem CS.25 *The arc length* $\mathbb{L}(\mathrm{cis}[0,k]) = \mathbb{L}(\mathrm{cis}[k,\pi])$ *so that* $\pi - k = k$ *and therefore* $k = \frac{\pi}{2}$.

Proof. For all numbers s, $(\cos' s)^2 + (\sin' s)^2 = (\sin s)^2 + (\cos s)^2 = 1$. Also by Theorem CS.19(B), if s and t are real numbers such that $0 \le s < t \le 2\pi$, then the arc length $\mathbb{L}(\mathrm{cis}[s,t])$ is $t - s$.

Let $\mathcal{L} = \overleftrightarrow{(0,0)(0,1)}$, and let $\mathcal{R}_{\mathcal{L}}$ be the mirror mapping (reflection) defined on $\mathbb{R}^2$ in *Specht* Ch.21 Definition LB.16 and further developed in Remark LC.22 and subsequently. Then $\mathcal{R}_{\mathcal{L}}$ maps each point $\mathrm{cis}\, t = (\cos t, \sin t)$, where $t \in [0,k]$, to the point $(-\cos t, \sin t) = (\cos \bar{t}, \sin \bar{t}) = \mathrm{cis}\, \bar{t}$ where $\bar{t} \in [k,\pi]$. Moreover, $\mathcal{R}_{\mathcal{L}}(\mathrm{cis}\, 0) = \mathrm{cis}\, \pi$; and because $\mathrm{cis}\, k \in \mathcal{L}$, $\mathcal{R}_{\mathcal{L}}(\mathrm{cis}\, k) = \mathrm{cis}\, k$. Then

$$\mathcal{R}_{\mathcal{L}}(\mathrm{cis}[0,k]) = (\mathrm{cis}[\pi,k]) = (\mathrm{cis}[k,\pi]).$$

Taking arc lengths and applying Lemma CS.23, we have

$$\mathbb{L}(\text{cis}[k,\pi]) = \mathbb{L}(\mathcal{R}_{\mathcal{L}}(\text{cis}[0,k])) = \mathbb{L}(\text{cis}[0,k]),$$

so that $\mathbb{L}(\text{cis}[k,\pi]) = \mathbb{L}(\text{cis}[0,k])$. Then by Theorem CS.19(B), the left-hand side is $\pi - k$, and the right-hand side is k, so that $k = \pi - k$ and $k = \frac{\pi}{2}$. $\square$

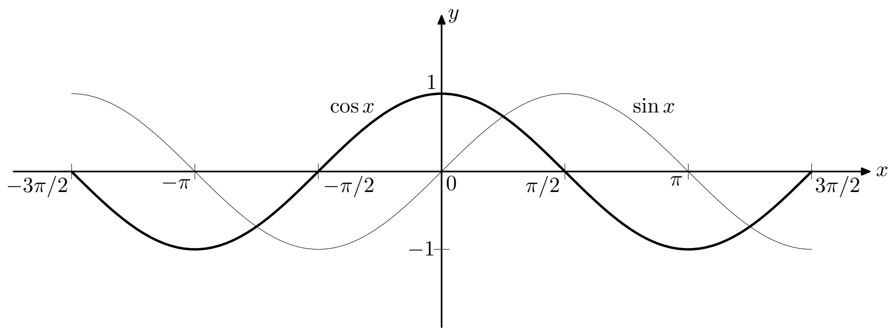


Fig. 4.4 Graphs of $\sin x$ and $\cos x$ for reference.

Theorem CS.26 (Summary) (A) $\cos$ *and* $\sin$ *are continuous functions, periodic of period* 2π*, mapping* $\mathbb{R}$ *onto* $[-1,1]$*; for every* $t \in \mathbb{R}$, $\cos t = \cos(-t)$ *and* $\sin t = -\sin(-t)$.

(B) *For every integer* n,

$\cos\frac{\pi}{2} = \cos(\frac{\pi}{2} + \pi n) = 0,$

$\cos 0 = \cos 2\pi n = 1,$ *and*

$\cos\pi = \cos(-\pi) = \cos(\pi + 2\pi n) = -1.$

(C) *For every integer* n,

$\sin 0 = \sin\pi n = 0,$

$\sin\frac{\pi}{2} = \sin(\frac{\pi}{2} + 2\pi n) = 1,$ *and*

$\sin(-\frac{\pi}{2}) = \sin(-\frac{\pi}{2} + 2\pi n) = -1.$

(D) $\text{cis}\, t = (\cos t, \sin t)$ *is a continuous function, periodic of period* 2π*, mapping* $[0, 2\pi[$ *onto the unit circle* $\mathcal{C}((0,0);1)$*. The restriction of* cis *to any interval* $[s,t]$ *where* $|t-s| < 2\pi$ *is one-to-one.*

Proof. See Figure 4.4. In Definition CS.2(C) and (D) we defined real numbers $h = \lim_{x\to+\infty} g(x)$ and $k = g(1)$; in Theorem CS.15(B) we showed that $h = \pi$ and in Theorem CS.25 that $k = \frac{\pi}{2}$. In this proof we will use these facts freely without further reference.

(A) By Theorem CS.12, cos and sin are continuous and periodic of period 2π; by Theorem CS.9 both cos and sin take on both values 1 and -1 and by the Intermediate Value Theorem for continuous functions, they both map onto $[-1,1]$; by Theorem CS.11, $\cos t = \cos(-t)$ and $\sin t = -\sin(-t)$.

(B) By Theorem CS.6(D), $\cos\frac{\pi}{2} = \cos(-\frac{\pi}{2}) = 0$; by periodicity, for all integers n, $\cos(\frac{\pi}{2} + \pi n) = 0$. By Theorem CS.9, $\cos 0 = 1$ and $\cos\pi = \cos(-\pi) = -1$; by periodicity, for all integers n, $\cos 2\pi n = 1$ and $\cos(\pi + 2\pi n) = -1$.

(C) By Theorem CS.6(D) and Definition/Remark CS.5, $\sin 0 = \sin\pi = \sin(-\pi) = 0$; by periodicity, for all integers n, $\sin(\pi + \pi n) = 0$. By Theorem CS.9, $\sin\frac{\pi}{2} = 1$ and $\sin -\frac{\pi}{2} = -1$; by periodicity, for all integers n, $\sin\frac{\pi}{2} = \sin(\frac{\pi}{2} + 2\pi n) = 1$; and $\sin(-\frac{\pi}{2}) = \sin(-\frac{\pi}{2} + 2\pi n) = -1$.

(D) This is Theorem CS.18. □

Corollary CS.27 *Let u and v be any real numbers; then there exists a number $t \in [0, 2\pi[$ and a real number $r \geq 0$ such that $u = r\cos t$ and $v = r\sin t$, that is, $(u, v) = r\,\mathrm{cis}\,t$.*

Proof. By Theorem CS.26(D), cis maps $[0, 2\pi]$ onto the unit circle. Let $\hat{u} = \frac{u}{\sqrt{u^2+v^2}}$ and $\hat{v} = \frac{v}{\sqrt{u^2+v^2}}$. Then $\hat{u}^2 + \hat{v}^2 = 1$ and since cis maps onto the unit circle, there exists a number $t \in [0, 2\pi[$ such that $\hat{u} = \cos t$ and $\hat{v}\sin t$; let $r = \sqrt{u^2 + v^2}$; then $u = r\cos t$ and $v = r\sin t$. □

4.5 Rotations; sum and difference formulas

It is intuitive to think of rotations as rigid motions about a center. In Chapter 10 of *Specht*, where we studied rotations, the closest we came to showing this was in Theorem ROT.22. In that theorem we showed that if α is a rotation about O, the angle $\angle AO(\alpha(A))$ is congruent to every other angle $\angle BO(\alpha(B))$.

On the coordinate plane $\mathbb{R}^2$ we can show that a rotation, as defined in Definition ROT.1, is indeed a rigid motion. We do this by showing that a rotation that takes $\mathrm{cis}\,0$ into $\mathrm{cis}\,s$ also takes $\mathrm{cis}(t-s)$ into $\mathrm{cis}(t-s+s) = \mathrm{cis}\,t$.

In the following, reflections are as in Chapter 21 of *Specht*, Definition LB.16; by Theorem LC.25 these preserve distance. Thus any reflection over a line through the center O of a unit circle maps the unit circle onto itself.

Theorem CS.28 (A rotation is a rigid motion) *Let s and t be distinct real numbers, $0 < s < t < 2\pi$, and $t - s < \pi$. There exists a unique rotation ρ of $\mathbb{R}^2$ about $O = (0,0)$ such that*

(A) $\rho(\operatorname{cis}[0, t-s]) = \operatorname{cis}[s,t]$, $\rho(\operatorname{cis} 0) = \operatorname{cis} s$, $\rho(\operatorname{cis}(t-s)) = \operatorname{cis} t$, *and*

(B) *the arclength* $\mathbb{L}(\operatorname{cis}[0, t-s]) = \mathbb{L}(\operatorname{cis}[s,t]) = t - s$.

Proof. By Theorem ROT.15, there exist unique rotations ρ and α about O such that $\rho(\operatorname{cis} 0) = \operatorname{cis} s$ and $\alpha(\operatorname{cis} 0) = \operatorname{cis}(t-s)$. By Theorem ROT.21,

$$\rho(\operatorname{cis}(t-s)) = \rho(\alpha(\operatorname{cis} 0)) = \alpha(\rho(\operatorname{cis} 0)) = \alpha(\operatorname{cis} s)$$

so that by Lemma CS.23, $\rho(\operatorname{cis}[0, t-s])$ is the arc on the unit circle with endpoints $\operatorname{cis} s$ and $\alpha(\operatorname{cis} s)$, and the arclength $\mathbb{L}(\rho(\operatorname{cis}[0, t-s])) = \mathbb{L}(\operatorname{cis}[0, t-s])$; by Theorem CS.19(B), this length is $t - s$.

The end points of $\operatorname{cis}[s,t]$ are $\operatorname{cis} s$ and $\operatorname{cis} t$, and $\mathbb{L}(\operatorname{cis}[s,t]) = t - s$; thus if we can show that $\alpha(\operatorname{cis} s)$ and $\operatorname{cis} t$ are on the same side of the line $\overleftrightarrow{O \operatorname{cis} s}$, it will follow from Theorem ARC.12 that $\rho(\operatorname{cis}(t-s)) = \alpha(\operatorname{cis} s) = \operatorname{cis} t$, proving the theorem.

First we dispose of the case where α is the point reflection $\mathcal{R}_O$ (and see Figure 4.5); in this case $\operatorname{cis}(t-s) = \alpha(\operatorname{cis} 0) = \operatorname{cis} \pi$ and $t - s = \pi$. The line $\overleftrightarrow{O(\operatorname{cis} s)}$ intersects the unit circle at the point $\operatorname{cis}(s+\pi) = \operatorname{cis}(s+t-s) = \operatorname{cis} t$. Therefore, since ρ is an isometry and a belineation, by Theorem NEUT.15,

$$\begin{aligned}\overleftrightarrow{(\rho(\operatorname{cis}(t-s)))(\rho(\operatorname{cis} 0))} &= \rho(\overleftrightarrow{(\operatorname{cis}(t-s))(\operatorname{cis} 0)}) = \rho(\overleftrightarrow{O(\operatorname{cis} 0)}) \\ &= \overleftrightarrow{\rho(O)(\rho(\operatorname{cis} 0))} = \overleftrightarrow{O(\operatorname{cis} s)} = \overleftrightarrow{(\operatorname{cis} t)(\operatorname{cis} s)}.\end{aligned}$$

By assumption, $\rho(\operatorname{cis} 0) = \operatorname{cis} s$, so that $\rho(\operatorname{cis}(t-s)) = \operatorname{cis} t$, and $\rho(\operatorname{cis}[0, t-s]) = \operatorname{cis}[s,t]$. Moreover, the arclengths of $\operatorname{cis}[0, t-s]$ and $\rho(\operatorname{cis}[0, t-s])$ are both π.

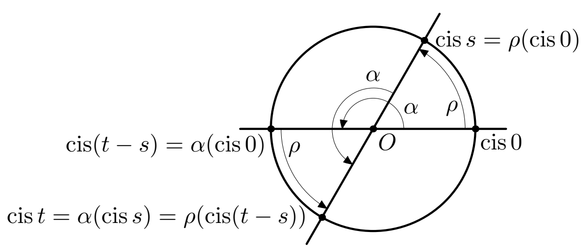


Fig. 4.5 The case where α is a point reflection.

Now suppose that α is not the point reflection. By Exercise ROT.4(D), if X and Y are any points on the unit circle, $\alpha(X) \in Y$-side of $\overleftrightarrow{OX}$ iff $\alpha(Y)$ is in the side of $\overleftrightarrow{OY}$ opposite X. We apply this to the present situation by making the following assignments: let $X = \operatorname{cis} 0$, $Y = \operatorname{cis} s$, so that $\alpha(X) =$

$\alpha(\operatorname{cis} 0) = \operatorname{cis}(t-s)$ and $\alpha(Y) = \alpha(\operatorname{cis} s) = \rho(\operatorname{cis}(t-s))$. Then there are two cases (I) and (II) as follows:

(Case I: $\alpha(\operatorname{cis} 0) = \operatorname{cis}(t-s) \in \operatorname{cis} s$-side of $\overleftrightarrow{O(\operatorname{cis} 0)}$ and $\alpha(\operatorname{cis} s)$ is in the side of $\overleftrightarrow{O(\operatorname{cis} s)}$ opposite $\operatorname{cis} 0$.) See Figure 4.6. Note that $s < \pi$; for if $s > \pi$, since $\operatorname{cis}(t-s)$ is in the $\operatorname{cis} s$-side of $\overleftrightarrow{O(\operatorname{cis} 0)}$, then $t-s > \pi$ so that $t = s+t-s > 2\pi$ contradicting our assumption that $t < 2\pi$.

Therefore $s < \pi$ and $t - s < \pi$, $0 < s < t = t - s + s < \pi + s$, and $\pi + s < 2\pi < 2\pi + s$. By Theorem CS.22, $\operatorname{cis} 0 = \operatorname{cis} 2\pi$ is on the opposite side of $\overleftrightarrow{O(\operatorname{cis} s)}$ from $\operatorname{cis} t$; from our assumption for Case (I), $\alpha(\operatorname{cis} s)$ is on the side of $\overleftrightarrow{O(\operatorname{cis} s)}$ opposite $\operatorname{cis} 0$, so that $\operatorname{cis} t$ and $\alpha(\operatorname{cis} s)$ are on the same side of $\overleftrightarrow{O(\operatorname{cis} s)}$.

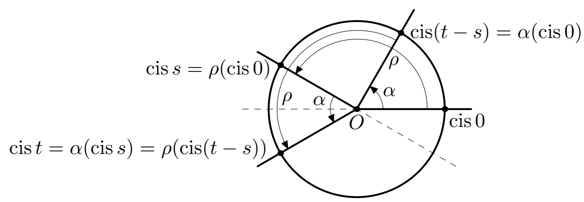


Fig. 4.6 Illustrating Case I.

(Case II: $\alpha(\operatorname{cis} 0) = \operatorname{cis}(t-s)$ is in the side of $\overleftrightarrow{O(\operatorname{cis} 0)}$ opposite $\operatorname{cis} s$ and $\alpha(\operatorname{cis} s) \in \operatorname{cis} 0$-side of $\overleftrightarrow{O(\operatorname{cis} s)}$.) The reader may wish to construct figures illustrating the two subcases.

(Subcase a: $0 < t - s < \pi$ and $\pi < s < 2\pi$.) Then $s < 2\pi < s + \pi$ and $s < t = s + t - s < s + \pi$. By Theorem CS.22, $\operatorname{cis} t$ and $\operatorname{cis} 0 = \operatorname{cis} 2\pi$ are on the same side of $\overleftrightarrow{O(\operatorname{cis} s)}$.

(Subcase b: $0 < s < \pi$ and $\pi < t - s < 2\pi$.) Then $t = t - s + s > \pi + s$ and $t < 2\pi$ so that $s < t < s + 2\pi$. Also, $s < 2\pi < s + 2\pi$ so that by Theorem CS.22, $\operatorname{cis} t$ and $\operatorname{cis} 0 = \operatorname{cis} 2\pi$ are on the same side of $\overleftrightarrow{O(\operatorname{cis} s)}$.

In either subcase, $\operatorname{cis} t \in \operatorname{cis} 0$-side of $\overleftrightarrow{O(\operatorname{cis} s)}$; by assumption in Case (II), $\alpha(\operatorname{cis} s)$ is also in this side, so $\alpha(\operatorname{cis} s)$ and $\operatorname{cis} t$ are on the same side of $\overleftrightarrow{O(\operatorname{cis} s)}$.

Thus in all cases $\rho(\operatorname{cis}(t-s)) = \alpha(\operatorname{cis} s) = \operatorname{cis} t$, proving part (A); part (B) follows immediately from Lemma CS.23.

See also Figure 4.7 and the following remark. □

Remark CS.28.1 To discern the inner structure and action of the rotation ρ in Theorem CS.28, let $\mathcal{L} = \overleftrightarrow{O(\operatorname{cis} 0)}$ (the "horizontal axis") and let $\mathcal{M}$ be the line of symmetry of the angle $\angle(\operatorname{cis} 0)O(\operatorname{cis} s)$. Then define

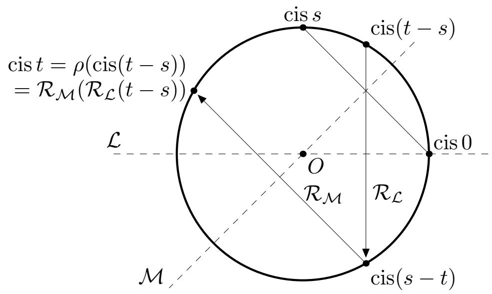


Fig. 4.7 For Theorem CS.28 and Remark CS.28.1, showing action of rotation ρ.

$\rho = \mathcal{R}_{\mathcal{M}} \circ \mathcal{R}_{\mathcal{L}}$. This is the rotation guaranteed by Theorem ROT.15, which maps $\text{cis}[0, t-s]$ to $\text{cis}[t, s]$. Refer also to Figure 4.7.

$\mathcal{R}_{\mathcal{L}}$ maps $\text{cis}[0, t-s]$ to the arc $\text{cis}[s-t, 0]$, and then $\mathcal{R}_{\mathcal{M}}$ maps this arc to $\text{cis}[t, s]$, carrying each point of $\text{cis}[s-t, 0]$ along a line parallel to $\overleftrightarrow{(\text{cis}\,0)(\text{cis}\,s)}$. If $\text{cis}\,x$ is a point of $\text{cis}[s-t, 0]$, then $\overleftrightarrow{(\text{cis}\,x)(\mathcal{R}_{\mathcal{M}}(\text{cis}\,x))}$ is a fixed line for $\mathcal{R}_{\mathcal{M}}$, and all such lines are parallel.

Corollary CS.29 *Let $\mathcal{E}$ be an arc of the unit circle having length less than π. Then there exist numbers s and t and a rotation ρ about O such that $0 \le s < 2\pi$, $0 \le t < 2\pi$, $\rho(\text{cis}[0, t-s]) = \mathcal{E}$, and $\rho(\text{cis}\,0)$ and $\rho(\text{cis}(t-s))$ are endpoints of $\mathcal{E}$.*

Proof. Suppose the endpoints of $\mathcal{E}$ are the points $\text{cis}\,u$ and $\text{cis}\,v$ where $0 \le u < 2\pi$ and $0 \le v < 2\pi$; we may choose the notation so that $0 \le u < v < 2\pi$. If $v < u + \pi < 2\pi$, the arc can be written as $\text{cis}[u, v]$ and the result follows from Theorem CS.28 by letting $t = v$ and $s = u$.

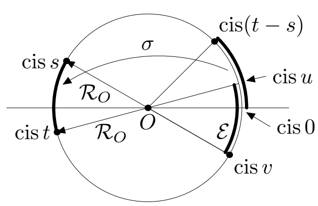


Fig. 4.8 Showing mapping of $\text{cis}[0, t-s]$ to $\mathcal{E}$.

Suppose, on the other hand, that the arc is "split" by the point $\text{cis}\,0$, as in Figure 4.8. That is, $0 \le u < \pi$ and $v > u + \pi$, so that the point $\text{cis}\,u$ lies on one side of $\mathcal{L} = \overleftrightarrow{O(1,0)}$ and $\text{cis}\,v$ lies on the other side, and $\text{cis}\,v$ is on the same side of the line $\overleftrightarrow{O(\text{cis}\,u)}$ as $\text{cis}\,O$. Then there are no points s and t in $[0, 2\pi[$ with $t > s$ such that $\mathcal{E} = \text{cis}[s, t]$. We can, however, represent the arc as $\mathcal{E} = \text{cis}[v, u + 2\pi]$, so its arc length is $u + 2\pi - v$. In this case we cannot directly apply Theorem CS.28, since $u + 2\pi \notin [0, 2\pi[$.

Let $\mathcal{R}_O$ be the point reflection, which is a rotation about O. (*cf* Definition ROT.1) Then $\text{cis}\, v$–O–$\mathcal{R}_O(\text{cis}\, v)$ and $\text{cis}\, u$–O–$\mathcal{R}_O(\text{cis}\, u)$. Let $s = v - \pi$ and $t = u + \pi$; then $\mathcal{R}_O(\text{cis}\, v) = \text{cis}\, s$ and $\mathcal{R}_O(\text{cis}\, u) = \text{cis}\, t$, so that $\mathcal{R}_O(\mathcal{E}) = \text{cis}[s, t]$, where $0 \le t < 2\pi$ and $0 \le s < 2\pi$.

Now $t - s = u + \pi - (v - \pi) = u - v + 2\pi < 2\pi$ since $u - v < 0$; since $u - v < -\pi$, $t - s < \pi$. By Theorem CS.28, there exists a rotation σ about O such that $\sigma(\text{cis}[0, t - s]) = \text{cis}[s, t]$, $\sigma(\text{cis}\, 0) = \text{cis}\, s$, and $\sigma(\text{cis}(t - s)) = \text{cis}\, t$. Define $\rho = \mathcal{R}_O \circ \sigma$, which by Theorem ROT.17, is a rotation. Then $\rho(\text{cis}\, 0) = \mathcal{R}_O(\sigma(\text{cis}\, 0)) = \mathcal{R}_O(\text{cis}\, s) = \text{cis}\, v$, $\rho(\text{cis}\, t - s) = \mathcal{R}_O(\sigma(\text{cis}(t-s))) = \mathcal{R}_O(\text{cis}\, t) = \text{cis}\, u$, and $\rho(\text{cis}[0, t-s]) = \mathcal{R}_O(\sigma(\text{cis}[0, t-s])) = \mathcal{R}_O(\text{cis}[s, t]) = \mathcal{E}$. Here we have used the fact that $\mathcal{R}_O$ is its own inverse. □

Corollary CS.30 *Let* $\text{cis}[s, t]$ *and* $\text{cis}[u, v]$ *be arcs of the unit circle* $\mathcal{C}(O; 1)$*, and that* $\pi > t - s = v - u > 0$*, so that the lengths of both arcs are the same; then there exists a rotation* ρ *of the unit circle such that* $\rho(\text{cis}[s, t]) = \text{cis}[u, v]$.

Proof. . By Corollary CS.29 there exist rotations σ and μ such that $\text{cis}[s, t] = \sigma(\text{cis}[0, t - s])$ and $\text{cis}[u, v] = \mu(\text{cis}[0, v - u]) = \mu(\text{cis}[0, t - s])$. Define $\rho = \mu \circ \sigma^{-1}$, which is a rotation by Theorem ROT.17. Then $\rho(\text{cis}[s, t]) = \mu(\sigma^{-1}(\text{cis}[s, t]) = \mu(\text{cis}[0, t - s]) = \text{cis}[u, v]$. □

Theorem CS.31 *If* s *and* t *are distinct real numbers* $0 \le s < 2\pi$ *and* $0 \le t < 2\pi$ *then* $\cos(t - s) = \cos t \cos s + \sin t \sin s$.

Proof. (Case 1: $t > s$.) By Theorem CS.28, there exists a rotation such that $\rho(\text{cis}\, 0) = \text{cis}\, s$ and $\rho(\text{cis}(t-s)) = \text{cis}\, t$. Since ρ is an isometry, thus preserving distance,

$$\text{dis}(\text{cis}\, 0, \text{cis}(t - s)) = \text{dis}(\text{cis}\, s, \text{cis}\, t).$$

Using the definition of distance and equating the squares of both sides of this, we have

$$\begin{aligned}(\cos 0 - \cos(t - s))^2 + (\sin 0 - \sin(t - s))^2 \\ = (1 - \cos(t - s))^2 + (\sin(t - s))^2 \\ = (\cos s - \cos t)^2 + (\sin s - \sin t)^2\end{aligned}$$

which reduces to

$$\begin{aligned}&1 - 2\cos(t - s) + \cos^2(t - s) + \sin^2(t - s) \\ &\quad = \cos^2 s + \sin^2 s + \cos^2 t + \sin^2 t - 2\cos s \cos t - 2\sin s \sin t,\end{aligned}$$

that is,

$$1 - 2\cos(t-s) + 1 = 1 + 1 - 2\cos s\cos t - 2\sin s\sin t.$$

Hence $\cos(t-s) = \cos s\cos t + \sin s\sin t$.

(Case 2: $t < s$.) The proof is Exercise CS.2. $\square$

Theorem CS.32 (Composite argument formulae for cosine and sine.) *For all real numbers s and t:*

(I) $\cos(s-t) = \cos s\cos t + \sin s\sin t$.

(II) $\sin(s-t) = \sin s\cos t - \cos s\sin t$.

(III) $\cos(s+t) = \cos s\cos t - \sin s\sin t$.

(IV) $\sin(s+t) = \sin s\cos t + \cos s\sin t$.

Proof. (I) (Case 1: $s = t \neq 0$.) $\cos(s-t) = \cos 0 = 1$, and

$$\cos s\cos t + \sin s\sin t = \cos^2 s + \sin^2 s = 1.$$

(Case 2: $s = 0$.) Then $\cos(t-s) = \cos t$, and $\cos s\cos t + \sin s\sin t = \cos t + 0$.

(Case 3: $t = 0$.) Then $\cos(t-s) = \cos -s = \cos s$, and

$$\cos s\cos t + \sin s\sin t = \cos s + 0.$$

(Case 4: $0 < t < 2\pi$ and $0 < s < 2\pi$ and $s \neq t$.) The proof is Theorem CS.24.

(Case 5: s and t are real numbers such that $s \neq t$.) Then by periodicity (*cf* Definition CS.10) there exist integers j and k and real numbers u and v such that $0 < u < 2\pi$, $0 < v < 2\pi$, $u \neq v$ and $s = 2j\pi + u$ and $t = 2k\pi + v$. By Theorem CS.31

$$\begin{aligned}\cos(s-t) &= \cos(u-v) = \cos u\cos v + \sin u\sin v\\ &= \cos(2j\pi + u)\cos(2k\pi + v) + \sin(2j\pi + u)\sin(2k\pi + v)\\ &= \cos s\cos t + \sin s\sin t.\end{aligned}$$

(II) For every real number u, $\cos(\frac{\pi}{2} - u) = \cos(\frac{\pi}{2})\cos u + \sin(\frac{\pi}{2})\sin u$; also, for every real number v, $\sin(\frac{\pi}{2} - v) = \cos v$; if we let $u = \frac{\pi}{2} - v$ then $v = \frac{\pi}{2} - u$ so that for every real number u, $\sin u = \cos v = \cos(\frac{\pi}{2} - u)$. Hence for all real numbers s and t,

$$\begin{aligned}\sin(s-t) &= \cos(\tfrac{\pi}{2} - (s-t)) = \cos((\tfrac{\pi}{2} - s) + t)\\ &= \cos(\tfrac{\pi}{2} - s)\cos t - \sin(\tfrac{\pi}{2} - s)\sin t = \sin s\cos t - \cos s\sin t.\end{aligned}$$

(III) Exercise CS.2.

(IV) By part (II), for all real numbers s and t, $\sin(s+t) = \sin(s-(-t)) = \sin s\cos t - \cos s\sin(-t) = \sin s\cos t + \cos s\sin t$. $\square$

Theorem CS.33 (Traditional angle definition of sin **and** cos**)** *Assume that:*

(1) *t is a real number such that* $0 \le t < 2\pi$*;*

(2) *φ is the mapping, applied to* $\mathcal{C}((0,0);r)$*, whose existence is established by Theorem ARC.12;*

(3) *Q is the point* $\operatorname{cis} t$*;*

(4) *$P = (x_1, x_2)$ is a member of* $\overrightarrow{OQ} \setminus \{Q\}$*; and*

(5) *$r > 0$, and* $r^2 = x_1^2 + x_2^2$.

Then $\cos t = \frac{x_1}{r}$ *and* $\sin t = \frac{x_2}{r}$.

Proof. (Case 1: $t = 0$.) $\cos 0 = \frac{1}{1} = 1$. $\sin 0 = \frac{0}{1} = 0$.

(Case 2: $0 < t < \frac{\pi}{2}$.) Let $R = \operatorname{ftpr}(Q, \overleftrightarrow{OU})$ and $S = \operatorname{ftpr}(P, \overleftrightarrow{OU})$. Then $\triangle OPQ \sim \triangle OSP$. Hence $\cos(t) = \frac{x_1}{r}$ and $\sin t = \frac{x_2}{r}$.

(Case 3: $t = \frac{\pi}{2}$.) Exercise CS.3.

(Case 4: $\frac{\pi}{2} < t < \frac{3\pi}{2}$.) Exercise CS.4.

(Case 5: $t = \frac{3\pi}{2}$.) Exercise CS.5.

(Case 6: $\frac{3\pi}{2} < t < 2\pi$.) Exercise CS.6. □

Remark CS.34 The next two theorems give analytic form to rotations we studied in Theorems CS.28 and CS.29. In particular, Theorem CS.36 shows that the line of symmetry of the angle $\angle(\operatorname{cis} 0)O(\operatorname{cis} s)$, referred to early in the proof of Theorem CS.28, is in fact $\overleftrightarrow{O(\operatorname{cis}\frac{s}{2})}$, as we might reasonably expect.

Theorem CS.35 *Let s be a real number such that $0 \le s < \pi$, and let $\mathcal{M} = \overleftrightarrow{O(\operatorname{cis} s)}$ be the line from the origin through the point $\operatorname{cis} s$ (which belongs to the unit circle $\mathcal{C}(O;1)$). Let $\mathcal{R}_{\mathcal{M}}$ be the mapping (Φ) defined over $\mathcal{M}$ in Definition LB.16. Then for every point (x_1, x_2) of the plane $\mathbb{R}^2$,*

$$\mathcal{R}_{\mathcal{M}}(x_1, x_2) = \begin{pmatrix} (\cos 2s)x_1 + (\sin 2s)x_2 \\ (\sin 2s)x_1 - (\cos 2s)x_2 \end{pmatrix}.$$

Proof. The equation of the line $\mathcal{M}$ is $ax_1 + bx_2 + c = 0$ where $a = \sin s$, $b = -\cos s$, and $c = 0$, as can be verified by substituting $x_1 = \cos s$ and $x_2 = \sin s$ into $(\sin s)x_1 - (\cos s)x_2$. From Defnition LB.16, for any point $(x_1, x_2) \in \mathbb{R}^2$,

$$\mathcal{R}_{\mathcal{M}}(x_1, x_2) = \begin{pmatrix} \frac{(b^2-a^2)x_1 - 2abx_2 - 2ac}{a^2+b^2} \\ \frac{-2abx_1 + (a^2-b^2)x_2 - 2bc}{a^2+b^2} \end{pmatrix};$$

since $a^2 + b^2 = \cos^2 s + \sin^2 s = 1$ and $c = 0$, this becomes

$$\begin{aligned} \mathcal{R}_{\mathcal{L}}(x_1, x_2) &= \begin{pmatrix} (b^2 - a^2)x_1 - 2abx_2 \\ -2abx_1 + (a^2 - b^2)x_2 \end{pmatrix} \\ &= \begin{pmatrix} (\cos^2 s - \sin^2 s)x_1 + 2(\sin s \cos s)x_2 \\ 2(\sin s \cos s)x_1 + (\sin^2 s - \cos^2 s)x_2 \end{pmatrix}. \end{aligned}$$

By Theorem CS.32(III), $\cos 2s = \cos^2 s - \sin^2 s$, and by Theorem CS.32(IV), $\sin 2s = 2 \sin s \cos s$, so that the last expression becomes

$$\mathcal{R}_{\mathcal{M}}(x_1, x_2) = \begin{pmatrix} (\cos 2s)x_1 + (\sin 2s)x_2 \\ (\sin 2s)x_1 - (\cos 2s)x_2 \end{pmatrix}. \quad \square$$

Theorem CS.36 *Let s be a real number such that $0 \le s < 2\pi$, and let $\mathcal{M} = \overleftrightarrow{O(\operatorname{cis} \frac{s}{2})}$ be the line from the origin through the point $\operatorname{cis} \frac{s}{2} = (\cos \frac{s}{2}, \sin \frac{s}{2})$, which point belongs to the unit circle $\mathcal{C}((0,0);1)$. Let $\mathcal{L} = \overleftrightarrow{(0,0)(1,0)}$. Let $\mathcal{R}_{\mathcal{M}}$ and $\mathcal{R}_{\mathcal{L}}$ be the reflections (Φ) defined in Definition LB.16 over the lines $\mathcal{M}$ and $\mathcal{L}$ respectively, and let $\rho = \mathcal{R}_{\mathcal{M}} \circ \mathcal{R}_{\mathcal{L}}$. By Definition ROT.1, ρ is a rotation.*

(A) *For every point $(x_1, x_2) \in \mathbb{R}^2$,*

$$\rho(x_1, x_2) = \mathcal{R}_{\mathcal{M}}(\mathcal{R}_{\mathcal{L}}(x_1, x_2) = \begin{pmatrix} (\cos s)x_1 - (\sin s)x_2 \\ (\sin s)x_1 + (\cos s)x_2 \end{pmatrix}.$$

(B) *ρ is the unique rotation of $\mathbb{R}^2$ about $O = (0,0)$ such that $\rho(1,0) = \rho(\operatorname{cis} 0) = (\cos s, \sin s) = \operatorname{cis} s$.*

(C) *The line $\mathcal{M}$ is the line of symmetry of $\angle(\operatorname{cis} 0)O(\operatorname{cis} s)$.*

Proof. (A) $\mathcal{R}_{\mathcal{L}}(x_1, x_2) = (x_1, -x_2)$ and by Theorem CS.28,

$$\begin{aligned} \rho(x_1, x_2) &= \mathcal{R}_{\mathcal{M}}(\mathcal{R}_{\mathcal{L}}(x_1, x_2)) = \mathcal{R}_{\mathcal{M}}(x_1, -x_2) \\ &= \begin{pmatrix} (\cos 2(\frac{s}{2}))x_1 + (\sin 2(\frac{s}{2}))(-x_2) \\ (\sin 2(\frac{s}{2}))x_1 - (\cos 2(\frac{s}{2}))(-x_2) \end{pmatrix} = \begin{pmatrix} (\cos s)x_1 - (\sin s)x_2 \\ (\sin s)x_1 + (\cos s)x_2 \end{pmatrix}. \end{aligned}$$

(B) $\rho(1,0) = \rho(\operatorname{cis} 0) = \begin{pmatrix} (\cos s) \cdot 1 - (\sin s) \cdot 0 \\ (\sin s) \cdot 1 + (\cos s) \cdot 0 \end{pmatrix} = (\cos s, \sin s) = \operatorname{cis} s$;

and by Theorem ROT.15(A), there can be only one rotation ρ about O such that $\rho(\operatorname{cis} 0) = \operatorname{cis} s$.

(C) $\mathcal{M}$ is the line of symmetry of $\angle(\operatorname{cis} 0)O(\operatorname{cis} s)$ because by part (B), $\mathcal{R}_{\mathcal{M}}$ maps $\operatorname{cis} 0$ to $\operatorname{cis} s$ and O is a fixed point for $\mathcal{R}_{\mathcal{M}}$. $\square$

4.6 Translations of $\mathbb{R}^2$.

Remark CS.37 (A) Recall from *Specht* Ch.3 Definition CAP.6 that a *translation* is a collineation α of the plane which has no fixed point, and such that for every line $\mathcal{L}$ either $\alpha(\mathcal{L}) \parallel \mathcal{L}$ or $\alpha(\mathcal{L}) = \mathcal{L}$.

Theorem CS.38 *A mapping α of $\mathbb{R}^2$ into itself is a translation iff for some point $(p_1, p_2) \neq (0,0)$ of $\mathbb{R}^2$, and every point $(x_1, x_2) \in \mathbb{R}^2$, $\alpha(x_1, x_2) = (x_1 + p_1, x_2 + p_2)$.*

Proof. (A) Suppose that there exists a point (p_1, p_2) such that for every point $(x_1, x_2) \in \mathbb{R}^2$,

$$\alpha(x_1, x_2) = (y_1, y_2) = (x_1 + p_1, x_2 + p_2).$$

Let $\mathcal{L}$ be a line having the equation $ax_1 + bx_2 + c = 0$, and let $d = c - ap_1 - bp_2$. Then $(x_1, x_2) \in \mathcal{L}$ iff

$$\begin{aligned} ay_1 + by_2 + d &= ax_1 + ap_1 + bx_2 + bp_2 + c - ap_1 - bp_2 \\ &= (ax_1 + bx_2 + c) + 0 = 0 \end{aligned}$$

which is true iff $ay_1 + by_2 + d = 0$, that is to say, $\alpha(x_1, x_2) = (y_1, y_2)$ is a member of the line $ax_1 + bx_2 + d = 0$. This is true iff the line $\mathcal{L}$ is either parallel to, or equal to $\alpha(\mathcal{L})$. Therefore α is a a translation.

(B) Conversely, suppose α is a translation. Then for every line $\mathcal{L}$ with equation $ax_1 + bx_2 + c = 0$, $\alpha(\mathcal{L})$ is a line parallel or equal to $\mathcal{L}$, that is, for some d, $\alpha(\mathcal{L})$ has equation $ay_1 + by_2 + d = 0$.

If $c = d$ then $\alpha(\mathcal{L}) = \mathcal{L}$ and we may let $p_1 = p_2 = 0$.

If $c \neq d$, then either $a \neq 0$ or $b \neq 0$ (or both). If $a \neq 0$, let $p_1 = \frac{-d+c-b}{a}$ and $p_2 = 1$. Then $(x_1, x_2) \in \mathcal{L}$ iff $ax_1 + bx_2 + c = 0$ which is true iff

$$\begin{aligned} ay_1 + by_2 + d &= a(x_1 + \tfrac{-d+c-b}{a}) + b(x_2 + 1) + d \\ &= ax_1 + bx_2 - d + c - b + b + d = ax_1 + bx_2 + c = 0. \end{aligned}$$

Thus (x_1, x_2) is on the line $ax_1 + bx_2 + c = 0$ iff $\alpha(x_1, x_2) = (y_1, y_2) = (x_1 + p_1, x_2 + p_2)$ is on the line $ay_1 + by_2 + d = 0$.

The case where $b \neq 0$ we leave to the reader as Exercise CS.7. □

Theorem CS.39 *Let $\overrightarrow{AB}$ be a closed ray on $\mathbb{R}^2$, and let α be a translation; then $\alpha(\overrightarrow{AB})$ is the ray $\overrightarrow{\alpha(A)\alpha(B)}$.*

Proof. By *Specht* Ch.21 Definition LA.1(3D), $\overrightarrow{AB} = \{A + t(B - A) \mid t \geq 0\}$. By Theorem CS.38, there exists a point P such that for all X, $\alpha(X) = X + P$. Then $Y \in \alpha(\overrightarrow{AB})$ iff for some $X \in \overrightarrow{AB}$,

$$\begin{aligned} Y = \alpha(X) = X + P &= A + P + t((B + P) - (A + P)) \\ &= \alpha(X) + t(\alpha(B) - \alpha(A)) \end{aligned}$$

which is true iff $Y \in \overrightarrow{\alpha(A)\alpha(B)}$. □

Remark CS.40 (A) Since an angle is the union of two non-opposite rays, (*cf Specht* Ch.5 Definition PSH.29), if α is a translation and if $\angle BAC$ is an angle with corner A, then $\alpha(\angle BAC) = \angle\alpha(B)\alpha(A)\alpha(C)$.

(B) We have seen in *Specht* Ch.12 Theorem ISM.4(B) that if $\mathcal{L}_1$ and $\mathcal{L}_2$ are parallel lines in a Euclidean plane, the mapping $\mathcal{R}_{\mathcal{L}_2} \circ \mathcal{R}_{\mathcal{L}_1}$ is a translation. The next theorem proves this for $\mathbb{R}^2$, using direct computation.

Theorem CS.41 *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be parallel lines on $\mathbb{R}^2$. Then $\mathcal{R}_{\mathcal{L}_2} \circ \mathcal{R}_{\mathcal{L}_1}$ is a translation.*

Proof. By Theorem LB.13 there exist real numbers a, b, c_1 and c_2 such that $c_1 \neq c_2$,

$$\mathcal{L}_1 = \{(x_1, x_2) \mid (x_1, x_2) \in \mathbb{R}^2 \text{ and } ax_1 + bx_2 + c_1 = 0\} \text{ and}$$
$$\mathcal{L}_2 = \{(x_1, x_2) \mid (x_1, x_2) \in \mathbb{R}^2 \text{ and } ax_1 + bx_2 + c_2 = 0\}.$$

Let (x_1, x_2) be any member of $\mathbb{R}^2$; by Definition LB.16,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathcal{R}_{\mathcal{L}_1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{b^2-a^2}{a^2+b^2} x_1 - \frac{2ab}{a^2+b^2} x_2 - \frac{2ac_1}{a^2+b^2} \\ \frac{-2ab}{a^2+b^2} x_1 + \frac{a^2-b^2}{a^2+b^2} x_2 - \frac{2bc_1}{a^2+b^2} \end{pmatrix}.$$

so that

$$\mathcal{R}_{\mathcal{L}_2}\left(\mathcal{R}_{\mathcal{L}_1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} \frac{b^2-a^2}{a^2+b^2} \left(\frac{b^2-a^2}{a^2+b^2} x_1 - \frac{2ab}{a^2+b^2} x_2 - \frac{2ac_1}{a^2+b^2}\right) \\ -\frac{2ab}{a^2+b^2} \left(\frac{-2ab}{a^2+b^2} x_1 + \frac{a^2-b^2}{a^2+b^2} x_2 - \frac{2bc_1}{a^2+b^2}\right) - \frac{2ac_2}{a^2+b^2} \\ \\ \frac{-2ab}{a^2+b^2} \left(\frac{b^2-a^2}{a^2+b^2} x_1 - \frac{2ab}{a^2+b^2} x_2 - \frac{2ac_1}{a^2+b^2}\right) \\ +\frac{a^2-b^2}{a^2+b^2} \left(\frac{-2ab}{a^2+b^2} x_1 + \frac{a^2-b^2}{a^2+b^2} x_2 - \frac{2bc_1}{a^2+b^2}\right) - \frac{2bc_2}{a^2+b^2} \end{pmatrix}$$

$$= \begin{pmatrix} \left(\left(\frac{b^2-a^2}{a^2+b^2}\right)^2 + \left(\frac{-2ab}{a^2+b^2}\right)^2\right) x_1 \\ + \left(\left(\frac{b^2-a^2}{a^2+b^2}\right)\left(\frac{-2ab}{a^2+b^2}\right) + \left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{a^2-b^2}{a^2+b^2}\right)\right) x_2 \\ + \left(\frac{b^2-a^2}{a^2+b^2}\right)\left(\frac{-2ac_1}{a^2+b^2}\right) + \left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{-2bc_1}{a^2+b^2}\right) - \frac{2ac_2}{a^2+b^2} \\ \\ \left(\left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{b^2-a^2}{a^2+b^2}\right) + \left(\frac{a^2-b^2}{a^2+b^2}\right)\left(-\frac{2ab}{a^2+b^2}\right)\right) x_1 \\ + \left(\left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{-2ab}{a^2+b^2}\right) + \frac{(a^2-b^2)^2}{(a^2+b^2)^2}\right) x_2 \\ + \left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{-2ac_1}{a^2+b^2}\right) + \left(\frac{a^2-b^2}{a^2+b^2}\right)\left(\frac{-2bc_1}{a^2+b^2}\right) - \frac{2bc_2}{a^2+b^2} \end{pmatrix}.$$

The coefficients of x_1 and x_2 (in order of appearance in the matrix just above) reduce to

$$\left(\left(\frac{b^2-a^2}{a^2+b^2}\right)^2 + \left(\frac{-2ab}{a^2+b^2}\right)^2\right) = 1,$$
$$\left(\left(\frac{b^2-a^2}{a^2+b^2}\right)\left(\frac{-2ab}{a^2+b^2}\right) + \left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{a^2-b^2}{a^2+b^2}\right)\right) = 0,$$
$$\left(\left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{b^2-a^2}{a^2+b^2}\right) + \left(\frac{a^2-b^2}{a^2+b^2}\right)\left(-\frac{2ab}{a^2+b^2}\right)\right) = 0, \text{ and}$$
$$\left(\left(\frac{-2ab}{a^2+b^2}\right)\left(\frac{-2ab}{a^2+b^2}\right) + \frac{(a^2-b^2)^2}{(a^2+b^2)^2}\right) = 1.$$

The constant terms reduce as follows:

$$\begin{aligned}
&\left(\tfrac{b^2-a^2}{a^2+b^2}\right)\left(\tfrac{-2ac_1}{a^2+b^2}\right)+\left(\tfrac{-2ab}{a^2+b^2}\right)\left(\tfrac{-2bc_1}{a^2+b^2}\right)-\tfrac{2ac_2}{a^2+b^2}\\
&=\left(\tfrac{(b^2-a^2)(-2ac_1)+(4ab^2c_1)}{(a^2+b^2)^2}\right)-\tfrac{2ac_2}{a^2+b^2}=\left(\tfrac{2ac_1(a^2-b^2+2b^2)}{(a^2+b^2)^2}\right)-\tfrac{2ac_2}{a^2+b^2}\\
&=\left(\tfrac{2ac_1(a^2+b^2)}{(a^2+b^2)^2}\right)-\tfrac{2ac_2}{a^2+b^2}=\left(\tfrac{2ac_1}{a^2+b^2}\right)-\tfrac{2ac_2}{a^2+b^2}=\left(\tfrac{2a(c_1-c_2)}{a^2+b^2}\right).
\end{aligned}$$

$$\begin{aligned}
&\left(\tfrac{-2ab}{a^2+b^2}\right)\left(\tfrac{-2ac_1}{a^2+b^2}\right)+\tfrac{(a^2-b^2)(-2bc_1)}{(a^2+b^2)^2}-\tfrac{2bc_2}{a^2+b^2}\\
&=\left(\tfrac{(4a^2bc_1)+(b^2-a^2)(2bc_1)}{(a^2+b^2)^2}\right)-\tfrac{2bc_2}{a^2+b^2}=\left(\tfrac{(a^2+b^2)(2bc_1)}{(a^2+b^2)^2}\right)-\tfrac{2bc_2}{a^2+b^2}\\
&=\left(\tfrac{2bc_1}{(a^2+b^2)}\right)-\tfrac{2bc_2}{a^2+b^2}=\left(\tfrac{2b(c_1-c_2)}{a^2+b^2}\right)
\end{aligned}$$

so that $\mathcal{R}_{\mathcal{L}_2}(\mathcal{R}_{\mathcal{L}_1}(x_1,x_2)) = (x_1+k_1, x_2+k_2)$ where $k_1 = \frac{2a(c_1-c_2)}{a^2+b^2}$ and $k_2 = \frac{2b(c_1-c_2)}{a^2+b^2}$.

At least one of a or b is non-zero, so that $a^2+b^2 \geq 0$; and $c_1 \neq c_2$; therefore at least one of k_1 or k_2 is non-zero. Thus the mapping $\mathcal{R}_{\mathcal{L}_1} \circ \mathcal{R}_{\mathcal{L}_1}$ is not the identity, and therefore by Theorem CS.38 is a translation . $\square$

4.7 Exercises for cosine and sine

Exercise CS.1 Prove part II of Theorem CS.9.

Exercise CS.2* Prove Case 2 of Theorem CS.32.

Exercise CS.3 Prove Case 3 of Theorem CS.33.

Exercise CS.4 Prove Case 4 of Theorem CS.33.

Exercise CS.5 Prove Case 5 of Theorem CS.33.

Exercise CS.6 Prove Case 6 of Theorem CS.33.

Exercise CS.7* Complete the proof of Theorem CS.38(B), the case where $b \neq 0$.

Exercise CS.8* Let $\mathcal{L} = \overleftrightarrow{(0,0)(\cos s, \sin s)}$; for each each $(x_1,x_2) \in \mathbb{R}^2$ define

$$\begin{gathered}
\alpha(x_1,x_2) = x_1\cos 2s + x_2 \sin 2s,\\
\beta(x_1,x_2) = x_1 \sin 2s + x_2\cos 2s, \text{ and}\\
\mathcal{R}_{\mathcal{L}}(x_1,x_2) = (\alpha(x_1,x_2), \beta(x_1,x_2)).
\end{gathered}$$

Prove that for every $(x_1,x_2) \in \mathbb{R}^2 \setminus \mathcal{L}$, $(\dfrac{\alpha(x_1,x_2)+x_1}{2}, \dfrac{\beta(x_1,x_2)+x_2}{2})$ is a point on the line $\overleftrightarrow{(0,0)(\cos s, \sin s)}$.

Exercise CS.9 Let a and b be real numbers such that $(a,b) \neq (0,0)$; for each $(x_1,x_2) \in \mathbb{R}^2$, define

$$\rho(x_1,x_2) = \left(\tfrac{ax_1}{\sqrt{a^2+b^2}} - \tfrac{bx_2}{\sqrt{a^2+b^2}}, \tfrac{bx_1}{\sqrt{a^2+b^2}} + \tfrac{ax_2}{\sqrt{a^2+b^2}}\right);$$

then ρ is a rotation about $(0,0)$.

Exercise CS.10* According to Theorem ISM.5 for any point $A \in \mathbb{R}^2$ there is a translation τ_A such that $\tau_A(O) = A$. Definition VEC.1 uses this fact to define addition on a Euclidean plane such as $\mathbb{R}^2$.

Let τ_A be a translation of $\mathbb{R}^2$, where $A = (a_1,a_2) \neq (0,0)$ is some point of that plane. Show that τ_A has an inverse τ_{-A} which is a translation.

Exercise CS.11* Show that a translation of $\mathbb{R}^2$ preserves distance, *i.e.* if $A = (a_1,a_2)$ is a member of $\mathbb{R}^2 \setminus \{(0,0)\}$ and τ_A is the translation such that for every member (x_1,x_2) of $\mathbb{R}^2$, $\tau_A(x_1,x_2) = (x_1+a_1, x_2+a_2)$, then for any two members (x_1,x_2) and (y_1,y_2) of $\mathbb{R}^2$,

$$\mathrm{dis}(\tau_A(x_1,x_2), \tau_A(y_1,y_2)) = \mathrm{dis}((x_1,x_2),(y_1,y_2)).$$

Exercise CS.12* If $\mathrm{cis}\, t = (\cos t, \sin t)$ is a point on the unit circle $\mathcal{C}(O;1)$ where $O = (0,0)$, then

$$\mathrm{cis}(t+\pi) = -\mathrm{cis}\, t = (-\cos t, -\sin t)$$

and $(\mathrm{cis}\, t)$–O–$(cis(t+\pi))$.

Exercise CS.13* Let $\mathcal{C}(O;r)$ be a circle in $\mathbb{R}^2$ with radius r and center $O = (0,0)$. Then (A) no line intersects the circle $\mathcal{C}(O;r)$ in more than two points; and (B) a line containing intersects $\mathcal{C}(O;r)$ in two points iff it it contains a point X such that $\mathrm{dis}(X,O) < r$. Moreover, if a line intersects the circle $\mathcal{C}(O;r)$ in two points A and B, the line of symmetry of $\angle AOB$ intersects the line $\overleftrightarrow{AB}$ at a point C such that $\mathrm{dis}\, OC < r$.

4.8 Selected answers for exercises cosine and sine

Exercise CS.2 Proof. From Case 1 and the fact that cos is an even function,

$$\begin{aligned}\cos(t-s) = \cos(s-t) &= \cos t \cos s + \sin t \sin s \\ &= \cos s \cos t + \sin s \sin t.\end{aligned}$$

Exercise CS.7 Proof. If $b \neq 0$, let $p_1 = 1$ and $p_2 = \frac{-d+c-a}{b}$; then for all $(x_1, x_2) \in \mathcal{L}$,

$$\alpha(x_1, x_2) = (y_1, y_2) = (x_1 + 1, x_2 + \tfrac{-d+c-a}{b})$$

and

$$\begin{aligned} &ay_1 + by_2 + d = a(x_1 + 1) + b(x_2 + \tfrac{-d+c-a}{b}) + d \\ &\qquad = ax_1 + bx_2 + a - d + c - a + d = ax_1 + bx_2 + c = 0, \end{aligned}$$

so every (x_1, x_2) in the line $ax_1 + bx_2 + c = 0$ maps to a point $(x_1 + p_1, x_2 + p_2)$ in the line $ay_1 + by_2 + d = 0$.

This completes the converse argument. □

Exercise CS.8 Proof.

$$\begin{aligned} &\sin s\left(\tfrac{x_1 \cos 2s + x_2 \sin 2s + x_1}{2}\right) - \cos s\left(\tfrac{x_1 \sin 2s + x_2 \cos 2s + x_2}{2}\right) \\ &= \tfrac{\cos 2s \sin s - \sin 2s \cos s}{2} x_1 + \tfrac{\sin s \sin 2s - \cos 2s \cos s}{2} x_2 + \tfrac{x_1 \sin s - x_2 \cos s}{2} \\ &= \tfrac{\sin s}{2} x_1 + \tfrac{-\cos s}{2} x_2 = 0. \quad \square \end{aligned}$$

Exercise CS.10 Proof. If $A = (a_1, a_2)$, then $-A = (-a_1, -a_2)$; For every member (x_1, x_2) of $\mathbb{R}^2$, $\tau_A(x_1, x_2) = (x_1 + a_1, x_2 + a_2)$, and ${\tau_A}^{-1}(x_1, x_2) = \tau_{-A}(x_1, x_2) = (x_1 - a_1, x_2 - a_2)$. Thus $\tau_{-A}(\tau_A(x_1, x_2)) = ((x_1 + a_1) - a_1, (x_2 + a_2) - a_2) = (x_1, x_2)$. □

Exercise CS.11 Proof.

$$\begin{aligned} &\operatorname{dis}(\tau_A(x_1, x_2), (y_1, y_2)) = \operatorname{dis}((x_1 + a_1, x_2 + a_2), (y_1 + a_1, y_2 + a_2)) \\ &\qquad = \sqrt{((x_1 + a_1) - (y_1 + a_1))^2 + ((x_2 + a_2) - (y_2 + a_2))^2} \\ &\qquad = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \\ &\qquad = \operatorname{dis}((x_1, x_2), (y_1, y_2)). \quad \square \end{aligned}$$

Exercise CS.12 Proof. By Theorem CS.32,

$$\begin{aligned} &\cos(t + \pi) = \cos t \cos(\pi) + \sin t \sin(\pi) \\ &\qquad = \cos t(-1) + \sin t \cdot 0 = -\cos t \end{aligned}$$

and

$$\begin{aligned} &\sin(t + \pi) = \sin t \cos(\pi) - \cos t \sin(\pi) \\ &\qquad = \sin t(-1) - \cos t \cdot 0 = -\sin t. \quad \square \end{aligned}$$

Exercise CS.13 Proof. If $X = (x_1, x_2)$ is any point on a line $\mathcal{L}$, then by Remark LB.2(C) there exists a point $Y = (y_1, y_2)$ such that $Z \in \mathcal{L}$ iff $Z = X + tY$ for some real number t. Thus, a point $Z = X + tY \in \mathcal{L}$ is a point of $\mathcal{C}(O; r)$ iff $(x_1 + ty_1, x_2 + ty_2) \in \mathcal{C}(O; r)$, that is, $(x_1 + ty_1)^2 + (x_2 + ty_2)^2 = r^2$. Expanding, we have

$$x_1^2 + t^2 y_1^2 + t(2x_1y_1) + x_2^2 + t^2 y_2^2 + t(2x_2y_2) = r^2,$$

and rearranging,

$$x_1^2 + x_2^2 + t^2(y_1^2 + y_2^2) + t(2(x_1y_1 + x_2y_2)) = r^2,$$

or

$$(x_1^2 + x_2^2 - r^2) + t(2(x_1y_1 + x_2y_2)) + t^2(y_1^2 + y_2^2) = 0.$$

This is a quadratic in t and by the quadratic formula, it has at most two solutions. This proves part (A).

To show part (B), assume that $\text{dis}(X, O) < r$, that is, $x_1^2 + x_2^2 < r^2$. Then we may state the above equation as $a + tb + t^2c = 0$ where $a = x_1^2 + x_2^2 - r^2$, $b = 2(x_1y_1 + x_2y_2)$, and $c = y_1^2 + y_2^2$. By the quadratic formula, this has two solutions iff the discriminant $b^2 - 4ac > 0$. That is,

$$(2(x_1y_1 + x_2y_2))^2 - 4(x_1^2 + x_2^2 - r^2)(y_1^2 + y_2^2) > 0.$$

Then $(2(x_1y_1 + x_2y_2))^2 > 0$; $(y_1^2 + y_2^2) > 0$, and $x_1^2 + x_2^2 - r^2 < 0$ so that the discriminant is greater than 0, and there are two solutions; hence the line $\mathcal{L}$ intersects the circle at two points.

Conversely, if $\mathcal{L}$ intersects the circle at two points $A = (a_1, a_2)$ and $B = (b_1, b_2)$, let $\mathcal{M}$ be the line of symmetry of the angle $\angle AOB$. Then let C be the point of intersection of $\mathcal{M}$ and $\mathcal{L}$; $C \in \text{ins}\,\angle AOB$ and $\mathcal{L}$ is a fixed line for $\mathcal{R}_{\mathcal{M}}$, so the two lines are perpendicular. It follows from the Pythagorean Theorem (Theorem VEC.26.5 or *Specht* Ch.15 Theorem SIM.23.1) that $\text{dis}(C, O) < 1$. □

Chapter 5
Angle Measure (AM)

Dependencies: *Euclidean Geometry and its Subgeometries (Specht); Chapters 3 and 4 of this supplement*
Acronym: *AM*
New terms defined: *(radian) measure of an angle*

We now develop the concept of angle measure, using the ideas of arc length as developed in Chapter 3 and sin and cos as developed in Chapter 4 of this Supplement. Again, in this chapter, we work in $\mathbb{R}^2$. Since it is shown in Chapter 21 of *Specht* that all the axioms of that development hold for $\mathbb{R}^3$ and $\mathbb{R}^2$, here we may use all the theorems from that book.

In chapters 3 and 4 we considered arcs on a unit circle having arc lengths up to (but not including) 2π. In this chapter we will shift our focus to arcs (on the unit circle) having length less than π, that is, which subtend *angles*, the definition of which (*Specht* Ch.5 Definition PSH.29) specifically excludes "straight angles" or anything "greater." There are no "270 degree angles" here. Note that when we write $\angle AOB = \angle COD$ we mean set equality; there is no implication that $\overrightarrow{OA} = \overrightarrow{OC}$ or $\overrightarrow{OA} = \overrightarrow{OD}$.

References in this chapter to items labeled VEC, ARC, CS, and AM are to this Supplement; all other references, such as "Theorem PSH.41" are to *Specht*. Again, we refer the reader to the note **Citations and references** at the end of the Preface to this Supplement, and to the abbreviated Table of Contents (with acronyms) included there.

5.1 Definitions and theorems for angle measure

Definition/Remark AM.1 (A) Let $O = (0,0)$ be a point of $\mathbb{R}^2$, and let r be a positive real number. In Definition CS.13 we defined $\mathcal{C}((0,0);r) = \mathcal{C}(O;r)$ to be the *circle with center* $O = (0,0)$ *and radius* r, and defined its inside, enclosure, and diameter. We also designated $\mathcal{C}(O;1)$ as the *unit circle*.

(B) Let C and D be points of $\mathbb{R}^2$ such that C, D, and O are noncollinear, so that the rays $\overrightarrow{OC}$ and $\overrightarrow{OD}$ form $\angle COD$. Let $r > 0$ be a real number, and let A and B be points on $\overrightarrow{OC}$ and $\overrightarrow{OD}$, respectively, such that $\mathrm{dis}(O,A) = \mathrm{dis}(O,B) = r$, so that $\{A\} = \mathcal{C}(O;r) \cap \overrightarrow{OC}$, and $\{B\} = \mathcal{C}(O;r) \cap \overrightarrow{OD}$, so that $\angle AOB = \angle COD$.

Define $\widehat{AB}$ to be $\mathcal{C}(O;r) \cap \mathrm{enc}\,\angle AOB$. $\widehat{AB}$ is called an **arc** of the circle $\mathcal{C}(O;r)$, and is said to **subtend** the angle $\angle AOB$, which is a **central angle** of $\mathcal{C}(O;r)$. Here $\mathrm{enc}\,\angle AOB$ means the *enclosure* of $\angle AOB$, that is, the union of $\angle AOB$ and its *inside* $\mathrm{ins}\,\angle AOB$ (*cf* Definition PSH.36).

(C) According to the definition of $\widehat{AB}$ just above, if F is a point of $\widehat{AB}$, which is neither A nor B, then $\widehat{AB}$ is a subset of $\overleftrightarrow{OA} \cup F$-side of $\overleftrightarrow{OA}$, and also of $\overleftrightarrow{OB} \cup F$-side of $\overleftrightarrow{OB}$ (*cf* Theorem CS.22).

By the definition of angle (*Specht* Ch.5 Definition PSH.29), $\overleftrightarrow{OA} \neq \overleftrightarrow{OB}$, that is, the rays $\overrightarrow{OA}$ and $\overrightarrow{OB}$ are not opposite. This definition does not allow for arcs which contain both endpoints of any diameter of the circle. In particular, on the unit circle, it does not allow arcs with length greater or equal to π.

(D) Let $\angle COD$ be any angle, and let A and B be points on $\overrightarrow{OC}$ and $\overrightarrow{OD}$, respectively, such that $\mathrm{dis}(O,A) = \mathrm{dis}(O,B) = 1$, so that both A and B are points of the unit circle $\mathcal{C}(O;1)$. Define the **(radian) angle measure of** $\angle COD = \angle AOB$ (denoted $\mathbf{meas}\angle AOB$) to be the arc length $\mathbb{L}(\widehat{AB})$ of $\widehat{AB}$.

Theorem AM.2 *Let A and B be distinct points on the circle $\mathcal{C}(O;r)$, where $r > 0$, for which it is false that A–O–B. Then $\widehat{AB}$ is an arc as defined in Chapter 3, Definition ARC.1.*

Proof. Let A' and B' respectively be the points of intersection of the unit circle with $\overleftrightarrow{OA}$ and $\overleftrightarrow{OB}$. By Theorem CS.22 the notation may be chosen (and the points possibly renamed) so that $A = rA' = r\,\mathrm{cis}\,a$ and $B = rB' = r\,\mathrm{cis}\,b$, $b > a$, and $b - a < \pi$. From part (C) of that theorem, $r\,\mathrm{cis}\,]a,b[\, \subseteq \mathrm{ins}\,\angle AOB$ and $r\,\mathrm{cis}\,]b, a+2\pi[\, \subseteq \mathrm{out}\,\angle AOB$. From Theorem PSH.41(B),

$$\mathcal{C}(O;r) \setminus \{A,B\} = (\mathrm{ins}\,\angle AOB \cap \mathcal{C}(O;r)) \cup (\mathrm{out}\,\angle AOB \cap \mathcal{C}(O;r)).$$

Since $r\,\mathrm{cis}$ maps $[a, a+2\pi[$ one-to-one onto $\mathcal{C}(O;r)$, it is also true that
$$\mathcal{C}(O;r) \setminus \{A, B\} = r\,\mathrm{cis}\,]a, b[\cup r\,\mathrm{cis}\,]b, a+2\pi[$$
By elementary set theory, $r\,\mathrm{cis}\,]a, b[= \mathrm{ins}\,\angle AOB \cap \mathcal{C}(O;r)$ and hence
$$r\,\mathrm{cis}[a, b] = \mathrm{enc}\,\angle AOB \cap \mathcal{C}(O;r);$$
therefore every arc $\widehat{AB}$ is an arc $\mathrm{cis}[a, b]$ as previously defined. □

Theorem AM.3 *Let O and O' be distinct points on $\mathbb{R}^2$, r a positive real number, and let T be a translation of $\mathbb{R}^2$ such that $T(O) = O'$. Let A and B be distinct points on the circle $\mathcal{C}(O;r)$ and let C and D be distinct points on the circle $\mathcal{C}(O';r)$, so that $\widehat{AB}$ is an arc on $\mathcal{C}(O;r)$ and $\widehat{A'B'}$ is an arc on $\mathcal{C}(O';r)$. Then $\widehat{AB} \cong \widehat{A'B'}$ iff $\widehat{T(A)T(B)} \cong \widehat{A'B'}$.*

Proof. If $\widehat{T(A)T(B)} \cong \widehat{A'B'}$ there exists an isometry φ such that
$$\varphi(\widehat{T(A)T(B)}) = \widehat{A'B'};$$
then
$$(\varphi \circ T)(\widehat{AB})) = \varphi(\widehat{T(A)T(B)})) = \widehat{A'B'},$$
so $\widehat{AB} \cong \widehat{A'B'}$.

Conversely, if $\widehat{AB} \cong \widehat{A'B'}$, there exists an isometry ψ such that $\psi(\widehat{AB}) = \widehat{A'B'}$; then $T \circ \psi^{-1}(\widehat{A'B'}) = \widehat{T(A)T(B)}$ so that $\widehat{T(A)T(B)} \cong \widehat{A'B'}$. □

Take note: for simplicity, several of the following theorems are stated and proved for the unit circle $\mathcal{C}(O;1)$; it is easy to extend them as needed to the circle $\mathcal{C}(O;r)$ where $r > 0$.

Theorem AM.4 *Let $\widehat{AB}$ and $\widehat{CD}$ be arcs on the unit circle $\mathcal{C}(O;1)$. Suppose there exists an isometry φ such that $\varphi(\widehat{CD}) = \widehat{AB}$, $\varphi(C) = A$ and $\varphi(D) = B$; then O is a fixed point for φ, and $\varphi(\angle COD) = \angle AOB$.*

Proof. Since all the points A, B, C, and D are on the unit circle with center O, the distance from each to O is 1. Suppose that O is not a fixed point for φ, and let $O' = \varphi(O) \neq O$.

Let $\mathcal{L}$ be the right-angle bisector of the segment $\overline{AB} = \overline{\varphi(C)\varphi(D)}$, and let $\{P\} = \overline{AB} \cap \mathcal{L}$; then $\mathcal{L}$ is the line of symmetry of $\angle AOB$. Because φ preserves distance, A and B are also on the unit circle centered at O', and $\mathcal{L}$ is also the line of symmetry of $\angle AO'B$. Therefore $\mathcal{L} = \overleftrightarrow{O\varphi(O)}$.

Let Q be the point of intersection of $\widehat{AB}$ and $\mathcal{L}$. Since φ maps the arc $\widehat{CD}$ into the unit circle $\mathcal{C}(O';1)$ and onto the arc $\widehat{AB}$, $Q \in \widehat{AB} \subseteq \mathcal{C}(O';1)$, and $\mathrm{dis}(Q, O') = 1$.

Case 1: If O' is on the O-side of Q, then $O' = O$, because there is only one point on $\mathcal{L}$ on each side of Q which is a distance 1 from Q.

Case 2: If O–Q–O'; by the Pythagorean Theorem (Theorem SIM.23), $\operatorname{dis}(P, O) < 1$ and $\operatorname{dis}(P, O') < 1$; since $\operatorname{dis}(Q, O) = 1$, O–P–Q. Since O–Q–O' it follows from Theorem PSH.8(A) that P–Q–O'; therefore, from $\operatorname{dis}(P, O') < 1$, we have $\operatorname{dis}(Q, O') < 1$; but $\operatorname{dis}(Q, O') = 1$ because $Q \in \mathcal{C}(O'; 1)$. This is a contradiction. $\square$

Theorem AM.5 *Let O be a point on $\mathbb{R}^2$, and let A, B, C, and D be points on the unit circle $\mathcal{C}(O; 1)$ such that $A \neq B$, $C \neq D$, and $\widehat{AB}$ and $\widehat{CD}$ are arcs on this circle.*

(1) *If φ is an isometry on $\mathbb{R}^2$; then $\varphi(\widehat{AB}) = \widehat{CD}$ iff $\varphi(\angle AOB) = \angle COD$. That is to say, $\widehat{AB} \cong \widehat{CD}$ iff $\angle AOB \cong \angle COD$.*

(2) *If ρ is a rotation of the plane; then $\rho(\widehat{AB}) = \widehat{CD}$ iff $\rho(\angle AOB) = \angle COD$.*

(3) *The following statements are equivalent:*

(a) *$\widehat{AB}$ and $\widehat{CD}$ have the same arc length; that is, $\mathbb{L}(\widehat{AB}) = \mathbb{L}(\widehat{CD})$;*

(b) *there exists an isometry on $\mathbb{R}^2$ such that $\varphi(\widehat{AB}) = \widehat{CD}$; and*

(c) *there exists a rotation ρ such that $\rho(\widehat{AB}) = \widehat{CD}$.*

Proof. (1) If φ is an isometry such that $\varphi(\widehat{AB}) = \widehat{CD}$, then either $\varphi(A) = C$ and $\varphi(B) = D$, or $\varphi(A) = D$ and $\varphi(B) = C$. By Theorem AM.4, $\varphi(O) = O$, so that either $\varphi(A) \in \overrightarrow{OC}$ and $\varphi(B) \in \overrightarrow{OD}$ or $\varphi(A) \in \overrightarrow{OD}$ and $\varphi(B) \in \overrightarrow{OC}$; in either case, $\varphi(\angle AOB) = \angle COD$.

Conversely, if φ is an isometry and $\varphi(\angle AOB) = \angle COD$, by Theorem NEUT.15(11) $\varphi(\operatorname{ins} \angle AOB) = \operatorname{ins} \angle COD$, so that $\varphi(\operatorname{enc} \angle AOB) = \operatorname{enc} \angle COD$. Also φ preserves distance so that it maps the unit circle onto itself. It follows that $\varphi(\operatorname{enc} \angle AOB) \cap \mathcal{C}(O; 1) = \operatorname{enc} \angle COD \cap \mathcal{C}(O; 1)$, that is to say, $\varphi(\widehat{AB}) = \widehat{CD}$.

(2) If ρ is a rotation, it is an isometry; then if $\rho(\widehat{AB}) = \widehat{CD}$, by part (1) $\rho(O) = O$ and $\rho(\angle AOB) = \angle COD$. Conversely, if ρ is a rotation such that $\rho(\angle AOB) = \angle COD$, by part (1) $\rho(\widehat{AB} = \widehat{CD}$.

(3) We shall prove (a)$\Rightarrow$(c)$\Rightarrow$(b)$\Rightarrow$(a).

(a)$\Rightarrow$(c): If $\mathbb{L}(\widehat{AB}) = \mathbb{L}(\widehat{CD})$, then by Corollary CS.30, there exists a rotation ρ such that $\rho(\widehat{AB}) = \widehat{CD}$.

(c)$\Rightarrow$(b): If ρ is a rotation then it is an isometry.

(b)$\Rightarrow$(a): If φ is an isometry such that $\varphi(\widehat{AB}) = \widehat{CD}$, by Theorem AM.4, $\varphi(O) = O$; by Lemma CS.23(B), it follows that $\mathbb{L}(\widehat{AB}) = \mathbb{L}(\widehat{CD})$. $\square$

Theorem AM.6 *Let A, B, O, A', B', and O' be points on $\mathbb{R}^2$ such that A, B, and O are noncollinear and A', B', and O' are noncollinear; then $\angle AOB \cong \angle A'O'B'$ iff* $\text{meas}\,\angle AOB = \text{meas}\,\angle A'O'B'$.

Proof. We may choose these points so that $\text{dis}(O, A) = \text{dis}(O, B) = 1$ and $\text{dis}(O', A') = \text{dis}(O', B') = 1$. Then by Definition AM.1, $\text{meas}\,\angle AOB = \text{meas}\,\angle A'O'B'$ iff the arc lengths $\mathbb{L}(\widehat{AB})$ and $\mathbb{L}(\widehat{A'B'})$ are the same.

Suppose $\angle AOB \cong \angle A'O'B'$; let φ be the isometry such that $\varphi(\angle AOB) = \angle A'0'B'$; then by *Specht* Ch.8 Theorem NEUT.15(11), $\varphi(\text{ins}\,\angle AOB) = \text{ins}\,\angle A'0'B'$. Since φ preserves distance, $\varphi(\mathcal{C}(O; 1)) = \mathcal{C}(O'; 1)$. Since φ is one-to-one, and using elementary set theory,

$$\begin{aligned}\varphi(\widehat{AB}) &= \varphi(\text{ins}\,\angle AOB \cap \mathcal{C}(0; 1)) = \varphi(\text{ins}\,\angle AOB) \cap \varphi(\mathcal{C}(0; 1)) \\ &= \text{ins}\,\angle A'O'B' \cap \mathcal{C}(0'; 1) = \widehat{A'B'}.\end{aligned}$$

Then by Lemma CS.23, $\mathbb{L}(\widehat{AB}) = \mathbb{L}(\widehat{A'B'})$, that is, the angles have the same measure.

Conversely, suppose that $\mathbb{L}(\widehat{AB}) = \mathbb{L}(\widehat{A'B'})$, Let $\mathcal{T}$ be the translation of $\mathbb{R}^2$ such that $\mathcal{T}(O) = O'$. Then $\mathcal{T}$ is an isometry, and preserves distance, so that $\mathcal{T}(\mathcal{C}(O; 1)) = \mathcal{C}(O'; 1)$. By Lemma CS.23, $\widehat{T(A)T(B)}$ is an arc on $\mathcal{C}(O'; 1)$ which has the same arc length as $\widehat{AB}$. Because $\mathbb{L}(\widehat{A'B'}) = \mathbb{L}(\widehat{AB}) = \mathbb{L}(\widehat{T(A)T(B)})$, by Theorem AM.5(3) there exists an isometry ψ such that $\psi(\widehat{T(A)T(B)}) = \widehat{A'B'}$; then $\psi \circ \varphi$ is an isometry mapping $\widehat{AB}$ onto $\widehat{A'B'}$. $\square$

Theorem AM.7 *(*A*) Let A, O, and B be noncollinear points on $\mathbb{R}^2$ and let C be a member of* $\text{ins}\,\angle AOB$*, then*

$$\text{meas}\,\angle AOC + \text{meas}\,\angle BOC = \text{meas}\,\angle AOB.$$

(B) *Let A, O, and B be points such that A–O–B and let C be a point off of $\overleftrightarrow{AB}$, then* $\text{meas}\,\angle AOC + \text{meas}\,\angle BOC = \pi$.

Proof. Without loss of generality we can assume that all points A, B, and C are on the unit circle $\mathcal{C}(O; 1)$ with center O and radius 1.

(A) Since by Theorem AM.2, $\widehat{AB}$ is the image under the mapping cis of some interval $[a, b]$ where $A = \text{cis}\, a$ and $B = \text{cis}\, b$, for some c with $a < c < b$, $C = \text{cis}\, c$. Then by Theorem ARC.4,

$$\text{meas}\,\angle AOC + \text{meas}\,\angle BOC = \mathbb{L}(\widehat{AC}) + \mathbb{L}(\widehat{CB}) = \mathbb{L}(\widehat{AB}) = \text{meas}\,\angle AOB.$$

(B) Suppose that $A = \text{cis}\, a$; then $B = \text{cis}(a+\pi) = \text{cis}(a-\pi)$. Then $C = \text{cis}\, c$ where either $a < c < a+\pi$ or $a > c > a-\pi$. If $a < c < a+\pi$ ($a > c > a-\pi$), let

$\mathcal{E} = \{A, B\} \cup (\mathcal{C}(O; 1) \cap C\text{-side of } \overleftrightarrow{AB})$. Then $\mathcal{E} = \operatorname{cis}[a, a+\pi] (= \operatorname{cis}[a, a-\pi])$, and $\mathcal{E} = \widehat{AC} \cup \widehat{CB}$. By Theorem ARC.4,

$$\mathbb{L}(\mathcal{E}) = \mathbb{L}(\widehat{AC}) + \mathbb{L}(\widehat{CB}) = \operatorname{meas}\angle AOC + \operatorname{meas}\angle COB;$$

By Theorem CS.19(B), $\mathbb{L}(\mathcal{E}) = \pi$. $\square$

Theorem AM.8 (Sum of measures of angles of a triangle is π.) *Let A, B, and C be noncollinear points on $\mathbb{R}^2$, then* $\operatorname{meas}\angle ABC + \operatorname{meas}\angle BCA + \operatorname{meas}\angle CAB = \pi$.

Proof. Let $\mathcal{L} = \operatorname{par}(B, \overleftrightarrow{CA})$ be the line parallel to $\overleftrightarrow{CA}$) which contains B. (cf Axiom PS) Let D and E be points of $\mathcal{L}$ such that D is on the A-side of $\overleftrightarrow{BC}$, E is on the C-side of $\overleftrightarrow{AB}$, and D–B–E. By Theorem EUC.11 $\angle ABD \cong \angle BAC$ and $\angle EBC \cong \angle ACB$. By Theorem AM.6 $\operatorname{meas}\angle ABD = \operatorname{meas}\angle BAC$ and $\operatorname{meas}\angle EBC = \operatorname{meas}\angle ACB$. By Theorem AM.7(A) $\operatorname{meas}\angle EBC + \operatorname{meas}\angle ABC = \operatorname{meas}\angle EBA$. Putting all this together and using Theorem AM.7(B),

$$\begin{aligned}
&\operatorname{meas}\angle ACB + \operatorname{meas}\angle ABC + \operatorname{meas}\angle BAC \\
&\quad = \operatorname{meas}\angle EBC + \operatorname{meas}\angle ABC + \operatorname{meas}\angle ABD \\
&\quad = \operatorname{meas}\angle EBA + \operatorname{meas}\angle ABD = \pi. \quad \square
\end{aligned}$$

Corollary AM.9 *Let A, B, and C be noncollinear points on $\mathbb{R}^2$, and let D be a point such that C–A–D; then* $\operatorname{meas}\angle ABC + \operatorname{meas}\angle BCA = \operatorname{meas}\angle BAD$.

Proof. Let E be a point such that $E \in \operatorname{par}(A, \overleftrightarrow{BC})$ and E and B are on the same side of $\overleftrightarrow{AC}$. By Theorem EUC.11 $\angle ABC \cong \angle BAE$ and $\angle BCA \cong \angle EAD$. By Theorem AM.6 $\operatorname{meas}\angle ABC = \operatorname{meas}\angle BAE$ and $\operatorname{meas}\angle ACB = \operatorname{meas}\angle DAE$. By Theorem AM.7 $\operatorname{meas}\angle BAE + \operatorname{meas}\angle DAE = \operatorname{meas}\angle BAD$. Therefore $\operatorname{meas}\angle ABC + \operatorname{meas}\angle ACB = \operatorname{meas}\angle BAD$. $\square$

We can restate Corollary AM.9 thusly: *The sum of the measures of any two angles of a triangle is equal to the measure of an exterior angle at the other corner of the triangle.*

Theorem AM.10 *Let r be a positive real number and let $\mathcal{C}(O; r)$ be the circle with center O and radius r on $\mathbb{R}^2$. Let A, B, and C be noncollinear points on $\mathcal{C}(O; r)$ such that C–O–A; then* $\operatorname{meas}\angle ACB = \operatorname{meas}\angle OBC = \frac{1}{2}\operatorname{meas}\angle AOB$.

Proof. Since $\operatorname{dis}(O,B) = \operatorname{dis}(O,C) = r$, $\overline{OB} \cong \overline{OC}$. By Theorem NEUT.40 (*Pons Asinorum*) $\angle OBC \cong \angle OCB$. By Theorem AM.6 $\operatorname{meas}\angle OBC = \operatorname{meas}\angle OCB$. By Theorems AM.6 and AM.9 $\operatorname{meas}\angle AOB = \operatorname{meas}\angle OBC + \operatorname{meas}\angle OCB = 2\operatorname{meas}\angle ACB$. Therefore $\operatorname{meas}\angle ACB = \operatorname{meas}\angle OBC = \frac{1}{2}\operatorname{meas}\angle AOB$. $\square$

Corollary AM.11 *Let r be a positive real number and let $\mathcal{C}(O;r)$ be the circle with center O and radius r on $\mathbb{R}^2$. Let A, B, and C be noncollinear points on $\mathcal{C}(O;r)$ such that C–O–A and A, B, and C are noncollinear, then* $\operatorname{meas}\angle ACB = \frac{\pi}{2}$.

Proof. By Theorem AM.10,

$$\operatorname{meas}\angle OBC = \tfrac{1}{2}\operatorname{meas}\angle AOB \text{ and } \operatorname{meas}\angle OBA = \tfrac{1}{2}\operatorname{meas}\angle BOC.$$

Then by Theorem AM.7(B), $\operatorname{meas}\angle AOB + \operatorname{meas}\angle BOC = \pi$, so that $\operatorname{meas}\angle ABC = \operatorname{meas}\angle OBC + \operatorname{meas}\angle OBA = \frac{\pi}{2}$. $\square$

Theorem AM.12 *Let r be a positive real number and let $\mathcal{C}(O;r)$ be the circle on $\mathbb{R}^2$; let A, B, and C be noncollinear points on $\mathcal{C}(O;r)$, then* $\operatorname{meas}\angle BAC = \frac{1}{2}\operatorname{meas}\angle BOC$.

Proof. (Case 1: A–O–C.) This is Theorem AM.10.
In the next two cases, let D be the point of intersection of $\overrightarrow{AO}$ and $\mathcal{C}(O;r)$.

(Case 2: $O \in \operatorname{ins}\angle BAC$.) By Theorem AM.7

$$\operatorname{meas}\angle BAC = \operatorname{meas}\angle BAD + \operatorname{meas}\angle CAD.$$

By Theorem AM.10,

$$\operatorname{meas}\angle CAD = \tfrac{1}{2}\operatorname{meas}\angle COD \text{ and } \operatorname{meas}\angle BAD = \tfrac{1}{2}\operatorname{meas}\angle BOD.$$

By Theorem AM.7 $\operatorname{meas}\angle BOD + \operatorname{meas}\angle COD = \operatorname{meas}\angle BOC$. Therefore

$$\begin{aligned}\tfrac{1}{2}\operatorname{meas}\angle BOC &= \tfrac{1}{2}\operatorname{meas}\angle BOD + \tfrac{1}{2}\operatorname{meas}\angle COD \\ &= \operatorname{meas}\angle BAD + \operatorname{meas}\angle CAD = \operatorname{meas}\angle BAC.\end{aligned}$$

(Case 3: $O \in \operatorname{out}\angle BAC$.) By Theorem AM.7(A)

$$\operatorname{meas}\angle BAC + \operatorname{meas}\angle CAD = \operatorname{meas}\angle BAD$$

and

$$\operatorname{meas}\angle BOC + \operatorname{meas}\angle COD = \operatorname{meas}\angle BOD.$$

Therefore,

$$\operatorname{meas}\angle BAC = \operatorname{meas}\angle BAD - \operatorname{meas}\angle CAD$$

and

$$\operatorname{meas}\angle BOC = \operatorname{meas}\angle BOD - \operatorname{meas}\angle COD.$$

By Case 1 (that is, Theorem AM.10),

$$\tfrac{1}{2}\operatorname{meas}\angle COD = \operatorname{meas}\angle CAD \text{ and } \tfrac{1}{2}\operatorname{meas}\angle BOD = \operatorname{meas}\angle BAD.$$

Then

$$\begin{aligned}\tfrac{1}{2}\,\mathrm{meas}\,\angle BOC &= \tfrac{1}{2}\,\mathrm{meas}\,\angle BOD - \tfrac{1}{2}\,\mathrm{meas}\,\angle COD\\ &= \mathrm{meas}\,\angle BAD - \mathrm{meas}\,\angle CAD = \mathrm{meas}\,\angle BAC. \quad \square\end{aligned}$$

Theorem AM.13 (Geometric proof of square roots) *Given a segment with length H, another segment can be constructed having length $\sqrt{H}$.*

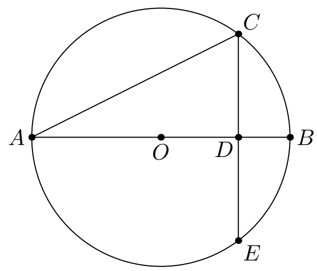


Fig. 5.1 For Theorem AM.13.

Proof. Refer to Figure 5.1. Let O be any point on $\mathbb{R}^2$ and let $\mathcal{C}(O; \frac{H+1}{2})$ be the circle on $\mathbb{R}^2$ with center O and radius $\frac{H+1}{2}$. Let A and B be points on this circle such that A–O–B, and let D be the point on $\overrightarrow{AO}$ such that $\mathrm{dis}(A, D) = H$ and $\mathrm{dis}(D, B) = 1$ (if $H = 1$ then $D = O$). Let $\mathcal{M}$ be the line through D which is perpendicular to $\overleftrightarrow{AB}$. Since $\mathrm{dis}(D, O) < \frac{H+1}{2}$, by Exercise CS.13 the line $\mathcal{M}$ intersects $\mathcal{C}(O; \frac{H+1}{2})$ at two points C and E. By Theorem AM.8

$$\mathrm{meas}\,\angle CAD + \mathrm{meas}\,\angle ACD + \mathrm{meas}\,\angle ADC = \pi.$$

By Exercise AM.1, $\mathrm{meas}\,\angle ADC = \frac{\pi}{2}$ so that

$$\mathrm{meas}\,\angle CAD + \mathrm{meas}\,\angle ACD = \tfrac{\pi}{2}.$$

By similar reasoning, $\mathrm{meas}\,\angle BCD + \mathrm{meas}\,\angle CBD = \frac{\pi}{2}$.

By Corollary AM.11, $\mathrm{meas}\,\angle ACB = \frac{\pi}{2}$; again by Theorem AM.8

$$\mathrm{meas}\,\angle ABC + \mathrm{meas}\,\angle ACB + \mathrm{meas}\,\angle BAC = \pi,$$

so that $\mathrm{meas}\,\angle ABC + \mathrm{meas}\,\angle BAC = \frac{\pi}{2}$.

Since $\triangle ABC$ and $\triangle ADC$ have $\angle BAC$ in common,

$$\mathrm{meas}\,\angle ACD = \mathrm{meas}\,\angle CBD.$$

Since $\triangle ABC$ and $\triangle CBD$ have $\angle ABC$ in common,

$$\mathrm{meas}\,\angle BAC = \mathrm{meas}\,\angle BCD.$$

By *Specht* Ch.15 Theorem SIM.18 $\triangle ACD \sim \triangle CBD$. Moreover $\dfrac{[\overline{CD}]}{[\overline{AD}]} = \dfrac{[\overline{BD}]}{[\overline{CD}]}$, *i.e.*, $[\overline{CD}]^2 = [\overline{AD}] \odot [\overline{DB}]$. Hence $(\mathrm{dis}(C,D))^2 = \mathrm{dis}(A,D) \cdot \mathrm{dis}(D,B)$. Since $\mathrm{dis}(D,B) = 1$, $\mathrm{dis}(C,D) = \sqrt{\mathrm{dis}(A,D)}$. $\square$

Remark AM.14 In algebra the ordered extension field $\mathcal{SF}$ of the rational numbers such that every positive (real) number belonging to it has a square root is called the **surd** field. It is the "smallest" ordered subfield of real numbers such that every positive number belonging to it has a square root belonging to it.

Theorem AM.13 shows that the underlying field in coordinate space must contain square roots in order to carry out the full development of this book in coordinate space.

That is, given a segment $\overline{AD}$ having length H, from Theorem AM.13 there exists a segment $\overline{CD}$ having length $\sqrt{H}$. Definition VEC.26.1 in Chapter 1 defines the length of this segment to be the real number c such that $\overline{OcU} = \overline{OX}$, where $\overline{CD} \cong \overline{OS}$ and the length of U is 1. This is impossible unless there is a real number $\sqrt{H}$.

5.2 Exercises for angle measure

Exercise AM.1* *The radian measure of a right angle is* $\frac{\pi}{2}$.

5.3 Selected answers for angle measure

Exercise AM.1 Proof. By Definition NEUT.41(C) an angle $\angle BAC$ is right iff it is congruent to a supplement of itself. Let $\angle DAE$ be a supplement of $\angle BAC$, so that $\angle DAE \cong \angle BAC$. By Theorem AM.7(B), $\mathrm{meas}\,\angle DAE + \mathrm{meas}\,\angle BAC = \pi$. By Theorem AM.6 $\mathrm{meas}\,\angle DAE = \mathrm{meas}\,\angle BAC$, therefore $\mathrm{meas}\,\angle DAE = \mathrm{meas}\,\angle BAC = \frac{\pi}{2}$. $\square$

Chapter 6

The Jordan Curve Theorem for Polygons

Dependencies: *Chapters 1, 4, 5, and 6 from Euclidean Geometry and its Subgeometries (Specht)*
Acronyms: *JCT, PLGN, SEP, CNV, CNT*
New terms defined: *polygon, inside, outside, enclosure, exclosure, side; polygonal path, subpath, j-corner, j-edge, adjacent edges, adjacent corners, endpoints, simple; polygonally connected, admissible ray, admissible angle, entering, exiting; parity, even, odd; separates the plane; support, supporting line, basic line, extremal point, normal point, bounded (set); regular corner, irregular corner*

This chapter is dependent only on Pasch geometry and ordering, as developed in Chapters 5 and 6 of *Specht*; it does not depend on, or refer to other chapters in this Supplement. In it, "plane" will mean "Pasch plane."

Items referenced by acronyms JCT, PLGN, SEP, CNV, or CNT (for instance "Lemma PLGN.3") will be internal references to this chapter. Other items (as, for instance "Theorem PSH.6") will be from chapters in *Specht* according to the following table, which is a subset of the abbreviated Table of Contents for *Specht* at the end of the Preface of this Supplement.

Acronym	Chapter	Title	Page
I	1	Preliminaries and Incidence Geometry	1
IB	4	Incidence and Betweenness	63
PSH	5	Pasch Geometry	79
ORD	6	Ordering a line in a Pasch Plane	139

The Jordan Curve Theorem says that the complement of a simple, closed curve is the union of two connected sets, one of them (the interior) being bounded, the other (the exterior) unbounded; the curve is the boundary of both the interior and the exterior, and any path from the interior to the exterior must intersect this boundary.

This is quite intuitively obvious; indeed, Professor F. E. Ulrich used to introduce it in his Complex Analysis course at Rice by saying "Every cow knows this theorem; she knows that if she's in the pasture and wants to get out, she has to cross the fence."

Camille Jordan (1838–1922) was the first to give a proof in the continuous case, which was published in his book *Cours d'analyse de l'École Polytechnique* [1]; but his proof was considered by many mathematicians to be faulty. Moreover, Jordan assumed the validity of the theorem for the case of a simple closed polygonal path.

It is this polygonal case that we prove in this chapter. Some say this case is quite easy to prove, but we have not sought a minimal path to it; we will, rather, embark on a rather leisurely exploration of the properties of polygons which will eventuate in Jordan's theorem, which we now state more precisely.

Theorem JCT.1 (Jordan Curve Theorem (JCT) for a simple polygon) *If $\mathcal{G}$ is a simple polygon in the Pasch plane $\mathcal{P}$, then*

(A) $\mathcal{P} = \mathcal{G} \cup \operatorname{ins}\mathcal{G} \cup \operatorname{out}\mathcal{G}$, *where* $\mathcal{G}$, $\operatorname{ins}\mathcal{G}$, *and* $\operatorname{out}\mathcal{G}$ *are pairwise disjoint sets;*

(B) *if* $P \in \operatorname{ins}\mathcal{G}$ *and* $Q \in \operatorname{out}\mathcal{G}$, *then* $\overline{PQ} \cap \mathcal{G} \neq \emptyset$*;*

(C) $\mathcal{G}$ *and* $\operatorname{ins}\mathcal{G}$ *are bounded sets, and* $\operatorname{out}\mathcal{G}$ *is unbounded; and*

(D) $\operatorname{ins}\mathcal{G}$ *and* $\operatorname{out}\mathcal{G}$ *are polygonally connected sets.*

Here we have violated our usual practice and used terms before they are defined, relying on your intuition for the meanings of several words. We have not defined the terms *simple polygon* or *polygonal connectedness* (to be defined in Definition PLGN.5), let alone *inside* ($\operatorname{ins}\mathcal{G}$) and *outside* ($\operatorname{out}\mathcal{G}$) of a polygon $\mathcal{G}$ (to be defined in Definition SEP.3). The terms *bounded* and *unbounded* sets are defined in Definition CNV.21.

Remark JCT.2 (Alternatives for reading the chapter) Conclusions (A) and (B) of Theorem JCT.1 are Theorem SEP.12; together they define what we mean by saying that the curve (or polygon) "is the boundary of both the interior and the exterior" of the plane: the *boundary* is what the cow has to cross to get out of the pasture.

Conclusion (C) is Theorem CNV.22; and conclusion (D) is Theorem CNT.3. This provides several alternatives for reading the chapter, or part of it.

The development leading to conclusion (C) includes Theorems SEP.13 through SEP.15, and the development leading to conclusion (D) includes at least Theorems CNV.22 through CNV.29. The acronyms do not correspond to the part of the JCT theorem being proved.

The reader who wishes to pursue the proof in the more general continuous case might well begin with G. T. Whyburn, *Analytic Topology*, AMS Colloquium Publications, Chapter VI, (Reprint) [4].

6.1 Segments and rays (PLGN)

This section develops basic facts about the intersections of a ray and an arbitrary collection of segments in the plane. In order to have the most general possible applicaton we use the notation CD^S to denote a "generic" segment with endpoints C and D. Thus, CD^S will denote either $\overset{[\,\text{—}\,]}{CD}, \overset{[\,\text{—}\,[}{CD}, \overset{]\,\text{—}\,]}{CD}$, or $\overset{]\,\text{—}\,[}{CD}$. In the case that $C = D$, either $CD^S = \overset{[\,\text{—}\,]}{CD} = \{C\} = \{D\}$ or $CD^S = \emptyset$. (Note that *Specht* Ch.4 Definition IB.3, where segments are defined, does not accommodate the case where both endpoints are the same and the segment is a single point, whereas here we may wish sometimes to regard that case as a "degenerate" segment.)

Lemma PLGN.1 (Intersections of a segment and a ray) *Let UV^S be any segment in the Pasch plane $\mathcal{P}$, with $U \neq V$, and let A and B be two distinct points of the plane, where $A \notin UV^S$. Then exactly one of the following is true:*

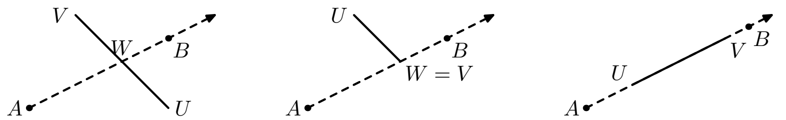
(0) $\overrightarrow{AB} \cap UV^S = \emptyset$,

(1) $\overrightarrow{AB} \cap UV^S = \{W\}$ *where* $U{-}W{-}V$,

(2) $\overrightarrow{AB} \cap UV^S = \{W\}$ *where* $W = U$ *or* $W = V$, *or*

(3) $\overrightarrow{AB} \cap UV^S = UV^S$ *and* $A, B, U,$ *and* V *are collinear.*

Proof. See Figure 6.1. If $\overrightarrow{AB} \cap UV^S \neq \emptyset$, either the intersection is a single point or is not. If it is the single point $\{W\}$, either $W = U$, $W = V$ or $U{-}W{-}V$. (Alternatives (1) and (2) in the statement of the Lemma)



(1) Intersecting at one point, not an endpoint

(2) Intersecting at one point, an endpoint

(3) Intersecting at more than one point

Fig. 6.1 Showing possibilities for intersection of a ray and a segment.

If the intersection contains more than one point, by Axiom I.1 the line $\overleftrightarrow{AB}$ is the same as the line $\overleftrightarrow{UV}$ so that A, B, U, and V must be collinear. If UV^S is not a subset of $\overrightarrow{AB}$, let X be a point of UV^S that is not a member of $\overrightarrow{AB}$; then X–A–B. Let $Y \in \overrightarrow{AB} \cap UV^S$, so that X–A–Y. Both X and Y belong to UV^S and therefore $A \in UV^S$, a contradiction to our hypothesis. Thus $UV^S \subseteq \overrightarrow{AB}$ and (3) is established. □

Corollary PLGN.1.1 (A) *If $A, B, U,$ and V are collinear points, either (0) or (3) holds.*

(B) *Conclusion (2) implies that either U or $V \in UV^S$, and is impossible if $UV^S = \overline{UV}$.*

Corollary PLGN.1.2 *If $A, B,$ and U are distinct points of the plane and $UV^S = \overline{UU}$, then conclusion (1) is impossible, and (2) and (3) are equivalent.*

In what follows, we adopt the following conventions: if $\mathcal{K}$ and $\mathcal{L}$ are two subsets of a set ordered by $<$ and A is a point of the set, then

(1) $\mathcal{K} < \mathcal{L}$ means that for every $K \in \mathcal{K}$ and every $L \in \mathcal{L}$, $K < L$;

(2) $A < \mathcal{L}$ means $\{A\} < \mathcal{L}$, that is, for every $L \in \mathcal{L}, A < L$; and

(3) $\mathcal{L} < A$ means $\mathcal{L} < \{A\}$, that is, for every $L \in \mathcal{L}, L < A$.

Lemma PLGN.2 (Order of disjoint collinear segments) *Let $\overline{PQ}$ be a segment in the plane which is ordered by an order relation $<$ according to Definition ORD.1 (as a subset of a line $\mathcal{L}$). Let $A, B, C,$ and D be distinct points on $\overline{PQ}$.*

(A) *If $A \notin CD^S$, and there exists a point $Y \in CD^S$ such that $A < Y$, then for every $Z \in CD^S$, $A < Z$, that is, $A < CD^S$.*

(B) *If $A \notin CD^S$, and there exists a point $Y \in CD^S$ such that $A > Y$, then for every $Z \in CD^S$, $A > Z$, that is, $A > CD^S$.*

(C) *If* $AB^S \cap CD^S = \emptyset$, *and there exists a point* $X \in AB^S$ *and a point* $Y \in CD^S$ *such that* $X < Y$, *then for every* $W \in AB^S$ *and* $Z \in CD^S$, $W < Z$, *that is,* $AB^S < CD^S$.

Proof. (A) Let Z be any point of CD^S; either $Y \leq Z$ or $Z < Y$. In the first case, $A < Y \leq Z$ so $A < Z$. In the second case, if $Z \leq A, Z \neq A$ because $A \notin CD^S$, so the only possibility is $Z < A < Y$; but then $A \in \overline{ZY} \subseteq CD^S$, so the only possibility is $Z < A < Y$; but then $A \in \overline{ZY} \subseteq CD^S$ which is impossible by hypothesis. A similar proof (with the inequalities reversed) shows part (B).

(C) Since $X < Y$ and $X \notin CD^S$, by part (A), $X < Z$ for every $Z \in CD^S$. Let Z be any member of CD^S; since $Z \notin AB^S$ and $Z > X$, for every $W \in AB^S, Z > W$ by part (B). Since Z is arbitrary, $AB^S < CD^S$. □

Lemma PLGN.3 *Let* P *and* Q *be any distinct points on the plane; let* $<$ *be the ordering on* $\overleftrightarrow{PQ}$ *given by Specht Ch.6 Definition ORD.1 with* $P < Q$, *and let* $\mathcal{E}$ *be any finite subset of* $\overrightarrow{PQ}$. *Then*

(1) *there exists a minimum (least, first) point* C *and a maximum (greatest, last) point* D *for* $\mathcal{E}$ *(see also Remark PLGN.5.1),*

(2) $\mathcal{E} \subseteq \overline{CD}$, *and*

(3) *if* $\mathcal{H}$ *is any segment and* $\mathcal{E} \subseteq \mathcal{H} \subseteq \overleftrightarrow{PQ}$, *then* $\overline{CD} \subseteq \mathcal{H}$ *(*$\overline{CD}$ *is the smallest segment containing* $\mathcal{E}$*).*

Proof. Conclusion (1) is immediate from Theorem ORD.10; (2) follows (Definition ORD.8) because for all $X \in \mathcal{E}, X \geq C$ and $X \leq D$. To show (3) suppose $\mathcal{E} \subseteq \mathcal{H}$, and there is a point $X \in \overline{CD}$ such that $X \notin \mathcal{H}$. Then $C \leq X \leq D$. $\mathcal{H}$ is a segment containing $\mathcal{E}$ so either $\mathcal{H} < X$ or $\mathcal{H} > X$. If $\mathcal{H} < X, \mathcal{E} < X$ and in particular $D < X$ which contradicts $X \leq D$. A similar proof gives a contradiction if $\mathcal{H} > X$. □

Theorem PLGN.4 *Let* $\mathcal{E} = \bigcup_{k=1}^{n} \mathcal{E}_k$, *where each* $\mathcal{E}_k$ *is a closed segment (possibly a single point) in the Pasch plane* $\mathcal{P}$. *Let* P *and* Q *be points on the Pasch plane* $\mathcal{P}$ *where* $P \notin \mathcal{E}$.

(A) *If* $\overleftrightarrow{PQ}$ *is ordered so that* $P < Q$ *and* $\overrightarrow{PQ} \cap \mathcal{E} \neq \emptyset$, *there exists a first point* $C \in \overrightarrow{PQ}$ *such that* $C \in \mathcal{E}$, *and a last point* $D \in \overrightarrow{PQ}$ *such that* $D \in \mathcal{E}$. *Moreover,* $\overrightarrow{PQ} \cap \mathcal{E} \subseteq \overline{CD}$.

(B) *There exists a point* $A \in \overrightarrow{PQ}$ *such that* $\overline{PA} \cap \mathcal{E} = \emptyset$.

Proof. (A) The intersection of every closed segment with $\overrightarrow{PQ}$ is a closed segment or the empty set. By elementary set theory, $\mathcal{E} \cap \overrightarrow{PQ} = \bigcup_{k=1}^{n} (\mathcal{E}_k \cap \overrightarrow{PQ})$ and each set $\mathcal{F}_k = \mathcal{E}_k \cap \overrightarrow{PQ}$ is also a closed segment which is a subset of $\overrightarrow{PQ}$.

Each segment $\mathcal{F}_k$ has endpoints C_k and D_k, where $C_k \leq D_k$ (again allowing the possibility that $C_k = D_k$, so that for each k, $\mathcal{F}_k = \overline{C_k D_k}$). Let $\mathcal{G}$ be the set of all end points C_k and D_k. Using Lemma PLGN.3(1), choose C and D to be the first and last points, respectively, of $\mathcal{G}$.

Let X be any point of $\bigcup_{k=1}^{n} \mathcal{F}_k$; then for some k, $X \in \mathcal{F}_k$, so that $C \leq C_k \leq X \leq D_k \leq D$. Therefore C is the first point of $\overrightarrow{PQ}$ such that $C \in \bigcup_{k=1}^{n} \mathcal{F}_k = \mathcal{E} \cap \overrightarrow{PQ}$, D is the last such point, and $\mathcal{E} \cap \overrightarrow{PQ} \subseteq \overline{CD}$.

(B) If $\overrightarrow{PQ} \cap \mathcal{E} = \emptyset$ then let A be any point of $\overrightarrow{PQ}$, and $\overline{PA} \cap \mathcal{E} = \emptyset$. If $\overrightarrow{PQ} \cap \mathcal{E} \neq \emptyset$, order $\overleftrightarrow{PQ}$ so that $P < Q$, and apply part (A) to get C, the first point of intersection of $\overrightarrow{PQ}$ and $\mathcal{E}$. Let A be any point of $\overrightarrow{PQ}$ such that P–A–C; then $\overline{PA} \cap \mathcal{E} = \emptyset$. $\square$

6.2 Polygons, polygonal paths, and rays (PLGN)

In the following theorems we will need some basic terminology about modular integers. We say that an integer a is **divisible** by an integer b iff there exists an integer c such that $a = bc$, that is, b divides a without a remainder. Let m be a natural number, and let a and b be integers. We say that a is **congruent to** b **mod** m (and write $a \equiv b \pmod{m}$) iff $a - b$ is divisible by m, which is called the **modulus**.

For any natural number m, the relation $a \equiv b \pmod{m}$ is an *equivalence relation* (*cf* Section 1.4). The equivalence class of any integer a is the set of all integers of the form $km + a$, where k is any integer and m is the modulus.

In modular numbering, two integers which differ by a multiple of the modulus are identified, as they belong to the same equivalence class (mod m). Telling time on an ordinary 12 hour clock is a good example of a use of modular numbers. There is a 48 hour difference between 10 AM on 11 May and 10 AM on 13 May, but the clock indicates the same at both times because the number of hours between these two times is a multiple of 12.

In this section we will arbitrarily choose some corner of a polygon[1] as the "first" one, and label it as X_1. As the polygon is traversed in some

[1] Here we are being *very* informal, as we have not yet defined the term *polygon*.

predetermined direction, each corner encountered in the traversal is named successively X_2, X_3, and so on, until the first corner is again reached. X_1 is then given a second name, X_{m+1}, where m is the number of corners of the polygon (the modulus of the numbering system). Additional traversals in the same direction will re-name X_1 as X_{1+2m}, X_{1+3m}, ... and so on. The polygon can also be traversed in the opposite direction, with successive corners being labeled X_0, X_{-1}, X_{-2}, and so on, until the first corner is reached, and is named X_{1-m}; another traversal in the same direction will name this point X_{1-2m}, and so on. Thus the first corner will end up having infinitely many names, $\ldots$, X_{1-3m}, X_{1-2m}, X_{1-m}, X_1, X_{1+m}, X_{1+2m}, X_{1+3m}, $\ldots$, each of which differs from another by a multiple of m. We formalize these ideas in the following definitions and proofs.

It is not our purpose to create a general theory of polygons, so in the next definition we define only *simple* polygons. We will sometimes use the word "polygon" alone, but it will always mean "simple polygon."

Definition PLGN.5 Let X be a mapping from the set $\mathbb{Z}$ of integers into a Pasch plane $\mathcal{P}$ (that is, $X : \mathbb{Z} \to \mathcal{P}$). Then X pairs each integer $i = \ldots, -m, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, , n, \ldots$ with a point X_i of $\mathcal{P}$ (here, instead of the usual notation $X(i)$ for the value of X at the integer i, we write X_i). X is a **labeling** function, and is necessarily *many-to-one*, as it pairs many integers with the same point on the plane.

(A) **Polygons**: Let $n \geq 3$, and suppose that for all integers i and j, $X_i = X_j$ iff $i \equiv j (\mathrm{mod}\ n)$. The set $\mathcal{G} = \bigcup_{k=1}^{n} \overline{X_k X_{k+1}}$ is a **simple polygon** (notation: $\langle X_1, \ldots, X_n \rangle$) iff both

(1) for all integers j,

$\overline{X_j X_{j+1}} \cap \overline{X_{j+1} X_{j+2}} = \{X_{j+1}\}$ and

X_j, X_{j+1}, and X_{j+2} are noncollinear, and

(2) for all integers j and k such that $j \not\equiv k (\mathrm{mod}\ n)$, $j \not\equiv k+1 (\mathrm{mod}\ n)$, and $j \not\equiv k-1 (\mathrm{mod}\ n)$, $\overline{X_j X_{j+1}} \cap \overline{X_k X_{k+1}} = \emptyset$.

For every integer j, the point X_j is called the j**-corner** and the segment $\overline{X_j X_{j+1}}$ is called the j**-edge**.

Two *integers* i and j are **adjacent** iff $i = j + 1$ or $j = i + 1$.

Two *corners* X_i and X_j of a polygon with n edges are **adjacent** iff either $i \equiv j + 1 (\mathrm{mod}\ n)$ or $j \equiv i + 1 (\mathrm{mod}\ n)$.

The i-edge $\overline{X_iX_{i+1}}$ and the j-edge $\overline{X_jX_{j+1}}$ of a polygon are **adjacent** iff either $i \equiv j+1 \pmod{n}$ or $j \equiv i+1 \pmod{n}$.

Note that any quadrilateral as defined by *Specht* Ch.5 Definition PSH.31 is a simple polygon, since opposite edges do not intersect, and any triangle as defined by Definition IB.7 is a simple polygon (vacuously, since condition (2) of Definition PLGN.5(A) is true for a triangle because its hypothesis is false).

(B) **Polygonal paths**: In this part of the definition, we do not use modular numbering. Let $\mathcal{P}$ be the Pasch plane and X be a mapping from the set $[1; m+1]$ of integers into $\mathcal{P}$ such that for all integers i and $j \in [1; m+1]$, $X_i = X_j$ iff $i = j$.

(1) Let $m \geq 3$; the set $\mathcal{J} = \bigcup_{k=1}^{m} \overline{X_kX_{k+1}}$ (notation: $\langle\langle X_1, \ldots, X_{m+1}\rangle\rangle$) is a **polygonal path** (with endpoints X_1 and X_{m+1}) iff for all integers $j \in [1; m-1]$,

$\overline{X_jX_{j+1}} \cap \overline{X_{j+1}X_{j+2}} = \{X_{j+1}\}$ and

X_j, X_{j+1}, and X_{j+2} are noncollinear.

In this case we may say that $\mathcal{J}$ is a polygonal path **joining**, or **connecting** X_1 and X_{m+1}.

For any $k \in [1; m+1]$, X_k is the *k-corner*; for any $k \in [1; m]$, $\overline{X_kX_{k+1}}$ is the *k-edge*; and X_1, X_{m+1} are the **endpoints** of $\langle\langle X_1, \ldots, X_{m+1}\rangle\rangle$.

(2) Let $\mathcal{J}$ be a polygonal path; then $\mathcal{I}$ is a **subpath** of $\mathcal{J}$ iff $\mathcal{I}$ is a polygonal path and $\mathcal{I} \subseteq \mathcal{J}$. This means that every point of $\mathcal{I}$ is also a point of $\mathcal{J}$.

(3) A polygonal path $\langle\langle X_1, \ldots, X_{m+1}\rangle\rangle$ is said to be **simple** iff for all members j and k of $[1; m]$, if $j+1 < k$, then $\overline{X_jX_{j+1}} \cap \overline{X_kX_{k+1}} = \emptyset$.

(4) A subset $\mathcal{E}$ of the plane $\mathcal{P}$ is **polygonally connected** iff for every pair $\{A, B\}$ of distinct points in $\mathcal{E}$ there exists a polygonal path $\mathcal{J}$ with endpoints A and B such that $\mathcal{J} \subseteq \mathcal{E}$.

Remark PLGN.5.1 In this chapter we shall frequently be dealing with a ray $\overrightarrow{AB}$ which intersects a set such as a polygon at a finite number of points. In such cases we will frequently speak of the *first* or *last* intersection of the ray with the other set. In such cases it will be understood that we are assuming an order relation $<$ to exist on the ray with $A < B$, and that the *first* and *last* points of intersection are those guaranteed by Lemma PLGN.3(1).

Remark PLGN.6 (A) The edges of a polygonal path intersect at the common endpoints of edges with adjacent indices, and if the path is simple, adjacent edges will intersect only at their endpoints (for if two adjacent edges intersect at two points, their lines will be the same and there will be three successive collinear corners). If the path is not simple there will be pairs of non-adjacent edges intersecting, possibly at endpoints and possibly at other points.

(B) Any subpath of a simple polygonal path $\mathcal{J}$ is also simple.

(C) Let $\mathcal{G} = \langle X_1, \ldots, X_n \rangle$ be a polygon and let A and B be distinct points of $\mathcal{G}$ where $A \in \overline{X_i X_{i+1}}$ and $B \in \overline{X_j X_{j+1}}$ where $i \leq j$. Then $\langle\langle A, X_{i+1}, \ldots, X_j, B \rangle\rangle$ is a polygonal path connecting A and B. Thus every polygon is polygonally connected.

(D) If $\mathcal{E}$ and $\mathcal{F}$ are polygonally connected subsets of the plane and $\mathcal{E} \cap \mathcal{F} \neq \emptyset$, then $\mathcal{E} \cup \mathcal{F}$ is polygonally connected.

(E) Every convex set is polygonally connected.

(F) If $\mathcal{G} = \langle X_1, \ldots, X_n \rangle$ and $\mathcal{F}$ are simple polygons and $\mathcal{F} \subseteq \mathcal{G}$, then $\mathcal{F} = \mathcal{G}$.

Assertions (A) through (E) do not need extensive proof. We prove assertion (F).

(i) First we need to prove that if $\mathcal{F} \subseteq \mathcal{G}$, every edge $\overline{AB}$ of $\mathcal{F}$ is a subset of some edge of $\mathcal{G}$. Since $\overline{AB}$ contains infinitely many points, and there are only finitely many edges of $\mathcal{G}$, some edge $\overline{CD}$ of $\mathcal{G}$ contains at least two points of $\overline{AB}$. Then $\overline{AB} \subseteq \overleftrightarrow{CD}$, and $\overline{AB} \cap \overline{CD}$ is a segment. If $\overline{AB} \not\subseteq \overline{CD}$, then A–C–B–D or A–C–D–B or C–A–D–B.

We will prove the result for either of the first two cases; the proof in the last case is similar. There exists exactly one other edge $\overline{CC'}$ of $\mathcal{G}$ which intersects $\overline{CD}$ at the point C, and C' cannot be collinear with C and D, so that $\overline{CC'} \cap \overleftrightarrow{CD} = \emptyset$. Let $\mathcal{G}'$ denote the union of all edges of $\mathcal{G}$ other than $\overline{CD}$ and $\overline{CC'}$. Then $C \notin \mathcal{G}'$. If the ray $\overrightarrow{CA}$ has empty intersection with $\mathcal{G}'$ then $A \notin \mathcal{G}'$ and, because $A \notin \overline{CD}$, $A \notin \mathcal{G}$; this contradicts our assumption that $\mathcal{F} \subseteq \mathcal{G}$. Otherwise, $\overrightarrow{CA}$ has a first point P of intersection with $\mathcal{G}'$ (by Theorem PLGN.4) and either A–P–C–D or P–A–C–D. Choose Y so that both A–Y–C and P–Y–C. Then $Y \notin \mathcal{G}'$; $Y \notin \overline{CD}$ so $Y \notin \mathcal{G}$ which is a contradiction because $Y \in \overline{AB} \subseteq \mathcal{G}$. Hence $\overline{AB} \subseteq \overline{CD}$.

(ii) If two edges of $\mathcal{F}$ intersect at some point, then the two edges of $\mathcal{G}$ that contain them must also intersect at the same point. Thus, every corner of $\mathcal{F}$ is a corner of $\mathcal{G}$, and since this is true, the endpoints of any edge of $\mathcal{F}$ are corners of $\mathcal{G}$ and every edge of $\mathcal{F}$ is an edge of $\mathcal{G}$.

(iii) It follows that the set of all edges belonging to both $\mathcal{F}$ and $\mathcal{G}$ is non-empty. If $\mathcal{G}$ has an edge $\overline{X_i X_{i+1}}$ that is not an edge of $\mathcal{F}$, let j be the smallest integer such that $j > i$ and $\overline{X_jX_{j+1}}$ is an edge of both $\mathcal{F}$ and $\mathcal{G}$. Now there must be a second edge $\overline{AX_j}$ of $\mathcal{F}$ intersecting $\overline{X_jX_{j+1}}$ at X_j, but this second edge must be an edge of $\mathcal{G}$ which must therefore be $\overline{X_{j-1}X_j}$. Then $\overline{X_{j-1}X_j}$ is an edge of both $\mathcal{F}$ and $\mathcal{G}$, contradicting our choice of j. □

Theorem PLGN.7 (Simplification of a polygonal path) *Let $\mathcal{M} = \langle\langle X_1, \ldots, X_{m+1}\rangle\rangle$ be any polygonal path joining X_1 and X_{m+1}. Then there exists a simple subpath $\mathcal{M}'$ joining X_1 and X_{m+1}.*

Proof. We first define an ordering of $\mathcal{M}$. If P and Q are distinct points of $\mathcal{M}$, $P < Q$ if and only if either

(a) $P \in \overline{X_iX_{i+1}}$ and $Q \in \overline{X_jX_{j+1}}$ and $i < j$ (where $\overline{X_iX_{i+1}}$ and $\overline{X_jX_{j+1}}$ are edges of $\mathcal{M}$), or

(b) P and Q belong to the same edge $\overline{X_iX_{i+1}}$ of $\mathcal{M}$ and either

X_i–P–Q–X_{i+1},
$X =P$–Q–X_{i+1},
X_i–P–$Q = X_{i+1}$, or
$X_i = P$ and $X_{i+1} = Q$

(that is, the edge $\overline{X_iX_{i+1}}$ has the ordering of ORD.1 where $X_i < X_{i+1}$).

From Definition PLGN.5(B)(3) we see that the polygonal path $\mathcal{M} = \langle\langle X_1, \ldots, X_{m+1}\rangle\rangle$ is not simple iff for some members j and k of $[1; m]$, $j + 1 < k$ and $\overline{X_jX_{j+1}} \cap \overline{X_kX_{k+1}} \neq \emptyset$. Clearly this intersection will be a closed segment $\overline{AB}$, which may be degenerate, that is, a single point. Define a **loop point** to be an end point of a segment so defined. There are only finitely many such segments because the set of edges if finite, hence only finitely many loop points.

If A is a loop point of $\mathcal{M}$, there exists a sub-path $\langle\langle A = Y_1, \ldots, Y_{r+1} = A\rangle\rangle$ of $\mathcal{M}$, called a **loop**, which joins A back to a. A loop is a polygon, but is not necessarily simple. There are only finitely many such loops since there are only finitely many edges in $\mathcal{M}$.

Let A_1 be the first loop point, and let B_1 be the last point (in the ordering $<$ of $\mathcal{M}$) such that $B_1 = A_1$. Then there exists a polygonal path $\mathcal{N}_1 = \langle\langle A_1 = Y_1, \ldots, Y_{r+1} = B_1 = A_1\rangle\rangle$ which is a subpath of $\mathcal{M}$, where at least $Y_2, \ldots, Y_r$ are successive corners of $\mathcal{M}$.

Let $\mathcal{M}_1 = (\mathcal{M} \setminus \mathcal{N}_1) \cup \{A_1\}$. $\mathcal{M}_1$ is a subpath of $\mathcal{M}$ connecting X_1 to X_{m+1} in which all loops connecting A_1 to A_1 have been eliminated. If $\mathcal{M}_1$ is simple, the proof is complete. If not, the argument may be repeated as needed to arrive at the conclusion by induction. □

Definition PLGN.8 Let $\mathcal{G}$ be a simple polygon $\langle X_1, \ldots, X_n \rangle$ and A and B be two distinct points of the plane with $A \notin \mathcal{G}$. Then $\overrightarrow{AB}$ is an **admissible ray (for** $\mathcal{G}$**)** if it contains no corner of $\mathcal{G}$. An angle $\angle BAC$ in the plane is an **admissible angle** if both $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are admissible rays.

Remark PLGN.9 Let $\mathcal{G}$ be a simple polygon.

(A) The categories of "admissible" and "non-admissible" are applied only to rays $\overrightarrow{AB}$ and angles $\angle BAC$ where $A \notin \mathcal{G}$. Where there is only one polygon $\mathcal{G}$ in view, it will be convenient to speak of an admissible ray or angle without naming the polygon. In contexts of more than one polygon, we may find rays that are admissible for one polygon but not for another.

(B) If $\overrightarrow{AB}$ is an admissible ray, no intersection of $\overrightarrow{AB}$ with an edge of $\mathcal{G}$ satisfies either condition (2) or (3) of Lemma PLGN.1, because these imply that $\overrightarrow{AB}$ contains a corner. Thus every edge of $\mathcal{G}$ either is disjoint from $\overrightarrow{AB}$ (condition (0) of Lemma PLGN.1) or intersects $\overrightarrow{AB}$ at exactly one point, as in condition (1) of this lemma. Since $\mathcal{G}$ has a finite number of edges, $\overrightarrow{AB} \cap \mathcal{G}$ contains at most a finite number of points of $\mathcal{G}$.

(C) Let A be any point of the plane. There are finitely many rays $\overrightarrow{AX}$ that contain a corner of $\mathcal{G}$ and infinitely many that do not. Thus every point $A \notin \mathcal{G}$ always has both admissible and non-admissible rays $\overrightarrow{AB}$, and the number of non-admissible rays is no greater than the number of corners of $\mathcal{G}$.

(D) The following fact will make the statement of some theorems slightly more compact. If A is any point and $\mathcal{E}$ is a subset of the plane, there exists a point B such that $\overline{AB} \cap \mathcal{E} = \emptyset$ iff there exists a point C such that $\overrightarrow{AC} \cap \mathcal{E} = \emptyset$ (for if $\overline{AB} \cap \mathcal{E} = \emptyset$ we may choose $C \in \overline{AB}$).

Remark PLGN.10 A main purpose of this and the next section (SEP) is to define the "inside" and "outside" of an arbitrary simple polygon $\mathcal{G}$. It will be shown that either all the admissible rays from a point A not on $\mathcal{G}$ have an odd number of points of intersection with $\mathcal{G}$ (in which case we will define the *parity* of A to be *odd*), or they all have an even number of intersections (in which case the *parity* of A will be *even*). In the first case we will say that A is *inside* $\mathcal{G}$, and in the second case it will be *outside*. This makes it possible to use a single "test" ray to determine whether A is inside or outside the polygon. What makes this work is that all intersections of an admissible ray with edges of a polygon are genuine *crossings* of the polygon.

This project could have been carried out by considering *arbitrary* rays emanating from the point A, not just admissible ones. That is, test rays might include corners of $\mathcal{G}$. This approach would add some complexity because it would necessitate deciding which intersections (including those cases where the ray contains a whole edge) are legitimate *crossings* of $\mathcal{G}$ and should be counted. Intuitively, it is not hard to see that moving the ray over a little (*cf* Theorem PLGN.13) so as to avoid containing any corners will not affect the parity.

The main complexity in our approach of using only admissible rays arises from the unfortunate possibility that the most convenient ray for testing the parity of a point A might not be admissible. Lemma PLGN.11 through Theorem PLGN.13 are technical arguments which are useful in situations where the obvious test ray is not admissible, and an alternate test ray must be constructed.

Lemma PLGN.11 *Let $\mathcal{G}$ be a simple polygon, and let A be any point of the plane.*

(A) *If $A \notin \mathcal{G}$ and $X \neq A$ is arbitrarily chosen, then there exists a point $Z \in \overrightarrow{AX}$ such that $\overline{AZ} \cap \mathcal{G} = \emptyset$.*

(B) *if $A \in \mathcal{G}$ and if X is chosen so that $\overrightarrow{AX}$ contains no corner of $\mathcal{G}$, then there exists a point $Z \in \overrightarrow{AX}$ such that $\overline{AZ} \cap \mathcal{G} = \emptyset$.*

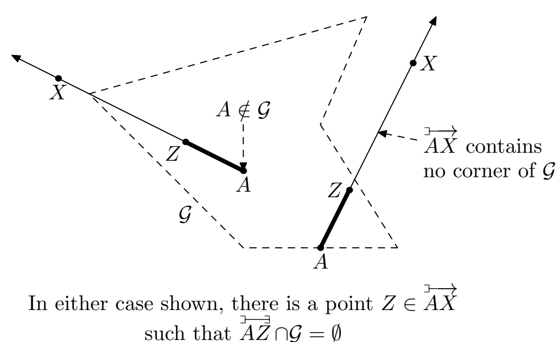


Fig. 6.2 For Lemma PLGN.11.

Proof. See Figure 6.2. (A) This is Theorem PLGN.4(B).

(B) Assume that A is a member of the edge $\overline{PQ} \subseteq \mathcal{G}$; we consider two cases:

(a) If P–A–Q, then $A \notin \mathcal{G} \setminus \overline{PQ}$. Then we may apply Theorem PLGN.4 to $\overrightarrow{AX}$ and the union of all the edges of $\mathcal{G}$ other than $\overline{PQ}$, that is, $\mathcal{G} \setminus \overline{PQ}$, and find that there exists a point Z such that $\overrightarrow{AZ} \cap \mathcal{G} \setminus \overline{PQ} = \emptyset$. Since $\overrightarrow{AX}$ contains no corner of $\mathcal{G}$, $X \notin \overleftrightarrow{PQ}$, and hence $\overrightarrow{AX} \cap \overline{PQ} = \{A\}$ so $\overrightarrow{AZ} \cap \overline{PQ} = \emptyset$, and the conclusion follows by elementary set theory.

(b) If $A = P$, let $\overline{RP}$ be the edge of $\mathcal{G}$ intersecting $\overline{PQ}$ at P; then

$$\overrightarrow{AX} \cap \overline{RP} = \overrightarrow{AX} \cap \overline{PQ} = \{P\}.$$

Applying Theorem PLGN.4 to $\mathcal{G} \setminus (\overline{RP} \cup \overline{PQ})$ we conclude that there exists a point Z such that

$$\overrightarrow{AZ} \cap (\mathcal{G} \setminus (\overline{RP} \cup \overline{PQ})) = \emptyset.$$

But $\overrightarrow{AZ} \cap (\overline{RP} \cup \overline{PQ}) = \emptyset$, since $\overrightarrow{AX}$ contains no corner of $\mathcal{G}$, and the conclusion follows. A similar proof holds if $A = Q$. $\square$

Theorem PLGN.12 *Let $\mathcal{E}$ be an edge of a simple polygon $\mathcal{G}$; let A, B, and C be noncollinear points of the plane, and suppose that no corner of $\mathcal{G}$ belongs to* $\text{ins}\,\triangle ABC$ *or to $\overline{AC}$. If $\mathcal{E}$ intersects $\overline{AC}$ in a singleton, then $\mathcal{E}$ intersects $\overline{AB}$ or $\overline{BC}$, but not both, and this intersection is a singleton.*

Proof. Suppose that $R \in \mathcal{E} \cap \overline{AC}$ and $\mathcal{L}$ is the line containing $\mathcal{E}$. Then $\mathcal{L} \cap \overline{AC} = \{R\}$. By Theorem PSH.6 and the Postulate of Pasch, either (a) $\overline{BC} \cap \mathcal{L} \neq \emptyset$, (b) $\overline{AB} \cap \mathcal{L} \neq \emptyset$, or (c) $B \in \mathcal{L}$.

In case (a) there is some point T such that $\mathcal{L} \cap \overline{BC} = \{T\}$. If $\mathcal{E} \cap \overline{BC} = \emptyset$, then there exists a corner S of $\mathcal{G}$ with R–S–T. Since $R \in \overrightarrow{AC}$ and $T \in \text{ins}\,\angle BAC$, it follows that $S \in \text{ins}\,\angle BAC$. Also, since $T \in \overline{BC}$, $S \in \text{ins}\,\angle ACB$, so that $S \in \text{ins}\,\triangle ABC$, a contradiction to the assumption that $\text{ins}\,\triangle ABC$ contains no corner of $\mathcal{G}$. Therefore $\mathcal{E}$ intersects $\overline{BC}$ at T; by *Specht* Ch.1 Exercise I.1, the intersection is a singleton since $\mathcal{L} \neq \overline{BC}$.

In cases (b) and (c) similar arguments show that $\mathcal{E}$ intersects $\overline{AB}$ or $\{B\}$ respectively, and the intersection is a singleton. $\square$

Corollary PLGN.12.1 *Let $\mathcal{E}$ be an edge of a simple polygon $\mathcal{G}$; let A, B, and C be noncollinear points of the plane, and suppose that no corner of $\mathcal{G}$ belongs to* $\text{ins}\,\triangle ABC$ *or to $\overline{AC}$. If $\overline{AC}$ intersects $\mathcal{E}$ but is not a subset of $\mathcal{E}$, then $\mathcal{E}$ intersects $\overline{AB}$ or $\overline{BC}$, but not both, and this intersection is a singleton.*

Proof. Let R be a point of intersection of $\mathcal{E}$ and $\overline{AC}$. If there are two points of intersection then by Exercise I.1 the line containing $\mathcal{E}$ is the same as $\overleftrightarrow{AC}$, so that $\mathcal{E} \subseteq \overleftrightarrow{AC}$. By *Specht* Ch.4 Definition IB.3 A–R–C.

By assumption, $\overline{AC} \not\subseteq \mathcal{E}$. Thus there exists a point $X \in \overline{AC}$ such that $X \notin \mathcal{E}$. Let E_1 and E_2 be the endpoints of $\mathcal{E}$ (which is a closed segment).

Neither E_1 or E_2 is a member of $\overline{AC}$, for by hypothesis no corner of $\mathcal{G}$ can belong to this set. Since R is a member of $\overline{AC}$, neither $R = E_1$ nor $R = E_2$. Hence E_1–R–E_2. Since $X \notin \mathcal{E}$, either X–E_1–R–E_2 or E_1–R–E_2–X. In the first case, since both X and R are members of $\overline{AC}$, $E_1 \in \overline{AC}$ which contradicts our assumption that $\overline{AC}$ contains no corner of $\mathcal{G}$. In the second case, $E_2 \in \overline{AC}$ which also yields a contradiction. It follows that there is only one point of intersection of $\overline{AC}$ and $\mathcal{E}$, so that Theorem PLGN.12 applies, and the result follows. $\square$

The following theorem allows the comparison of the number of intersections (with a simple polygon) of two specific admissible rays from two points, based on the intersections of the segment connecting them. It gives meaning to the idea of "moving a ray over a little" in order to get an admissible ray, and provides a basic tool needed to define the parity of a ray, and hence of a point.

Theorem PLGN.13 (Fudge theorem) *Let $\mathcal{G}$ be a simple polygon, and let $A \notin \mathcal{G}$ and B be distinct points of the plane, and let P be a point not in $\overleftrightarrow{AB}$. Then by Remark PLGN.9(C), there exists a point $Q \in$* the P-side of $\overleftrightarrow{AB}$ *such that $\overrightarrow{BQ}$ contains no corner of $\mathcal{G}$.*

Whenever this is true, there exists a point $D \in \overrightarrow{BQ}$ such that the following are all true:

(A) $\overline{BD} \cap \mathcal{G} = \emptyset$,

(B) *the ray $\overrightarrow{AD}$, and indeed, every ray $\overrightarrow{AX}$ with $X \in \overline{BD}$, is admissible, and no corner of $\mathcal{G}$ belongs to* $\operatorname{ins}\angle DAB \cup \overrightarrow{AD}$, *or to* $\operatorname{ins}\triangle DAB \cup \overline{AD}$, *its subset;*

(C) *every edge $\mathcal{E}$ of $\mathcal{G}$ that intersects $\overline{AD}$ does so in a singleton, and also intersects $\overline{AB}$.*

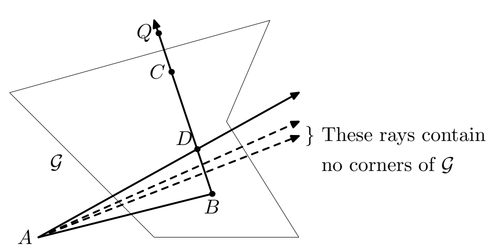


Fig. 6.3 For Theorem PLGN.13.

Proof. See Figure 6.3. By Lemma PLGN.11(A), there exists a point $Z \in \overrightarrow{BQ}$ such that $\overline{BZ} \cap \mathcal{G} = \emptyset$. The set of all corners of $\mathcal{G}$ on the P-side of $\overleftrightarrow{AB}$ is finite (again by Remark PLGN.9(C)). Therefore the set $\mathcal{C} = \{Y \in \overrightarrow{BQ} \mid \overrightarrow{AY}$ contains a corner of $\mathcal{G}\}$ is a finite set and may be ordered by Definition ORD.1 with $B < Q$. By Theorem ORD.10 we may choose C as the least point belonging to $\mathcal{C}$; choose $D < \min\{C, Z\}$. Then $\overline{BD} \subseteq \overline{BZ}$, which we have already seen to be disjoint from $\mathcal{G}$, proving (A).

By our choice of D, neither $\overrightarrow{AD}$ nor any ray $\overrightarrow{AY}$ where $Y \in \overline{BD}$ can contain a corner of $\mathcal{G}$. For any point $W \in \operatorname{ins}\angle DAB$, by Theorem PSH.29 (Crossbar), $\overrightarrow{AW} \cap \overline{BD} \neq \emptyset$, so $\overrightarrow{AW}$ cannot contain a corner of $\mathcal{G}$. Therefore, $\operatorname{ins}\angle DAB \cup \overrightarrow{AD}$ cannot contain a corner of $\mathcal{G}$, proving conclusion (B).

If some edge $\mathcal{E}$ of $\mathcal{G}$ were to intersect $\overline{AD}$ in more than one point, it would be collinear with A and D, and $\overrightarrow{AD}$ would contain at least one endpoint of $\mathcal{E}$ (a corner of $\mathcal{G}$) which is impossible by definition of D; thus all the hypotheses of Theorem PLGN.12 are satisfied, and $\mathcal{E}$ intersects $\overline{AB}$, proving (C). $\square$

Theorem PLGN.14 *Let $\overline{UV}$ be a segment in the plane, and let A, B, and C be noncollinear points of the plane where $A \notin \overline{UV}$ and neither U nor V belongs to either ray of $\angle BAC$. Let $<$ be an ordering of $\overline{UV}$ defined by ORD.1 with $U < V$. Then exactly one of the following holds:*

(1) *$\overline{UV}$ has empty intersection with $\angle BAC$, in which case both U and V belong to $\operatorname{ins}\angle BAC$ or both belong to $\operatorname{out}\angle BAC$;*

(2) *$\overline{UV}$ has a single point P of intersection with $\angle BAC$, in which case*

$$\overline{UP} \subseteq \operatorname{ins}\angle BAC \text{ and } \overline{PV} \subseteq \operatorname{out}\angle BAC, \text{ or}$$
$$\overline{UP} \subseteq \operatorname{out}\angle BAC \text{ and } \overline{PV} \subseteq \operatorname{ins}\angle BAC; \text{ or}$$

(3) *$\overline{UV}$ has exactly two points P and Q of intersection with $\angle BAC$, in which case both U and V belong to $\operatorname{out}\angle BAC$ and $\overline{PQ} \subseteq \operatorname{ins}\angle BAC$. In this case if R is some point of $\overline{PQ}$, $\overline{PR}$ and $\overline{RQ}$ are both subsets of $\operatorname{ins}\angle BAC$, while $\overline{UP}$ and $\overline{QV}$ are both subsets of $\operatorname{out}\angle BAC$.*

Proof. First note that $\overline{UV}$ cannot intersect the angle $\angle BAC$ in more than two points, because then it would intersect one of the rays in two points and that ray would contain either U or V. This shows that these alternatives are the only possibilities.

(1) If U and V are on "opposite" sides of $\angle BAC$ (that is, one is inside and the other outside), Theorem PSH.44 says that $\overline{UV}$ must intersect the angle; thus, if there is no intersection, the two end points must both be inside or both be outside.

(2) From Theorem PSH.44(B), U and V have to be on "opposite sides" of $\angle BAC$; if $U \in \operatorname{ins}\angle BAC$ and $X \in \overline{UP}$, by part (1) X also belongs to $\operatorname{ins}\angle BAC$ so $\overline{UP} \subseteq \operatorname{ins}\angle BAC$; the remaining assertions follow by similar arguments.

(3) See Figure 6.4. If there are exactly two points of intersection P and Q, they must lie on different rays of the angle (otherwise there would be more than two points of intersection). From Theorem PSH.37 the segment $\overline{PQ} \subseteq \operatorname{ins}\angle BAC$; let R be a point with P–R–Q; then apply part (2) to the segments $\overline{UR}$ and $\overline{RV}$ separately. $\square$

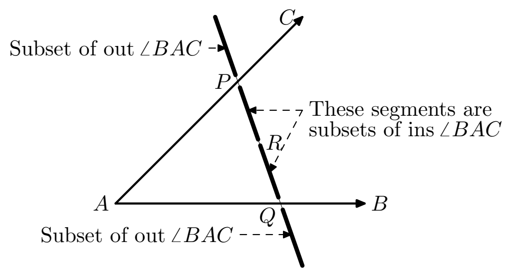


Fig. 6.4 For Theorem PLGN.17 Alternative (3).

6.3 Separation (SEP)

In this section we begin to show that a simple polygon separates the plane into two regions. To signal this development we change our acronym to SEP.

Theorem SEP.1 *Let $\mathcal{G} = \bigcup_{j=1}^{n} \overline{X_jX_{j+1}}$ be a simple polygon on a Pasch plane; let $A \notin \mathcal{G}$ and let $\angle BAC$ be an admissible angle for $\mathcal{G}$. Then the number of intersections of $\angle BAC$ and $\mathcal{G}$ is even.*

Proof. We start by constructing another set of segments $\overline{Y_iY_{i+1}}$ whose union is $\mathcal{G}$.

For each segment $\overline{X_jX_{j+1}}$ which has two points P and Q of intersection with $\angle BAC$ (as in case (3)), we may choose the notation so that X_j–P–Q–X_{j+1}. By Theorem PLGN.14 there exists a point R_j such that P–R_j–Q, so that X_j–P–R_j–Q–X_{j+1}. Then $\overline{X_jR_j} \cap \angle BAC = \{P\}$ and $\overline{R_jX_{j+1}} \cap \angle BAC = \{Q\}$.

Let $\mathcal{E}$ be the union of the set of points R_j so defined and the set of corners X_j; $\mathcal{E}$ is a finite subset of $\mathcal{G}$. We rename the points of $\mathcal{E}$ by an inductive process as follows:

(1) Let $Y_1 = X_1$; if the segment $\overline{Y_1Y_2}$ does not intersect $\angle BAC$, or if it intersects $\angle BAC$ in exactly one point, define $Y_2 = X_2$. If $\overline{X_1X_2} \cap \angle BAC$ contains two points P and Q, define $Y_2 = R_1$ (as defined above), and $Y_3 = X_2$.

(2) Suppose that for some $k \leq n$, X_k has been renamed as Y_i. If $\overline{X_kX_{k+1}}$ is disjoint from $\angle BAC$, or if it intersects $\angle BAC$ in exactly one point, define $Y_{i+1} = X_{k+1}$. If $\overline{X_kX_{k+1}} \cap \angle BAC$ contains two points P and Q, define $Y_{i+1} = R_k$ and $Y_{i+2} = X_{k+1}$; then both the resulting segments intersect $\angle BAC$ in exactly one point.

Continue this renaming process until for some $m > 1$, $X_{n+1} = X_1$ has been renamed Y_{m+1}. Since $Y_1 = X_1$, $Y_{m+1} = Y_1$. Then $\mathcal{G} = \bigcup_{i=1}^{m} \overline{Y_iY_{i+1}}$, and each interval $\overline{Y_iY_{i+1}}$ intersects $\angle BAC$ in at most one point. Also, for each $i \in [1; m]$ define $Y_{i+m} = Y_i$, so that each point Y_i has two labels. The reason for making this second label will become apparent shortly.

We say a segment $\overline{Y_iY_{i+1}}$ is a **passing** segment iff it does not intersect $\angle BAC$. By Theorem PLGN.14(1), both end points of a passing segment are on the same side of $\angle BAC$—either both are in $\text{ins}\,\angle BAC$ or both are in $\text{out}\,\angle BAC$.

If a segment $\overline{Y_iY_{i+1}}$ is not a passing segment, it intersects $\angle BAC$ in exactly one point. We shall call such a segment a **crossing** segment. If $\overline{Y_iY_{i+1}}$ is a crossing segment, by Theorem PLGN.14(2) either $Y_i \in \text{out}\,\angle BAC$ and $Y_{i+1} \in \text{ins}\,\angle BAC$, in which case we will call it an **entering** segment, or $Y_i \in \text{ins}\,\angle BAC$ and $Y_{i+1} \in \text{out}\,\angle BAC$, in which case we call it an **exiting** segment. Thus, every segment $\overline{Y_iY_{i+1}}$ is either a *passing* or a *crossing* segment, and *crossing* segments come in two flavors, *entering* or *exiting*.

Now let $\overline{Y_iY_{i+1}}$ (where $i \in [1; m]$) be a crossing segment. If there is no other crossing segment $\overline{Y_kY_{k+1}}$, all other segments are passing, and both Y_i and Y_{i+1} are on the same side of $\angle BAC$, a contradiction to Theorem PLGN.14(2). Therefore there exists at least one crossing segment $\overline{Y_kY_{k+1}}$ where $k \leq m$ such that $k \neq i$. Either $i < k$ or $k < i$; in the latter case, $k + m > i$; in either case there exists an integer $j \in [1; 2m]$ such that $i < j$ and $\overline{Y_jY_{j+1}}$ is a crossing segment. Let j be the smallest integer such that $i < j \leq 2m$ and $\overline{Y_jY_{j+1}}$ is crossing.

Informally, one may think of $\overline{Y_jY_{j+1}}$ as the first crossing segment encountered in a traversal from $\overline{Y_iY_{i+1}}$ in the direction of increasing indices.

If $\overline{Y_iY_{i+1}}$ is exiting, $Y_{i+1} \in \operatorname{out}\angle BAC$; if there are passing segments $\overline{Y_kY_{k+1}}$ with $i < k < j$, both Y_k and Y_{k+1} are in $\operatorname{out}\angle BAC$, so that $Y_j \in \operatorname{out}\angle BAC$ and $\overline{Y_jY_{j+1}}$ must be an entering segment. By similar reasoning, if $\overline{Y_iY_{i+1}}$ is entering, $Y_{i+1} \in \operatorname{ins}\angle BAC$, so that $Y_j \in \operatorname{ins}\angle BAC$, and $\overline{Y_jY_{j+1}}$ must be exiting.

Thus, as the polygon is traversed in the direction of increasing indices of Y_j, each crossing segment is entering (exiting) iff both the next prior and next succeeding crossing segment are exiting (entering). Therefore the number of exiting segments is the same as the number of entering segments, and the total number of intersections of $\mathcal{G}$ with $\angle BAC$ is even. $\square$

We have "buried" the definitions of *passing*, *crossing*, *entering*, and *exiting* segments in the above proof because these notions are not used elsewhere in the development.

Definition SEP.2 Let $\mathcal{G}$ be a simple polygon and A a point not on $\mathcal{G}$.

(A) Let $\overrightarrow{AB}$ be an admissible ray for $\mathcal{G}$; if $\overrightarrow{AB} \cap \mathcal{G}$ contains an even number of members, the ray has **even parity** (relative to $\mathcal{G}$). If it has an odd number of elements, the ray has **odd parity**. A ray that does not intersect $\mathcal{G}$ has **even parity**, because 0 is an even number.

(B) A point A has **odd (even) parity** with respect to $\mathcal{G}$ iff all admissible rays with endpoint A have odd (even) parity with respect to $\mathcal{G}$.

(C) The **inside of** $\mathcal{G}$ (notation: $\operatorname{ins}\mathcal{G}$) is the set of points not on $\mathcal{G}$ which have odd parity with respect to $\mathcal{G}$. The **outside of** $\mathcal{G}$ (notation: $\operatorname{out}\mathcal{G}$) is the set of points not on $\mathcal{G}$ which have even parity with respect to $\mathcal{G}$.

(D) The **enclosure of** $\mathcal{G}$ (notation: $\operatorname{enc}\mathcal{G}$) is the union of $\mathcal{G}$ and $\operatorname{ins}\mathcal{G}$, or $\mathcal{G} \cup \operatorname{ins}\mathcal{G}$.

(E) The **exclosure of** $\mathcal{G}$ (notation: $\operatorname{exc}\mathcal{G}$) is the union of $\mathcal{G}$ and $\operatorname{out}\mathcal{G}$, or $\mathcal{G} \cup \operatorname{out}\mathcal{G}$.

(F) The inside or the outside of $\mathcal{G}$ is called a **side of** $\mathcal{G}$. If Q is any point not on G, that is, $Q \in \mathcal{P} \setminus \mathcal{G}$, then the side of $\mathcal{G}$ to which Q belongs is called the **Q-side of** $\mathcal{G}$. The inside and the outside of $\mathcal{G}$ are **opposite** sides of $\mathcal{G}$.

The next theorem (which is really a corollary of Theorem SEP.1) shows that every point $A \notin \mathcal{G}$ has a parity according to Definition SEP.2(B) above. It also shows that to determine the parity of a point, it is only necessary to determine the parity of a single ray starting from that point. Such a ray, used to determine parity for A, may be referred to as a **test** ray for A.

Theorem SEP.3 *Let $\mathcal{G}$ be a simple polygon on a Pasch plane, and let $A \notin \mathcal{G}$; then the parity of any admissible ray $\overrightarrow{AB}$ is the same as the parity of any other admissible ray $\overrightarrow{AC}$.*

Proof. If the number of intersections of one ray of $\angle BAC$ with $\mathcal{G}$ is even while the number of intersections of the other ray with $\mathcal{G}$ is odd, the total number of intersections with $\angle BAC$ would be odd, contradicting Theorem SEP.1. Therefore the number of intersections of $\overrightarrow{AB}$ with $\mathcal{G}$ is even (or odd) iff the number of intersections of $\overrightarrow{AC}$ with $\mathcal{G}$ is even (or odd). $\square$

The following theorem will be invoked freely without reference in the proofs that follow.

Theorem SEP.4 *Let $\mathcal{G}$ be a simple polygon and let A and B be points not on $\mathcal{G}$.*

(A) *If the segment $\overline{AB}$ contains no points of $\mathcal{G}$ then the parity of A equals the parity of B.*

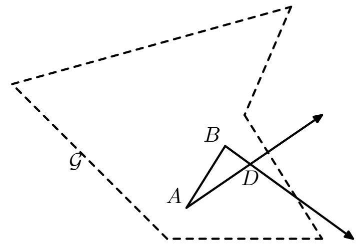
(B) *If $\overline{AB} \cap \mathcal{G} = \{R\}$ is a singleton, and R is not a corner of $\mathcal{G}$, then the parities of A and B are different.*

(C) *If $\overline{AB} \cap \mathcal{G} = \{R\}$ is a singleton, and $R = X_i$ is a corner of $\mathcal{G}$, then A and B have the same parity if and only if $\overline{AB} \cap \operatorname{ins} \angle X_{i-1}X_iX_{i+1} = \emptyset$.*

(D) *Suppose $\overline{AB} \cap \mathcal{G} = \{R\}$ is a singleton. If $A \in \operatorname{ins} \mathcal{G}\,(\operatorname{out} \mathcal{G})$ then $\overline{AR} \subseteq \operatorname{ins} \mathcal{G}\,(\operatorname{out} \mathcal{G})$ and similarly for B.*

Proof. (A) (Case I) If $\overrightarrow{AB}$ is admissible, the number of intersections it has with $\mathcal{G}$ is the same as the number of intersections of the ray $\overrightarrow{AB} \setminus \overline{AB}$ so the parities of A and B are the same. A similar proof holds if $\overrightarrow{BA}$ is admissible.

(Case II) See Figure 6.5 for a visualization. If neither $\overrightarrow{AB}$ nor $\overrightarrow{BA}$ is admissible, let P be a point not on $\overleftrightarrow{AB}$ such that $\overrightarrow{BP}$ is admissible. Then by Theorem PLGN.13 there exists a point $D \in \overrightarrow{BP}$ such that $\overrightarrow{AD}$ is admissible, $\overline{BD} \cap \mathcal{G} = \emptyset$, and every edge $\mathcal{E}$ of $\mathcal{G}$ that intersects $\overline{AD}$ must also intersect $\overline{AB}$; but there are no edges intersecting $\overline{AB}$, so there are no edges of $\mathcal{G}$ that intersect $\overline{AD}$, and it follows that $\overline{AD} \cap \mathcal{G} = \emptyset$. By Case I, A and D have the same parity, and D and B have the same parity.

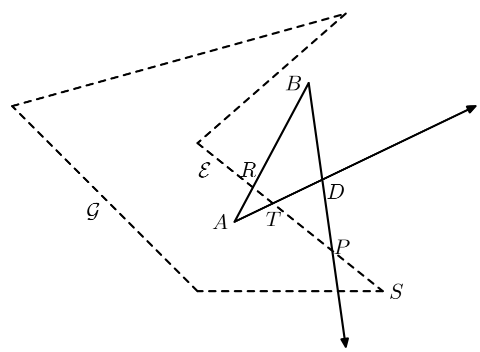


If neither $\overrightarrow{AB}$ or $\overrightarrow{BA}$ is admissible;
construct admissible rays $\overrightarrow{AD}$ and $\overrightarrow{BD}$ so that
neither $\overline{AD}$ or $\overline{BD}$ contain points of $\mathcal{G}$
Then by Case I, A, B, and D have the same parity.

Fig. 6.5 For Theorem SEP.4 (A) Case II.

(B) Let $\mathcal{E}$ be the edge of $\mathcal{G}$ containing R. If the line containing $\mathcal{E}$ were the same as $\overleftrightarrow{AB}$, then since A and B do not belong to $\mathcal{G}$, $\overline{AB}$ would contain an endpoint of $\mathcal{E}$, and the intersection would not be a singleton. Thus $\mathcal{E} \not\subseteq \overleftrightarrow{AB}$.

(Case I) If $\overrightarrow{AB}$ is admissible then it has one more intersection with $\mathcal{G}$ than does $\overrightarrow{AB} \setminus \overline{AB}$, so that the parities of A and B are different. A similar proof holds if $\overrightarrow{BA}$ is admissible.



If neither $\overrightarrow{AB}$ or $\overrightarrow{BA}$ is admissible, construct admissible rays
$\overrightarrow{AD}$ and $\overrightarrow{BD} = \overrightarrow{BP}$ so that $\overline{AD}$ intersects only the
edge $\mathcal{E}$ and $\overline{BD}$ contains no points of $\mathcal{G}$. Then A and D
have different parities, while B and D have the same parity.

Fig. 6.6 For Theorem SEP.4(B) Case II.

(Case II) See Figure 6.6. If neither $\overrightarrow{AB}$ nor $\overrightarrow{BA}$ is admissible, pick S to be an endpoint of the edge $\mathcal{E}$ containing R. Note that $S \notin \overleftrightarrow{AB}$.

Let $P \in \overline{RS}$ be such that $\overrightarrow{BP}$ is admissible. Then by Theorem PLGN.13 choose $D \in \overline{BP}$ so that $\overline{BD} \cap \mathcal{G} = \emptyset$, the ray $\overrightarrow{AD}$ is admissible, and every edge $\mathcal{E}'$ that intersects $\overline{AD}$ also intersects $\overline{AB}$. Since there are no edges other than $\mathcal{E}$ which intersect $\overline{AB}$, no edge other than $\mathcal{E}$ can intersect $\overline{AD}$.

Claim: $\mathcal{E}$ intersects $\overline{AD}$. Note that by Theorem PSH.37, $D \in \operatorname{ins} \angle BAP$, so that by the Crossbar Theorem PSH.39, $\overrightarrow{AD}$ intersects $\overline{RP}$ at some point $T \in \mathcal{E}$. Now $R \in A$-side of $\overleftrightarrow{BP}$, and hence by Theorem IB.14, $\overrightarrow{PR} \subseteq A$-side of $\overleftrightarrow{BP}$, so T is also on the A-side of $\overleftrightarrow{BP}$ and $T \in \overline{AD}$. This proves the claim.

Therefore exactly one edge, namely $\mathcal{E}$, intersects $\overline{AD}$, and since $\overrightarrow{AD}$ is admissible, by case I the parities of A and D are different. Since $\overline{BD} \cap \mathcal{G} = \emptyset$, by part (A) the parities of B and D are the same; hence the parities of A and B are different.

(C) For this case we leave the construction of figures to the reader. Suppose that there is a point $Q \in \overline{AB} \cap \operatorname{ins} \angle X_{i-1}X_iX_{i+1}$. Then $\overrightarrow{X_iQ}$ is a subset of the X_{i-1}-side of $\overleftrightarrow{X_iX_{i+1}}$ and of the X_{i+1}-side of $\overleftrightarrow{X_iX_{i-1}}$ and hence $\overrightarrow{X_iQ} \subseteq \operatorname{ins} \angle X_{i-1}X_iX_{i+1}$. Either A or B belongs to $\overrightarrow{X_iQ}$ and without loss of generality we may assume $A \in \operatorname{ins} \angle X_{i-1}X_iX_{i+1}$. Then A, X_i and X_{i+1} are not collinear and $X_{i+1} \notin \overleftrightarrow{AB}$.

Choose $P \in \overline{X_iX_{i+1}}$ so that $\overrightarrow{BP}$ is admissible. By Theorem PLGN.13 there exists a point $D \in \overrightarrow{BP}$ such that $\overline{BD} \cap \mathcal{G} = \emptyset$, $\overrightarrow{AD}$ is admissible, and every edge that intersects $\overline{AD}$ also must intersect $\overline{AB}$.

By Theorem PSH.37, $D \in \operatorname{ins} \angle BAP$, so that by the Crossbar Theorem PSH.39, $\overrightarrow{AD}$ intersects $\overline{X_iP}$ at some point T, and X_i–T–P–X_{i+1}, so that T is not a corner of $\mathcal{G}$. The only edges of $\mathcal{G}$ that intersect $\overline{AB}$ are $\overline{X_{i-1}X_i}$ and $\overline{X_iX_{i+1}}$; $\overline{X_{i-1}X_i}$ is on the side of $\overleftrightarrow{AB}$ opposite to D; therefore $\overline{X_iX_{i+1}}$ is the only edge intersecting $\overline{AD}$.

Since $\overrightarrow{AD}$ is admissible, by part (B) A and D have different parities. By part (A) B and D have the same parity since $\overline{BD} \cap \mathcal{G} = \emptyset$. Thus A and B have different parities. By the contrapositive, if A and B have the same parity, $\overline{AB} \cap \operatorname{ins} \angle X_{i-1}X_iX_{i+1} = \emptyset$.

Conversely, assume that $\overline{AB} \cap \operatorname{ins} \angle X_{i-1}X_iX_{i+1} = \emptyset$. If X_{i-1} and X_{i+1} are on opposite sides of $\overleftrightarrow{AB}$, then $\overline{X_{i-1}X_{i+1}}$ intersects $\overleftrightarrow{AB}$ at a single point C by Axiom PSA, and $C \in \operatorname{ins} \angle X_{i-1}X_iX_{i+1}$ by Theorem PSH.37. This contradicts our initial assumption, and X_{i-1} and X_{i+1} must be on the same side of $\overleftrightarrow{AB}$.

Let $\mathcal{H}$ be the side of $\overleftrightarrow{AB}$ opposite X_{i-1} and X_{i+1}, and let $P \in \mathcal{H}$ be such that $\overrightarrow{BP}$ is admissible. By Theorem PLGN.13 there exists a point $D \in \overrightarrow{BP}$ such that $\overline{BD} \cap \mathcal{G} = \emptyset$ and every edge of $\mathcal{G}$ intersecting $\overline{AD}$ also intersects $\overline{AB}$.

By part (A) the parities of B and D are the same. There are two edges of $\mathcal{G}$ intersecting $\overline{(AB]}$ at X_i, but both of these are disjoint from $\mathcal{H}$. Hence no edge of $\mathcal{G}$ intersects $\overline{(AD)}$ and by part (A) the parities of A and D are the same, so the parities of A and B are the same.

(D) $\overline{[AR)}$ and $\overline{[BR)}$ contain no points of $\mathcal{G}$, so the result follows immediately from part (A). □

Corollary SEP.4.1 *Let $\mathcal{G}$ be a simple polygon and let A and B be distinct points not on $\mathcal{G}$. Then if the parities of A and B are different, $\overline{[AB]} \cap \mathcal{G} \neq \emptyset$.*

Proof. Contrapositive of Theorem SEP.4(A). □

Corollary SEP.4.2 *Let $\mathcal{G}$ be a simple polygon and let A and B be distinct points not on $\mathcal{G}$. Then if the parities of A and B are the same, $\overline{[AB]} \cap \mathcal{G}$ is either a singleton which is a corner of $\mathcal{G}$ or is not a singleton. (The intersection could be empty, or it could contain two points, or a whole segment.)*

Proof. Contrapositive of Theorem SEP.4(B). □

Remark SEP.5 Let $\mathcal{G}$ be a simple polygon or a polygonal path.

(A) Let A and B be distinct points not on $\mathcal{G}$. Then if $\overline{[AB]} \cap \mathcal{G} = \{P\}$ is a singleton which is a corner of $\mathcal{G}$, it is possible for $\overline{[AP)}$ and $\overline{[BP)}$ to both be subsets of the inside, or both subsets of the outside of $\mathcal{G}$.

(B) Let A, B, and C be any points with A–C–B. Then $\overline{[AB]} \cap \mathcal{G} = \{C\}$ if and only if $C \in \mathcal{G}$ and $\overline{[AC)} \cap \mathcal{G} = \emptyset$ and $\overline{[BC)} \cap \mathcal{G} = \emptyset$.

Theorem SEP.6 *Let C and D be distinct points, with $C \in \overline{[X_iX_{i+1}]}$, an edge of a simple polygon $\mathcal{G}$. If either*

(1) C is not a corner of $\mathcal{G}$ and $D \notin \overleftrightarrow{X_iX_{i+1}}$, or

(2) $C = X_i$, a corner of $\mathcal{G}$, and $D \notin \angle X_{i-1}X_iX_{i+1}$,

then there exists a point $E \in \overline{[CD]}$ such that $E \neq C$ and $\overline{[EC]} \cap \mathcal{G} = \{C\}$.

Proof. (1) If $\overline{(CD]} \cap \overleftrightarrow{X_iX_{i+1}} = \emptyset$, let $E = D$. Otherwise apply Theorem PLGN.4 to the ray $\overrightarrow{[CD}$ with $C < D$; let F be the first intersection of $\overrightarrow{[CD}$ and $\mathcal{G} \setminus \overline{(X_iX_{i+1})}$, and let $E < \min\{F, D\}$. Then $\overline{[EC]} \cap \mathcal{G} = \{C\}$.

(2) The proof is the same except that we choose F to be the first intersection of $\overrightarrow{[CD}$ and $\mathcal{G} \setminus (\overline{(X_{i-1}X_i]} \cup \overline{[X_iX_{i+1})})$. □

Theorem SEP.7 *Let C be any point on the edge $\overline{X_iX_{i+1}}$ of a simple polygon $\mathcal{G}$, and let $\mathcal{L}$ be a line through C. If either*

(1) C is not a corner of $\mathcal{G}$ and $\mathcal{L}$ is different from $\overleftrightarrow{X_iX_{i+1}}$, or

(2) $C = X_i$, a corner of $\mathcal{G}$ (endpoint of $\overline{X_iX_{i+1}}$), and $\mathcal{L}$ contains at least one point of $\operatorname{ins}\angle X_{i-1}X_iX_{i+1}$;

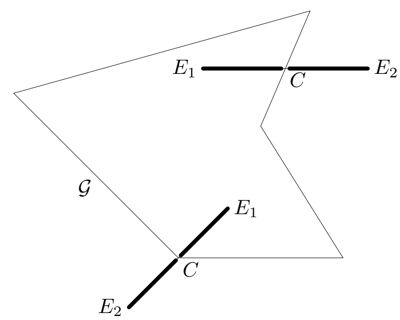
then there exist points E_1 and $E_2 \in \mathcal{L}$ such that

(a) E_1–C–E_2,

(b) $\overline{E_1E_2} \cap \mathcal{G} = \{C\}$,

(c) $\overline{E_1C} \subseteq \operatorname{ins}\mathcal{G}$, *and*

(d) $\overline{E_2C} \subseteq \operatorname{out}\mathcal{G}$.



Every point C of a polygon $\mathcal{G}$ is the end point of two segments $\overline{E_1C}$ and $\overline{E_2C}$ lying respectively in $\operatorname{ins}\mathcal{G}$ and $\operatorname{out}\mathcal{G}$

Fig. 6.7 For Theorem SEP.7.

Proof. See Figure 6.7. (1) If $\overline{X_iX_{i+1}} \cap \mathcal{L}$ contains more than one point, then $\overleftrightarrow{X_iX_{i+1}} = \mathcal{L}$ by Axiom I.1; therefore $\overline{X_iX_{i+1}} \cap \mathcal{L} = \{C\}$, a singleton. Let P and Q be points of $\mathcal{L}$ such that P–C–Q. Then since P and Q are not in $\overleftrightarrow{X_iX_{i+1}}$ by Theorem SEP.6(1) there exist points E_1 and E_2 such that C–E_1–P, C–E_2–Q, $\overline{E_1C} \cap \mathcal{G} = \{C\}$ and $\overline{E_2C} \cap \mathcal{G} = \{C\}$. This shows conclusions (a) and (b).

Theorem SEP.4(A) says that every point in $\overline{CE_1}$ has the same parity, and likewise for $\overline{CE_2}$. By Theorem SEP.4(B) the parity of E_1 is different from the parity of E_2, so that with appropriate re-labeling, $E_1 \in \operatorname{ins}\mathcal{G}$ and conclusions (c) and (d) follow.

(2) The proof is similar to that for part (1). Let P be a point on $\mathcal{L}$ belonging to $\operatorname{ins}\angle X_{i-1}X_iX_{i+1}$ and let Q–C–P. Then Q is on the side of $\overleftrightarrow{X_iX_{i+1}}$ opposite X_{i-1} and $Q \in \operatorname{out}\angle X_{i-1}X_iX_{i+1}$ by Theorem PSH.41(C). Since neither P nor Q is on $\angle X_{i-1}X_iX_{i+1}$ we may apply Theorem SEP.6(2) to show conclusions (a) and (b).

As before, every point in $\overline{E_1C}$ has the same parity, and likewise for $\overline{E_2C}$. By SEP.4(C) the parity of E_1 is different from the parity of E_2, so that with appropriate re-labeling, $E_1 \in \text{ins}\,\mathcal{G}$ and conclusions (c) and (d) follow. $\square$

Corollary SEP.7.1 *For any point C of a simple polygon $\mathcal{G}$, points E_1 and E_2 can be chosen so that conclusions (a) through (d) of Theorem SEP.7 are satisfied and the line $\overleftrightarrow{E_1E_2}$ contains no corner of the polygon $\mathcal{G}$, other than the point C in the case that C is a corner of $\mathcal{G}$.*

Proof. Suppose $C \in \overline{X_iX_{i+1}}$, an edge of $\mathcal{G}$. Infinitely many lines $\overleftrightarrow{CP}$ may be generated where $P \in \text{ins}\,\angle X_{i-1}X_iX_{i+1}$, and since there are only finitely many corners, we may choose P so that $\overleftrightarrow{CP}$ contains no corner other than possibly C. Then let $\mathcal{L} = \overleftrightarrow{CP}$, and apply Theorem SEP.7. $\square$

Corollary SEP.7.2 *For any simple polygon $\mathcal{G}$, both* $\text{ins}\,\mathcal{G} \neq \emptyset$ *and* $\text{out}\,\mathcal{G} \neq \emptyset$.

Theorem SEP.8 *Suppose S is a nonempty subset of the Pasch plane $\mathcal{P}$, and $\mathcal{T}_1$ and $\mathcal{T}_2$ are subsets of $\mathcal{P}$ such that $\mathcal{P} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{S}$ and $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{S}$ are pairwise disjoint. Then the following are equivalent statements:*

(a) *For every $M_1 \in \mathcal{T}_1$ and every $M_2 \in \mathcal{T}_2$, $\overline{M_1M_2} \cap \mathcal{S} \neq \emptyset$.*

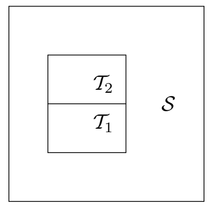
(b) *For every $M_1 \in \mathcal{T}_1$ and every $M_2 \in \mathcal{T}_2$, every simple polygonal path $\langle\langle M_1, X_2, \ldots, X_m, M_2\rangle\rangle$ connecting M_1 and M_2 intersects $\mathcal{S}$.*

Proof. Assume (a) is true. Let $N_1 \in \mathcal{T}_1$, $N_2 \in \mathcal{T}_2$, and suppose further that $\langle\langle N_1, X_2, \ldots, X_m, N_2\rangle\rangle$ is any simple polygonal path connecting N_1 and N_2. For convenience rename these corners by letting $Y_1 = N_1$, $Y_i = X_i$ for all $i \in [2; m]$, and $Y_{m+1} = N_2$.

If some corner $Y_2, \ldots, Y_m \in \mathcal{S}$ then the path intersects $\mathcal{S}$. If none of the corners $Y_2, \ldots, Y_m \in \mathcal{S}$ then all the corners belong either to $\mathcal{T}_1$ or $\mathcal{T}_2$. Since $N_1 \in \mathcal{T}_1$ and $N_2 \in \mathcal{T}_2$, the set of all corners belonging to $\mathcal{T}_1$ and the set of all corners belonging to $\mathcal{T}_2$ are both non-empty. Let Y_i be that corner belonging to $\mathcal{T}_2$ having the smallest index. Then $Y_i \neq N_1$, $i \geq 2$, and $Y_{i-1} \in \mathcal{T}_1$. Since $Y_{i-1} \in \mathcal{T}_1$ and $Y_i \in \mathcal{T}_2$, by assumption $\overline{Y_{i-1}Y_i} \cap \mathcal{S} \neq \emptyset$, so that $\langle\langle N_1, X_2, \ldots, X_m, N_2\rangle\rangle$ intersects $\mathcal{S}$.

The converse is obvious since the segment $\overline{M_1M_2}$ is a simple polygonal path joining M_1 and M_2. $\square$

Remark SEP.9 The statement "for any $M_1 \in \mathcal{T}_1$ and any $M_2 \in \mathcal{T}_2$, some polygonal path $\langle\langle M_1, X_2, \ldots, X_m, M_2\rangle\rangle$ connecting M_1 and M_2 intersects $\mathcal{S}$" is not equivalent to the equivalent statements (a) and (b) in Theorem SEP.8. A counterexample is shown in Figure 6.8.



For any $M_1 \in \mathcal{T}_1$ and any $M_2 \in \mathcal{T}_2$ there is a polygonal path connecting M_1 and M_2 that intersects $\mathcal{S}$, but not every such path does so.

Fig. 6.8 For Remark SEP.9.

Definition SEP.10 A nonempty subset $\mathcal{S}$ of the Pasch plane $\mathcal{P}$ **separates $\mathcal{P}$ into two subsets $\mathcal{T}_1$ and $\mathcal{T}_2$** if and only if

(1) $\mathcal{P} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{S}$,

(2) $\mathcal{T}_1$ and $\mathcal{T}_2$ and $\mathcal{S}$ are pairwise disjoint, and

(3) for all $M_1 \in \mathcal{T}_1$ and $M_2 \in \mathcal{T}_2$, $\overline{M_1M_2} \cap \mathcal{S} \neq \emptyset$.

Informally, we may say that $\mathcal{S}$ **separates the plane into two parts $\mathcal{T}_1$ and $\mathcal{T}_2$**, or more briefly, $\mathcal{S}$ **separates the plane**.

Remark SEP.11 By Theorem SEP.8, $\mathcal{S}$ separates the plane into the two parts $\mathcal{T}_1$ and $\mathcal{T}_2$ iff conditions (1) and (2) (of SEP.8) hold and every simple polygonal path joining two points $M_1 \in \mathcal{T}_1$ and $M_2 \in \mathcal{T}_2$ intersects $\mathcal{S}$.

Theorem SEP.12 (Proof of Theorem JCT.1, parts (A) and (B)) *A simple polygon $\mathcal{G}$ separates the plane $\mathcal{P}$ into two parts,* $\operatorname{ins}\mathcal{G}$ *and* $\operatorname{out}\mathcal{G}$*.*

Proof. By Definition SEP.3(C) $\operatorname{ins}\mathcal{G}$ and $\operatorname{out}\mathcal{G}$ are disjoint sets, both of which are disjoint from $\mathcal{G}$, and $\mathcal{P} = \operatorname{ins}\mathcal{G} \cup \operatorname{out}\mathcal{G} \cup \mathcal{G}$, thus satisfying conditions (1) and (2) of Definition SEP.10. Let $M_1 \in \operatorname{ins}\mathcal{G}$ and $M_2 \in \operatorname{out}\mathcal{G}$. Then M_1 and M_2 have different parities, and hence $\overline{M_1M_2}$ must intersect $\mathcal{G}$, by Corollary SEP.4,1, so that condition (3) follows. □

Theorem SEP.13 (A) *If $\mathcal{E}$ is a polygonally connected set and $\mathcal{G}$ is a simple polygon such that $\mathcal{E} \cap \mathcal{G} = \emptyset$, then either $\mathcal{E} \subseteq \operatorname{ins}\mathcal{G}$ or $\mathcal{E} \subseteq \operatorname{out}\mathcal{G}$. In* *other words, all points of* $\mathcal{E}$ *have the same parity relative to* $\mathcal{G}$. *Notice that this is a generalization of Theorem SEP.4(A).*

(B) *If* $\mathcal{E}$ *is a convex set and* $\mathcal{G}$ *is a simple polygon such that* $\mathcal{E} \cap \mathcal{G} = \emptyset$, *then either* $\mathcal{E} \subseteq \operatorname{ins} \mathcal{G}$ *or* $\mathcal{E} \subseteq \operatorname{out} \mathcal{G}$.

(C) *If* $\mathcal{F}$ *and* $\mathcal{G}$ *are simple polygons and* $\mathcal{F} \cap \mathcal{G} = \emptyset$, *then* $\mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$ *or* $\mathcal{F} \subseteq \operatorname{out} \mathcal{G}$.

(D) *If* $\mathcal{F}$ *and* $\mathcal{G}$ *are simple polygons and* $\mathcal{F} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$ *and* $\mathcal{F} \cap \operatorname{out} \mathcal{G} \neq \emptyset$, *then* $\mathcal{F} \cap \mathcal{G} \neq \emptyset$.

Proof. (A) Let P and Q be two points of $\mathcal{E}$ which are connected by a polygonal path $\langle\langle P = X_1, X_2, \ldots, X_m, X_{m+1} = Q \rangle\rangle \subseteq \mathcal{E}$. Then by Theorem SEP.4(A), $P = X_1$ has the same parity as X_2, and a simple induction argument shows that P has the same parity as Q.

(B) Immediate from part (A) since every convex set is polygonally connected.

(C) Immediate from part (A) and Remark PLGN.6(C).

(D) Since $\mathcal{F}$ is polygonally connected, there is a polygonal path in $\mathcal{F}$ connecting a point of $\operatorname{ins} G$ with a point of $\operatorname{out} G$, and this polygonal path must contain a point of $\mathcal{G}$ by Theorem SEP.12. □

Theorem SEP.14 *If* $\mathcal{F}$ *and* $\mathcal{G}$ *are simple polygons and* $\mathcal{F} \cap \mathcal{G} = \emptyset$, *exactly one of the following holds:*

(1) $\mathcal{G} \subseteq \operatorname{ins} \mathcal{F}$ *and* $\mathcal{F} \subseteq \operatorname{out} \mathcal{G}$, *in which case*
   $\operatorname{exc} \mathcal{F} = \mathcal{F} \cup \operatorname{out} \mathcal{F}$ *is a proper subset of* $\operatorname{out} \mathcal{G}$, *and*
   $\operatorname{enc} \mathcal{G} = \mathcal{G} \cup \operatorname{ins} \mathcal{G}$ *is a proper subset of* $\operatorname{ins} \mathcal{F}$; *or*

(2) $\mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$ *and* $\mathcal{G} \subseteq \operatorname{out} \mathcal{F}$, *in which case*
   $\operatorname{exc} \mathcal{G} = \mathcal{G} \cup \operatorname{out} \mathcal{G}$ *is a proper subset of* $\operatorname{out} \mathcal{F}$, *and*
   $\operatorname{enc} \mathcal{F} = \mathcal{F} \cup \operatorname{ins} \mathcal{F}$ *is a proper subset of* $\operatorname{ins} \mathcal{G}$; *or*

(3) $\mathcal{G} \subseteq \operatorname{out} \mathcal{F}$ *and* $\mathcal{F} \subseteq \operatorname{out} \mathcal{G}$, *in which case*
   $\operatorname{enc} \mathcal{G} = \mathcal{G} \cup \operatorname{ins} \mathcal{G}$ *is a proper subset of* $\operatorname{out} \mathcal{F}$ *and*
   $\operatorname{enc} \mathcal{F} = \mathcal{F} \cup \operatorname{ins} \mathcal{F}$ *is a proper subset of* $\operatorname{out} \mathcal{G}$.

The remaining logical possibility, (4) $\mathcal{G} \subseteq \operatorname{ins} \mathcal{F}$ *and* $\mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$, *is impossible.*

Proof. By Theorem SEP.13(C) each of $\mathcal{F}$ and $\mathcal{G}$ must be a subset of either the inside or the outside of the other. The alternatives (1) through (4) above are the only logical possibilities. Figure 6.9 below illustrates the proof that alternative (4) is impossible.

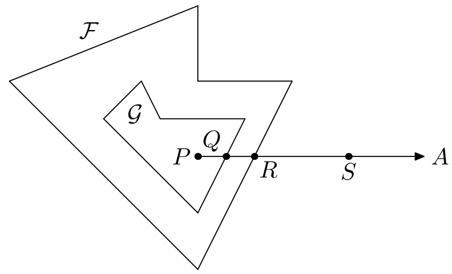
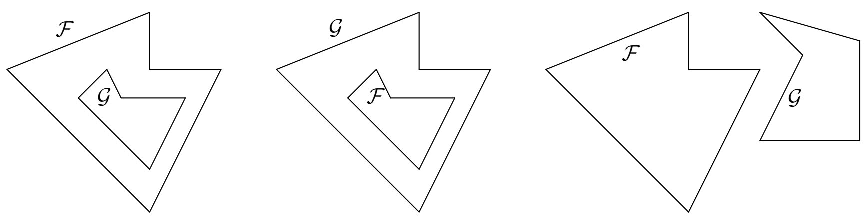


Fig. 6.9 For proof that alternate (4) is impossible.

Proof that (4) is impossible. Suppose that $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{ins}\mathcal{G}$. Let $P \in \operatorname{ins}\mathcal{G}$, and let A be a point such that $\overleftrightarrow{PA}$ contains no corner of $\mathcal{F}$ or $\mathcal{G}$. Order $\overrightarrow{PA}$ by Definition ORD.1 with $P < A$. By Definition SEP.3 $\overrightarrow{PA}$ has a non-empty intersection with $\mathcal{G}$ and by Theorem PLGN.4(A) there is a last intersection Q of $\mathcal{G}$ with $\overrightarrow{PA}$. Since $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$, $Q \in \operatorname{ins}\mathcal{F}$ so that the ray $\overrightarrow{PA} \setminus \overrightarrow{PQ}$ intersects $\mathcal{F}$ and there is a last intersection R of $\mathcal{F}$ with $\overrightarrow{PA} \setminus \overrightarrow{PQ}$, and $Q \neq R$ since $Q \in \operatorname{ins}\mathcal{F}$, and $R \in \mathcal{F}$, whence $Q < R$.

Since $R \in \mathcal{F} \subseteq \operatorname{ins}\mathcal{G}$, the ray $\overrightarrow{PA} \setminus \overrightarrow{PR}$ intersects $\mathcal{G}$ at some point S. $S \neq R$ since $R \in \operatorname{ins}\mathcal{G}$ and $S \in \mathcal{G}$ whence $Q < R < S$. This contradicts the fact that Q is the last intersection of $\overrightarrow{PA}$ with $\mathcal{G}$, so that alternative (4) is impossible.

To visualize the other alternatives, see Figure 6.10 below.



Alternative (1)
$\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$

Alternative (2)
$\mathcal{F} \subseteq \operatorname{ins}\mathcal{G}$ and $\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$

Alternative (3)
$\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$

Fig. 6.10 For Theorem SEP.14, alternatives (1), (2), and (3).

If alternative (1) is true: We assume that $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$.

(a) First we show that $\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$. Let M be any point of $\operatorname{out}\mathcal{F}$ and let H be any other point such that $\overrightarrow{MH}$ contains no corner of $\mathcal{F}$ or $\mathcal{G}$.

Note that $M \notin \mathcal{G}$ since $M \in \operatorname{out}\mathcal{F}$, which is disjoint from $\operatorname{ins}\mathcal{F} \supseteq \mathcal{G}$. Suppose now $\overrightarrow{MH}$ does not intersect $\mathcal{F}$. If $\overrightarrow{MH}$ should intersect $\mathcal{G}$ at M', then the point $M' \in \operatorname{ins}\mathcal{F}$ and $M \in \operatorname{out}\mathcal{F}$, so by Theorem SEP.12 there is a point of $\mathcal{F}$ belonging to $\overline{MM'} \subseteq \overrightarrow{MH}$ which is impossible. Hence $\overrightarrow{MH} \cap \mathcal{G} = \emptyset$, the parity of M (relative to $\mathcal{G}$) is even and $M \in \operatorname{out}\mathcal{G}$.

Now suppose that $\overrightarrow{MH}$ intersects $\mathcal{F}$. Order $\overrightarrow{MH}$ by Definition ORD.1 with $M < H$. By Theorem PLGN.4(A) there exists a first point N of intersection of $\overrightarrow{MH}$ with $\mathcal{F}$, and since $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$, $N \in \operatorname{out}\mathcal{G}$. Now $\overline{MN} \cap \mathcal{F} = \emptyset$ so that $\overline{MN} \subseteq \operatorname{out}\mathcal{F}$ (by Theorem SEP.4(A)), and $\overline{MN} \cap \mathcal{G} = \emptyset$ because $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$. Since neither M nor N belongs to $\mathcal{G}$, and $N \in \operatorname{out}\mathcal{G}$, it follows from SEP.4(A) that $M \in \operatorname{out}\mathcal{G}$. Therefore $\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$ and since $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$, $\mathcal{F} \cup \operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$.

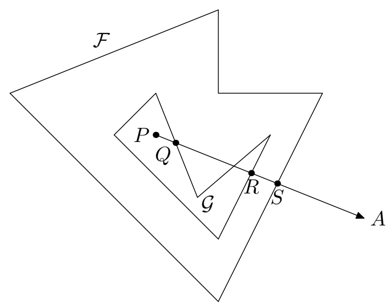


Fig. 6.11 For Theorem SEP.14, alternative (1) parts (b) and (c).

(b) See Figure 6.11. We prove next that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$. Let P be any point of $\operatorname{ins}\mathcal{G}$ and let A be any other point such that $\overleftrightarrow{PA}$ contains no corner of $\mathcal{F}$ or $\mathcal{G}$. We know that $\overrightarrow{PA}$ must intersect $\mathcal{G}$, and ordering $\overrightarrow{PA}$ with $P < A$, there exists a first point Q of intersection of $\mathcal{G}$ and $\overrightarrow{PA}$.

$\overline{PQ}$ contains no point of $\mathcal{G}$ other than Q, so $\overline{PQ} \subseteq \operatorname{ins}\mathcal{G}$ (SEP.4) which is disjoint from $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$. $Q \in \operatorname{ins}\mathcal{F}$ because $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$, so $Q \notin \mathcal{F}$, hence $\overline{PQ}$ contains no point of $\mathcal{F}$. Since $Q \in \mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$, it follows from SEP.4(A) that $P \in \operatorname{ins}\mathcal{F}$. This proves that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$, and since $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$ it follows that $\mathcal{G} \cup \operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$.

(c) Finally, we show the inclusions $\mathcal{G} \cup \operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$ and $\mathcal{F} \cup \operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$ are proper. Continue the construction of the immediately previous paragraphs (illustrated by the figure) as follows: let R be the last point of intersection of $\mathcal{G}$ and $\overrightarrow{PA}$, so that $R \in \mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$, and by Theorem PLGN.4 (since the parity of R relative to $\mathcal{F}$ is odd) there exists a first point S of intersection of $\mathcal{F}$ and the ray $\overrightarrow{PA} \setminus \overline{PR}$.

Since $\overline{RS} \cap \mathcal{F} = \overline{RS} \cap \mathcal{G} = \emptyset$, all points of the segment $\overline{RS}$ have the same parity with respect to $\mathcal{F}$ as does R (by SEP.4), which belongs to $\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$, so $\overline{RS} \subseteq \operatorname{ins}\mathcal{F}$. Also the ray $\overrightarrow{PQ} \setminus \overline{PR}$ is disjoint from $\mathcal{G}$ so that $\overline{RS} \subseteq \operatorname{out}\mathcal{G}$.

This shows that there are points of $\operatorname{ins}\mathcal{F}$ which are not in $\operatorname{ins}\mathcal{G}$, and also that there are points of $\operatorname{out}\mathcal{G}$ that are not points of $\operatorname{out}\mathcal{F}$. This completes the proof of (1).

If alternative (2) is true: The proof is the same as that for (1) with $\mathcal{F}$ and $\mathcal{G}$ interchanged.

If alternative (3) is true: We assume that $\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$.

First we prove that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$. Let P be any point of $\operatorname{ins}\mathcal{G}$, and let A be any point such that $\overrightarrow{PA}$ does not contain a corner of $\mathcal{F}$ or $\mathcal{G}$. $P \in \operatorname{ins}\mathcal{G}$ so that every ray from P must intersect $\mathcal{G}$ at least once. Order $\overrightarrow{PA}$ with $P < A$, and let Q be the first point and R the last point of intersection of $\overrightarrow{PA}$ with $\mathcal{G}$.

(a) If the ray $\overrightarrow{PA} \setminus \overline{PR}$ intersects $\mathcal{F}$ let S be the first point of intersection of $\overrightarrow{PA} \setminus \overline{PR}$ with $\mathcal{F}$, otherwise (b) let S be any point of $\overrightarrow{PA} \setminus \overline{PR}$. Then $\overline{PQ}$ contains no point of $\mathcal{G}$ other than Q so that $\overline{PQ} \subseteq \operatorname{ins}\mathcal{G}$ and hence contains no point of $\mathcal{F}$ since $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$. Now $Q \in \mathcal{G} \subseteq \operatorname{out}\mathcal{F}$, so $P \in \operatorname{out}\mathcal{F}$ by Theorem SEP.4(A). Therefore $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$.

Finally we show that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$ is a proper inclusion. The construction is similar to that shown in the figure for part (1) above, the only difference being that now $\mathcal{F}$ does not "enclose" $\mathcal{G}$. In this construction, the segment $\overline{RS}$ contains no point of $\mathcal{G}$ since R is the last point of $\mathcal{G}$, and no point of $\mathcal{F}$ since S is the first point of $\mathcal{F}$ with $R < S$. By Theorem SEP.4(A), since $R \in \mathcal{G} \subseteq \operatorname{out}\mathcal{F}$, $\overline{RS} \subseteq \operatorname{out}\mathcal{F}$; similarly since $S \in \operatorname{out}\mathcal{G}$, $\overline{RS} \subseteq \operatorname{out}\mathcal{G}$. Thus points of $\overline{RS}$ are in $\operatorname{out}\mathcal{F}$ but not in $\mathcal{G} \cup \operatorname{ins}\mathcal{G}$, and hence the inclusion $\mathcal{G} \cup \operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$ is proper.

By exactly the same argument, with the roles of $\mathcal{F}$ and $\mathcal{G}$ interchanged, $\mathcal{F} \cup \operatorname{ins}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$ is also a proper inclusion. $\square$

Theorem SEP.15 *If $\mathcal{F}$ and $\mathcal{G}$ are simple polygons and $\mathcal{F} \cap \mathcal{G}$ is a segment $\overline{CD}$, then both $\mathcal{F} \setminus \overline{CD}$ and $\mathcal{G} \setminus \overline{CD}$ are polygonally connected sets. By Theorem SEP.13*

(A) *either* $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$ *or* $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{F}$, *and*

(B) *either* $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{G}$ *or* $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{G}$.

Thus there are four mutually exclusive logical possibilities as follows:

(1) $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$ *and* $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{G}$, *in which case*
   $\operatorname{out}\mathcal{F}$ *is a proper subset of* $\operatorname{out}\mathcal{G}$, *and*
   $\operatorname{ins}\mathcal{G}$ *is a proper subset of* $\operatorname{ins}\mathcal{F}$; *or*

(2) $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{G}$ *and* $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{F}$, *in which case*
   $\operatorname{out}\mathcal{G}$ *is a proper subset of* $\operatorname{out}\mathcal{F}$, *and*

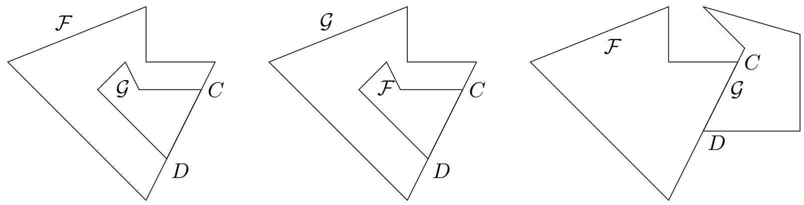
$\operatorname{ins}\mathcal{F}$ *is a proper subset of* $\operatorname{ins}\mathcal{G}$; *or*

(3) $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{F}$ *and* $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{G}$, *in which case*

$\operatorname{ins}\mathcal{G}$ *is a proper subset of* $\operatorname{out}\mathcal{F}$, *and*

$\operatorname{ins}\mathcal{F}$ *is a proper subset of* $\operatorname{out}\mathcal{G}$.

The remaining logical possibility (4) $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$ *and* $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{G}$, *is impossible.*



Alternative (1) $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$ and $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{G}$

Alternative (2) $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{G}$ and $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{F}$

Alternative (3) $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{G}$ and $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{F}$

Fig. 6.12 For Theorem SEP.15.

Proof. See Figure 6.12 above. Most parts of the proof are essentially the same as the corresponding parts of the proof of Theorem SEP.14, with the following exceptions:

(a) Since $\mathcal{F}$ and $\mathcal{G}$ are not disjoint, we cannot prove the same set of inclusions as in SEP.14, *e.g.* $\mathcal{G} \cup \operatorname{ins}\mathcal{G}$ is a proper subset of $\operatorname{ins}\mathcal{F}$—we can only show that $\operatorname{ins}\mathcal{G}$ is a proper subset of $\operatorname{ins}\mathcal{F}$, etc.

(b) When constructing a ray $\overrightarrow{PA}$, say from a point $P \in \operatorname{ins}\mathcal{G}$, we must choose A so that $\overrightarrow{PA}$ is disjoint from $\overline{CD}$ and intersections of the ray with either $\mathcal{F}$ or $\mathcal{G}$ will belong to $\mathcal{F} \setminus \overline{CD}$ or to $\mathcal{G} \setminus \overline{CD}$. We formalize this idea as follows.

Lemma *Let C, D, P, and A be points such that $P \notin \overleftrightarrow{CD}$. If the point $A \in \operatorname{out}\angle CPD$, then*

(1) $\overrightarrow{PA} \cap \overline{CD} = \emptyset$;

(2) *if $\mathcal{G}$ is a simple polygon and $\overline{CD} \subseteq \mathcal{G}$, $\overrightarrow{PA} \cap \mathcal{G} \subseteq \mathcal{G} \setminus \overline{CD}$; and*

(3) *a point X of $\overrightarrow{PA} \notin \mathcal{G}$ if and only if $X \notin \mathcal{G} \setminus \overline{CD}$.*

Proof. (1) By Theorem PSH.41(C) either A belongs to the side of $\overleftrightarrow{PD}$ opposite C or to the side of $\overleftrightarrow{PC}$ opposite D, and by Theorem IB.14 $\overrightarrow{PA}$ will be on the same side and hence $\overrightarrow{PA} \subseteq \operatorname{out}\angle CPD$. Therefore no point of $\overrightarrow{PA}$ belongs to $\overline{CD}$, since $\overline{CD} \subseteq \operatorname{ins}\angle CPD$ by Theorem PSH.37.

Conclusions (2) and (3) are obvious consequences of conclusion (1). □

We now return to the proof of Theorem SEP.15.

Proof that alternative (4) is impossible. We repeat the proof that alternative (4) of Theorem SEP.14 is impossible, altering that proof by choosing the point A to lie in out $\angle CPD$. Then by the Lemma, all the intersections of the ray $\overrightarrow{PA}$ with $\mathcal{F}$ and $\mathcal{G}$ lie in the sets $\mathcal{F} \setminus \overline{CD}$ and $\mathcal{G} \setminus \overline{CD}$ respectively, so that the proof from SEP.14 suffices.

If alternative (1) is true: $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins} \mathcal{F}$ and $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out} \mathcal{G}$.

(a) First we show that $\operatorname{out} \mathcal{F} \subseteq \operatorname{out} \mathcal{G}$. Let M be any point of $\operatorname{out} \mathcal{F}$ and let H be any point with $H \neq M$ with $H \in \operatorname{out} \angle CMD$, such that $\overrightarrow{MH}$ contains no corner of $\mathcal{F}$ or $\mathcal{G}$.

Note that $M \notin \mathcal{G} \subseteq \operatorname{ins} \mathcal{F}$ since $M \in \operatorname{out} \mathcal{F}$, which is disjoint from $\operatorname{ins} \mathcal{F} \supseteq \mathcal{G}$. If $\overrightarrow{MH}$ does not intersect $\mathcal{F}$ the proof is identical to the proof in SEP.14.

If $\overrightarrow{MH}$ intersects $\mathcal{F}$, we order $\overrightarrow{MH}$ by Definition ORD.1 with $M < H$. By Theorem PLGN.4(A) there exists a first point N of intersection of $\overrightarrow{MH}$ with $\mathcal{F}$, and since $N \notin \overline{CD}$ by the Lemma, and $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out} \mathcal{G}$, $N \in \operatorname{out} \mathcal{G}$.

Now $\overline{MN} \cap \mathcal{F} = \emptyset$ so that $\overline{MN} \subseteq \operatorname{out} \mathcal{F}$ (by Theorem SEP.4(A)), and $\overline{MN} \cap \mathcal{G} = \emptyset$ because $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins} \mathcal{F}$ and $\overrightarrow{MH} \cap \overline{CD} = \emptyset$. Since neither M nor N belongs to $\mathcal{G}$, and $N \in \operatorname{out} \mathcal{G}$, it follows from Theorem SEP.4(A) that $M \in \operatorname{out} \mathcal{G}$. Therefore $\operatorname{out} \mathcal{F} \subseteq \operatorname{out} \mathcal{G}$.

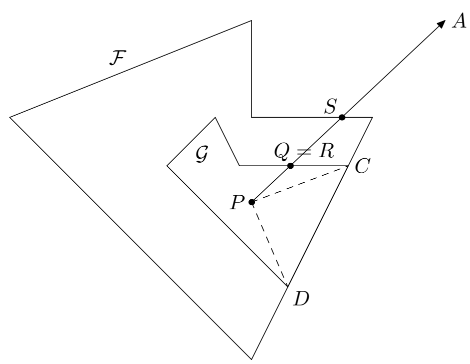


Fig. 6.13 For the construction for part (1)(b) of Theorem SEP.15.

(b) Next we prove that $\operatorname{ins} \mathcal{G} \subseteq \operatorname{ins} \mathcal{F}$. (See Figure 6.13 above for an illustration of this case.) Let P be any point of $\operatorname{ins} \mathcal{G} \setminus \overleftrightarrow{CD}$ and let A be any other point where $A \in \operatorname{out} \angle CPD$ (so that $\overrightarrow{PA} \cap \overline{CD} = \emptyset$ by the Lemma) and $\overleftrightarrow{PA}$ contains no corner of $\mathcal{F}$ or $\mathcal{G}$. Ordering $\overrightarrow{PA}$ with $P < A$, there exists a first point Q of intersection of $\mathcal{G}$ and $\overrightarrow{PA}$. $Q \notin \overline{CD}$ and since $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$, $Q \in \operatorname{ins}\mathcal{F}$. Now the segment $\overline{PQ}$ contains no point of $\mathcal{G}$ so $\overline{PQ} \subseteq \operatorname{ins}\mathcal{G}$ (SEP.4) which is disjoint from $\mathcal{F} \setminus \overline{CD}$ because $\mathcal{F} \setminus \overline{CD} \subseteq \operatorname{out}\mathcal{G}$, and hence $\overline{PQ}$ contains no point of $\mathcal{F}$. It follows from SEP.4(A) that $P \in \operatorname{ins}\mathcal{F}$, proving that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$.

(c) Finally, we show that the inclusions $\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$, and $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$ are proper. Continue the construction of the previous paragraphs (illustrated by Figure 6.13) as follows: let R be the last point of intersection of $\mathcal{G}$ and $\overrightarrow{PA}$, and since $R \notin \overline{CD}$ and thus $R \in \mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$, by Theorem PLGN.4(A) there exists a first point S of intersection of $\mathcal{F} \setminus \overline{CD}$ and the ray $\overrightarrow{PA} \setminus \overline{PR}$.

Since $\overline{RS} \cap \mathcal{F} = \overline{RS} \cap \mathcal{G} = \emptyset$, all points of the segment $\overline{RS}$ have the same parity with respect to $\mathcal{F}$ as does R (by SEP.4), which belongs to $\mathcal{G} \setminus \overline{CD} \subseteq \operatorname{ins}\mathcal{F}$, so $\overline{RS} \subseteq \operatorname{ins}\mathcal{F}$. Also the ray $\overrightarrow{PA} \setminus \overline{PR}$ is disjoint from $\mathcal{G}$ so that $\overline{RS} \subseteq \operatorname{out}\mathcal{G}$. This shows that there are points of $\operatorname{ins}\mathcal{F}$ which are not in $\operatorname{ins}\mathcal{G}$, and also that there are points of $\operatorname{out}\mathcal{G}$ that are not points of $\operatorname{out}\mathcal{F}$. This completes the proof of part (1).

If alternative (2) is true: Interchange $\mathcal{G}$ and $\mathcal{F}$ in the proof for alternative (1).

If alternative (3) is true: The proof is left to the reader as Exercise SEP.1. □

6.4 Rotundity and convexity (CNV)

Theorem CNV.1 *Let $\mathcal{L}$ be a line and $\mathcal{G}$ a simple polygon in the Pasch plane $\mathcal{P}$. Let $\mathcal{C}$ be the set of all corners of $\mathcal{G}$ that do not belong to $\mathcal{L}$ ($\mathcal{G} \cap \mathcal{L}$ may or may not be empty). Then for either side $\mathcal{H}$ of $\mathcal{L}$, the following statements are equivalent:*

(1) $\mathcal{C} \subseteq \mathcal{H}$,

(2) $\mathcal{G} \setminus \mathcal{L} \subseteq \mathcal{H}$, *that is, every edge of $\mathcal{G}$ is a subset of $\mathcal{L} \cup \mathcal{H}$, and*

(3) $\operatorname{ins}\mathcal{G} \subseteq \mathcal{H}$.

Proof. (1)⇒(2) If $\mathcal{C} \subseteq \mathcal{H}$, let $\mathcal{E}$ be any edge of $\mathcal{G}$. If both endpoints of $\mathcal{E}$ are members of $\mathcal{C}$, then $\mathcal{E} \subseteq \mathcal{H}$. If both ends of $\mathcal{E}$ belong to $\mathcal{L}$, then $\mathcal{E} \subseteq \mathcal{L}$. All other edges are of the form $\overline{PQ}$ where $P \in \mathcal{L}$ and $Q \in \mathcal{C}$. By Theorem IB.14, $\overline{PQ} \subseteq \mathcal{H}$ so that for every edge $\mathcal{E}$, $\mathcal{E} \subseteq \mathcal{L} \cup \mathcal{H}$.

(2)⇒(1) If $\mathcal{G} \setminus \mathcal{L} \subseteq \mathcal{H}$, then $\mathcal{C} \subseteq \mathcal{G} \setminus \mathcal{L}$, so $\mathcal{C} \subseteq \mathcal{H}$.

(2)⇒(3) Suppose $\mathcal{G} \setminus \mathcal{L} \subseteq \mathcal{H}$. Let $\mathcal{K}$ be the side of $\mathcal{L}$ opposite $\mathcal{H}$, and suppose $P \in \operatorname{ins}\mathcal{G}$. Now P could belong to $\mathcal{L}$, $\mathcal{H}$, or $\mathcal{K}$. If $P \in \mathcal{L}$, let Q be any point of $\mathcal{K}$ such that $\overrightarrow{PQ}$ is admissible, and $\overrightarrow{PQ}$ must intersect $\mathcal{G}$ at some point $R \notin \mathcal{L}$. But $\overrightarrow{PQ} \subseteq \mathcal{K}$ by Theorem IB.14, so $R \in \mathcal{K}$ which contradicts the assumption that $\mathcal{G} \setminus \mathcal{L} \subseteq \mathcal{H}$.

If $p \in \mathcal{K}$, let P' be a point of $\mathcal{L}$ such that $\overrightarrow{P'P} \subseteq \mathcal{K}$ and $\overrightarrow{P'P}$ is admissible. Then since $P \in \operatorname{ins}\mathcal{G}$, the ray $\overrightarrow{P'P} \setminus \overline{P'P}$ must intersect $\mathcal{G}$ at some point R; this ray is also a subset of $\mathcal{K}$, so $R \in \mathcal{K}$, again a contradiction to $\mathcal{G} \setminus \mathcal{L} \subseteq \mathcal{H}$. Therefore, no point of $\operatorname{ins}\mathcal{G}$ can be in $\mathcal{L}$ or $\mathcal{K}$, and $\operatorname{ins}\mathcal{G} \subseteq \mathcal{H}$.

(3)⇒(2) Suppose $\operatorname{ins}\mathcal{G} \subseteq \mathcal{H}$, and let $\mathcal{K}$ be the side of $\mathcal{L}$ opposite $\mathcal{H}$. Let $P \in \mathcal{G} \setminus \mathcal{L}$ and suppose $P \notin \mathcal{H}$, so that $P \in \mathcal{K}$. Then P belongs to some edge $\mathcal{E} = \overline{X_iX_{i+1}}$ of $\mathcal{G}$. If $P \in \overline{X_iX_{i+1}}$ choose $Q \notin \overleftrightarrow{X_iX_{i+1}}$. If P is an endpoint of $\mathcal{E}$, say $P = X_i$ choose $Q \in \operatorname{ins}\angle X_{i-1}X_iX_{i+1}$. In either case, by Theorem SEP.7, the line $\mathcal{M} = \overleftrightarrow{PQ}$ contains a point E such that $\overline{PE} \subseteq \operatorname{ins}\mathcal{G}$.

$P \notin \mathcal{L}$ so if $\overrightarrow{PE} \cap \mathcal{L} \neq \emptyset$, there is only one point F in the intersection. Choose C so that P–C–E and P–C–F. If $\overrightarrow{PE} \cap \mathcal{L} = \emptyset$ let $C = E$. Then $C \in \operatorname{ins}\mathcal{G}$ and $C \in \mathcal{K}$ because $\overline{PC} \cap \mathcal{L} = \emptyset$, a contradiction to the assumption that $\operatorname{ins}\mathcal{G} \subseteq \mathcal{H}$. □

Definition CNV.2 Let $\mathcal{G}$ be a simple polygon in the Pasch plane $\mathcal{P}$.

(A) A line $\mathcal{L}$ is a **supporting line** of $\mathcal{G}$, or **supports** $\mathcal{G}$ if and only if $\mathcal{G} \cap \mathcal{L} \neq \emptyset$ and $\operatorname{ins}\mathcal{G}$ is contained in a halfplane with edge $\mathcal{L}$ (that is, $\operatorname{ins}\mathcal{G}$ lies entirely within one side of $\mathcal{L}$).

(B) A segment $\overline{PQ}$ where P and Q are corners of $\mathcal{G}$, is said to **support** $\mathcal{G}$ iff the line $\overleftrightarrow{PQ}$ is a supporting line of $\mathcal{G}$. If $\overline{PQ}$ is an edge of $\mathcal{G}$ then we say that it is a **supporting edge** of $\mathcal{G}$. If $\mathcal{G}$ is a quadrilateral, a diagonal may or may not support $\mathcal{G}$.

(C) A polygon $\mathcal{E}$ is **rotund** iff for every line $\mathcal{L}$ containing an edge of $\mathcal{E}$, the corners of $\mathcal{E}$ not on $\mathcal{L}$ are on the same side of $\mathcal{L}$. This is an extension of Definition PSH.31, which applies to quadrilaterals.

Theorem CNV.3 (A) *A line $\mathcal{L}$ is a supporting line of a simple polygon $\mathcal{G}$ if and only if $\mathcal{G} \cap \mathcal{L} \neq \emptyset$ and all the corners of $\mathcal{G}$ not lying on $\mathcal{L}$ belong to the same side of $\mathcal{L}$.*

(B) *A simple polygon $\mathcal{G}$ is rotund if and only if every edge supports $\mathcal{G}$.*

(C) *An edge $\overline{PQ}$ supports a simple polygon $\mathcal{G}$ if and only if all the corners of $\mathcal{G}$ other than P and Q belong to the same side of $\overleftrightarrow{PQ}$.*

(D) *For any simple polygon* $\mathcal{G} = \langle X_1, \dots, X_n \rangle$, *the following statements are equivalent:*

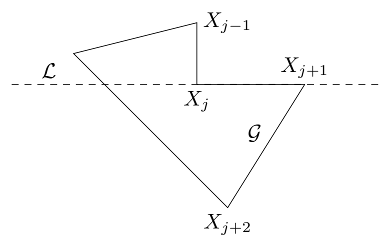
(1) $\mathcal{G}$ *is rotund,*

(2) *for each integer* $j \in [1; n]$, *all the corners* X_k *where* $X_k \neq X_j$ *and* $X_k \neq X_{j+1}$ *belong to the same side* $\mathcal{H}_j$ *of* $\overleftrightarrow{X_j X_{j+1}}$, *and* $\mathcal{G} \setminus \overline{X_j X_{j+1}} \subseteq \mathcal{H}_j$,

(3) *for each integer* $j \in [1; n]$, $\operatorname{ins} \mathcal{G}$ *is a subset of one side of* $\overleftrightarrow{X_j X_{j+1}}$, *and*

(4) *for every integer* j *the corners different from* X_{j-1}, X_j, *and* X_{j+1} *belong to* $\operatorname{ins} \angle X_{j-1} X_j X_{j+1}$. *(Note: if* $j = 1$ *then* $j - 1 = 0 \equiv n (\operatorname{mod} n)$ *so that* $X_{j-1} = X_n$, *and if* $j = n$ *then* $j + 1 \equiv 1 (\operatorname{mod} n)$ *so that* $X_{j+1} = X_1$.*)*

(E) *Every triangle is rotund.*



The polygon $\mathcal{G}$ is not rotund since X_{j-1} and X_{j+2} are on opposite sides of $\mathcal{L}$ and $\mathcal{L}$ is not supporting.

Fig. 6.14 For Theorem CNV.3(A).

Proof. See Figure 6.14. (A) is immediate from Definition CNV.2(A) and Theorem CNV.1. (B) follows immediately from Definition CNV.3(C) and part (A) just above. (C) follows immediately from Definition CNV.2(B) and part (A) just above.

In part (D), statement (2) is a restatement of Definition CNV.3(C), (the last part follows immediately from Theorem CNV.1) and this is equivalent to statement (3) by Theorem CNV.1. To prove (4) is equivalent to (1), suppose $\mathcal{G}$ is rotund and let Q be any corner of $\mathcal{G}$ different from X_{j-1}, X_j, and X_{j+1}, where $j \in [1; n]$. Then by part (D)(2) above, Q and X_{j+1} are on the same side of $\overleftrightarrow{X_{j-1} X_j}$ and Q and X_{j-1} are on the same side of $\overleftrightarrow{X_j X_{j+1}}$. By Definition PSH.36, $Q \in \operatorname{ins} \angle X_{j-1} X_j X_{j+1}$.

Conversely, let $j \in [1; n]$ and let Q be any corner of $\mathcal{G}$ other than X_{j-1}, X_j, and X_{j+1}, so that $Q \in \operatorname{ins} \angle X_{j-1} X_j X_{j+1}$. Then by Definition PSH.36, Q and X_{j-1} are on the same side of $\overleftrightarrow{X_j X_{j+1}}$; therefore all corners other than X_j

and X_{j+1} are on the same side of $\overleftrightarrow{X_jX_{j+1}}$ and since j was chosen arbitrarily, $\mathcal{G}$ is rotund by part (D)(2) above.

(E) Let A, B, and C be noncollinear points on the plane. Since C is the only corner of $\triangle ABC$ not on $\overleftrightarrow{AB}$, B is the only corner of $\triangle ABC$ not on $\overleftrightarrow{AC}$ and A is the only corner of $\triangle ABC$ not on $\overleftrightarrow{BC}$, condition (D)(2) above is vacuously satisfied and $\triangle ABC$ is rotund. $\square$

Theorem CNV.4 *Let $\mathcal{G} = \langle X_1, \ldots, X_n \rangle$ be a rotund polygon, and for each $i \in [1;n]$ let $\mathcal{H}_i$ be the side of $\overleftrightarrow{X_iX_{i+1}}$ containing all corners other than X_i and X_{i+1}. Then*

(A) *for every point $P \in \bigcap_{i=1}^{n} \mathcal{H}_i$, every ray $\overrightarrow{PQ}$ intersects $\mathcal{G}$ in exactly one point (so that the parity of P is odd);*

(B) $\operatorname{ins}\mathcal{G} = \bigcap_{i=1}^{n} \mathcal{H}_i$*, and is convex; and*

(C) *for every $j \in [1;n]$, $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\angle X_{j-1}X_jX_{j+1}$.*

(If $j = 1$, $X_{j-1} = X_n$; if $j = n$ then $X_{j+1} = X_1$.)

Proof. (A) Let $P \in \bigcap_{i=1}^{n} \mathcal{H}_i$ and suppose $\overrightarrow{PQ} \cap \mathcal{G}$ contains two points A and B. If these are on the same edge of $\mathcal{G}$, then by Axiom I.1, P, Q, A, and B are collinear which is impossible since P belongs to a side of $\overleftrightarrow{AB}$.

Now suppose A is the first of two points of intersection A and B ($\overrightarrow{PQ}$ is ordered with $P < Q$), and that $A \in \overline{X_jX_{j+1}}$ and $B \in \overline{X_kX_{k+1}}$, where $k \neq j$. Now $P \in \mathcal{H}_j$, and the ray $\overrightarrow{PA}$ intersects $\overleftrightarrow{X_jX_{j+1}}$ in only one point, so that the ray $\overrightarrow{PA} \setminus \overline{PA} \subseteq \mathcal{K}_j$, the side of $\overleftrightarrow{X_jX_{j+1}}$ opposite $\mathcal{H}_j$. Therefore $B \in \mathcal{K}_j$. But by Theorem CNV.1, $\mathcal{G} \subseteq \overleftrightarrow{X_jX_{j+1}} \cup \mathcal{H}_j$ which is disjoint from $\mathcal{K}_j$ so we have a contradiction. Note that in this part of the proof it does not matter if A or B is a corner of $\mathcal{G}$.

(B) By Theorem CNV.1, $\operatorname{ins}\mathcal{G} \subseteq \mathcal{H}_j$ for every $i \in [1;n]$, and therefore $\operatorname{ins}\mathcal{G} \subseteq \bigcap_{i=1}^{n} \mathcal{H}_i$. Conversely, by part (A), every point of $\bigcap_{i=1}^{n} \mathcal{H}_i$ has odd parity and $\bigcap_{i=1}^{n} \mathcal{H}_i \subseteq \operatorname{ins}\mathcal{G}$. Since each $\mathcal{H}_i$ is a convex set (Corollary to Theorem PSH.7) and the intersection of any collection of convex sets is convex (Exercise IB.15), $\operatorname{ins}\mathcal{G}$ is convex.

(C) $\operatorname{ins}\mathcal{G} = \bigcap_{i=1}^{n} \mathcal{H}_i \subseteq \mathcal{H}_{j-1} \cap \mathcal{H}_j = \operatorname{ins}\angle X_{j-1}X_jX_{j+1}$. $\square$

Corollary CNV.4.1 *If $\mathcal{G}$ is a rotund polygon and $P \in \operatorname{ins}\mathcal{G}$, then every ray from P intersects $\mathcal{G}$ in a singleton Q. Furthermore, $\overline{PQ} \subseteq \operatorname{ins}\mathcal{G}$ and $\overrightarrow{PQ} \setminus \overline{PQ} \subseteq \operatorname{out}\mathcal{G}$.*

Proof. From parts (A) and (B) every ray from P (admissible or not) intersects $\mathcal{G}$ only once, at a point Q, and if P–X–Q, $\overrightarrow{XQ}$ intersects $\mathcal{G}$ only once so that $\overline{PQ} \subseteq \operatorname{ins}\mathcal{G}$. For any X, if P–Q–X then $\overrightarrow{PQ} \setminus \overline{PX}$ is a ray originating from X which has no intersection with $\mathcal{G}$ and hence $X \in \operatorname{out}\mathcal{G}$. $\square$

Corollary CNV.4.2 *Let $\mathcal{G}$ be a rotund polygon and let $P \in \operatorname{ins}\mathcal{G}$ and $Q \in \operatorname{out}\mathcal{G}$. Then $\mathcal{G}$ intersects $\overline{PQ}$ at exactly one point.*

Proof. By Corollary CNV.4.1, $\overrightarrow{PQ}$ intersects $\mathcal{G}$ at a single point A, and by Corollary SEP.4.1, $\overline{PQ}$ must intersect $\mathcal{G}$, and since A is the only possible point of intersection, $A \in \overline{PQ}$. $\square$

Corollary CNV.4.3 *If $\mathcal{G}$ is a triangle, then the definitions of* $\operatorname{ins}\mathcal{G}$ *in PSH.36 and in SEP.3 have the same meaning.*

Proof. Theorem CNV.3(E) says that a triangle is rotund; part (B) above states that if $\mathcal{G}$ is rotund, $\operatorname{ins}\mathcal{G} = \bigcap_{i=1}^{n} \mathcal{H}_i$. The left hand side is the definition of inside as in Definition SEP.3, using parity; the right hand side is the definition of inside as in PSH.36. $\square$

Theorem CNV.5 (Generalization of Theorem PSH.50)
Let $\mathcal{G} = \langle X_1, \ldots, X_n \rangle$ be a rotund polygon and $\mathcal{L}$ a line. If $\mathcal{G} \cap \mathcal{L} \neq \emptyset$, then

(A) $\mathcal{L} \cap \operatorname{enc}\mathcal{G}$ *is either a single point or a segment, and*

(B) $\mathcal{L} \cap \mathcal{G}$ *is exactly one of the following:*

- (1) *a single point S, in which case*
  - (a) *S is a corner X_j of $\mathcal{G}$,*
  - (b) $\mathcal{L} \cap \operatorname{ins}\mathcal{G} = \emptyset$ *and* $\mathcal{L} \cap \operatorname{ins}\angle X_{j-1}X_jX_{j+1} = \emptyset$
    (if $j = 1$, $X_{j-1} = X_n$; if $j = n$ then $X_{j+1} = X_1$),
  - (c) *the sets $\mathcal{G} \setminus \{X_j\}$, $\overrightarrow{X_jX_{j-1}}$, $\overrightarrow{X_jX_{j+1}}$,* $\operatorname{ins}\mathcal{G}$ *and*
    $\operatorname{ins}\angle X_{j-1}X_jX_{j+1}$ *are all subsets of the X_{j-1}-side (= X_{j+1}-side) of $\mathcal{L}$;*
- (2) *exactly two points P and Q, in which case*
  - (a) *no edge of $\mathcal{G}$ contains both P and Q,*
  - (b) $\mathcal{L} \cap \operatorname{ins}\mathcal{G} = \overline{PQ} \neq \emptyset$,
    $\mathcal{L} \cap \operatorname{out}\mathcal{G} = \{X | X\text{–}P\text{–}Q\} \cup \{X | P\text{–}Q\text{–}X\}$,
  - (c) $\mathcal{L} = \overleftrightarrow{PQ}$
    $= \{X | X\text{–}P\text{–}Q\} \cup \{P\} \cup \overline{PQ} \cup \{Q\} \cup \{X | P\text{–}Q\text{–}X\}$
    $= \mathcal{L} \setminus \overline{PQ} \cup \{P\} \cup \overline{PQ} \cup \{Q\}$;
- (3) *more than two points, in which case*
  - (a) $\mathcal{L} \cap \operatorname{ins}\mathcal{G} = \emptyset$, *and*
  - (b) *$\mathcal{L}$ contains an edge $\mathcal{E}$ of $\mathcal{G}$.*

Proof. (A) Follows immediately from (B). The following Lemma will facilitate the rest of the proof.

Lemma *Let $\mathcal{G} = \langle X_1, \ldots, X_n \rangle$ be a rotund polygon and $\mathcal{L}$ a line.*

(A) *If $\mathcal{L} \cap \mathcal{G}$ contains two points P and Q and $\overline{PQ} \cap \mathcal{G} = \emptyset$, then $\overline{PQ} \subseteq \operatorname{ins} \mathcal{G}$; and*

(B) *if $\mathcal{L} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$, then there exist exactly two points P and Q such that $\mathcal{L} \cap \mathcal{G} = \{P, Q\}$ and $\overline{PQ} \cap \mathcal{G} = \emptyset$.*

Proof. (A) Let j be any member of $[1; n]$, and denote by $\mathcal{H}_j$ the side of $\mathcal{L}$ containing all corners of $\mathcal{G}$ that are not on $\mathcal{L}$. P and Q cannot both be members of $\overline{X_j X_{j+1}}$ for then $\overline{PQ} \cap \mathcal{G} = \overline{PQ}$ would not be empty. Therefore either P or Q fails to belong to $\overline{X_j X_{j+1}}$.

Let $O \in \overline{PQ}$ and suppose that $P \notin \overline{X_j X_{j+1}}$; by Theorem CNV.3(D)(2) $P \in \mathcal{H}_j$, and $\overline{OP}$ contains no point of $\mathcal{G}$ so that $O \in \mathcal{H}_j$ by Definition IB.11. A similar proof holds if $Q \notin \overline{X_j X_{j+1}}$. Thus $O \in \bigcap_{i=1}^{n} \mathcal{H}_i = \operatorname{ins} \mathcal{G}$ by Theorem CNV.4(B).

Conversely, if $R \in \mathcal{L} \cap \operatorname{ins} \mathcal{G}$, then by Theorem CNV.4(A) every ray from R intersects $\mathcal{G}$ at exactly one point. Let X and Y be such that X–R–Y; then $\overrightarrow{RX}$ intersects $\mathcal{G}$ at exactly one point P and $\overrightarrow{RY}$ intersects $\mathcal{G}$ at exactly one point Q, $\mathcal{L} \cap \mathcal{G} = \{P, Q\}$ and $\overline{PQ} \cap \mathcal{G} = \emptyset$. □

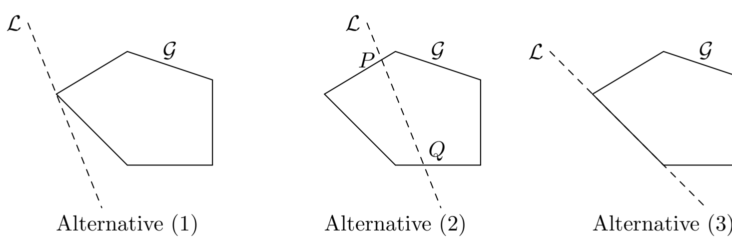


Fig. 6.15 For Theorem CNV.5(B).

(B) Clearly exactly one of the alternatives (1), (2), or (3) holds. See Figure 6.15.

(1) (a) Suppose $\mathcal{L} \cap \mathcal{G} = \{S\}$ and S is not a corner of $\mathcal{G}$. Then pick points A and B on $\mathcal{L}$ such that A–S–B. Then $\overline{AB} \cap \mathcal{G}$ is a singleton and by Theorem SEP.4(B) the parities of A and B are different so one of them, say B, belongs to $\operatorname{ins} \mathcal{G}$. Then choose X so that A–S–B–X. By Theorem CNV.4(A), $\overrightarrow{BX}$, and

hence $\mathcal{L}$ must intersect $\mathcal{G}$ at some point C. This contradicts the assumption that $\mathcal{L} \cap \mathcal{G} = \{S\}$, and therefore S must be a corner X_j of $\mathcal{G}$.

(b) By the Lemma, part (2), if $\mathcal{L} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$ then there are exactly two points in the intersection $\mathcal{L} \cap \mathcal{G}$, contradicting the assumption that $\mathcal{L} \cap \mathcal{G} = \{S\}$. Hence $\mathcal{L} \cap \operatorname{ins} \mathcal{G} = \emptyset$.

If $\mathcal{L}$ contains a point B of $\operatorname{ins} \angle X_{j-1}X_jX_{j+1}$, choose A so that A–X_j–B. Since $\mathcal{L} \cap \mathcal{G}$ is a singleton, so is $\overleftrightarrow{AB} \cap \mathcal{G}$, and by Theorem SEP.4(C), A and B have different parities, so one of them belongs to $\operatorname{ins} \mathcal{G}$. Suppose $B \in \mathcal{L} \cap \operatorname{ins} \mathcal{G}$, and let X be such that A–X_j–B–X. The ray $\overrightarrow{BX}$ must intersect $\mathcal{G}$ at some point C (by CNV.4), so there are two points in $\mathcal{L} \cap \mathcal{G}$, a contradiction. Thus $\mathcal{L} \cap \operatorname{ins} \angle X_{j-1}X_jX_{j+1} = \emptyset$.

This fact furnishes an alternative proof that $\mathcal{L} \cap \operatorname{ins} \mathcal{G} = \emptyset$, since $\operatorname{ins} \mathcal{G} \subseteq \operatorname{ins} \angle X_{j-1}X_jX_{j+1}$ (CNV.4(C)).

(c) Suppose X_{j-1} and X_{j+1} are on different sides of $\mathcal{L}$. Then by Axiom PSA, $\overline{X_{j-1}X_{j+1}}$ intersects $\mathcal{L}$ at some point P, and by Theorem PSH.37 $P \in \operatorname{ins} \angle X_{j-1}X_jX_{j+1}$, in contradiction to (b). Thus both X_{j-1} and X_{j+1} belong to the same side of $\mathcal{L}$, which we will call $\mathcal{H}$. By Theorem IB.14, $\overrightarrow{X_jX_{j-1}} \subseteq \mathcal{H}$ and $\overrightarrow{X_jX_{j+1}} \subseteq \mathcal{H}$.

If $Q \in \operatorname{ins} \angle X_{j-1}X_jX_{j+1}$, then $\overrightarrow{X_jQ}$ intersects $\overline{X_{j-1}X_{j+1}}$ at some point R by the Crossbar Theorem PSH.39 $\overline{QR} \subseteq \overrightarrow{X_jQ} \subseteq Q$-side of $\mathcal{L}$ and hence $\overline{QR}$ does not intersect $\mathcal{L}$, and $\overline{X_{j-1}X_{j+1}}$ is also disjoint from $\mathcal{L}$, since X_{j-1} is on the same side of $\mathcal{L}$ as X_{j+1}. Therefore all the points (in particular Q) of $\overline{QR} \cup \overline{X_{j-1}X_{j+1}}$ are on the same side, which must be $\mathcal{H}$.

Finally, by CNV.4(C), $\operatorname{ins} \mathcal{G} \subseteq \operatorname{ins} \angle X_{j-1}X_jX_{j+1} \subseteq \mathcal{H}$, and by Theorem CNV.1, $\mathcal{G} \setminus \{X_j\} \subseteq \mathcal{H}$, completing the proof of this part.

(2) (a) If a single edge $\mathcal{E}$ of $\mathcal{G}$ contains two points of $\mathcal{L}$, then $\mathcal{E} \subseteq \mathcal{L}$ and $\mathcal{E}$ contains more than two points of $\mathcal{L}$ which contradicts the hypothesis. Therefore, P and Q cannot be on the same edge.

(b) If $\mathcal{L} \cap \mathcal{G} = \{P, Q\}$, then $\overline{PQ} \cap \mathcal{L} = \emptyset$ and by the Lemma, $\overline{PQ} \subseteq \operatorname{ins} \mathcal{G}$. By Corollary CNV.4.1, both $\{X | X\text{–}P\text{–}Q\}$ and $\{X | P\text{–}Q\text{–}X\}$ are subsets of $\operatorname{out} \mathcal{G}$, which proves (b); the assertions of (c) follow easily.

(3) (a) If $\mathcal{L} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$, then by the Lemma (2) there exist exactly two points P and Q such that $\mathcal{L} \cap \mathcal{G} = \{P, Q\}$ and $\overline{PQ} \cap \mathcal{G} = \emptyset$. This contradicts the assumption that there are more than two points of intersection.

(b) First order $\mathcal{L}$ by Definition ORD.1. If $\mathcal{L}$ does not contain any edge of $\mathcal{G}$, then the intersection of $\mathcal{L}$ with every edge is a single point, and the set of all such intersections, being a finite set, has a least element P. Let Q be the second (next) element of the intersection. Then $\overline{PQ} \cap \mathcal{G} = \emptyset$, and by the

Lemma, part (A), $\overline{PQ} \subseteq \text{ins}\,\mathcal{G}$, which is impossible by part (a). □

In order to make the statements and proofs of the following Lemma as clear as possible, we will state each part separately with its proof and any illustrations.

Lemma CNV.6(A) *A quadrilateral is not rotund if and only if exactly one of its corners belongs to the inside of the triangle whose corners are the other three corners of the quadrilateral.*

Proof. (A) Lemma CNV.6(A) is a repetition of Theorem PSH.53, (*q.v.*). There are four possibilities ((i)–(iv))listed at the end of the proof of Theorem PSH.53, for the quadrilateral $\langle A, B, C, D\rangle$ to be non-rotund. One of these figures illustrates alternative (i), and we reproduce again here as Figure 6.16 for the reader's convenience. It is quite easy to construct figures of the other possibilities, and in each case it will be seen that there are two supporting edges and two edges that are not supporting. □

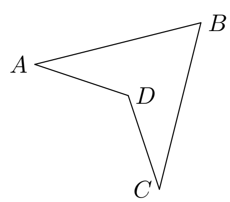


Fig. 6.16 For one case of Lemma CNV.6(A) (see also Theorem PSH.53).

Lemma CNV.6(B) *If $\mathcal{G}$ is a quadrilateral with three supporting edges, then $\mathcal{G}$ is rotund.*

Proof. (B) If $\mathcal{G} = \langle A, B, C, D\rangle$ is not rotund, then by part (A) exactly two edges are not supporting edges. □

Lemma CNV.6(C) *For every quadrilateral $\mathcal{F} = \langle A, B, C, D\rangle$, $\overline{AC}$ and $\overline{BD}$ are its diagonals. Then*

(1) *each side of a line containing a non-supporting diagonal must contain a corner of $\mathcal{F}$;*

(2) *$\mathcal{F}$ is not rotund iff exactly one of its diagonals supports $\mathcal{F}$;*

(3) *every quadrilateral has at least one non-supporting diagonal, and is rotund iff neither diagonal supports $\mathcal{F}$;*

(4) (a) $\overline{AC}$ *is the common edge of* $\triangle ABC$ *and* $\triangle ADC$*; if it is non-supporting,* $\mathring{\overline{AC}}$*,* $\operatorname{ins}\triangle ABC$*, and* $\operatorname{ins}\triangle ADC$ *are all subsets of* $\operatorname{ins}\mathcal{F}$*, and*

(b) $\overline{BD}$ *is the common edge of* $\triangle BAD$ *and* $\triangle BCD$*; if it is non-supporting,* $\mathring{\overline{BD}}$*,* $\operatorname{ins}\triangle BAD$*,* $\operatorname{ins}\triangle BAD$*, and* $\operatorname{ins}\triangle BCD$ *are all subsets of* $\operatorname{ins}\mathcal{F}$*.*

Proof. (C) Result (1) follows directly from Definition CNV.2(C).

(2) This is essentially a re-statement of Lemma CNV.6(A). To say that the corner A, for instance, of $\mathcal{F}$ belongs to the inside of the triangle $\triangle BCD$ is the same as saying that both A and C are on the same side of $\overleftrightarrow{BD}$, that is, $\overleftrightarrow{BD}$ is a supporting line containing the diagonal $\overline{BD}$.

(3) If $\mathcal{F}$ is not rotund, it has exactly one supporting diagonal by part (2); the other is non-supporting. If $\mathcal{F}$ is rotund, then by Theorem PSH.54 the diagonals $\overline{AC}$ and $\overline{BD}$ intersect. If the intersection were a corner, then three corners would be collinear which would violate the definition of quadrilateral. Therefore by Definition IB.11, B and D are on opposite sides of $\overleftrightarrow{AC}$ and A and C are on opposite sides of $\overleftrightarrow{BD}$. Hence neither diagonal is supporting.

(4) Suppose a diagonal, say $\overline{AC}$ is not supporting, so that the corners B and D are on opposite sides of $\overleftrightarrow{AC}$. Then if $X \in \mathring{\overline{AC}}$, let $Y \in \mathring{\overline{BC}}$; the ray $\mathring{\overrightarrow{XY}} \subseteq B$-side of $\overleftrightarrow{AC}$, so cannot intersect either $\overline{AD}$ or $\overline{CD}$, which are on the D-side of $\overleftrightarrow{AC}$, and cannot intersect $\overline{AB}$ which is a subset of the side of $\overleftrightarrow{XY}$ opposite C. Therefore Y is the only point of intersection of $\mathring{\overrightarrow{XY}}$ with $\mathcal{F}$, and since Y contains no corner of $\mathcal{F}$, $\mathring{\overrightarrow{XY}}$ is admissible, and the parity of X is odd.

If $P \in \operatorname{ins}\triangle ABC$, let $Q \in \mathring{\overline{AC}}$ be a point such that $\overrightarrow{PQ}$ contains no corner of $\mathcal{F}$. The segment $\mathring{\overline{PQ}}$ contains no point of $\triangle ABC$, and since it lies entirely on the B-side of $\overleftrightarrow{AC}$, it is disjoint from $\overline{AD}$ and $\overline{CD}$, and hence from $\mathcal{F}$. The ray $\overrightarrow{PQ}$ contains the ray $\overrightarrow{PQ} \setminus \overline{PQ}$ which contains an odd number of points of $\mathcal{F}$, since $Q \in \operatorname{ins}\mathcal{F}$. Hence P has odd parity, and belongs to $\operatorname{ins}\mathcal{F}$. A similar argument for $\triangle ADC$ shows that both $\operatorname{ins}\triangle ABC \subseteq \operatorname{ins}\mathcal{F}$ and $\operatorname{ins}\triangle ADC \subseteq \operatorname{ins}\mathcal{F}$.

A similar argument holds if the diagonal $\overline{BD}$ is not supporting. $\square$

Lemma CNV.6(D) *Let* $\mathcal{G}$ *be a simple polygon, and let* A*,* B*,* C*, and* D *be noncollinear points where* $\mathcal{F} = \langle A, B, C, D\rangle$ *is a quadrilateral such that*

(1) $\mathring{\overline{AD}} \cap \mathcal{G} = \mathring{\overline{BC}} \cap \mathcal{G} = \emptyset$*,*

(2) $\operatorname{ins}\langle A, B, C, D\rangle$ *contains no corner of* $\mathcal{G}$*, and*

(3) $\mathring{\overline{CD}}$ *contains no corner of* $\mathcal{G}$*.*

Then every edge $\mathcal{E}$ of $\mathcal{G}$ that intersects $\overline{CD}$ at a single point must also intersect $\overline{AB}$, and the intersection is a singleton.

Proof. (D) By Lemma CNV.6(C)(3), at least one of $\overline{AC}$ or $\overline{BD}$ is a diagonal that supports $\mathcal{F}$. Without loss of generality we may assume that $\overline{AC}$ does not support $\mathcal{F}$. By Lemma CNV.6(C)(4), $\overline{AC}$ forms the common edge of the two triangles $\triangle ABC$ and $\triangle ADC$, $\overline{AC} \subseteq \operatorname{ins}\mathcal{F}$, $\operatorname{ins}\triangle ABC \subseteq \operatorname{ins}\mathcal{F}$ and $\operatorname{ins}\triangle ADC \subseteq \operatorname{ins}\mathcal{F}$. Therefore there are no corners of $\mathcal{G}$ in $\operatorname{ins}\triangle ABC$, $\operatorname{ins}\triangle ADC$, or $\overline{AC}$.

Suppose the edge $\mathcal{E}$ (of $\mathcal{G}$) intersects $\overline{CD}$ at a point P and no other point. Applying Lemma PLGN.12 to $\triangle ADC$, we find that $\mathcal{E}$ must intersect the diagonal $\overline{AC}$ in a single point; if that point is A, we are done; if the intersection is a point of $\overline{AC}$ then we can apply the same theorem to the triangle $\triangle ABC$, concluding that $\mathcal{E}$ intersects $\overline{AB}$ (and hence $\overline{AB}$), again in a singleton. □

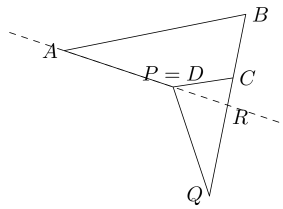
Lemma CNV.6(E) *If a quadrilateral $\mathcal{G} = \langle A, B, Q, P\rangle$ is such that P and Q are on the same side of $\overleftrightarrow{AB}$ (that is, $\overline{AB}$ is a supporting edge), then there exist points C and D such that $D \in \overline{AP}$, $C \in \overline{BQ}$, and $\langle A, B, C, D\rangle$ is rotund.*

Proof. (E) If $\mathcal{G}$ is rotund the result follows immediately by choosing $C = Q$ and $D = P$. If $\mathcal{G}$ is not rotund, from Lemma CNV.6(A) there are two possibilities where P and Q are on the same side of $\overleftrightarrow{AB}$:

(i) In alternative (i) (illustrated by Figure 6.17 overleaf), $\overleftrightarrow{AP}$ intersects $\overline{BQ}$ at R between B and Q; take C to be any point between R and B and take $D = P$. P and Q are on the same side of $\overleftrightarrow{AB}$ and by Theorem CNV.1, so are all points of $\mathcal{G} \setminus \overline{AB}$, including D and C. A and D ($= P$) are on the same side of $\overleftrightarrow{BC}(= \overleftrightarrow{BQ})$. Since $C \in \overline{RB}$ and $D \in \overline{RA}$, by Theorem PSH.4 $\overleftrightarrow{CD}$ does not intersect $\overline{AB}$ and hence A and B are on the same side of $\overleftrightarrow{CD}$ by Definition IB.11. Finally, since R–D–A, A and D are on the same side of $\overleftrightarrow{BC} = \overleftrightarrow{RB}$, and $\langle A, B, C, D\rangle$ is rotund.

(ii) In alternative (ii), $\overleftrightarrow{BQ}$ intersects $\overline{AP}$ at S between A and P; take D to be any point between S and A and take $C = Q$. By a similar argument $\langle A, B, C, D\rangle$ is rotund. We leave the construction of a figure illustrating alternative (ii) to the reader. □

Lemma CNV.6(F) *If a quadrilateral $\langle A, B, C, D\rangle$ is rotund, J is a point between A and D, and K is a point between B and C, then $\langle A, B, K, J\rangle$ is rotund.*



A and B are on opposite sides of $\overleftrightarrow{QP}$
B and Q are on opposite sides of $\overleftrightarrow{AP}$
$D \in \text{ins}\,\triangle ABQ$
Q and P are on the same side of $\overleftrightarrow{AB}$
A and P are on the same side of $\overleftrightarrow{BQ}$

Fig. 6.17 For Lemma CNV.6(E) alternative (i).

Proof. (F) By Theorem CNV.1 both J and K (which belong to $\mathcal{G} \setminus \overline{AB}$) are on the same side of $\overleftrightarrow{AB}$; both K and B (which belong to $\mathcal{G} \setminus \overline{AD}$) are on the same side of $\overleftrightarrow{AD}$ $(= \overleftrightarrow{AJ})$; and both A and J (which belong to $\mathcal{G} \setminus \overline{BC}$) are on the same side of $\overleftrightarrow{BC}$ $(= \overleftrightarrow{BK})$. Then by Lemma CNV.6(B), $\mathcal{G}$ is rotund, since it has three supporting edges. $\square$

Lemma CNV.6(G) *If $\mathcal{G} = \langle A, B, C, D\rangle$ is a rotund quadrilateral and $\mathcal{E}$ is a nonempty finite set of points which contains no point of $\overline{AB}$, then there exist points R and S such that A–R–D, B–S–C, and $\text{enc}\langle A, B, S, R\rangle \cap \mathcal{E} = \emptyset$ (and by Lemma CNV.6(F) $\langle A, B, S, R\rangle$ is rotund).*

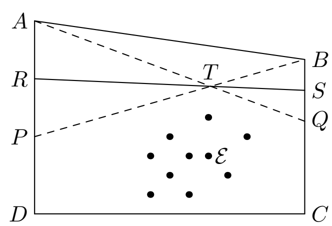


Fig. 6.18 For proof of Lemma CNV.6(G).

Proof. (G) See Figure 6.18. Let $\mathcal{C} = \{X | A\text{–}X\text{–}D \text{ and } \overline{XB} \cap \mathcal{E} \neq \emptyset\}$. $\mathcal{C}$ is a finite set which does not contain A, so by Theorem ORD.10 we may let P' be the first point of $\mathcal{C}$ (where $A < D$), and let P be any point with A–P–P'. If $Y \in \text{enc}\,\triangle ABP \cap \mathcal{E}$, then either $Y \in \overline{BP}$ or $Y \in \text{ins}\,\angle ABP$ in which case $\overrightarrow{BY}$ intersects $\overline{AP}$ at some point Z, by the Crossbar Theorem PLGN.39, so that either P or Z is a point of $\mathcal{C}$, a contradiction to the definition of P'.

Thus enc $\triangle ABP$ contains no point of $\mathcal{E}$. Similarly we may find $Q \in \overline{BC}$ such that enc $\triangle ABQ$ contains no point of $\mathcal{E}$.

Now $Q \in \operatorname{ins}\angle BAD$ and $B \in \overline{AB}$ and $P \in \overline{AD}$ so by the Crossbar Theorem PSH.39, $\overrightarrow{AQ} \cap \overline{BP} \neq \emptyset$ and the intersection is a point $\{T\}$. A similar construction can be done for $P \in \operatorname{ins}\angle ABC$, the ray $\overrightarrow{BP}$, and the segment $\overline{BP}$, and the point of intersection is again $\{T\}$ since distinct lines can intersect in only one point.

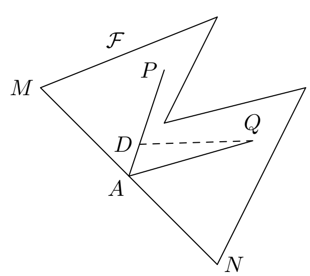
Pick S with B–S–Q. Then by Theorem PSH.37 $S \in \operatorname{ins}\angle BTQ$, that is, $S \in \overrightarrow{PBQ} \cap \overrightarrow{AQB}$ (see Definition PSH.36). $\overrightarrow{ST}$ intersects both $\overline{AQ}$ and $\overline{BP}$ at the point T and contains points on the other side of both, hence any point X with X–T–S belongs to $\overrightarrow{PBA} \cap \overrightarrow{AQP} = \operatorname{ins}\angle PTA$. Again by the Crossbar Theorem, $\overrightarrow{ST}$ intersects $\overline{AP}$ at a point R with A–R–P.

Then $\langle A, B, S, R\rangle$ is a quadrilateral where

$$\operatorname{enc}\langle A, B, S, T\rangle \subseteq \operatorname{enc}\triangle ABP \cup \operatorname{enc}\triangle ABQ,$$

both of which are disjoint from $\mathcal{E}$. By Lemma CNV.6(F) $\langle A, B, S, R\rangle$ is rotund. $\square$

Lemma CNV.6(H) *If*
(1) $\overline{MN}$ *is an edge of a simple polygon* $\mathcal{F}$,
(2) A *is a point such that* M–A–N,
(3) P *and* Q *are points on the same side of* $\overleftrightarrow{MN}$, *and*
(4) $\overline{AP} \cap \mathcal{F} = \overline{AQ} \cap \mathcal{F} = \{A\}$,
then P *and* Q *have the same parity with respect to* $\mathcal{F}$ *(either both* P *and* Q *belong to* ins $\mathcal{F}$ *or both belong to* out $\mathcal{F}$*).*



Here both P and Q belong to ins $\mathcal{F}$

Fig. 6.19 For proof of Lemma CNV.6(H).

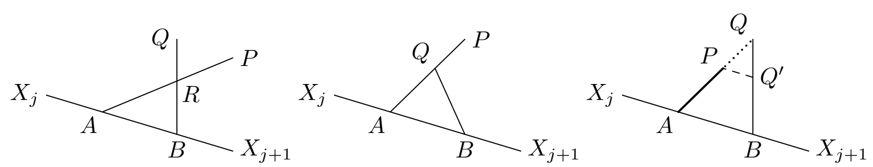
Proof. (H) See Figure 6.19. If P, Q, and A are collinear, the result is immediate from Theorem SEP.4(A). Otherwise, $P \notin \overleftrightarrow{AQ}$ and since P is a point such that $\overline{AP} \cap \mathcal{F} = \emptyset$, the points Q, A, and P satisfy the hypotheses of Theorem PLGN.13. Thus there exists a point $D \in \overline{AP}$ such that no corner of $\mathcal{F}$ is in either $\text{ins}\,\triangle AQD$ or $\overline{QD}$, and every edge that intersects $\overline{QD}$ must also intersect $\overline{QA}$. But there is only one edge intersecting $\overline{QA}$, that is $\overline{MN}$, and it does not intersect $\overline{QD}$ since both D and Q are on the same side of $\overleftrightarrow{MN}$. Since neither D nor $Q \in \mathcal{F}$, $\overline{QD} \cap \mathcal{F} = \emptyset$, and $\langle\langle P, D, Q\rangle\rangle$ is a polygonal path connecting P and Q which does not intersect $\mathcal{F}$. Hence by Theorem SEP.13(A), P and Q have the same parity. $\square$

Lemma CNV.6(I) *Suppose $\overline{X_jX_{j+1}}$ is an edge of a simple polygon $\mathcal{F}$, A and B are distinct members of $\overline{X_jX_{j+1}}$, P and Q are points on the same side of $\overleftrightarrow{X_jX_{j+1}}$, $\overline{PA} \cap \mathcal{F} = \{\mathcal{A}\}$, $\overline{QB} \cap \mathcal{F} = \{B\}$, and one of the following hypotheses holds:*

(1) X_j–A–B–X_{j+1},
(2) $A = X_j$, $B \neq X_{j+1}$, *and* $X_{j-1} \in \text{out}\,\angle PAB$,
(3) $A \neq X_j$, $B = X_{j+1}$, *and* $X_{j+2} \in \text{out}\,\angle QBA$, *or*
(4) $A = X_j$, $B = X_{j+1}$, $X_{j-1} \in \text{out}\,\angle PAB$ *and* $X_{j+2} \in \text{out}\,\angle QBA$.

Then P and Q have the same parity with respect to $\mathcal{F}$ (either both P and Q belong to $\text{ins}\,\mathcal{F}$ or both belong to $\text{out}\,\mathcal{F}$).

Proof. (I) For cases (i)–(iii), the first three figures given as Figure 6.20 are drawn as if hypothesis (1) above holds—the points A and B are not end points; however, they can serve to illustrate hypotheses (2)–(4), where A or B is an end point (we have to draw the figures *some* way). For case (iv), Figure 6.21 illustrates hypothesis (1), and Figure 6.22 illustrates hypotheses (2), (3), and (4).



Case (i) Case (ii) Case (iii)

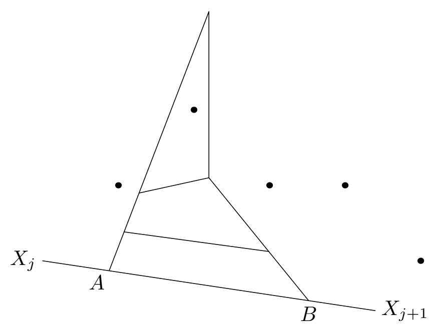
Fig. 6.20 For proof of Lemma CNV.6(I), cases (i)–(iii).

(Case i) If $\overline{AP} \cap \overline{BQ} \neq \emptyset$, let $\{R\} = \overline{AP} \cap \overline{BQ}$. Then $\overline{AR} \cup \overline{RB}$ is a polygonal path connecting P and Q which is disjoint from $\mathcal{F}$ and by Theorem SEP.13(A), P and Q have the same parity with respect to $\mathcal{F}$.

(Case ii) If either A–Q–P or B–P–Q, then by Theorem SEP.4(A) P and Q have the same parity since $\overline{PQ}$ is a subset of either $\overline{PA}$ or $\overline{QB}$ and therefore $\overline{PQ} \cap \mathcal{F} = \emptyset$.

(Case iii) If $\{A, P, Q\}$ is collinear and A–P–Q then choose Q' so that B–Q'–Q, and in this case both $\{A, P, Q'\}$ and $\{B, Q', P\}$ are noncollinear. Similarly if $\{B, Q, P\}$ is collinear and B–Q–P choose P' so that A–P'–P, so that both $\{A, P', Q\}$ and $\{B, Q, P'\}$ are noncollinear. The following argument applied to either $\langle A, B, Q', P\rangle$ or $\langle A, B, Q, P'\rangle$ as the case may be will show that P and Q', or P' and Q', and thus P and Q have the same parity.

(Case iv) Suppose $\{A, P, Q\}$ and $\{B, Q, P\}$ are noncollinear, and that $\overline{AP}$ is disjoint from $\overline{BQ}$. Since P and Q are on the same side of $\overleftrightarrow{X_j X_{j+1}} = \overleftrightarrow{AB}$, $\overline{PQ} \cap \overline{AB} = \emptyset$ and condition (2) of Definition PSH.31 is satisfied. The triples $\{A, B, P\}$ and $\{Q, A, B\}$ are noncollinear, so that $\langle A, B, P, Q\rangle$ is a quadrilateral, but not necessarily rotund (see Figure 6.21).



Case (iv): the dots suggest the set $\mathcal{H}$ of ,
all corners of $\mathcal{F}$ other than X_j and X_{j+1}

Fig. 6.21 For Lemma CNV.6(I), hypothesis (1), case (iv).

By Lemma CNV.6(E) above, we can find $C \in \overline{BQ}$ and $D \in \overline{AP}$ such that $\langle A, B, C, D\rangle$ is rotund. Now let $\mathcal{H}$ be the set of all corners of $\mathcal{F}$ other than X_j and X_{j+1}. By Lemma CNV.6(G) above we may find points $F \in \overline{AD}$ and $E \in \overline{BC}$ such that if $\mathcal{G} = \langle A, B, E, F\rangle$, then $\operatorname{ins}\mathcal{G} \cup \mathcal{G} = \operatorname{enc}\mathcal{G}$ contains no corners of $\mathcal{F}$ other than possibly A or B (which would be the case if A or B were the same as X_j or X_{j+1}). Thus $\overline{FE}$ contains no corner of $\mathcal{F}$, is a subset of $\operatorname{enc}\mathcal{G}$ and does not contain either X_j or X_{j+1}.

Since both $\overline{AF}$ and $\overline{BE}$ are disjoint from $\mathcal{F}$ we may apply Lemma CNV.6(D) to conclude that every edge which intersects $\overline{FE}$ in only one point must also intersect $\overline{AB}$. But no edge of $\mathcal{F}$ contains two points of $\overline{FE}$, for if this segment contained two points of an edge, that edge would be a subset of it (since neither E nor $F \in \mathcal{F}$) and $\overline{FE}$ would contain a corner of $\mathcal{F}$, a contradiction. Thus all edges intersecting $\overline{FE}$ must also intersect $\overline{AB}$, so that $\overline{X_{j-1}X_j}$, $\overline{X_jX_{j+1}}$, and $\overline{X_{j+1}X_{j+2}}$ are the only edges that might intersect $\overline{FE}$. We now show that none of these are possible.

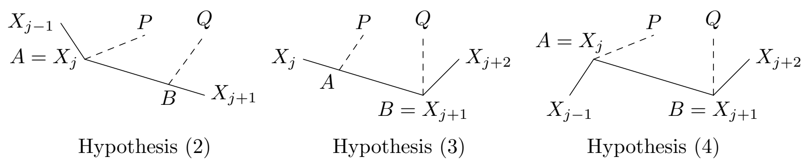


Fig. 6.22 For Lemma CNV.6(I), hypotheses (2)–(4), case (iv).

(a) $\overline{X_jX_{j+1}}$ cannot intersect $\overline{FE}$ since both F and E are on the same side of $\overleftrightarrow{AB} = \overleftrightarrow{X_jX_{j+1}}$.

(b) Neither can the edge $\overline{X_{j-1}X_j}$ intersect $\overline{FE}$: if hypotheses (1) or (3) hold, $A \neq X_j$, so that if $\overline{X_{j-1}X_j}$ intersects $\overline{FE}$ (which is in the side of $\overleftrightarrow{AP}$ opposite X_j), $\overline{X_{j-1}X_j}$ would contain points on both sides of $\overleftrightarrow{AP}$ and hence would intersect $\overleftrightarrow{AP}$ by Axiom PSA, and because this last intersection would be on the X_j-side of $\overleftrightarrow{FE}$ and also on the F-side of $\overleftrightarrow{AB}$, it would belong to $\overline{AP}$, contradicting $\overline{AP} \cap \mathcal{F} = \emptyset$.

If hypotheses (2) or (4) hold, $X_{j-1} \in \operatorname{out} \angle PAB = \operatorname{out} \angle PX_jB$, so that $\overline{X_{j-1}X_j} \subseteq \operatorname{out} \angle PX_jB$, whereas $\overline{FE} \subseteq \operatorname{ins} \angle PAB$ (as it lies on the B-side of $\overleftrightarrow{AP}$ and on the P-side of $\overleftrightarrow{AB}$), so that $\overline{X_{j-1}X_j}$ cannot intersect $\overline{FE}$.

(c) An argument similar to (b) shows that the edge $\overline{X_{j+1}X_{j+2}}$ cannot intersect $\overline{FE}$. Hence $\overline{FE} \cap \mathcal{F} = \emptyset$, and therefore $\langle\langle P, F, E, Q \rangle\rangle$ is a polygonal path joining P and Q, which does not intersect $\mathcal{F}$, so by Theorem SEP.13(A), P and Q have the same parity. $\square$

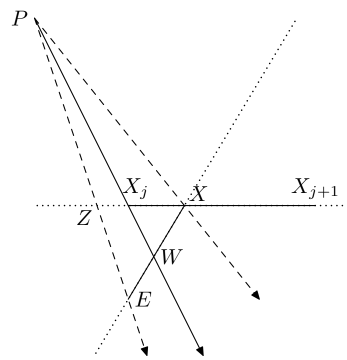
Theorem CNV.7 *Let* $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ *be a simple polygon. Then* $\operatorname{ins} \mathcal{F}$ *is convex iff* $\mathcal{F}$ *is rotund iff* $\operatorname{enc} \mathcal{F}$ *is convex.*

Proof. (A) From Theorem CNV.4(B), if $\mathcal{F}$ is rotund, then $\operatorname{ins} \mathcal{F}$ is convex. We show the converse, that if $\operatorname{ins} \mathcal{F}$ is convex, $\mathcal{F}$ is rotund. See Figure 6.23.

Suppose $\mathcal{F}$ is not rotund. Then by Theorem CNV.3(D)(3) for some integer $j \in [1; n]$ there are points of $\operatorname{ins} \mathcal{F}$ on both sides of $\overleftrightarrow{X_jX_{j+1}}$. Let X be a point such that X_j–X–X_{j+1}. Then by Theorem SEP.7(A) there is a point $E \notin \overleftrightarrow{X_jX_{j+1}}$ such that $\overline{EX} \subseteq \operatorname{ins} \mathcal{F}$.

Let $P \in \operatorname{ins}\mathcal{F}$ be on the side of $\overleftrightarrow{X_jX_{j+1}}$ opposite E. Without loss of generality we assume P is on the X_j-side of $\overleftrightarrow{EX}$. Then $\overrightarrow{PE}$ intersects $\overleftrightarrow{X_jX_{j+1}}$ at a point Z where P–Z–E. If $Z \in \overline{X_jX_{j+1}}$ then $\overline{PE}$ is a segment joining P and E which also contains a point of $\mathcal{F}$, a contradiction to the assumption that $\operatorname{ins}\mathcal{F}$ is convex.

If $Z \notin \overline{X_jX_{j+1}}$, Z–X_j–X_{j+1}, since P is on the X_j-side of $\overleftrightarrow{EX}$. Then since $X_j \in \overline{ZX}$, and both $\overline{ZX}$ and $\overline{EX}$ are subsets of $\operatorname{ins}\angle EPX$, by the Crossbar Theorem the ray $\overrightarrow{PX_j}$ intersects $\overline{EX}$ at some point W which belongs to $\operatorname{ins}\mathcal{F}$. Thus $\overline{PW}$ is a segment joining P and W which contains a point X_j of $\mathcal{F}$, contradicting the assumption that $\operatorname{ins}\mathcal{F}$ is convex. We have shown that $\operatorname{ins}\mathcal{F}$ is convex iff $\mathcal{F}$ is rotund.



The case in (A) where $\overrightarrow{PE}$ intersects $\overleftrightarrow{X_jX_{j+1}}$ outside $\overline{X_jX_{j+1}}$

Fig. 6.23 For proof of Theorem CNV.7(A).

(B) Now we show that $\operatorname{ins}\mathcal{F}$ is convex iff $\operatorname{enc}\mathcal{F}$ is convex. Suppose $\operatorname{ins}\mathcal{F}$ is convex; by Lemma CNV.6(A) $\mathcal{F}$ is rotund. If P and $Q \in \operatorname{enc}\mathcal{F}$, either

(i) both P and $Q \in \operatorname{ins}\mathcal{F}$,

(ii) both P and $Q \in \mathcal{F}$,

(iii) $Q \in \operatorname{ins}\mathcal{F}$ and $P \in \mathcal{F}$, or

(iv) $P \in \operatorname{ins}\mathcal{F}$ and $Q \in \mathcal{F}$.

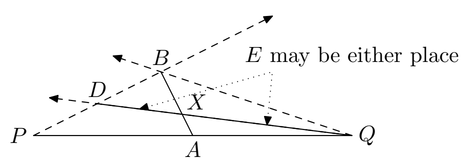
In case (i) $\overline{PQ} \subseteq \operatorname{ins}\mathcal{F} \subseteq \operatorname{enc}\mathcal{F}$ by assumption; in case (ii), if P and Q belong to the same edge of $\mathcal{F}$ then $\overline{PQ}$ is a subset of that edge and hence of $\mathcal{F}$ and hence of $\operatorname{enc}\mathcal{F}$; if P and Q belong to different edges, by Theorem CNV.5(2) $\overline{PQ} \subseteq \operatorname{ins}\mathcal{F}$ and hence $\overline{PQ} \subseteq \operatorname{enc}\mathcal{F}$.

In case (iii), since $\overleftrightarrow{PQ} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$, $\overleftrightarrow{PQ}$ intersects $\mathcal{F}$ in exactly two points, P and some point R, and $Q \in \overline{PR}$ by Theorem CNV.5. Thus $\overline{PQ} \subseteq \overline{PR} \subseteq \operatorname{enc}\mathcal{F}$. Similarly for case (iv).

To prove the converse, assume $\operatorname{enc}\mathcal{F}$ is convex, and let P and Q be any two points of $\operatorname{ins}\mathcal{F}$. Then $\overline{PQ} \subseteq \operatorname{enc}\mathcal{F}$ because $\operatorname{enc}\mathcal{F}$ is convex.

Now suppose $\overline{PQ}$ is not a subset of $\operatorname{ins}\mathcal{F}$; then every point of $\overline{PQ}$ that does not belong to $\operatorname{ins}\mathcal{F}$ belongs to $\mathcal{F}$. Let A be the first point (where $P < Q$) such that $A \in \overline{PQ} \cap \mathcal{F}$. Then A is a corner and there exists at least one edge $\overline{AB}$ which is not a subset of $\overleftrightarrow{PQ}$.

Let D' be the first point of $\overrightarrow{PB} \cap \mathcal{F}$ (where $P < B$) and let D be such that P–D–D'–B. Then since $P \in \operatorname{ins}\mathcal{F}$, $D \in \operatorname{ins}\mathcal{F}$ by theorem SEP.4(A). Note that both $\overline{PB}$ and $\overline{AB}$ are subsets of $\operatorname{ins}\angle PQB$. Then $D \in \operatorname{ins}\angle PQB$ so by the Crossbar Theorem, there exists a point X such that $\overrightarrow{QD} \cap \overline{AB} = \{X\}$. By Theorem SEP.7 there exists a point E on $\overleftrightarrow{QD}$ such that $\overline{EX} \subseteq \operatorname{out}\mathcal{F}$. Thus we have a segment ($\overline{DQ}$) connecting two points D and Q of $\operatorname{ins}\mathcal{F}$ which contains points of $\operatorname{out}\mathcal{F}$ so that $\overline{DQ} \not\subseteq \operatorname{enc}\mathcal{F}$, contradicting the convexity of $\operatorname{enc}\mathcal{F}$. See Figure 6.24. $\square$



P, Q and D all belong to $\operatorname{ins}\mathcal{F}$ but there are points of $\overline{QD}$ that belong to $\operatorname{out}\mathcal{F}$

Fig. 6.24 For Theorem CNV.7, contradicting the convexity of $\operatorname{enc}\mathcal{F}$

Theorem CNV.8 (A) *Let $\mathcal{F}$ be a simple polygon, $\mathcal{L}$ a supporting line of $\mathcal{F}$, and let $\mathcal{H}$ be the half-plane with edge $\mathcal{L}$ such that* $\operatorname{ins}\mathcal{F} \subseteq \mathcal{H}$. *Then* $\operatorname{enc}\mathcal{F} \subseteq \mathcal{H} \cup \mathcal{L}$.

(B) *Let $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ be a simple polygon and let $\mathcal{L}$ be a line. Then* $\mathcal{L} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$ *iff there exist corners of $\mathcal{F}$ on opposite sides of $\mathcal{L}$.*

(C) *Let $\mathcal{F}$ be a simple polygon and $\mathcal{L}$ be a line. $\mathcal{L}$ is a supporting line of $\mathcal{F}$ if and only if $\mathcal{F} \cap \mathcal{L} \neq \emptyset$ and* $\mathcal{L} \cap \operatorname{ins}\mathcal{F} = \emptyset$.

Proof. (A) By Theorem CNV.1(2), every edge of $\mathcal{F}$ is a subset of $\mathcal{L} \cup \mathcal{H}$.

(B) Assume that $\mathcal{L} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$. Then $\mathcal{L}$ must intersect $\mathcal{F}$, for otherwise $\mathcal{L} \subseteq \operatorname{out}\mathcal{F}$ which is a contradiction to $\mathcal{L} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$. If all the corners not in $\mathcal{L}$ are on one of its sides $\mathcal{H}$, then $\operatorname{ins}\mathcal{F} \subseteq \mathcal{H}$ by Theorem CNV.1 which contradicts $\mathcal{L} \cap \operatorname{ins} \mathcal{F} \neq \emptyset$. Therefore at least one corner must be on each side of $\mathcal{L}$.

Conversely, assume that X_j and X_k are corners of $\mathcal{F}$ on opposite sides of $\mathcal{L}$, and choose the notation so that $1 \leq j < k \leq n$ (if necessary re-index the corners). Let r be the greatest integer less than k such that X_r is on the X_j-side of $\mathcal{L}$, and let s be the least integer greater than r such that X_s is on the X_k-side of $\mathcal{L}$.

Then for all i with $r < i < s$, $X_i \in \mathcal{L}$. Then $s \leq r+3$ because otherwise there would be three adjacent collinear corners of $\mathcal{F}$ which is impossible. Thus one of three possibilities holds:

(1) $s = r+1$, in which case X_r is on the opposite side of $\mathcal{L}$ from X_s and $\overline{X_rX_s} \cap \mathcal{L} \neq \emptyset$ by Axiom PSA, and the intersection is a singleton P since $\mathcal{L} \neq \overleftrightarrow{X_rX_s}$. Then by Theorem SEP.7 there is a point of $\operatorname{ins} \mathcal{F}$ on $\mathcal{L}$.

(2) $s = r+2$, in which case $X_{s-1} = X_{r+1} \in \mathcal{L}$. X_r, X_{r+1}, and X_s are noncollinear because $\mathcal{F}$ is a polygon. Let A and B be points on $\mathcal{L}$ with A–X_{r+1}–B. Then one of the points A or B is on the X_r-side of $\overleftrightarrow{X_{r+1}X_s}$ and also on the x_s-side of $\overleftrightarrow{X_rX_{r+1}}$ and thus belongs to $\operatorname{ins} \angle X_rX_{r+1}X_s$. By Theorem SEP.7 $\mathcal{L} \cap \operatorname{ins} \mathcal{F} \neq \emptyset$.

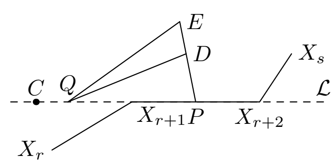


Fig. 6.25 For proof of Theorem CNV.8(B) alternative (3).

(3) $s = r+3$, in which case X_{r+1} and X_{r+2} both belong to $\mathcal{L}$, while X_r and X_s are on opposite sides of $\mathcal{L}$. Pick a point P with X_{r+1}–P–X_{r+2}; by Theorem SEP.7 there exists $E \notin \mathcal{L}$ with $\overline{PE} \subseteq \operatorname{ins} \mathcal{F}$. See Figure 6.25.

Now E is either on the X_s-side or on the X_r-side of $\mathcal{L}$. Without loss of generality we may assume the former; pick a point $C \in \mathcal{L}$ to be the first intersection of the ray $\overrightarrow{X_{r+2}X_{r+1}} \setminus \overline{X_{r+2}X_{r+1}}$ with $\mathcal{F}$, and let Q be such that C–Q–X_{r+1}.

Since $\overline{PE}$ contains no point of $\mathcal{F}$ by Theorem PLGN.13 there is also a point $D \in \overline{PE}$ such that every edge $\mathcal{E}$ that intersects $\overline{QD}$ also intersects $\overline{QP}$; but the only edges that intersect $\overline{QP}$ are (a) $\overline{X_{r+1}X_{r+2}} \subseteq \mathcal{L}$ which cannot intersect $\overline{QD}$ because D is on a side of $\mathcal{L}$, and (b) $\overline{X_rX_{r+1}}$ which cannot intersect $\overline{QD}$ because it is on the opposite side of $\mathcal{L}$ from D. Thus $\overline{QD}$ does not intersect any edge of $\mathcal{F}$ and by Theorem SEP.4 Q and D have the same parity. Since $D \in \operatorname{ins}\mathcal{F}$, $Q \in \operatorname{ins}\mathcal{F}$, hence $\mathcal{L} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$.

(C) If $\mathcal{L}$ is a supporting line, by Definition CNV.2(A) $\mathcal{F} \cap \mathcal{L} \neq \emptyset$ and $\operatorname{ins}\mathcal{F}$ is entirely on one side of $\mathcal{L}$, and so disjoint from $\mathcal{L}$. Conversely, suppose $\mathcal{F} \cap \mathcal{L} \neq \emptyset$ and $\mathcal{L} \cap \operatorname{ins}\mathcal{F} = \emptyset$. By part (B), if $\mathcal{L} \cap \operatorname{ins}\mathcal{F} = \emptyset$ then all the corners not on $\mathcal{L}$ are on the same side $\mathcal{H}$, so that by Theorem CNV.1, $\operatorname{ins}\mathcal{F} \subseteq \mathcal{H}$ and $\mathcal{L}$ is a supporting line. $\square$

Theorem CNV.9 *Let $\mathcal{F}$ be a simple polygon and $\mathcal{L}$ be a supporting line of $\mathcal{F}$. If there exist distinct corners A and B of $\mathcal{F}$ such that $A \in \mathcal{L}$, $B \in \mathcal{L}$, and $\overleftrightarrow{AB}$ is not an edge of $\mathcal{F}$, then* $\operatorname{enc}\mathcal{F}$ *and* $\operatorname{ins}\mathcal{F}$ *are both nonconvex.*

Proof. If every point between A and B belonged to $\mathcal{F}$, then $\overleftrightarrow{AB}$ would be an edge of $\mathcal{F}$. Therefore some point C with A–C–B fails to belong to $\mathcal{F}$, and since $\operatorname{ins}\mathcal{F} \cap \mathcal{L} = \emptyset$ (Definition CNV.2(A)) $C \in \operatorname{out}\mathcal{F}$. Therefore $\operatorname{enc}\mathcal{F}$ is nonconvex and by Theorem CNV.7 $\operatorname{ins}\mathcal{F}$ is nonconvex. $\square$

Corollary CNV.9.1 *Let $\mathcal{F}$ be a simple polygon and $\mathcal{L}$ be a supporting line of $\mathcal{F}$. If* $\operatorname{ins}\mathcal{F}$ *(or* $\operatorname{enc}\mathcal{F}$*) is convex, then for any distinct corners A and B of $\mathcal{F}$ such that $A \in \mathcal{L}$ and $B \in \mathcal{L}$, $\overleftrightarrow{AB}$ is an edge of $\mathcal{F}$.*

Proof. The corollary is the contrapositive of Theorem CNV.9. $\square$

Definition CNV.10 Let $\mathcal{E}$ be any nonempty subset of the Pasch plane. The **convex hull** of $\mathcal{E}$ (notation: $\operatorname{coh}\mathcal{E}$) is the set $\mathcal{T}$ such that $\mathcal{T}$ is convex, $\mathcal{E} \subseteq \mathcal{T}$, and if $\mathcal{H}$ is any convex set containing $\mathcal{E}$, then $\mathcal{T} \subseteq \mathcal{H}$.

Theorem CNV.11 (A) *If $\mathcal{E} \subseteq \mathcal{F}$ are nonempty subsets of a plane, then* $\operatorname{coh}\mathcal{E} \subseteq \operatorname{coh}\mathcal{F}$.

(B) *If $\mathcal{E}$ is any nonempty subset of a plane, then $\mathcal{E}$ is convex iff* $\operatorname{coh}\mathcal{E} \subseteq \mathcal{E}$ *iff* $\operatorname{coh}\mathcal{E} = \mathcal{E}$.

(C) *If $\mathcal{E}$ is any nonempty subset of a plane,* $\operatorname{coh}(\operatorname{coh}\mathcal{E}) = \operatorname{coh}\mathcal{E}$.

(D) *If $\mathcal{F}$ is a simple polygon, then* $\operatorname{ins}\mathcal{F} \subseteq \operatorname{coh}\mathcal{F}$, $\operatorname{enc}\mathcal{F} \subseteq \operatorname{coh}\mathcal{F}$, *and* $\operatorname{coh}(\operatorname{enc}\mathcal{F}) = \operatorname{coh}\mathcal{F}$.

(E) *If $\mathcal{F}$ is a simple polygon, then $\mathcal{F}$ is rotund iff* $\operatorname{ins}\mathcal{F}$ *is convex iff* $\operatorname{coh}\mathcal{F} = \operatorname{enc}\mathcal{F}$.

Proof. (A) $\operatorname{coh}\mathcal{F}$ is a convex set containing $\mathcal{E}$ so $\operatorname{coh}\mathcal{E} \subseteq \operatorname{coh}\mathcal{F}$.

(B) If $\mathcal{E}$ is convex then $\operatorname{coh}\mathcal{E} \subseteq \mathcal{E}$, since $\mathcal{E}$ is a convex set containing $\mathcal{E}$. Conversely, if $\operatorname{coh}\mathcal{E} \subseteq \mathcal{E}$, we know already that $\mathcal{E} \subseteq \operatorname{coh}\mathcal{E}$, so $\operatorname{coh}\mathcal{E} = \mathcal{E}$ and $\mathcal{E}$ is convex.

(C) $\operatorname{coh}\mathcal{E}$ is convex, so by part (B) $\operatorname{coh}(\operatorname{coh}\mathcal{E}) = \operatorname{coh}\mathcal{E}$.

(D) Let $P \in \operatorname{ins}\mathcal{F}$, and let A and B be two points such that A–P–B and $\overleftrightarrow{AB}$ contains no corner of $\mathcal{F}$; then both $\overrightarrow{PA}$ and $\overrightarrow{PB}$ intersect $\mathcal{F}$ at points C and D respectively, since the parity of P is odd, and $P \in \overline{CD} \subseteq \operatorname{coh}\mathcal{F}$ because $\operatorname{coh}\mathcal{F}$ is convex. Since $\mathcal{F} \subseteq \operatorname{coh}\mathcal{F}$ it follows that $\operatorname{enc}\mathcal{F} \subseteq \operatorname{coh}\mathcal{F}$, and from (A) and (C) $\operatorname{coh}(\operatorname{enc}\mathcal{F}) \subseteq \operatorname{coh}(\operatorname{coh}\mathcal{F}) = \operatorname{coh}\mathcal{F}$.

(E) If $\operatorname{ins}\mathcal{F}$ is convex, then $\operatorname{enc}\mathcal{F}$ is convex by Theorem CNV.7, so $\operatorname{enc}\mathcal{F} = \operatorname{coh}\mathcal{F}$ by part (B). If $\operatorname{coh}\mathcal{F} = \operatorname{enc}\mathcal{F}$ then $\operatorname{enc}\mathcal{F}$ is convex so that by Theorem CNV.7 $\operatorname{ins}\mathcal{F}$ is convex. $\square$

Theorem CNV.12 *Let $\mathcal{F}$ be a simple polygon, A be any member of* $\operatorname{ins}\mathcal{F}$*, P a corner of $\mathcal{F}$, and $\mathcal{H}$ a given halfplane with edge $\overleftrightarrow{AP}$. Then there exists a corner Q of $\mathcal{F}$ belonging to $\mathcal{H}$ such that no corner of $\mathcal{F}$ belongs to* $\operatorname{ins}\angle PAQ$.

Proof. Since the line $\overleftrightarrow{AP}$ has non-empty intersection with $\operatorname{ins}\mathcal{F}$, by Theorem CNV.8(B) there is a corner R of $\mathcal{F}$ with $R \in \mathcal{H}$ (because there must be corners on both sides of $\overleftrightarrow{AP}$). If there is no corner of $\mathcal{F}$ belonging to $\operatorname{ins}\angle PAR$ take $Q = R$ and we are done.

On the other hand, if there is a corner of $\mathcal{F}$ belonging to $\operatorname{ins}\angle PAR$, then the set $\mathcal{D} = \{X | X \text{ is a corner of } \mathcal{F} \text{ and } X \in \operatorname{ins}\angle PAR\}$ is non-empty. Define an ordering $<$ on $\mathcal{D}$ as follows: if S and $T \in \mathcal{D}$, $S < T$ if and only if $S \in \operatorname{ins}\angle PAT$. Let Q be the minimal element of $\mathcal{D}$ with respect to this ordering. Then no corner of $\mathcal{F}$ belongs to $\operatorname{ins}\angle PAQ$. $\square$

Corollary CNV.12.1 *Let $\mathcal{F}$ be a simple polygon, Q be any member of* $\operatorname{ins}\mathcal{F}$*, and let P be any corner of $\mathcal{F}$. Then there exists a corner Q of $\mathcal{F}$ such that no corner of $\mathcal{F}$ belongs to* $\operatorname{ins}\angle PAQ$.

Proof. Let $\mathcal{H}$ be either halfplane with edge $\overleftrightarrow{AP}$. $\square$

Theorem CNV.13 *Let $\mathcal{F}$ be a simple polygon, $\mathcal{D}$ the set of all segments both of whose endpoints are corners of $\mathcal{F}$, $\mathcal{E}$ the union of all segments in $\mathcal{D}$, and let A be an arbitrary point of* $\operatorname{ins}\mathcal{F}$.

(A) *Let P and Q be corners of $\mathcal{F}$ such that no corner of $\mathcal{F}$ belongs to* $\operatorname{ins}\angle PAQ$*, B be any member of* $\operatorname{ins}\angle PAQ$*, and let C be the last intersection* *of $\mathcal{E}$ and $\overrightarrow{AB}$. Then the line $\overleftrightarrow{EF}$ containing the member $\overline{EF}$ of $\mathcal{D}$ which intersects $\overrightarrow{AB}$ at C is a supporting line of $\mathcal{F}$.*

(B) *Let $\mathcal{L}$ be a supporting line of $\mathcal{F}$ which contains at least two distinct corners of $\mathcal{F}$. Then there exist distinct corners P and Q of $\mathcal{F}$ such that no corner of $\mathcal{F}$ belongs to* $\text{ins}\,\angle PAQ$*, and if B is any member of* $\text{ins}\,\angle PAQ$*, then the last intersection of the set $\mathcal{E}$ and $\overrightarrow{AB}$ is on $\mathcal{L}$.*

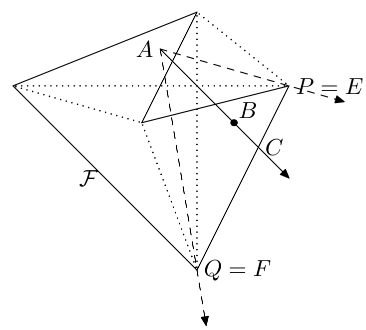


Fig. 6.26 For Theorem CNV.13(A),(B); the dotted lines are the lines of $\mathcal{D}$ which are not edges of $\mathcal{F}$.

Proof. See Figure 6.26. (A) It suffices to show that there is no corner of $\mathcal{F}$ on the side of $\overleftrightarrow{EF}$ opposite the A-side. If there were such a corner R, then since R is not on $\overleftrightarrow{AB}$ either E and R would be on the same side of $\overleftrightarrow{AB}$ or F and R would be on the same side of $\overleftrightarrow{AB}$. Furthermore, $\overline{FR}$ and $\overline{ER}$ would be on the side of $\overleftrightarrow{EF}$ opposite the A-side. Hence $\overrightarrow{AB}$ would intersect exactly one member of $\{\overline{FR}, \overline{ER}\}$ at M such that A–C–M. This would contradict the fact that C is the last point of intersection of $\mathcal{E}$ and $\overrightarrow{AB}$.

(B) Let the given corners on $\mathcal{L}$ be J and K. If we let $P = J$, then by Theorem CNV.12 there is a corner Q on the K-side of $\overleftrightarrow{AP}$ such that there are no corners of $\mathcal{F}$ in $\text{ins}\,\angle PAQ$. If $\overrightarrow{AQ}$ intersects $\mathcal{L}$ in a point M, then P–M–K or $M = K$ because if P–K–M, then $K \in \text{ins}\,\angle PAQ$ which contradicts the definition of Q. From this it follows that if $B \in \text{ins}\,\angle PAQ$, then $\overrightarrow{AB}$ intersects $\mathcal{L}$ at N such that P–N–M–K (possibly $M = K$). Let C be the last intersection of $\overrightarrow{AB}$ with $\mathcal{E}$. Now C cannot be on the A-side of $\mathcal{L}$ because $\overline{PK} \subseteq \mathcal{L}$ and $\overrightarrow{AB}$ intersects L, and C cannot be on the side of $\mathcal{L}$ opposite the A-side because then there would be corners of $\mathcal{F}$ on that side of $\mathcal{L}$, contradicting the definition of a supporting line. Hence $C \in \mathcal{L}$. $\square$

Definition CNV.14 Let $\mathcal{F}$ be a simple polygon. A supporting line $\mathcal{L}$ of $\mathcal{F}$ is **basic** if and only if $\mathcal{L}$ contains at least two corners of $\mathcal{F}$.

Remark CNV.15 (A) By Theorem CNV.13(A) every polygon has at least one basic supporting line.

(B) Let $\mathcal{L}$ be a basic supporting line of the simple polygon $\mathcal{F}$. The set $\mathcal{C}$ of corners of $\mathcal{F}$ belonging to $\mathcal{L}$ is finite, and may be ordered by Definition ORD.1. Hence there exist two (distinct) corners, P and Q, of $\mathcal{F}$ on $\mathcal{L}$ such that all other corners of $\mathcal{F}$ on $\mathcal{L}$ are between P and Q.

Definition CNV.16 The corners P and Q of the above remark are the **extremal** corners of $\mathcal{F}$ with respect to the basic supporting line $\mathcal{L}$.

Thus all the corners (other than extremal corners) of $\mathcal{F}$ lying on a basic supporting line $\mathcal{L}$ are between the extremal corners.

Theorem CNV.17 *If $\mathcal{F}$ is a simple polygon, $\mathcal{L}$ is a basic supporting line of $\mathcal{F}$, and V is an extremal corner of $\mathcal{F}$ with respect to $\mathcal{L}$, then V belongs to exactly one other basic supporting line $\mathcal{M}$ of $\mathcal{F}$, and V is an extremal corner of $\mathcal{F}$ with respect to $\mathcal{M}$.*

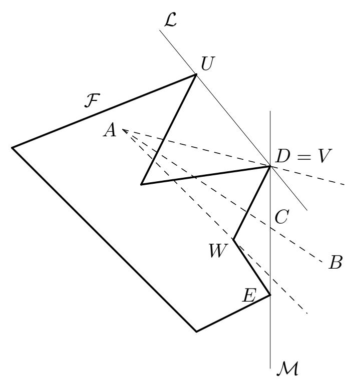


Fig. 6.27 For proof of Theorem CNV.17(1).

Proof. See Figure 6.27. (I: Existence of $\mathcal{M}$) Let U be any corner of $\mathcal{F}$ on $\mathcal{L}$ different from V and let A be any member of ins $\mathcal{F}$. By Theorem CNV.12 there exists a corner W of $\mathcal{F}$ such that U and W are on opposite sides of $\overleftrightarrow{AV}$ and there is no corner of $\mathcal{F}$ which belongs to ins $\angle VAW$.

Let B be any member of $\operatorname{ins}\angle VAW$, $\mathcal{D}$ be the set of segments joining corners of $\mathcal{F}$, $\mathcal{E}$ be the union of the members of $\mathcal{D}$, C be the last intersection of $\mathcal{E}$ with $\overrightarrow{AB}$; and let $\overline{DE}$ be the member of $\mathcal{D}$ such that D–C–E.

Suppose the notation has been chosen so that E is the endpoint of $\overline{DE}$ on the W-side of $\overleftrightarrow{AB}$. Therefore D lies on the V-side of $\overleftrightarrow{AB}$ and $D \notin \operatorname{ins}\angle VAW$. If $D \neq V$, there are three possibilities:

(i) $D \in \mathcal{L} \cap U$-side of $\overleftrightarrow{VA} \subseteq A$-side of $\overleftrightarrow{VE}$;

(ii) $D \in A$-side of $\mathcal{L} \cap U$-side of $\overleftrightarrow{VA} \subseteq A$-side of $\overleftrightarrow{VE}$; or

(iii) V–D–A.

In any of these cases, $\overrightarrow{ED} \subseteq A$-side of $\overleftrightarrow{VE}$ so C would lie between A and the intersection of $\overline{VE}$ with $\overrightarrow{AB}$, contradicting the definition of C. Thus $D = V$. By Theorem CNV.13(A) $\overleftrightarrow{VE} = \mathcal{M}$ is a supporting line of $\mathcal{F}$ and since $V \in \mathcal{M}$ it is a basic supporting line different from $\mathcal{L}$.

(II: V is an extremal corner) Since no corner of $\mathcal{F}$ belongs to $\mathcal{H}$, V is an extremal corner of $\mathcal{F}$ with respect to $\mathcal{M}$.

(III: Uniqueness of $\mathcal{M}$) All of the corners of $\mathcal{F}$ not on $\mathcal{L}$ or on $\mathcal{M}$ lie in $\operatorname{ins}\angle UVE$. If $\mathcal{N}$ were a line containing V and some other corner not on $\overrightarrow{VU}$ or $\overrightarrow{VE}$, that corner would belong to $\operatorname{ins}\angle UVE$ and U and E would be on opposite sides of $\mathcal{N}$, which could not be a supporting line. $\square$

Definition CNV.18 Let $\mathcal{F}$ be a simple polygon. A corner V of $\mathcal{F}$ is **normal** if and only if V is an extremal corner of $\mathcal{F}$ with respect to some basic supporting line of $\mathcal{F}$.

Remark CNV.19 (A) By Theorem CNV.17 and Remark CNV.15(A) every polygon has at least three normal corners and at least three basic supporting lines.

(B) If a simple polygon $\mathcal{G}$ is rotund (*i.e.*, $\operatorname{ins}\mathcal{G}$ is convex), then every edge of $\mathcal{G}$ is contained in a basic supporting line and every corner of $\mathcal{G}$ is normal.

Theorem CNV.20 *Let $\mathcal{F}$ be any simple polygon. There is a simple polygon $\mathcal{G}$ whose corners are the normal corners of $\mathcal{F}$, every edge of $\mathcal{G}$ is contained in a basic supporting line of $\mathcal{F}$, and every basic supporting line of $\mathcal{F}$ contains an edge of $\mathcal{G}$.*

(A) *The polygon $\mathcal{G}$ is rotund, and* $\operatorname{enc}\mathcal{G} = \operatorname{coh}\mathcal{G}$.

(B) *There is no simple polygon different from $\mathcal{G}$ whose corners are the normal corners of $\mathcal{F}$.*

(C) *Every corner of $\mathcal{F}$ not on $\mathcal{G}$ belongs to* $\operatorname{ins}\mathcal{G}$, *and* $\mathcal{F} \subseteq \operatorname{enc}\mathcal{G}$.

(D) $\operatorname{coh}\mathcal{F} = \operatorname{enc}\mathcal{G}$.

Proof. by Remark CNV.15(B) every basic supporting line of $\mathcal{F}$ contains exactly two normal corners. By Theorem CNV.17 every normal corner is contained in exactly two basic supporting lines.

Choose any normal corner of $\mathcal{F}$ and call it X_1. There are exactly two basic supporting lines containing X_1. Let $\mathcal{L}_1$ be either one of them and let X_2 be the other normal corner on $\mathcal{L}_1$. Then there is exactly one other basic supporting line containing X_2; name this line $\mathcal{L}_2$; $\mathcal{L}_2$ contains exactly one other normal corner; name this corner X_3. Continuing in this manner we define a mapping X of some set $[1; n]$ onto the set of all normal corners of $\mathcal{F}$.

Let $\mathcal{L}_j$ be the basic supporting line containing the normal corners X_j and X_{j+1}, $\mathcal{L}_k$ the basic supporting line containing the normal corners X_k and X_{k+1}, and suppose $\overline{X_jX_{j+1}}$ and $\overline{X_kX_{k+1}}$ intersect at some point C such that X_j–C–X_{j+1} and X_k–C–X_{k+1}. Then X_k and X_{k+1} are on opposite sides of $\mathcal{L}_j$, which by Theorem CNV.8 contradicts the fact that it is a supporting line of $\mathcal{F}$. Hence no such C exists. If, on the other hand, we were to have X_j–X_k–X_{j+1}, then these three normal corners would be collinear and this is impossible by definition of extremal point.

This shows that $\bigcup_{k=1}^{n} \overline{X_kX_{k+1}}$ is a simple polygon $\mathcal{G}$. That it has the properties claimed for it in (A) follows from the manner in which it was constructed.

(A) Since each edge of $\mathcal{G}$ is contained in a basic supporting line $\mathcal{L}$ of $\mathcal{F}$ it follows by Theorem CNV.8 that all of the corners of $\mathcal{F}$ except those on $\mathcal{L}$ lie on the same side of $\mathcal{L}$, so $\mathcal{G}$ is rotund. By CNV.11(E), $\operatorname{enc}\mathcal{G} = \operatorname{coh}\mathcal{G}$.

(B) The assertion follows immediately from the more general fact that a rotund polynomial is completely determined by its corners. We state this here as a lemma:

Lemma *Given a set of corners for a rotund polygon $\mathcal{G} = \langle X_1, X_2, \ldots, X_n\rangle$, there is no simple polygon different from $\mathcal{G}$ having the same set of corners.*

Proof. If there were a simple polygon $\mathcal{G}'$ having the same set of corners as $\mathcal{G}$, there would be a corner X_j of $\mathcal{G}$ (with $\overline{X_{j-1}X_j}$ and $\overline{X_jX_{j+1}}$ its adjacent edges in $\mathcal{G}$), such that one of the edges of $\mathcal{G}'$ containing X_j would be $\overline{X_jX_k}$ where $X_k \neq X_{j+1}$ and $X_k \neq X_{j-1}$.

Then by Theorem CNV.3(D) $X_k \in \operatorname{ins}\angle X_{j-1}X_jX_{j+1}$ so X_{j+1} and X_{j-1} are on opposite sides of $\overleftrightarrow{X_jX_k}$. Since X_{j-1} and X_{j+1} are both corners of $\mathcal{G}'$, there must exist a polygonal path $\mathcal{Q} = \langle\langle Y_1, \ldots, Y_{m+1}\rangle\rangle$ where $Y_1 = X_{j-1}$,

$Y_{m+1} = X_{j+1}$, each of the segments $\overline{Y_i Y_{i+1}}$ is an edge of $\mathcal{G}'$, and no segment in $\mathcal{Q}$ is $\overline{X_j X_k}$. (There is another polygonal path joining X_{j-1} and X_{j+1} in $\mathcal{G}'$ which includes $\overline{X_j X_k}$).

Let $p \in [1;m]$ be the greatest index such that Y_p belongs to the X_{j-1}-side of $\overleftrightarrow{X_j X_k}$; then $Y_{p+1} \notin \overleftrightarrow{X_j X_k}$ for otherwise there would be three collinear corners of $\mathcal{G}$, which is impossible since $\mathcal{G}$ is rotund. Thus Y_{p+1} belongs to the X_{j+1}-side of $\overleftrightarrow{X_j X_k}$, and by Axiom PSA, $\overline{Y_p Y_{p+1}} \cap \overleftrightarrow{X_j X_k} = \{C\}$ for some point C. By Theorem CNV.7, $\operatorname{enc} \mathcal{G}$ is convex, so that $\overline{Y_p Y_{p+1}} \subseteq \operatorname{enc} \mathcal{G}$ and hence $C \in \operatorname{enc} \mathcal{G}$, and thus $C \in \overline{X_j X_k}$, which is impossible since $\mathcal{G}'$ is simple. Therefore no such polygon $\mathcal{G}'$ as postulated above can exist. □

We return now to the proof of Theorem CNV.20.

(C) Let $\mathcal{L}_j$ be the basic supporting line containing the normal corners X_j and X_{j+1}, and let $\mathcal{H}_j$ be the side of $\mathcal{L}_j$ that contains all corners of $\mathcal{F}$ other than X_j and X_{j+1}. Then $\operatorname{ins} \mathcal{G} = \bigcap_{j=1}^{n} \mathcal{H}_j$ contains all the corners of $\mathcal{F}$ that are not on $\mathcal{G}$, and $\operatorname{enc} \mathcal{G} = \operatorname{ins} \mathcal{G} \cup \mathcal{G}$ contains all corners of $\mathcal{F}$. Since $\operatorname{enc} \mathcal{G}$ is convex, it contains every edge of $\mathcal{F}$.

(D) Continuing from part (C): from $\mathcal{F} \subseteq \operatorname{enc} \mathcal{G}$ and Theorem CNV.11(A), $\operatorname{coh} \mathcal{F} \subseteq \operatorname{coh}(\operatorname{enc} \mathcal{G}) = \operatorname{enc} \mathcal{G}$ since $\operatorname{enc} \mathcal{G}$ is convex. Now let $\mathcal{H}$ be any convex set containing $\mathcal{F}$. Every corner of $\mathcal{G}$ is a corner of $\mathcal{F}$, so $\operatorname{enc} \mathcal{G} = \operatorname{coh} \mathcal{G} \subseteq \mathcal{H}$ (by Definition CNV.10) and hence $\operatorname{enc} \mathcal{G} \subseteq \operatorname{coh} \mathcal{F}$. □

Definition CNV.21 (A) A convex subset $\mathcal{E}$ of a plane is **bounded** if and only if for each line $\mathcal{L}$, $\mathcal{E} \cap \mathcal{L}$ is contained in a segment (that is, a *bounded* set) of $\mathcal{L}$.

(B) A subset $\mathcal{E}$ of a plane is **bounded** if and only if there exists a bounded convex subset of the plane containing $\mathcal{E}$.

Theorem CNV.22 (Proof of Theorem JCT.1, part (C)) *Let $\mathcal{F}$ be a simple polygon. Then $\mathcal{F}$ and* $\operatorname{ins} \mathcal{F}$ *are bounded but* $\operatorname{out} \mathcal{F}$ *is not bounded.*

Proof. By Theorem CNV.20 there exists a rotund polygon $\mathcal{G}$ such that $\operatorname{coh} \mathcal{F} = \operatorname{enc} \mathcal{G}$. Since $\mathcal{G}$ is rotund, by Theorem CNV.5 $\mathcal{G}$ and $\operatorname{enc} \mathcal{G}$ are bounded. By Theorem CNV.11(D) $\mathcal{F}$ and $\operatorname{ins} \mathcal{F}$ are subsets of $\operatorname{coh} \mathcal{F}$ and hence are bounded.

Let $\overline{AB}$ be any edge of $\mathcal{G}$, C be any point on the side of $\overleftrightarrow{AB}$ opposite the side containing the corners of $\mathcal{G}$ different from A and B, and D be any point such that A–C–D. By the extension property of betweenness (*cf* Definition

IB.1(B.3) and Definitions IB.3 and IB.4) no segment contains $\overrightarrow{CD}$, and hence no segment contains $\operatorname{out}\mathcal{F} \cap \overleftrightarrow{CD}$ because $\overrightarrow{CD} \subseteq \operatorname{out}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$. Thus $\operatorname{out}\mathcal{F}$ is unbounded. $\square$

Remark CNV.22.1 It is quite easy to prove that any simple polygon $\mathcal{F}$ is bounded. Let $\mathcal{L}$ be any line that intersects $\mathcal{F}$; let $P \in \mathcal{L}$ be a point which is not on $\mathcal{F}$ and let Q and Q' be points on $\mathcal{L}$ such that Q–P–Q'; first order the line so that $P < Q$; by Theorem PLGN.4(A) there exists a first point C and a last point D in $\overrightarrow{PQ} \cap \mathcal{L}$; similarly, we may order the line so that $P < Q'$, and by the same theorem there exists a first point C' and a last point D' in $\overrightarrow{PQ'} \cap \mathcal{L}$. Moreover, $\mathcal{F} \cap \mathcal{L} \subseteq \overline{DD'}$, so that $\mathcal{F}$ is bounded.

This, however, does not prove that $\operatorname{ins}\mathcal{F}$ is bounded, although it seems intuitively obvious. But then, the Jordan Curve Theorem itself is intuitively obvious.

Theorem CNV.23 *Let $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ be a simple polygon.*

(A) *The inside* $\operatorname{ins}\mathcal{F}$ *is nonconvex if and only if there exist corners A and B of $\mathcal{F}$ such that $\overleftrightarrow{AB}$ is a supporting line of $\mathcal{F}$ and every point on this line between A and B belongs to* $\operatorname{out}\mathcal{F}$.

(B) *Suppose* $\operatorname{ins}\mathcal{F}$ *is nonconvex. Let A and B be the corners whose existence is assured by part (A). Re-index $\mathcal{F}$ (if necessary) so that $X_i = A$, $X_j = B$, and X_{i-1} is on the side of $\overleftrightarrow{X_i X_{i+1}}$ opposite B, where $\{i, j\} \subseteq [1; n]$ and $i < j$. For simplicity of notation, we will use A and B in place of X_i and X_j, respectively.*

Let
$$\mathcal{G} = \left(\bigcup\nolimits_{k=j-n}^{i-1} \overline{X_k X_{k+1}}\right) \cup \overline{AB},$$
and
$$\mathcal{H} = \left(\bigcup\nolimits_{k=i}^{j-1} \overline{X_k X_{k+1}}\right) \cup \overline{AB}.$$
Then $\mathcal{G}$ and $\mathcal{H}$ are simple polygons,
$$\begin{aligned} &\operatorname{ins}\mathcal{H} \subseteq \operatorname{ins}\mathcal{G},\ \operatorname{ins}\mathcal{F} \subseteq \operatorname{ins}\mathcal{G}, \\ &\operatorname{ins}\mathcal{H} \subseteq \operatorname{out}\mathcal{F},\ \operatorname{out}\mathcal{G} \subseteq \operatorname{out}\mathcal{F},\ \textit{and} \\ &\operatorname{ins}\mathcal{G} \setminus ((\mathcal{H} \setminus \overline{AB}) \cup \operatorname{ins}\mathcal{H}) \subseteq \operatorname{ins}\mathcal{F}, \end{aligned}$$
that is to say,
$$\operatorname{ins}\mathcal{G} \cap \operatorname{out}\mathcal{H} \subseteq \operatorname{ins}\mathcal{F}.$$
Furthermore, $\operatorname{out}\mathcal{F} = \operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \overline{AB}$.

Proof. (A) Suppose $\operatorname{ins}\mathcal{F}$ is nonconvex. By Theorem CNV.20 there exists a simple polygon $\mathcal{T}$ such that $\operatorname{coh}\mathcal{F} = \operatorname{enc}\mathcal{T}$ and every corner of $\mathcal{T}$ is a corner of

$\mathcal{F}$. Since $\operatorname{ins}\mathcal{F}$ is not convex, by Theorem CNV.11(E) $\operatorname{enc}\mathcal{T} = \operatorname{coh}\mathcal{F} \neq \operatorname{enc}\mathcal{F}$, so $\operatorname{enc}\mathcal{F}$ is a proper subset of $\operatorname{enc}\mathcal{T}$. Since every edge of $\mathcal{T}$ is contained in a supporting line of $\mathcal{F}$, $\mathcal{T}$ contains no points of $\operatorname{ins}\mathcal{F}$. Hence $\mathcal{T}$ contains only points of $\mathcal{F}$ and $\operatorname{out}\mathcal{F}$. Suppose it contains only points of $\mathcal{F}$. Then by Remark PLGN.6(F) $\mathcal{F} = \mathcal{T}$, $\operatorname{ins}\mathcal{F} = \operatorname{ins}\mathcal{T}$, and hence $\operatorname{ins}\mathcal{F}$ is convex which contradicts our original assumption, so that $\mathcal{T}$ must contain some points of $\operatorname{out}\mathcal{F}$.

Let $\overline{CD}$ be an edge of $\mathcal{T}$ which contains a point E in $\operatorname{out}\mathcal{F}$. Then the set of corners of $\mathcal{F}$ belonging to $\overline{CE}$ can be ordered by Definition ORD.1 with $C < D$. Let A be the corner of $\mathcal{F}$ on $\overline{CE}$ closest to E by this ordering.

Similarly, let B be the corner of $\mathcal{F}$ on $\overline{ED}$ which is closest to E. Then every point between A and B belongs to $\operatorname{out}\mathcal{F}$.

Conversely, if there exist corners A and B of $\mathcal{F}$ such that every point between A and B belongs to $\operatorname{out}\mathcal{F}$, then $\operatorname{enc}\mathcal{F}$ is nonconvex and by Theorem CNV.7 $\operatorname{ins}\mathcal{F}$ is nonconvex.

(B) A routine check of the definition of a simple polygon shows that $\mathcal{G}$ and $\mathcal{H}$ satisfy the definition, inasmuch as $\overline{AB}$ does not intersect $\mathcal{F}$. It should be noted, however, that if B lies on $\overleftrightarrow{X_{i-1}A}$, then A is not a corner of $\mathcal{G}$ and $\overline{X_{i-1}A}$ is an edge of $\mathcal{G}$. Likewise, if A lies on $\overleftrightarrow{BX_{j+1}}$, then B is not a corner of $\mathcal{G}$ and $\overline{AX_{j+1}}$ is an edge of $\mathcal{G}$. And if both these are true, neither A nor B is a corner of $\mathcal{G}$ and $\overline{X_{i-1}X_{j+1}}$ is an edge of $\mathcal{G}$. Note also that it is not possible, given the definition of A and B, for either X_{i+1} or X_{j-1} to be collinear with A and B, for that would force $A = X_{i+1}$ or $B = X_{j-1}$ or both.

Since $\mathcal{G} \cap (\mathcal{H} \setminus \overline{AB}) = \emptyset$ and $\mathcal{H} \setminus \overline{AB}$ is polygonally connected, by Theorem SEP.13(A) either $\mathcal{H} \setminus \overline{AB} \subseteq \operatorname{ins}\mathcal{G}$ or $\mathcal{H} \setminus \overline{AB} \subseteq \operatorname{out}\mathcal{G}$.

Now let Z be a point on the side of $\overleftrightarrow{AB}$ (a supporting line) opposite the corners of $\mathcal{F}$. Then $\overline{AZ} \cap \mathcal{G} = \emptyset$; then by Theorem PLGN.13 there exists a point $D \in \overline{AZ}$ such that every edge of $\mathcal{G}$ that intersects $\overline{X_{i+1}D}$ also intersects $\overline{X_{i+1}A}$; $\overline{AX_{i-1}}$ intersects $\overline{X_{i+1}A}$ but lies on the side of $\overleftrightarrow{X_{i+1}A}$ opposite D, so cannot intersect $\overline{X_{i+1}D}$; and $\overline{AB}$ is an edge of $\mathcal{G}$ that intersects both $\overline{X_{i+1}D}$ and $\overline{X_{i+1}A}$. Hence the ray $\overrightarrow{X_{i+1}D}$ has only one intersection with $\mathcal{G}$, and X_{i+1} has odd parity. It follows that $\mathcal{H} \setminus \overline{AB} \subseteq \operatorname{ins}\mathcal{G}$.

Then by Theorem SEP.15, case (2), we have

$$\begin{aligned}&\mathcal{H} \setminus \overline{AB} \subseteq \operatorname{ins}\mathcal{G} \text{ and } \mathcal{G} \setminus \overline{AB} \subseteq \operatorname{out}\mathcal{H},\\ &\operatorname{out}\mathcal{G} \text{ is a proper subset of } \operatorname{out}\mathcal{H}, \text{ and}\\ &\operatorname{ins}\mathcal{H} \text{ is a proper subset of } \operatorname{ins}\mathcal{G}.\end{aligned}$$

Now let P be any member of $\operatorname{out}\mathcal{G}$ and let Q be any member of $\operatorname{out}\angle APB$ such that $\overrightarrow{PQ}$ contains no corners of $\mathcal{F}$, $\mathcal{G}$, or $\mathcal{H}$. $\overrightarrow{PQ}$ intersects $\mathcal{G}$ in exactly the points where it intersects $\mathcal{F}$ and $\mathcal{H}$; the number of intersections with $\mathcal{G}$ is even, and since $\operatorname{out}\mathcal{G} \subseteq \operatorname{out}\mathcal{H}$ the number of intersections with $\mathcal{H}$ is even, and hence the number of intersections with $\mathcal{F}$ is even. Thus $\operatorname{out}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$. Taking complements with respect to the plane, this last relation gives us $\mathcal{F} \cup \operatorname{ins}\mathcal{F} \subseteq \mathcal{G} \cup \operatorname{ins}\mathcal{G}$, so that $\operatorname{ins}\mathcal{F} \subseteq \mathcal{G} \cup \operatorname{ins}\mathcal{G}$. But $\mathcal{G} \subseteq \mathcal{F} \cup \overline{AB}$, and both $\mathcal{F}$ and $\overline{AB}$ are disjoint from $\operatorname{ins}\mathcal{F}$, so that $\operatorname{ins}\mathcal{F} \cap \mathcal{G} = \emptyset$. Thus $\operatorname{ins}\mathcal{F} \subseteq \operatorname{ins}\mathcal{G}$.

Let R be any member of $\operatorname{ins}\mathcal{H}$ and S be any member of $\operatorname{out}\angle ARB$ such that $\mathring{\overrightarrow{RS}}$ contains no corners of $\mathcal{F}$, $\mathcal{G}$, or $\mathcal{H}$. Then $\mathring{\overrightarrow{RS}} \cap \overline{AB} = \emptyset$ and $\overrightarrow{RS}$ intersects $\mathcal{G}$ in exactly the points where it intersects $\mathcal{F}$ and $\mathcal{H}$. The number of intersections with $\mathcal{H}$ is odd, and the number of intersections with $\mathcal{G}$ is odd because $\operatorname{ins}\mathcal{H} \subseteq \operatorname{ins}\mathcal{G}$, so the number of intersections with $\mathcal{F}$ must be even. Hence $\operatorname{ins}\mathcal{H} \subseteq \operatorname{out}\mathcal{F}$.

Let T be any point of $\operatorname{ins}\mathcal{G} \setminus ((\mathcal{H} \setminus \overline{AB}) \cup \operatorname{ins}\mathcal{H})$ and U be any point of $\operatorname{out}\angle ATB$ such that $\overrightarrow{TU}$ contains no corners of $\mathcal{F}$, $\mathcal{G}$, or $\mathcal{H}$. Then since $T \in \operatorname{ins}\mathcal{G}$, $\overrightarrow{TU}$ has an odd number of intersections with $\mathcal{G}$.

Since $\overleftrightarrow{AB}$ is a supporting line of $\mathcal{G}$, $\operatorname{ins}\mathcal{G} \cap \overline{AB} = \emptyset$ and $T \notin \overline{AB}$; by its definition, $T \notin ((\mathcal{H} \setminus \overline{AB}) \cup \operatorname{ins}\mathcal{H})$, so $T \in \operatorname{out}\mathcal{H}$, and has an even number of intersections with $\mathcal{H}$. Thus T has an odd number of intersections with $\mathcal{F}$, $T \in \operatorname{ins}\mathcal{F}$, and $\operatorname{ins}\mathcal{G} \setminus ((\mathcal{H} \setminus \overline{AB}) \cup \operatorname{ins}\mathcal{H}) \subseteq \operatorname{ins}\mathcal{F}$. This reduces to $\operatorname{ins}\mathcal{G} \cap \operatorname{out}\mathcal{H} \subseteq \operatorname{ins}\mathcal{F}$ since $\operatorname{out}\mathcal{H}$ is the complement of $(\mathcal{H} \setminus \overline{AB}) \cup \operatorname{ins}\mathcal{H})$.

The final assertion follows from taking complements of the relation

$$\operatorname{ins}\mathcal{G} \setminus ((\mathcal{H} \setminus \overline{AB}) \cup \operatorname{ins}\mathcal{H}) \subseteq \operatorname{ins}\mathcal{F}$$

to get

$$\mathcal{F} \cup \operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \mathcal{G} \cup (\mathcal{H} \setminus \overline{AB}),$$

and since

$$\mathcal{G} \cup (\mathcal{H} \setminus \overline{AB}) = \mathcal{F} \cup \mathring{\overline{AB}},$$

this can be written

$$\mathcal{F} \cup \operatorname{out}\mathcal{F} \subseteq \mathcal{F} \cup \operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \mathring{\overline{AB}}.$$

Now

$$\mathcal{F} \cap \operatorname{out}\mathcal{F} = \mathcal{F} \cap (\operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \mathring{\overline{AB}}) = \emptyset,$$

so we have

$$\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \mathring{\overline{AB}}.$$

But

$$\operatorname{out}\mathcal{G} \subseteq \operatorname{out}\mathcal{F} \text{ and } \operatorname{ins}\mathcal{H} \subseteq \operatorname{out}\mathcal{F},$$

and by hypothesis

$$\mathring{\overline{AB}} \subseteq \operatorname{out}\mathcal{F},$$

so that $\operatorname{out}\mathcal{F} = \operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \mathring{\overline{AB}}$. $\square$

Remark CNV.24 Let $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ be a simple polygon. The edges $\overline{X_{j-1}X_j}$ and $\overline{X_jX_{j+1}}$ intersect at X_j. Let M and N be points such that M–X_j–X_{j+1} and N–X_j–X_{j-1}, and let $A \in \operatorname{ins}\angle NX_jX_{j+1}$ Then the line $\overleftrightarrow{AX_j}$ contains no point of $\operatorname{ins}\angle X_{j-1}X_jX_{j+1}$. Let A–X_j–B. Let A' be the first point of intersection of $\overrightarrow{X_jA}$ with $\mathcal{F}$ (if there is no intersection, let A' be any point of $\overrightarrow{X_jA}$), and let B' the first point of intersection of $\overrightarrow{X_jB}$ with $\mathcal{F}$,(again, if there is no intersection, let B' be any point of $\overrightarrow{X_jB}$). Finally, let P and Q be such that A'–P–X_j–Q–B'.

Then by Theorem SEP.4(C), since $\overline{PQ} \cap \operatorname{ins}\angle X_{j-1}X_jX_{j+1} = \emptyset$ both P and $Q \in \operatorname{out}\angle X_{j-1}X_jX_{j+1}$, and X_{j-1} and X_{j+1} are on the same side of $\overleftrightarrow{PQ}$ (if they were on different sides, then $\overline{PQ}$ would contain points of $\operatorname{ins}\angle X_{j-1}X_jX_{j+1}$, a contradiction). Also, by SEP.4, P and Q have the same parity relative to $\mathcal{F}$.

Definition CNV.25 Given the setup of Remark CNV.24 above, the corner X_j is **regular** iff the parity of P and Q is even, and **irregular** if their parity is odd.

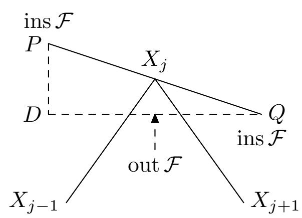


Fig. 6.28 For Remark CNV.26.

Remark CNV.26 (A) If a simple polygon $\mathcal{F}$ has an irregular corner X_j, then $\operatorname{ins}\mathcal{F}$ is nonconvex. To see this, let P and Q be points such that P–X_j–Q and both P and $Q \in \operatorname{ins}\mathcal{F}$; let D be a member of the X_{j-1}-side (X_{j+1}-side) of $\overleftrightarrow{PQ}$ be such that $\overline{PD} \cap = \emptyset$; then $D \in \operatorname{ins}\mathcal{F}$ and by Theorem PSH.6 $\overline{DQ}$ intersects both $\overline{X_jX_{j+1}}$ and $\overline{X_{j-1}X_j}$. By Theorem SEP.7(A), $\overline{DQ}$ contains a point of $\operatorname{out}\mathcal{F}$, so that $\operatorname{ins}\mathcal{F}$ is not convex.

(B) By part (A), if $\operatorname{ins}\mathcal{F}$ is convex, $\mathcal{F}$ has no irregular corners. By Theorem CNV.4(B), if $\mathcal{F}$ is rotund, $\operatorname{ins}\mathcal{F}$ is convex, hence every corner of a rotund polygon is regular. See Figure 6.28.

Theorem CNV.27 *Every simple polygon $\mathcal{F}$ has at least three regular corners.*

Proof. By Remark CNV.19(A) it suffices to show that every normal corner of $\mathcal{F}$ is a regular corner of $\mathcal{F}$. Let A be a normal corner of $\mathcal{F}$. By Theorem CNV.20 there exists a simple rotund polygon $\mathcal{G}$ such that A is a corner of $\mathcal{G}$; let $\overline{AB}$ and $\overline{AC}$ be edges of $\mathcal{G}$, and let D and E be the corners of $\mathcal{F}$ such that $\overline{AD}$ and $\overline{AE}$ are edges of $\mathcal{F}$. Then D and E both belong to $\angle BAC \cup \operatorname{ins} \angle BAC$. Since A is a regular corner of $\mathcal{G}$ ($\mathcal{G}$ is rotund—see CNV.26(B)) there exist points P and $Q \in \operatorname{out} \mathcal{G}$ such that A is between P and Q, $\overline{PQ} \cap \mathcal{G} = \{\mathcal{A}\}$, and P and Q both belong to $\operatorname{out} \angle BAC$. Since by Theorem PSH.41(D), $\operatorname{out} \angle BAC \subseteq \operatorname{out} \angle DAE$, both P and Q belong to $\operatorname{out} \mathcal{F}$. Hence A is a regular corner of $\mathcal{F}$. $\square$

Theorem CNV.28 *Let $\mathcal{F}$ be a simple polygon. Then $\mathcal{F}$ has an irregular corner if and only if* $\operatorname{ins} \mathcal{F}$ *is nonconvex.*

Proof. By Remark CNV.26, if $\mathcal{F}$ has an irregular corner, $\operatorname{ins} \mathcal{F}$ is not convex. Conversely, suppose $\operatorname{ins} \mathcal{F}$ is nonconvex and let X_i, X_j, $\mathcal{G}$ and $\mathcal{H}$ be as in Theorem CNV.23(B). By Theorem CNV.27 $\mathcal{H}$ has a regular corner X_k different from X_i and X_j so there exist points P and Q in $\operatorname{out} \mathcal{H}$ such that P–X_k–Q, $\overline{PQ} \cap \mathcal{F} = \{X_k\}$ and P and Q are both members of $\operatorname{out} \angle X_{k-1} X_k X_{k+1}$. Since $X_k \in ((\mathcal{H} \setminus \overline{X_i X_j}) \subseteq \operatorname{ins} \mathcal{G}$, and P and Q can be chosen so that $\overline{PQ} \cap \overline{X_i X_j} = \emptyset$, it follows that both P and Q belong to $\operatorname{ins} \mathcal{G}$. Hence by the last inclusion in Theorem CNV.23(B) both P and Q belong to $\operatorname{ins} \mathcal{F}$ and X_k is an irregular corner of $\mathcal{F}$. $\square$

Theorem CNV.29 (A) *Let $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ be a simple polygon with $n \geq 4$. Then there exist corners A and B of $\mathcal{F}$ such that $\overline{AB} \subseteq \operatorname{ins} \mathcal{F}$.*

(B) *If $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ is any simple polygon with $n \geq 4$, which has two corners X_i and X_j with $1 \leq i < j \leq n$, such that $\overline{X_i X_j} \subseteq \operatorname{ins} \mathcal{F}$, let*

$$\mathcal{G} = \left(\bigcup_{k=j-n}^{i-1} \overline{X_k X_{k+1}} \right) \cup \overline{X_i X_j},$$

and

$$\mathcal{H} = \left(\bigcup_{k=i}^{j-1} \overline{X_k X_{k+1}} \right) \cup \overline{X_i X_j}.$$

Then $\mathcal{G}$ and $\mathcal{H}$ are simple polygons and $\operatorname{ins} \mathcal{G} \subseteq \operatorname{out} \mathcal{H}$, $\operatorname{ins} \mathcal{H} \subseteq \operatorname{out} \mathcal{G}$, and $\operatorname{ins} \mathcal{F} = \operatorname{ins} \mathcal{G} \cup \operatorname{ins} \mathcal{H} \cup \overline{X_i X_j}$. Note that k may take on values $k < 1$ or $k > n$.

Proof. See Figure 6.29.

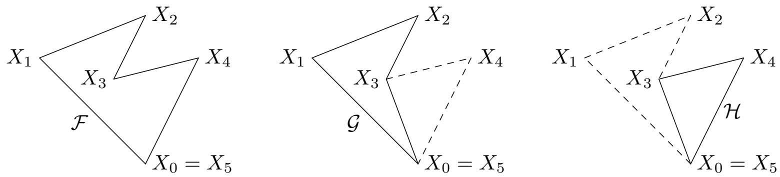
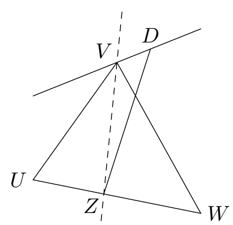


Fig. 6.29 For Theorem CNV.29, where $X_i = X_3$ and $X_j = X_5$.

(A) Let V be a regular corner of $\mathcal{F}$ and U and W be the corners of $\mathcal{F}$ such that $\overline{UV}$ and $\overline{WV}$ are edges of $\mathcal{F}$. Let $\mathcal{E} = \operatorname{enc}\langle U, V, W\rangle \setminus (\overline{UV} \cup \overline{WV})$, and $\mathcal{D}$ be the set of corners of $\mathcal{F}$ belonging to $\mathcal{E}$.

(Case I) See Figure 6.30 below. If $\mathcal{D} = \emptyset$, neither the inside of the triangle $\langle U, V, W\rangle$, or $\overline{UW}$ contains a corner of $\mathcal{F}$. Moreover, $\overline{UW}$ is not a subset of any edge of $\mathcal{F}$ because $\mathcal{F}$ has at least 4 edges. Then by Corollary PLGN.12.1, any edge that intersects $\overline{UW}$ must intersect either $\overline{UV}$, $\overline{WV}$, or $\{V\}$. But there are no edges that intersect $\overline{UV}$ and $\overline{WV}$, other than themselves, and we know they don't intersect $\overline{UW}$. Therefore no edge $\mathcal{E}$ can intersect $\overline{UW}$, and by Theorem SEP.4(A) all points of $\overline{UW}$ have the same parity.



$D \in \operatorname{out}\mathcal{F}$ hence $Z \in \operatorname{ins}\mathcal{F}$
and $\overline{UW} \subseteq \operatorname{ins}\mathcal{F}$

Fig. 6.30 For proof of Theorem CNV.29 Case (I).

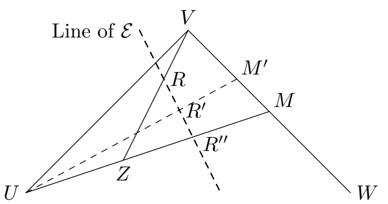
Since V is a regular corner there are points P and Q such that $\overline{PQ} \cap \mathcal{F} = \{V\}$ and both $\overline{PV}$ and $\overline{VQ} \subseteq \operatorname{out}\mathcal{F}$. Let $Z \in \overline{UW}$; without loss of generality we may assume that P is on the U-side of $\overleftrightarrow{ZV}$ and Q is on the W-side.

By Theorem PLGN.13 there exists a point $D \in \overline{VQ}$ such that every edge that intersects $\overline{ZD}$ must also intersect $\overline{ZV}$. There are two edges that do so: $\overline{UV}$ and $\overline{WV}$; $\overline{UV}$ is a subset of the side of $\overleftrightarrow{ZV}$ opposite W, so $\overline{UV}$ cannot intersect $\overline{ZD}$; therefore $\overline{WV}$ is the only edge that can intersect $\overline{ZD}$, and it does so in a singleton by the Crossbar Theorem PSH.39, because $W \in \operatorname{ins}\angle ZVQ$.

Then by Theorem SEP.4(A) and (B), the parity of Z is different from the parity of D (which is even) and hence $Z \in \operatorname{ins}\mathcal{F}$ and $\overline{UW} \subseteq \operatorname{ins}\mathcal{F}$. Let $A = U$ and $B = W$, so that $\overline{AB} \subset \operatorname{ins}\mathcal{F}$.

(Case II) If $\mathcal{D} \neq \emptyset$ let $\mathcal{C} = \{X | V\text{–}X\text{–}W$ and $\overline{UX}$ contains a corner of $\mathcal{F}\}$. Ordering $\overline{WV}$ with $V < W$ choose M to be the first element of $\mathcal{C}$, and let $Z \in \overline{UM}$ be a corner.

Claim (a) See Figure 6.31. If $\mathcal{E}$ is any edge that intersects $\overline{ZV}$ at a point R, then $\mathcal{E}$ must intersect $\overline{UV}$ or $\overline{VM}$.



Note that R' and R'' could lie on the U-side of $\overleftrightarrow{VZ}$ rather than as shown here.

Fig. 6.31 For proof of Theorem CNV.29 Case (II) Claim (a).

Note first that $\mathcal{E}$ cannot intersect $\overline{ZV}$ at two points, for then $\overline{ZV}$ would contain a corner. By Theorem PLGN.12 $\mathcal{E}$ must intersect $\overline{VM}$ or $\overline{ZM}$, but no edge can intersect $\overline{VM}$ because $\mathcal{F}$ is simple, so $\mathcal{E}$ must intersect $\overline{ZM}$ at some point R''. Let M' be such that $V\text{–}M'\text{–}M$; then $\overline{UM'}$ contains no corner and intersects $\overline{RR''}$ at some point R', by the Crossbar Theorem. We may apply Theorem PLGN.12 again to $\triangle UM'V$ to conclude that $\mathcal{E}$ must intersect $\overline{UV}$ or $\overline{VM'}$ ($\mathcal{E}$ cannot contain V because $\mathcal{F}$ is simple). This proves Claim (a).

But no edge intersects $\overline{UV}$ or $\overline{M'V} \subseteq \overline{WV}$ because $\mathcal{F}$ is simple. Therefore no edge intersects $\overline{ZV}$, and all points of $\overline{ZV}$ have the same parity.

Claim (b) $\overline{ZV} \subseteq \operatorname{ins}\mathcal{F}$. See Figure 6.32 below.

Let N be a point of $\overline{ZV}$. Without loss of generality we may assume that P is on the U-side of $\overleftrightarrow{ZV}$ and Q is on the M-side.

By Theorem PLGN.13 there exists a point $D \in \overline{VQ}$ such that every edge that intersects $\overline{ND}$ must also intersect $\overline{NV}$. There are two edges that do so: $\overline{UV}$ and $\overline{WV}$; $\overline{UV}$ is a subset of the side of $\overleftrightarrow{ZV}$ opposite D, so $\overline{UV}$ cannot intersect $\overline{ND}$; $\overline{WV}$, therefore, is the only edge that can intersect $\overline{ND}$, and it does so in a singleton by the Crossbar Theorem PSH.39, because M (and W) belong to $\operatorname{ins}\angle ZVQ$.

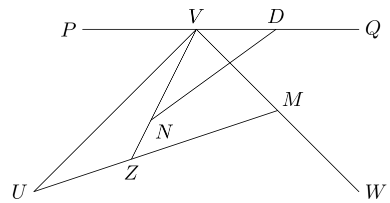


Fig. 6.32 For proof of Theorem CNV.29 Case (II) Claim (b).

Then by Theorem SEP.4(A) and (B), the parity of N is different from the parity of D (which is even) and hence $N \in \operatorname{ins}\mathcal{F}$ and $\overline{ZV} \subseteq \operatorname{ins}\mathcal{F}$. This proves Claim (b).

Let $A = Z$ and $B = V$, so that $\overline{AB} \subseteq \operatorname{ins}\mathcal{F}$.

(B) A routine check of the definition of a simple polygon shows that $\mathcal{G}$ and $\mathcal{H}$ are simple polygons. However, as in Theorem CNV.23, either X_i or X_j or both may fail to be corners of either $\mathcal{G}$ or $\mathcal{H}$. If X_i fails to be a corner of $\mathcal{G}$, then it must be a corner of $\mathcal{H}$ and vice versa. A similar statement holds for X_j.

Now $\mathcal{G}$ and $\mathcal{H}$ conveniently intersect in a segment $\overline{X_iX_j}$ so we may apply Theorem SEP.15 to them, and exactly one of its alternatives (1), (2), or (3) holds.

We first show that there exist points of $\operatorname{ins}\mathcal{H}$ that are not in $\operatorname{ins}\mathcal{G}$. Let $O \in \overline{X_iX_j}$, and choose a point R such that $R \notin \overleftrightarrow{X_iX_j}$ and $\overleftrightarrow{OR}$ contains no corner of $\mathcal{F}$. Since $\overline{X_iX_j}$ is an edge of $\mathcal{H}$, by Theorem SEP.7 there exist points P and Q such that P–O–Q and $\overline{PO} \subseteq \operatorname{out}\mathcal{H}$ and $\overline{OQ} \subseteq \operatorname{ins}\mathcal{H}$. Since $O \in \operatorname{ins}\mathcal{F}$, we may choose P and Q so that $\overline{PQ} \subseteq \operatorname{ins}\mathcal{F}$.

Let S be a point such that $S \in \operatorname{out}\angle X_iQX_j$ and $\overrightarrow{QS}$ contains no corner of $\mathcal{F}$. For this ray, the number of intersections with $\mathcal{F}$ is the total of the number of intersections with $\mathcal{G}$ and with $\mathcal{H}$.

$\overrightarrow{QS}$ has an odd number of intersections with $\mathcal{F}$ since $Q \in \operatorname{ins}\mathcal{F}$, an odd number of intersections with $\mathcal{H}$ since $Q \in \operatorname{ins}\mathcal{H}$, and hence an even number of intersections with $\mathcal{G}$, so that $Q \in \operatorname{out}\mathcal{G}$.

Therefore Q is a point of $\operatorname{ins}\mathcal{H}$ but not a point of $\operatorname{ins}\mathcal{G}$, and $\operatorname{ins}\mathcal{H} \subseteq \operatorname{ins}\mathcal{G}$ is false. Similarly, $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{H}$ is false, thus ruling out alternatives (1) and (2) of Theorem SEP.15, and alternative (3) is valid so that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{H}$ and $\operatorname{ins}\mathcal{H} \subseteq \operatorname{out}\mathcal{G}$. See Figure 6.33.

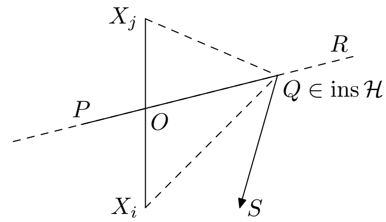


Fig. 6.33 Showing $\operatorname{ins}\mathcal{H} \not\subseteq \operatorname{ins}\mathcal{G}$.

We now show that $\operatorname{ins}\mathcal{F} \subseteq \operatorname{ins}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \overline{X_iX_j}$. If $X \in \operatorname{ins}\mathcal{F}$ and $X \notin \overline{X_iX_j}$ and $X \notin \operatorname{ins}\mathcal{H}$, it follows that $X \in \operatorname{ins}\mathcal{G}$.

Again, let S be a point such that $S \in \operatorname{out}\angle X_iXX_j$ and $\overrightarrow{XS}$ contains no corner of $\mathcal{F}$. As above, the number of intersections with $\mathcal{F}$ is the total of the number of intersections with $\mathcal{G}$ and with $\mathcal{H}$.

$\overrightarrow{XS}$ has an odd number of intersections with $\mathcal{F}$ since $X \in \operatorname{ins}\mathcal{F}$; since $X \notin \operatorname{ins} H$ (and does not belong to $\mathcal{H}$) it belongs to $\operatorname{out}\mathcal{H}$ and has an even number of intersections with $\mathcal{H}$. Hence it has an odd number of intersections with $\mathcal{G}$, so that $X \in \operatorname{ins}\mathcal{G}$. By a similar argument, if $X \in \operatorname{ins}\mathcal{F}$ and $X \notin \overline{X_iX_j}$ and $X \notin \operatorname{ins}\mathcal{G}$ then $X \in \operatorname{ins}\mathcal{H}$. Note also that it has now been proved that $\overline{PO} \subseteq \operatorname{ins}\mathcal{G}$.

Finally, we show that $\operatorname{ins}\mathcal{F} \supseteq \operatorname{ins}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \overline{X_iX_j}$. By alternative (3) of SEP.15 $\operatorname{ins}\mathcal{G} \cap \operatorname{ins}\mathcal{F} = \emptyset$. Let $X \in \operatorname{ins}\mathcal{H}$ and $X \notin \operatorname{ins}\mathcal{G}$ and $X \notin \overline{X_iX_j}$ so that $X \in \operatorname{out}\mathcal{G}$. Again, let S be a point such that $S \in \operatorname{out}\angle X_iXX_j$ and $\overrightarrow{XS}$ contains no corner of $\mathcal{F}$. Then $\overrightarrow{XS}$ intersects $\mathcal{H}$ an odd number of times, $\mathcal{G}$ an even number of times, so that it intersects $\mathcal{F}$ an odd number of times, so that $X \in \operatorname{ins}\mathcal{F}$. A similar argument shows that if $X \in \operatorname{ins}\mathcal{G}$ but does not belong to $\overline{X_iX_j}$ or to $\operatorname{ins}\mathcal{H}$, then $X \in \operatorname{ins}\mathcal{F}$. We already know that $\overline{X_iX_j} \subseteq \operatorname{ins}\mathcal{F}$, so it follows that $\operatorname{ins}\mathcal{F} = \operatorname{ins}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \overline{X_iX_j}$. $\square$

Definition CNV.30 Let $\mathcal{S}$ be a finite noncollinear set of points. (Meaning, according to Definition I.0, that there is no line $\mathcal{L}$ such that $\mathcal{S} \subseteq \mathcal{L}$.) A **supporting line** of $\mathcal{S}$ is a line $\mathcal{L}$ such that $\mathcal{S} \cap \mathcal{L} \neq \emptyset$ and all points of $\mathcal{S}$ not on $\mathcal{L}$ are on the same side of $\mathcal{L}$. A **basic supporting line** of $\mathcal{S}$ is a supporting line which has at least two members of $\mathcal{S}$ on it.

Theorem CNV.31 *Every finite noncollinear set $\mathcal{S}$ has a basic supporting line.*

Proof. Let $\mathcal{D}$ be the set of segments whose endpoints are members of $\mathcal{S}$ and let $\mathcal{E}$ be the union of the members of $\mathcal{D}$. Let A, P, and U be noncollinear members of $\mathcal{S}$. If any of the lines $\overleftrightarrow{AP}$, $\overleftrightarrow{AU}$, or $\overleftrightarrow{PU}$ is a supporting line of $\mathcal{S}$, then the proposition is true. Hence, assume none of these lines is a supporting line of $\mathcal{S}$.

If $(\text{ins}\,\angle PAU) \cap \mathcal{S} = \emptyset$, then let $Q = U$. If $(\text{ins}\,\angle PAU) \cap \mathcal{S} \neq \emptyset$, then there exists a member Q of $\text{enc}\,\angle PAU$ such that $\text{ins}\,\angle PAQ \cap \mathcal{S} = \emptyset$. Let B be any member of $\text{ins}\,\angle PAQ$ and let C be the last point of intersection of $\mathcal{E}$ and $\overrightarrow{AB}$. Then the line $\overleftrightarrow{EF}$ containing the member $\overline{EF}$ of $\mathcal{D}$ which intersects $\overrightarrow{AB}$ is a basic supporting line of $\mathcal{S}$. To show this it suffices to show that no member of $\mathcal{S}$ is on the side of $\overleftrightarrow{EF}$ opposite the A-side. If there were a member R of $\mathcal{S}$ on this side, then since R is not on $\overleftrightarrow{AB}$, either E and R are on the same side of $\overleftrightarrow{AB}$, or they are on opposite sides. Furthermore $\overline{ER}$ and $\overline{FR}$ are on the side of $\overleftrightarrow{EF}$ opposite the A-side. Hence by the Plane Separation Axiom $\overrightarrow{AB}$ intersects exactly one member of $\{\overline{ER}, \overline{FR}\}$ at M such that A–C–M. This contradicts the fact that C is the last intersection of $\mathcal{E}$ and $\overrightarrow{AB}$. $\square$

Definition CNV.32 Let $\mathcal{S}$ be a finite noncollinear set of points and let $\mathcal{L}$ be a basic supporting line of $\mathcal{S}$. Using Definition ORD.1 to order the points of $\mathcal{L}$, then by Theorem ORD.10, $\mathcal{L} \cap \mathcal{S}$ has a maximum and a minimum point. These points are the **extremal** points of $\mathcal{S}$ with respect to $\mathcal{L}$.

Theorem CNV.33 *Let $\mathcal{S}$ be a finite noncollinear set of points, $\mathcal{L}$ be a basic supporting line of $\mathcal{S}$, and V be an extremal point of $\mathcal{S}$ with respect to $\mathcal{L}$. Then V belongs to one and only one other basic supporting line $\mathcal{M}$ of $\mathcal{S}$ and V is an extremal point of $\mathcal{S}$ with respect to $\mathcal{M}$.*

Proof. (Existence of $\mathcal{M}$) Let U be any point of $\mathcal{S} \cap \mathcal{L}$ which is different from V, A be any member of $\mathcal{S}$ not on $\mathcal{L}$, and let $\mathcal{H}$ be the side of $\mathcal{L}$ opposite the A-side.

If $\overleftrightarrow{AV}$ is a supporting line of $\mathcal{S}$, let $\mathcal{M} = \overleftrightarrow{AV}$. Since $\mathcal{L}$ is a supporting line of $\mathcal{S}$, by Definition CNV.30 there is no point of $\mathcal{S}$ in $\mathcal{H}$. Therefore V is an extremal point of $\mathcal{S}$ with respect to $\mathcal{M}$.

If $\overleftrightarrow{AV}$ is not a supporting line of $\mathcal{S}$, then there exists a member W of $\mathcal{S}$ such that W and U are on opposite sides of $\overleftrightarrow{AV}$ and no point of $\mathcal{S}$ belongs to $\text{ins}\,\angle VAW$. (To see this choose any member W' of $\mathcal{S}$ where W' and U are on opposite sides of $\overleftrightarrow{AV}$, then let $\mathcal{C} = \{X | X \in \overline{VW'}$ and $\overrightarrow{AX}$ contains a point of $\mathcal{S}\}$. Ordering $\mathcal{C}$ with $V < W'$ let W be the least member of $\mathcal{C}$. Then there is no point of $\mathcal{S}$ belonging to $\operatorname{ins}\angle VAW$.)

Let $\mathcal{D}$ be the set of segments whose endpoints are members of $\mathcal{S}$ and let $\mathcal{E}$ be the union of the members of $\mathcal{D}$. Let$B$ be any member of $\operatorname{ins}\angle VAW$, C the last intersection of $\mathcal{E}$ with $\overrightarrow{AB}$, and let $\overline{DE}$ be a member of $\mathcal{D}$ such that D–C–E.

Choose the notation so that E is the endpoint of $\overline{DE}$ which is on the W-side of $\overleftrightarrow{AB}$; then D is on the V-side of $\overleftrightarrow{AB}$ and $D \notin \operatorname{ins}\angle VAW$. By the same argument as in Theorem CNV.17, $D = V$. (For once, even the notation is the same, so the arguments are letter-for-letter the same.)

As in the proof of Theorem CNV.31, $\mathcal{M} = \overleftrightarrow{VE}$ is a basic supporting line S. (This time the notation is not quite the same–substitute D for E and E for F in the original argument, and it works.)

Since by Defnition CNV.30 no point of $\mathcal{S}$ belongs to $\mathcal{H}$, V is an extremal point of V with respect to $\mathcal{S}$.

(Uniqueness of $\mathcal{M}$) Since $\mathcal{S}$ is contained in $\operatorname{enc}\angle UVE$, if $\mathcal{N}$ is a line through V and some other point of $\mathcal{S}$, then $\mathcal{S} \cap \operatorname{ins}\angle UVE$ is nonempty and $\mathcal{N}$ is not a supporting line of $\mathcal{S}$. □

Definition CNV.34 Let $\mathcal{S}$ be a finite noncollinear set of points. A point of $\mathcal{S}$ is **normal** if and only if it is an extremal point of $\mathcal{S}$ with respect to a basic supporting line of $\mathcal{S}$.

Remark CNV.35 By Theorems CNV.31 and CNV.33 every finite noncollinear set of points has at least three normal points and at least three basic supporting lines.

Theorem CNV.36 *Let $\mathcal{S}$ be a finite noncollinear set of points.*

(A) *There is a simple polygon $\mathcal{G}$ whose corners are the normal points of $\mathcal{S}$, every edge of $\mathcal{G}$ is contained in a basic supporting line of $\mathcal{S}$ and every basic supporting line of $\mathcal{S}$ contains an edge of $\mathcal{G}$.*

(B) *The polygon $\mathcal{G}$ is rotund.*

(C) *There is no simple polygon different from $\mathcal{G}$ whose corners are normal points of $\mathcal{S}$.*

(D) *Every member of $\mathcal{S}$ not on $\mathcal{G}$ belongs to* $\operatorname{ins}\mathcal{G}$, *and* $\mathcal{S} \subseteq \operatorname{enc}\mathcal{G}$.

(E) $\operatorname{coh}\mathcal{S} = \operatorname{enc}\mathcal{G}$.

Proof. A slight modification of the proof of Theorem CNV.20 is left to the reader as Exercise CNV.1. □

6.5 Connectedness (CNT)

Lemma CNT.1 *Suppose $\mathcal{F}$ is a simple polygon and $\mathcal{J}$ is a polygonal path such that $\mathcal{F} \cap \mathcal{J} = \emptyset$; by Theorem SEP.13(A) either $\mathcal{J} \subseteq \operatorname{ins}\mathcal{F}$ or $\subseteq \operatorname{out}\mathcal{F}$. Then if $\mathcal{I}$ is a closed segment $\overline{AB}$ and $\mathcal{J}$ is a simple polygonal path which is not a segment, then there exists a one-to-one mapping φ of $\mathcal{I}$ onto $\mathcal{J}$ such that the image of each endpoint of $\mathcal{I}$ is an endpoint of $\mathcal{J}$.*

Proof. Using Definition PLGN.5(B) let $\mathcal{J} = \bigcup_{k=1}^{m} \overline{Z_k Z_{k+1}}$, where m is a natural number greater than 1. Using the denseness property of a segment (*cf* Theorem PSH.22) we can find a set $\{A = W_1, W_2, \ldots, W_{m+1} = B\} \subseteq \overline{AB}$ such that for each $k \in [1;m]$, $W_k < W_{k+1}$ (where $\overline{AB}$ is ordered by ORD.1 with $A < B$). Using Theorem PSH.56, for each $k \in [1;m]$, let φ_k be a one-to-one mapping of $\overline{W_k W_{k+1}}$ onto $\overline{Z_k Z_{k+1}}$ such that $\varphi_k(W_k) = Z_k$ and $\varphi_k(W_{k+1}) = Z_{k+1}$. Let $\varphi = \bigcup_{k=1}^{m} \varphi_k$; then φ is a one-to-one mapping of $\mathcal{I}$ onto $\mathcal{J}$ such that $\varphi(W_1) = Z_1$ and $\varphi(W_{m+1}) = Z_{m+1}$. □

Theorem CNT.2 *The outside of a rotund polygon is polygonally connected.*

Proof. Let $\mathcal{F} = \langle X_1, \ldots, X_n \rangle$ be a rotund polygon and let $A \in \operatorname{ins}\mathcal{F}$. For each $i \in [1; n-1]$ define Y_i recursively as follows: let Y_1 and Y_2 be any points such that A–X_1–Y_1 and A–X_2–Y_2.

For each $i \in [3; n-1]$, if $\overrightarrow{AX_i} \cap \overleftrightarrow{Y_{i-2}Y_{i-1}}$ contains some point Z, let Y_i be any point such that A–X_i–Y_i–Z, otherwise let A–X_i–Y_i; if $\overrightarrow{AX_n} \cap \overleftrightarrow{Y_1 Y_2}$ contains some point Z, let Y_n be any point such that A–X_n–Y_n–Z, otherwise let A–X_n–Y_n.

No two rays $\overrightarrow{AX_i}$ and $\overrightarrow{AX_j}$ can intersect, for if they did, they would be the same, and would contain both $X_i \neq X_j$ in contradiction of Theorem CNV.4(A), so that $Y_i \neq Y_j$ if $i \neq j$. Also, by the construction, no three adjacent points Y_i can be collinear. Therefore $\mathcal{G} = \langle Y_1, \ldots, Y_n \rangle$ is a polygon.

$\mathcal{G}$ is a subset of $\operatorname{out}\mathcal{F}$ because every edge $\overline{Y_i Y_{i+1}}$ is on the side of $\overleftrightarrow{X_i X_{i+1}}$, and $\overline{Y_n Y_1}$ is on the side of $\overleftrightarrow{X_n X_1}$ opposite A. Finally, $\mathcal{G}$ is simple, for if there were integers j and k such that $j \equiv k (\operatorname{mod} n)$, $j \equiv k+1 (\operatorname{mod} n)$ and $j \equiv$

$k-1 (\mathrm{mod}\, n)$ and a point D such that $D \in \overline{Y_jY_{j+1}} \cap \overline{Y_kY_{k+1}}$, then $\overrightarrow{AD}$ would intersect both $\overline{X_jX_{j+1}}$ and $\overline{X_kX_{k+1}}$, contradicting Theorem CNV.4(A).

Now let P and Q be any members of $\operatorname{out}\mathcal{F}$. The Crossbar Theorem PSH.39 assures that $\mathcal{G}$ intersects each of $\overrightarrow{AP}$ and $\overrightarrow{AQ}$ at exactly one point R or S respectively. If $R = S$ let $\mathcal{J} = \overline{PQ}$. Without loss of generality we may assume there exist $i \le j$ belonging to $[1; n]$ such that $R \in \overline{Y_iY_{i+1}}$ and $S \in \overline{Y_jY_{j+1}}$. If $i = j$ let $\mathcal{J} = \overline{PR} \cup \overline{RS} \cup \overline{SQ}$. If $i \ne j$ let

$$\mathcal{J} = \overline{PR} \cup \overline{RY_{i+1}} \cup \left(\bigcup_{k=i+1}^{j-1} \overline{Y_kY_{k+1}}\right) \cup \overline{SY_j} \cup \overline{QS}.$$

Since A–P–R, $A \in \operatorname{ins}\mathcal{F}$, $P \in \operatorname{out}\mathcal{F}$, $R \in \operatorname{out}\mathcal{F}$, and $\overrightarrow{AP}$ has only one intersection with $\mathcal{F}$, $\overline{PR} \subseteq \operatorname{out}\mathcal{F}$; a similar argument shows $\overline{QS} \subseteq \operatorname{out}\mathcal{F}$; all the other segments in the construction of $\mathcal{J}$ are known already to be subsets of $\operatorname{out}\mathcal{F}$, so $\mathcal{J} \subseteq \operatorname{out}\mathcal{F}$. $\square$

Theorem CNT.3 (Proof of Theorem JCT.1, part (D); the inside and outside of a simple polygon are polygonally connected.) *Given any simple polygon* $\mathcal{F} = \langle X_1, \ldots, X_m \rangle$, *both* (A) *the inside* $\operatorname{ins}\mathcal{F}$ *and* (B) *the outside* $\operatorname{out}\mathcal{F}$ *are polygonally connected.*

Proof. (A) First note that the inside of a triangle is polygonally connected since it is convex. We now show that $\operatorname{ins}\mathcal{F}$ is polygonally connected whenever it is true that *for every* k *with* $3 \le k < m$, *the inside of every polygon with* k *edges is polygonally connected*.

This shows, by the "strong form" of mathematical induction, that the inside of every simple polygon is polygonally connected. The italicized statement is called the "induction hypothesis."

By Theorem CNV.29(A) there exist corners A and B of $\mathcal{F}$ such that $\overline{AB} \subseteq \operatorname{ins}\mathcal{F}$. Let $\mathcal{G}$ and $\mathcal{H}$ be the simple polygons described in Theorem CNV.29(B). Since $\mathcal{G}$ and $\mathcal{H}$ each have fewer corners than $\mathcal{F}$ has, both $\operatorname{ins}\mathcal{G}$ and $\operatorname{ins}\mathcal{H}$ are polygonally connected by the induction hypothesis.

As in the proof of CNV.29(B), let $I \ne J$ be points such that $\overline{IJ} \subseteq \operatorname{ins}\mathcal{F}$, $\overline{IJ} \cap \overline{AB} = \{O\}$, $\overline{IO} \subseteq \operatorname{ins}\mathcal{G}$ and $\overline{JO} \subseteq \operatorname{ins}\mathcal{H}$. Let $P \ne Q$ be any points of $\operatorname{ins}\mathcal{F}$. By CNV.29(B),

$$\{P, Q\} \subseteq \operatorname{ins}\mathcal{F} = \operatorname{ins}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \overline{X_iX_j}.$$

Then one of the following cases will hold:

(i) P and Q both belong to $\operatorname{ins}\mathcal{G}$ (or $\operatorname{ins}\mathcal{H}$);
(ii) $P \in \operatorname{ins}\mathcal{G}$ (or $\operatorname{ins}\mathcal{H}$) and A–Q–B;
(iii) $P \in \operatorname{ins}\mathcal{G}$ and $Q \in \operatorname{ins}\mathcal{H}$; or
(iv) P and Q both belong to $\overline{AB}$.

In case (iv) $\overline{[PQ]} \subseteq \overline{]AB[} \subseteq \operatorname{ins}\mathcal{F}$ and $\overline{[PQ]}$ is a polygonal path joining P and Q, and in case (i) there is a polygonal path joining P and Q in $\operatorname{ins}\mathcal{G}$ (or $\operatorname{ins}\mathcal{H}$) (hence in $\operatorname{ins}\mathcal{F}$) by the induction hypothesis.

In case (ii) there exists a polygonal path $\mathcal{J} \subseteq \operatorname{ins}\mathcal{G}$ ($\operatorname{ins}\mathcal{H}$) with endpoints P and I (P and J) so that $\mathcal{J} \cup \overline{[QO]} \cup \overline{[IO[}$ ($\mathcal{J} \cup \overline{[QO]} \cup \overline{[JO[}$) is a polygonal path contained in $\operatorname{ins}\mathcal{F}$ connecting P and Q.

In case (iii) there exist polygonal paths $\mathcal{I} \subseteq \operatorname{ins}\mathcal{G}$ and $\mathcal{J} \subseteq \operatorname{ins}\mathcal{H}$ such that $\mathcal{I}$ connects P to I and $\mathcal{J}$ connects Q to J. Then $\mathcal{I} \cup \overline{[IJ]} \cup \mathcal{J} \subseteq \operatorname{ins}\mathcal{F}$ is a polygonal path joining P and Q. It follows that $\operatorname{ins}\mathcal{F}$ is polygonally connected.

(B) If $\operatorname{ins}\mathcal{F}$ is convex, then by Theorem CNV.7, $\mathcal{F}$ is rotund and by Theorem CNT.2 $\operatorname{out}\mathcal{F}$ is polygonally connected. If $\operatorname{ins}\mathcal{F}$ is not convex, again we use induction on m to show that $\operatorname{out}\mathcal{F}$ is polygonally connected. We initiate the induction by noting that the outside of a triangle is polygonally connected since every triangle is convex. Now assume the induction hypothesis: *for every k with $3 \leq k < m$, the outside of every polygon with k edges is polygonally connected.*

By Theorem CNV.23(A) there exist corners A and B such that $\overleftrightarrow{AB}$ is a supporting line of $\mathcal{F}$ and every point between A and B belongs to $\operatorname{out}\mathcal{F}$. Let $\mathcal{G}$ and $\mathcal{H}$ be the polygons defined in Theorem CNV.23(B). By Theorem SEP.7, we may let $I \neq J$ be points such that $\overline{[IJ]} \subseteq \operatorname{ins}\mathcal{F}$, $\overline{[IJ]} \cap \overline{]AB[} = \{O\}$, $\overline{[IO[} \subseteq \operatorname{out}\mathcal{G}$ and $\overline{[JO[} \subseteq \operatorname{ins}\mathcal{H}$.

Also, from Theorem CNV.23(B) we get $\operatorname{out}\mathcal{F} = \operatorname{out}\mathcal{G} \cup \operatorname{ins}\mathcal{H} \cup \overline{]AB[}$. We will refer to this without reference in the rest of the proof.

Let $P \neq Q \in \operatorname{out}\mathcal{F}$. We then have the following cases:

(i) P and Q both belong to $\operatorname{out}\mathcal{G}$ (or $\operatorname{ins}\mathcal{H}$);

(ii) $P \in \operatorname{out}\mathcal{G}$ (or $\operatorname{ins}\mathcal{H}$) and A–Q–B;

(iii) $P \in \operatorname{out}\mathcal{G}$ and $Q \in \operatorname{ins}\mathcal{H}$; or

(iv) P and Q both belong to $\overline{]AB[}$.

In case (iv) $\overline{[PQ]} \subseteq \overline{]AB[} \subseteq \operatorname{out}\mathcal{F}$ and $\overline{[PQ]}$ joins P and Q; in case (i) there is a polygonal path joining P and Q in $\operatorname{out}\mathcal{G}$ (or $\operatorname{ins}\mathcal{H}$)(hence in $\operatorname{out}\mathcal{F}$) by the induction hypothesis.

In case (ii) there exists a polygonal path $\mathcal{J} \subseteq \operatorname{out}\mathcal{G}$ ($\operatorname{ins}\mathcal{H}$) with endpoints P and I (P and J) so that $\mathcal{J} \cup \overline{[QO]} \cup \overline{[IO[}$ ($\mathcal{J} \cup \overline{[QO]} \cup \overline{[JO[}$) is a polygonal path contained in $\operatorname{out}\mathcal{F}$ connecting P and Q.

In case (iii) there exist polygonal paths $\mathcal{I} \subseteq \operatorname{out}\mathcal{G}$ and $\mathcal{J} \subseteq \operatorname{ins}\mathcal{H}$ such that $\mathcal{I}$ connects P to I and $\mathcal{J}$ connects Q to J. Then $\mathcal{I} \cup \overline{[IJ]} \cup \mathcal{J} \subseteq \operatorname{out}\mathcal{F}$ is a polygonal path joining P and Q. It follows that $\operatorname{out}\mathcal{F}$ is polygonally connected. $\square$

Corollary CNT.3.1 *If $\mathcal{F}$ is a simple polygon, then* $\operatorname{enc}\mathcal{F}$ *and* $\operatorname{exc}\mathcal{F}$ *are polygonally connected.*

Proof. Let P and Q be any members of $\operatorname{enc}\mathcal{F}$. If both P and $Q \in \operatorname{ins}\mathcal{F}$ the proof is complete. By Theorem SEP.7, if $P \in \mathcal{F}$ and $Q \in \operatorname{ins}\mathcal{F}$ there exists a point $I \in \operatorname{ins}\mathcal{F}$ such that $\overline{PI} \subseteq \operatorname{ins}\mathcal{F}$; if $\mathcal{I}$ is a path in $\operatorname{ins}\mathcal{F}$ joining I and Q, $\overline{PI} \cup \mathcal{I}$ is a polygonal path in $\operatorname{enc}\mathcal{F}$ joining P and Q. Similar arguments show the other assertions. $\square$

Theorem CNT.4 *If $\mathcal{F}$ and $\mathcal{G}$ are simple polygons such that $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$, then* $\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$ *and* $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$. *Moreover, if $\mathcal{G} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$, then* $\operatorname{ins}\mathcal{G}$ *is a proper subset of* $\operatorname{ins}\mathcal{F}$.

Proof. Let $Q \in \operatorname{out}\mathcal{F}$, let $A \in \operatorname{ins}\mathcal{F}$, let C be the last point of intersection of $\mathcal{F}$ and $\overrightarrow{AQ}$, and let P be any point such that A–C–P. Then $P \in \operatorname{out}\mathcal{F}$. Let R be any point such that A–P–R. Since $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$, $\mathcal{G} \cap \overrightarrow{PR} = \emptyset$ and hence $P \in \operatorname{out}\mathcal{G}$. Since by Theorem CNT.3 $\operatorname{out}\mathcal{F}$ is polygonally connected there exists a polygonal path $\mathcal{J} \subseteq \operatorname{out}\mathcal{F}$ with endpoints P and Q. Since $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$, $\mathcal{J} \cap \mathcal{G} = \emptyset$. Then by Theorem SEP.13(A) either $\mathcal{J} \subseteq \operatorname{ins}\mathcal{G}$ or $\mathcal{J} \subseteq \operatorname{out}\mathcal{G}$. Since $P \in \mathcal{J}$ and $P \in \operatorname{out}\mathcal{G}$, $\mathcal{J} \subseteq \operatorname{out}\mathcal{G}$. Hence $Q \in \operatorname{out}\mathcal{G}$ and we have shown $\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$. Taking complements with respect to the plane gives $\operatorname{enc}\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$.

Now suppose $\mathcal{F} \cap \operatorname{ins}\mathcal{G} \neq \emptyset$. By Theorem SEP.7 there exist points C and F such that $C \in \mathcal{F} \cap \operatorname{ins}\mathcal{G}$, $\overline{CF} \cap \mathcal{F} = \{\mathcal{C}\}$ and $\overline{CF} \subseteq \operatorname{out}\mathcal{F}$. Let G be the first intersection of $\overrightarrow{CF}$ with $\mathcal{G}$. Since $C \in \operatorname{ins}\mathcal{F}$, $\overline{CG} \subseteq \operatorname{ins}\mathcal{G}$. Then $\emptyset \neq \overline{CF} \cap \overline{CG} \subseteq \operatorname{out}\mathcal{F} \cap \operatorname{ins}\mathcal{G}$, so there is a point of $\operatorname{out}\mathcal{F}$ not in $\operatorname{out}\mathcal{G}$ contradicting $\operatorname{out}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$. Hence $\mathcal{F} \cap \operatorname{ins}\mathcal{G} = \emptyset$ so $\operatorname{ins}\mathcal{G} \subseteq \operatorname{ins}\mathcal{F}$.

If $\mathcal{G} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$, and argument similar to the one just given shows that there exists a point of $\operatorname{out}\mathcal{G}$ (hence not in $\operatorname{ins}\mathcal{G}$) that is in $\operatorname{ins}\mathcal{F}$ showing that $\operatorname{ins}\mathcal{G}$ is a proper subset of $\operatorname{ins}\mathcal{F}$. $\square$

Theorem CNT.5 *Let $\mathcal{F}$ and $\mathcal{G}$ be simple polygons in the Pasch plane $\mathcal{P}$.*

(A) *If $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$, then $\mathcal{F} \subseteq \operatorname{exc}\mathcal{G}$.*

(B) *If $\mathcal{G} \subseteq \operatorname{exc}\mathcal{F}$, then either*

(1) $\operatorname{ins}\mathcal{F} \subseteq \operatorname{ins}\mathcal{G}$, $\operatorname{out}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$, $\operatorname{enc}\mathcal{F} \subseteq \operatorname{enc}\mathcal{G}$, *and* $\mathcal{F} \subseteq \operatorname{enc}\mathcal{G}$, *or*

(2) $\operatorname{ins}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$, $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$, $\operatorname{enc}\mathcal{F} \subseteq \operatorname{exc}\mathcal{G}$, *and* $\mathcal{F} \subseteq \operatorname{exc}\mathcal{G}$*;*

moreover, if (1) holds and $\mathcal{G} \cap \operatorname{out} \mathcal{F} \neq \emptyset$*, then both inclusions* $\operatorname{ins} \mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$ *and* $\operatorname{out} \mathcal{G} \subseteq \operatorname{out} \mathcal{F}$ *are proper.*

Proof. (A) The second part of the proof of Theorem CNT.4 shows that $\mathcal{F} \cap \operatorname{ins} \mathcal{G} = \emptyset$ so that $\mathcal{F} \subseteq \operatorname{exc} \mathcal{G} = \mathcal{G} \cup \operatorname{out} \mathcal{G}$.

(B) By Theorem CNT.2 $\operatorname{ins} \mathcal{F}$ is polygonally connected. Since $\mathcal{G} \subseteq \operatorname{exc} \mathcal{F}$, $\mathcal{G} \cap \operatorname{ins} \mathcal{F} = \emptyset$, so by Theorem SEP.13(A) either $\operatorname{ins} \mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$ or $\operatorname{ins} \mathcal{F} \subseteq \operatorname{out} \mathcal{G}$.

If (1) holds, and $\operatorname{ins} \mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$, then $\operatorname{exc} \mathcal{G} = \mathcal{P} \setminus \operatorname{ins} \mathcal{G} \subseteq \mathcal{P} \setminus \operatorname{ins} \mathcal{F} = \operatorname{exc} \mathcal{F}$, so $\operatorname{out} \mathcal{G} \subseteq \operatorname{exc} \mathcal{F}$. Suppose now that for some point C, $\operatorname{out} \mathcal{G} \cap \mathcal{F} = \{C\}$. By Theorem SEP.7 there exists a point E such that $\overrightarrow{CE} \subseteq \operatorname{ins} \mathcal{F}$; let D be the first point of intersection of $\overrightarrow{CE}$ with $\mathcal{G}$; then $\overline{CD} \subseteq \operatorname{out} \mathcal{G}$ (since $C \in \operatorname{out} \mathcal{G}$) and if we pick F so that C–F–D and C–F–E, $\overline{CF} \subseteq \operatorname{out} \mathcal{G} \cap \operatorname{ins} \mathcal{F}$. But this contradicts $\operatorname{ins} \mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$. Thus $\operatorname{out} \mathcal{G} \cap \mathcal{F} = \emptyset$, and since $\operatorname{out} \mathcal{G} \subseteq \operatorname{exc} \mathcal{F}$, $\operatorname{out} \mathcal{G} \subseteq \operatorname{out} \mathcal{F}$; taking complements we have $\operatorname{enc} \mathcal{F} \subseteq \operatorname{enc} \mathcal{G}$ so that in particular, $\mathcal{F} \subseteq \operatorname{enc} \mathcal{G}$.

If (2) holds, and $\operatorname{ins} \mathcal{F} \subseteq \operatorname{out} \mathcal{G}$, then $\operatorname{enc} \mathcal{G} = \mathcal{P} \setminus \operatorname{out} \mathcal{G} \subseteq \mathcal{P} \setminus \operatorname{ins} \mathcal{F} = \operatorname{exc} \mathcal{F}$, so that $\operatorname{ins} \mathcal{G} \subseteq \operatorname{exc} \mathcal{F}$. Again, if $\mathcal{F} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$, by similar reasoning to that of the previous paragraph, we can find points that belong both to $\operatorname{ins} \mathcal{G}$ and to $\operatorname{ins} \mathcal{F}$, which contradicts $\operatorname{ins} \mathcal{F} \subseteq \operatorname{out} \mathcal{G}$, so that $\mathcal{F} \cap \operatorname{ins} \mathcal{G} = \emptyset$. Hence $\operatorname{ins} \mathcal{G} \subseteq \operatorname{out} \mathcal{F}$; taking complements, $\operatorname{enc} \mathcal{F} \subseteq \operatorname{exc} \mathcal{G}$ so that $\mathcal{F} \subseteq \operatorname{exc} \mathcal{G}$.

We now show the final assertion of the Theorem. Let $O \in \mathcal{G} \cap \operatorname{out} \mathcal{F}$, let Q' be such that $\mathcal{L} = \overleftrightarrow{OQ'}$ contains no corners of either $\mathcal{F}$ or $\mathcal{G}$, and let Q'–O–P'. Choose Q and P to be the first intersections of $\overrightarrow{OQ'}$ and $\overrightarrow{OP'}$, respectively, with $\mathcal{F}$. Then no point of $\overline{QP}$ belongs to $\mathcal{F}$ and $\overline{QP} \subseteq \operatorname{out} \mathcal{F}$, by Theorem SEP.4.

Now by Theorem SEP.7 there exist points E and $F \in \mathcal{L}$ such that $\overline{EO} \subseteq \operatorname{ins} \mathcal{G}$ and $\overline{FO} \subseteq \operatorname{out} \mathcal{G}$. E is on either the P-side or on the Q-side of O. In the former case, choose E' so that P–E'–O and E–E'–O, so that $\overline{E'O} \subseteq \operatorname{ins} \mathcal{G} \cap \operatorname{out} \mathcal{F}$. A similar proof is valid in the other case. Thus both $\operatorname{ins} \mathcal{F} \subseteq \operatorname{ins} \mathcal{G}$ and $\operatorname{out} \mathcal{G} \subseteq \operatorname{out} \mathcal{F}$ are proper inclusions. □

Theorem CNT.6 *Let* $\mathcal{F}$ *and* $\mathcal{G}$ *be simple polygons in the Pasch plane* $\mathcal{P}$*. If* $\mathcal{F} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$ *and* $\mathcal{F} \cap \operatorname{out} \mathcal{G} \neq \emptyset$*, then*

(A) $\mathcal{G} \cap \operatorname{ins} \mathcal{F} \neq \emptyset$,
(B) $\mathcal{G} \cap \operatorname{out} \mathcal{F} \neq \emptyset$,
(C) $\operatorname{ins} \mathcal{F} \cap \operatorname{ins} \mathcal{G} \neq \emptyset$,
(D) $\operatorname{out} \mathcal{F} \cap \operatorname{out} \mathcal{G} \neq \emptyset$,
(E) $\operatorname{ins} \mathcal{F} \cap \operatorname{out} \mathcal{G} \neq \emptyset$*, and*

(F) $\operatorname{out}\mathcal{F} \cap \operatorname{ins}\mathcal{G} \neq \emptyset$.

Proof. If $\mathcal{G}$ were contained in $\operatorname{enc}\mathcal{F}$, then by Theorem CNT.5(A), $\mathcal{F}$ would be contained in $\operatorname{exc}\mathcal{G}$, contradicting $\mathcal{F} \cap \operatorname{ins}\mathcal{G} \neq \emptyset$. Hence $\mathcal{G} \cap \operatorname{out}\mathcal{F} \neq \emptyset$, proving (B).

If $\mathcal{G}$ were contained in $\operatorname{exc}\mathcal{F}$, then in Theorem CNT.5(B) alternate (2) is ruled out, so $\mathcal{F}$ would be contained in $\operatorname{enc}\mathcal{G}$, contradicting $\mathcal{F} \cap \operatorname{out}\mathcal{G} \neq \emptyset$. Hence $\mathcal{G} \cap \operatorname{ins}\mathcal{F} \neq \emptyset$, proving (A).

Let $C \in \mathcal{F} \cap \operatorname{ins}\mathcal{G}$. By Corollary SEP.7.1 there exists a point E such that $\overline{CE} \subseteq \operatorname{ins}\mathcal{F}$. Let Q be the first point of intersection of $\overrightarrow{CE}$ and $\mathcal{G}$. If E is between C and Q, then $\overline{CE} \subseteq \operatorname{ins}\mathcal{G}$; if Q is between C and E let P be any point between C and Q; then $\overline{CP} \subseteq \operatorname{ins}\mathcal{G}$. In either case, $\overline{CP} \subseteq \operatorname{ins}\mathcal{G} \cap \operatorname{ins}\mathcal{F}$, proving (C).

The proofs of the other cases are similar and left to the reader as Exercise CNT.1. $\square$

Theorem CNT.7 *Let $\mathcal{F}$ and $\mathcal{G}$ be simple polygons. If $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{enc}\mathcal{G}$, then $\mathcal{F} = \mathcal{G}$.*

Proof. If $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{enc}\mathcal{G}$, then by Theorem CNT.5(A), $\mathcal{F} \subseteq \operatorname{exc}\mathcal{G}$ and $\mathcal{G} \subseteq \operatorname{exc}\mathcal{F}$. But then $\mathcal{F} \subseteq \operatorname{enc}\mathcal{G} \cap \operatorname{exc}\mathcal{G} = \mathcal{G}$ and $\mathcal{G} \subseteq \operatorname{enc}\mathcal{F} \cap \operatorname{enc}\mathcal{F} = \mathcal{F}$. $\square$

Theorem CNT.8 (Re-statement of the Jordan Curve Theorem JCT.1 for simple polygons) *If $\mathcal{G}$ is a simple polygon in the Pasch plane $\mathcal{P}$, then*

(A) *$\mathcal{P} = \mathcal{G} \cup \operatorname{ins}\mathcal{G} \cup \operatorname{out}\mathcal{G}$, where $\mathcal{G}$, $\operatorname{ins}\mathcal{G}$, and $\operatorname{out}\mathcal{G}$ are pairwise disjoint sets;*

(B) *if $P \in \operatorname{ins}\mathcal{G}$ and $Q \in \operatorname{out}\mathcal{G}$, then $\overline{PQ} \cap \mathcal{G} \neq \emptyset$;*

(C) *$\mathcal{G}$ and $\operatorname{ins}\mathcal{G}$ are bounded sets, and $\operatorname{out}\mathcal{G}$ is unbounded; and*

(D) *$\operatorname{ins}\mathcal{G}$ and $\operatorname{out}\mathcal{G}$ are polygonally connected sets.*

Proof. Parts (A) and (B) follow immediately from Theorem SEP.12, part (C) from Theorem CNV.22, and part (D) from Theorem CNT.3. $\square$

There is a more extensive discussion of the Jordan Curve Theorem at the beginning of this chapter.

6.6 Exercises for Jordan Curve Theorem

Exercise PLGN.1* Prove part (A) of Theorem PLGN.15; that is, that an O-ordering of a simple polygon has the trichotomy property.

Exercise SEP.1* Prove Theorem SEP.15 in the case that alternative (3) is true.

Exercise CNV.1 Prove Theorem CNV.36 by modifying the proof of Theorem CNV.20.

Exercise CNT.1* Complete the proof of Theorem CNT.6.

6.7 Selected answers for Jordan Curve Theorem

Exercise PLGN.1 Proof. We show that for any P and $Q \in \mathcal{G} \setminus \{O\}$, exactly one of $P < Q$, $P = Q$, or $P > Q$ is true (Trichotomy). In this proof we will refer to the various parts of Definition PLGN.14 as "rule (A)," "rule (B)," etc. We will assume that the polygon $\mathcal{G}$ consists of n edges $\overline{X_i X_{i+1}}$ where $i = \alpha + 1, ...\alpha + n$, its corners being $X_{\alpha+1}, ..., X_{\alpha+n} = X_\alpha$, and that $O \in \overline{X_\alpha X_{\alpha+1}}$.

By rule (A), if both P and Q are corners of $\mathcal{G}$, their ordering is the same as the ordering of the integers $\{\alpha + 1, \alpha + 2, ..., \alpha + n\}$, for which trichotomy holds; therefore trichotomy holds for P and Q.

If we prove trichotomy in the case where P is not a corner of $\mathcal{G}$, this will also (by interchanging P and Q) show it is true if Q is not a corner of $\mathcal{G}$. Moreover, it will suffice to prove that if $P \neq Q$, then exactly one of $P < Q$ or $Q < P$ is true; for this implies that if neither $P < Q$ nor $Q < P$, then $P = Q$.

(Case 1: $P \neq Q$ and $P \in \overline{X_i X_{i+1}}$ where $\alpha < i < \alpha + n$.) Then either

(a) $Q \in \overline{X_i X_{i+1}}$,

(b) $Q \in \overline{X_j X_{j+1}}$ where $j \neq i$ and $\alpha < j < \alpha + n$, or

(c) $Q \in \overline{X_{\alpha+n} X_{\alpha+1}} \setminus \{O\} = \overline{X_\alpha X_{\alpha+1}} \setminus \{O\}$.

If (a) holds, then since trichotomy holds on $\overline{X_i X_{i+1}}$, either $P < Q$ or $Q < P$. If (b) holds, either

(1) $\alpha < i < j < \alpha + n$, in which case $X_j \leq Q$ and $X_{i+1} \leq X_j$ and $P < X_{i+1}$ so $P < X_{i+1} \leq X_j \leq Q$ and by rule (D) $P < Q$; or

(2) $\alpha < j < i < \alpha + n$, in which case $Q \le X_{j+1} \le X_i < P$ so $Q < P$.

If (c) holds, either

(1) O–Q–$X_{\alpha+1}$ or $Q = X_{\alpha+1}$, in which case, since $\alpha < i$, $\alpha+1 \le i$ and $Q \le X_{\alpha+1} \le X_i < P$ so by rule (D) $Q < P$; or

(2) $X_{\alpha+n}$–Q–O or $Q = X_{\alpha+n}$, in which case since $i < \alpha + n$, $X_{i+1} \le X_{\alpha+n}$ and $P < X_{i+1} \le X_{\alpha+n} \le Q$ and $P < Q$.

(Case 2: $P \ne Q$ and $P \in \overline{X_\alpha X_{\alpha+1}} \setminus \{O\}$. Then either

(a) $X_{\alpha+n}$–P–O or

(b) O–P–$X_{\alpha+1}$.

If (a) holds, then $X_\alpha = X_{\alpha+n} < P$; if $Q \in \overline{X_\alpha X_{\alpha+1}} \setminus \{O\}$ and $Q \ne P$ then either

(1) $X_{\alpha+n}$–Q–O or $Q = X_{\alpha+n}$ in which case either $P < Q$ or $Q < P$, by rule (C); or

(2) O–Q–$X_{\alpha+1}$ or $Q = X_{\alpha+1}$ in which case $Q < X_{\alpha+1}$ by rule (C), and $Q \le X_{\alpha+1} < X_{\alpha+n} < P$ so that $Q < P$.

Moreover, if (a) holds, and $Q \in \overline{X_i X_{i+1}}$ where $\alpha + 1 < i < \alpha + n$, then $Q \le X_{i+1} \le X_{\alpha+n} < P$ so $Q < P$.

If (b) holds, then $P < X_{\alpha+1}$; if $Q \ne P$ and $Q \in \overline{X_\alpha X_{\alpha+1}} \setminus \{O\}$ then either

(1) $X_{\alpha+n}$–Q–O or $Q = X_{\alpha+n}$ so that $Q \ge X_{\alpha+n}$ and $Q \ge X_{\alpha+n} > X_{\alpha+1} > P$ and $Q > P$; or

(2) O–Q–$X_{\alpha+1}$ or $Q = X_{\alpha+1}$ in which case either $P < Q$ or $Q < P$ by rule (C).

Moreover, if (b) holds, and $Q \in \overline{X_i X_{i+1}}$ where $\alpha < i < \alpha + n$, then $P < X_{\alpha+1} \le X_i \le Q$ so $P < Q$. □

Exercise SEP.1 Proof. If alternative (3) of Theorem SEP.15 holds, we assume that $\mathcal{G} \subseteq \operatorname{out} \mathcal{F}$ and $\mathcal{F} \subseteq \operatorname{out} \mathcal{G}$. We show that $\mathcal{G} \cup \operatorname{ins} \mathcal{G}$ is a proper subset of $\operatorname{out} \mathcal{F}$.

First we prove that $\operatorname{ins} \mathcal{G} \subseteq \operatorname{out} \mathcal{F}$. Let P be any point of $\operatorname{ins} \mathcal{G}$, and let A be any point such that $\overrightarrow{PA}$ does not contain a corner of $\mathcal{F}$ or $\mathcal{G}$, or intersect the segment $\overline{CD}$. Then by the Lemma in the proof of Theorem SEP.15, $\overrightarrow{PA} \cap \mathcal{G} = \overrightarrow{PA} \cap (\mathcal{G} \setminus \overline{CD})$.

Since $P \in \operatorname{ins} \mathcal{G}$ every ray from P must intersect $\mathcal{G}$ at least once. Order $\overrightarrow{PA}$ with $P < A$, and let Q be the first point and R the last point of intersection of $\overrightarrow{PA}$ with $\mathcal{G}$. If the ray $\overrightarrow{PA} \setminus \overline{PR}$ intersects $\mathcal{F}$ let S be the first point of intersection of $\overrightarrow{PA} \setminus \overline{PR}$ with $\mathcal{F}$, otherwise let S be any point of $\overrightarrow{PA} \setminus \overline{PR}$. Then $\overline{PQ}$ contains no point of $\mathcal{G}$ other than Q so that by Theorem SEP.4(A),

$\overline{PQ} \subseteq \operatorname{ins}\mathcal{G}$ and hence contains no point of $\mathcal{F}$ since $\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$. Now $Q \in \mathcal{G} \subseteq \operatorname{out}\mathcal{F}$, so $P \in \operatorname{out}\mathcal{F}$ by Theorem SEP.4(A), so that $P \in \operatorname{out}\mathcal{F}$. Therefore $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$.

Finally we show that $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$ is a proper inclusion. The segment $\overline{RS}$ contains no point of $\mathcal{G} \setminus \overline{CD}$ since R is the last point of $\mathcal{G} \setminus \overline{CD}$, and no point of $\mathcal{F} \setminus \overline{CD}$ since S is the first point of $\mathcal{F} \setminus \overline{CD}$ with $R < S$. By Theorem SEP.4(A), since $R \in \mathcal{G} \subseteq \operatorname{out}\mathcal{F}$, $\overline{RS} \subseteq \operatorname{out}\mathcal{F}$; similarly since $S \in \operatorname{out}\mathcal{G}$, $\overline{RS} \subseteq \operatorname{out}\mathcal{G}$. Thus points of $\overline{RS}$ are in $\operatorname{out}\mathcal{F}$ but not in $\operatorname{ins}\mathcal{G}$, and hence the inclusion $\operatorname{ins}\mathcal{G} \subseteq \operatorname{out}\mathcal{F}$ is proper.

By the same argument, with the roles of $\mathcal{F}$ and $\mathcal{G}$ interchanged, $\operatorname{ins}\mathcal{F} \subseteq \operatorname{out}\mathcal{G}$ is also a proper inclusion. $\square$

Exercise CNT.1 Proof. We prove parts (D), (E), and (F) of Theorem CNT.6.

(D) Let $C \in \mathcal{F} \cap \operatorname{out}\mathcal{G}$. By Corollary SEP.7.1 there exists a point E such that $\overline{CE} \subseteq \operatorname{out}\mathcal{F}$. Let Q be the first point of intersection of $\overrightarrow{CE}$ and $\mathcal{G}$ (or if $\overrightarrow{CE}$ does not intersect $\mathcal{G}$, let Q be any point of the ray.) If E is between C and Q, then $\overline{CE} \subseteq \operatorname{out}\mathcal{G}$; if Q is between C and E let P be any point between C and Q; then $\overline{CP} \subseteq \operatorname{out}\mathcal{G}$. In either case, $\overline{CP} \subseteq \operatorname{out}\mathcal{G} \cap \operatorname{out}\mathcal{F}$, proving (D).

(E) Let $C \in \mathcal{F} \cap \operatorname{out}\mathcal{G}$. By Corollary SEP.7.1 there exists a point E such that $\overline{CE} \subseteq \operatorname{ins}\mathcal{F}$. Let Q be the first point of intersection of $\overrightarrow{CE}$ and $\mathcal{G}$ (or if $\overrightarrow{CE}$ does not intersect $\mathcal{G}$, let Q be any point of the ray.) If E is between C and Q, then $\overline{CE} \subseteq \operatorname{out}\mathcal{G}$; if Q is between C and E let P be any point between C and Q; then $\overline{CP} \subseteq \operatorname{out}\mathcal{G}$. In either case, $\overline{CP} \subseteq \operatorname{out}\mathcal{G} \cap \operatorname{ins}\mathcal{F}$, proving (E).

(F) Let $C \in \mathcal{F} \cap \operatorname{ins}\mathcal{G}$. By Corollary SEP.7.1 there exists a point E such that $\overline{CE} \subseteq \operatorname{out}\mathcal{F}$. Let Q be the first point of intersection of $\overrightarrow{CE}$ and $\mathcal{G}$; there must be such a point because $C \in \operatorname{ins}\mathcal{G}$. If E is between C and Q, then $\overline{CE} \subseteq \operatorname{ins}\mathcal{G}$; if Q is between C and E let P be any point between C and Q; then $\overline{CP} \subseteq \operatorname{ins}\mathcal{G}$. In either case, $\overline{CP} \subseteq \operatorname{ins}\mathcal{G} \cap \operatorname{out}\mathcal{F}$, proving (F). $\square$

Chapter 7

Property PE on a Pasch Plane with Property LUB (LUPE)

Dependencies: *Chapters 1, 4, 5, and 6 of Euclidean Geometry and its Subgeometries (Specht); Axiom LUB*
Acronym: *LUPE*

This chapter might be considered something of a curiosity, or as an addendum to Chapter 6 of *Specht*, as the proof depends on ordering. In it we prove *Property PE* on a Pasch plane on which the LUB property holds, as defined in *Specht* Chapter 18. In that work, *Property PE* is proved as Theorem NEUT.48(B), part of neutral geometry.

Other than the LUB property, we invoke only the results of Chapters 1, 4, 5, and 6 of *Specht*. All references in this chapter are to *Specht*, and there are no references to other chapters of this Supplement. We begin by restating the LUB and PE properties.

Property LUB: Let $\mathcal{P}$ be a Pasch plane, and let $\mathcal{L}$ be a line on $\mathcal{P}$ which is equipped with an ordering $<$ by *Specht* Ch.6 Definition ORD.1. If $\mathcal{E} \subseteq \mathcal{L}$ is a set that is bounded above, then the set of all upper bounds has a minimum, called the **least upper bound** of $\mathcal{E}$, and denoted $\text{lub}(\mathcal{E})$.

Property PE: Given a Pasch plane $\mathcal{P}$, a line $\mathcal{L}$ on $\mathcal{P}$, for every point Q belonging to $\mathcal{P} \setminus \mathcal{L}$, there exists a line $\mathcal{M}$ through Q which is parallel to $\mathcal{L}$.

The designation "Property PE" is so named to suggest "Parallel Existence." There is no claim here of uniqueness—PE falls short of Axiom PS, the "strong form" of the parallel axiom. But if PE is joined with PW (the "weak form"), then we get PS. We name the single theorem in this chapter with the acronym *LUPE* to suggest both LUB and PE.

Theorem LUPE *If $\mathcal{P}$ is a Pasch plane on which property LUB holds, then property PE holds on $\mathcal{P}$.*

Proof. We begin with the assumption that property PE is false on $\mathcal{P}$: that is, there exists a line $\mathcal{L}$ on $\mathcal{P}$ and a point Q belonging to $\mathcal{P} \setminus \mathcal{L}$ such that there is no line $\mathcal{M}$ parallel to $\mathcal{L}$ containing Q. We prove the theorem by a construction and a series of claims; the last of these will show a contradiction with property LUB, thus showing that property PE is true.

The construction: See Figure 7.1 below. Let A and B be distinct points of $\mathcal{L}$, and order $\mathcal{L}$ so that $A < B$. By *Specht* Ch.5 Theorem PSH.22 there exists a point $C \in \mathcal{L}$ such that A–C–B. By property B.3 of *Specht* Ch.4 Definition IB.1, there exists a point T such that B–Q–T.

Using Definition IB.11, since B–Q–T, T and B are on opposite sides of $\overleftrightarrow{AQ}$; since A–C–B, C and B are on the same side of $\overleftrightarrow{AQ}$; therefore T and C are on opposite sides of $\overleftrightarrow{AQ}$, and by *Specht* Ch.5 Theorem PSH.12 there exists a point S such that $\overleftrightarrow{AQ} \cap \overline{CT} = \{S\}$, that is, $\overleftrightarrow{AQ} \cap \overleftrightarrow{CT} = \{S\}$. Finally, by property B.3 of Definition IB.1, there exists a point D such that S–T–D, that is, C–S–T–D.

Since $A \neq B$, the lines $\overleftrightarrow{AQ}$ and $\overleftrightarrow{BQ}$ are distinct and have only the point Q in common. The point T belongs to a side of $\overleftrightarrow{AQ}$ so that S, T, and Q are noncollinear and are the corners of a triangle.

Ordering: Using the machinery of Chapter 6, order line $\mathcal{L}$ so that $A < C < B$, and order the line $\overleftrightarrow{CT}$ so that $C < S < T < D$.

Definition of Φ: For every $X \in \mathcal{L}$ such that the line $\overleftrightarrow{XQ}$ intersects $\overline{ST}$ define $\Phi(X)$ so that $\{\Phi(X)\} = \overleftrightarrow{XQ} \cap \overline{ST}$. By this definition, $\Phi(A) = S$ and $\Phi(B) = T$.

Claim 1: $\overline{ST} \subseteq \operatorname{ins}\angle SQT$, and therefore is a subset of both the T-side of $\overleftrightarrow{SQ}$ and the S-side of $\overleftrightarrow{TQ}$. This follows immediately from Definition PSH.36 and Theorem PSH.37.

Claim 2: If $X \in \mathcal{L}$ and $X < A$, the ray $\overrightarrow{QX}$ intersects $\overline{ST}$, so that $\Phi(X)$ is defined, S–$\Phi(X)$–T, and $S < \Phi(X) < T$.

See Figure 7.1. If $X < A$, X–A–C–B. Since T–Q–B, Q and B are on the same side of $\overleftrightarrow{ST} = \overleftrightarrow{CD}$; hence X and Q are on opposite sides of $\overleftrightarrow{ST}$ so by

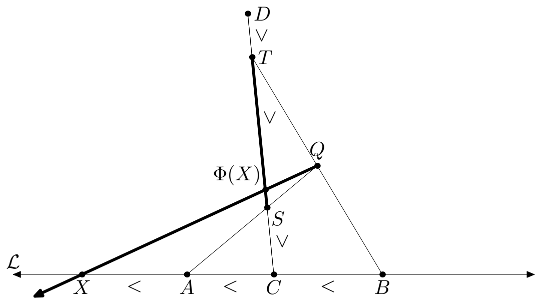


Fig. 7.1 For Claim 2 in Theorem LUPE.

Theorem PSH.12 $\overleftrightarrow{ST} \cap \overrightarrow{XQ} \neq \emptyset$. This point of intersection belongs to $\overleftrightarrow{XQ}$, and is therefore $\Phi(X)$, by definition.

Since X–A–C–B, X is on the C-side of $\overleftrightarrow{BT} = \overleftrightarrow{TQ}$, which by C–S–T is the S-side. Also, X is on the side opposite the C-side of $\overleftrightarrow{AQ} = \overleftrightarrow{SQ}$. Since C–S–T, this means $X \in$ the T-side of $\overleftrightarrow{SQ}$. By Definition PSH.36, $X \in \text{ins}\,\angle SQT$; by Theorem PSH.39 (Crossbar) $\overrightarrow{QX} \cap \overline{ST} \neq \emptyset$. As observed just above, this point of intersection is $\Phi(X)$. Then S–$\Phi(X)$–T, and $S < \Phi(X) < T$ since $S < T$.

Claim 3: If $X > B$, by property B.3 of Definition IB.1, there exists a point X' such that X–Q–X'; then the ray $\overrightarrow{QX'}$ intersects $\overline{ST}$, so that $\Phi(X)$is defined, S–$\Phi(X)$–T, and $S < \Phi(X) < T$.

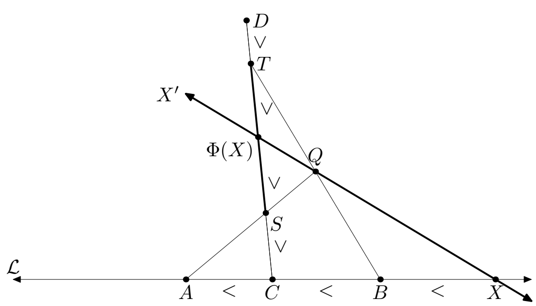


Fig. 7.2 For Claim 3 in Theorem LUPE.

See Figure 7.2. If $X > B$, A–C–B–X. Then X is on the B-side (C-side) of $\overleftrightarrow{AQ} = \overleftrightarrow{SQ}$. X is also on the side of $\overleftrightarrow{BQ} = \overleftrightarrow{TQ}$ opposite the A-side (the S-side since A–S–Q).

Since $\overleftrightarrow{XQ}$ intersects $\overleftrightarrow{AQ} = \overleftrightarrow{SQ}$ at the point Q, X' is on the side of $\overleftrightarrow{SQ}$ opposite the B-side; since B–Q–T this means that $X' \in$ the T-side of $\overleftrightarrow{SQ}$. Since $\overleftrightarrow{XQ}$ intersects $\overleftrightarrow{BQ} = \overleftrightarrow{TQ}$ at the point Q, X' is on the S-side of $\overleftrightarrow{TQ}$.

Then by Definition PSH.36, $X' \in \text{ins}\,\angle SQT$; by Theorem PSH.39 (Crossbar), $\overrightarrow{QX'} \cap \overline{ST} \neq \emptyset$, and therefore since $\overrightarrow{QX'} \subseteq \overleftrightarrow{QX}$, $\overleftrightarrow{QX} \cap \overline{ST} \neq \emptyset$. By definition this point of intersection is $\Phi(X)$. Then S–$\Phi(X)$–T; since $S < T$, by Theorem ORD.6 $S < \Phi(X) < T$.

Since $\Phi(A) = S$ and $\Phi(B) = T$, it follows from Claims 2 and 3 that for every X such that $X \leq A$ or $X \geq B$, $\Phi(X)$ is defined and belongs to $\overline{ST}$.

Claim 4: If A–X–B then $\overleftrightarrow{XQ} \cap \overline{ST} = \emptyset$, so that $\Phi(X)$ is not defined.

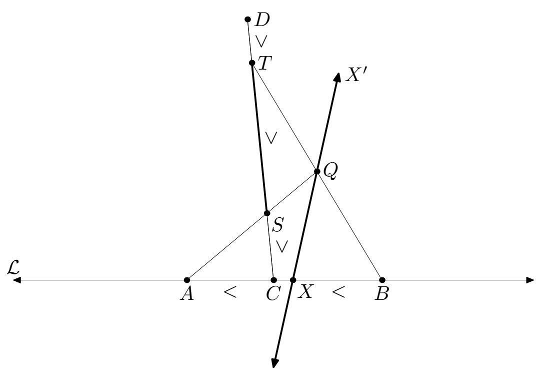


Fig. 7.3 For Claim 4 in Theorem LUPE.

See Figure 7.3. Since $X \in \overline{AB}$, by Theorem PSH.37 $X \in \text{ins}\,\angle AQB$. By Theorem PSH.38(B) $\overrightarrow{QX} \in \text{ins}\,\angle AQB$, and by Definition PSH.36 $\overrightarrow{QX}$ is a subset of the A-side of $\overleftrightarrow{BQ}$, and also of the B-side of $\overleftrightarrow{AQ}$. Now $S \in \overrightarrow{QA} \subseteq \overleftrightarrow{AQ}$, $T \in$ the T-side of $\overleftrightarrow{AQ}$, and $\overline{ST} \subseteq \text{ins}\,\angle SQT \subseteq$ the T-side of $\overleftrightarrow{AQ}$, so that $\overline{ST}$ is a subset of $\overleftrightarrow{AQ} \cup$ the A-side of $\overleftrightarrow{AQ}$, this last being the side opposite the B-side. It follows that $\overrightarrow{QX} \cap \overline{ST} = \emptyset$.

Let X' be a point such that X'–Q–X; since $\overleftrightarrow{QX}$ and $\overleftrightarrow{BQ}$ intersect at Q, X' belongs to the side of $\overleftrightarrow{BQ}$ opposite A, and by Theorem IB.14, this side contains $\overrightarrow{QX'}$. By reasoning similar to that just above, $\overline{ST} \subseteq \overleftrightarrow{BQ} \cup$the A-side of $\overleftrightarrow{BQ}$. Therefore $\overrightarrow{QX'} \cap \overline{ST} = \emptyset$.

Now $\overleftrightarrow{X'X'} = \overleftrightarrow{QX} = \overrightarrow{QX} \cup \overrightarrow{QX'} \cup \{Q\}$. Since $Q \notin \overline{ST}$ it follows that $\overleftrightarrow{XX'} \cap \overline{ST} = \emptyset$.

Claim 5: If X and Y are members of $\mathcal{L}$ and $X < Y < A$, then S–$\Phi(Y)$–$\Phi(X)$–T, and $S < \Phi(Y) < \Phi(X) < T$.

First we note that from Claim 2,S–$\Phi(Y)$–T and $S < \Phi(Y) < T$, and also S–$\Phi(X)$–T and $S < \Phi(X) < T$.

If, in the statement and proof of Claim 2, we substitute Y for A, so that $S = \Phi(A)$ becomes $\Phi(Y)$, we get a proof that $\Phi(Y)$–$\Phi(X)$–T; combining this with S–$\Phi(Y)$–T yields the desired result.

Claim 6: If X and Y are members of $\mathcal{L}$ and $B < X < Y$, then S–$\Phi(Y)$–$\Phi(X)$–T, and $S < \Phi(Y) < \Phi(X) < T$.

Again we note that from Claim 3, S–$\Phi(Y)$–T and $S < \Phi(Y) < T$; also S–$\Phi(X)$–T and $S < \Phi(X) < T$.

If, in the statement and proof of Claim 3, we substitute Y for X, and substitute X for B, so that $T = \Phi(B)$ becomes $\Phi(X)$, we get a proof that S–$\Phi(Y)$–$\Phi(X)$; combining this with S–$\Phi(X)$–T yields the desired result.

Claim 7: For every $X \leq A$ and every $Y \geq B$, $\Phi(X) < \Phi(Y)$, so that the set $\{\Phi(Y) \mid Y \geq B\}$ is the set of upper bounds for $\{\Phi(X) \mid X \leq A\}$.

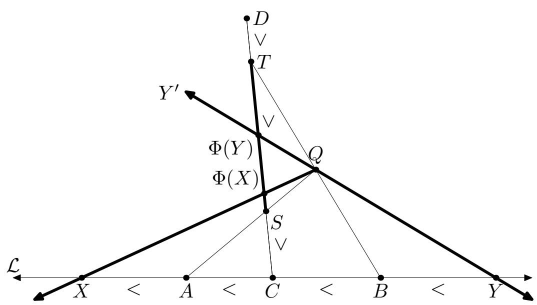


Fig. 7.4 For Claim 7 in Theorem LUPE.

See Figure 7.4. If $X = A$ and $Y = B$, then $\Phi(X) = S < T = \Phi(Y)$.
If $X = A$ and $Y > B$ then by Claim 3, $\Phi(X) = \Phi(A) = S < \Phi(Y)$.
If $Y = B$ and $X < A$ then by Claim 2, $\Phi(X) < T = \Phi(B) = \Phi(Y)$.

The only case needing more proof is where $X < A$ and $B < Y$. By Claim 3, $\Phi(Y)$ is the intersection of $\overleftrightarrow{QY}$ with $\overline{ST}$ and S–$\Phi(Y)$–T; by Claim 2, $\Phi(X)$ is the point of intersection of $\overrightarrow{QX}$ with $\overline{ST}$, and S–$\Phi(X)$–T.

Since X–A–C–B–Y, the X-side of $\overleftrightarrow{QY}$ is also the C-side; by the construction and Claim 3, C–S–$\Phi(Y)$–T, so that the C-side is also the S-side of $\overleftrightarrow{QY}$, and this is opposite the T-side. Therefore $\Phi(X) \in$ the S-side of $\overleftrightarrow{QY}$ which is opposite the T-side.

It follows from Theorem PSH.12 that $\overline{\Phi(X)T}$ intersects $\overleftrightarrow{QY}$ and since $\overline{\Phi(X)T}$ is a subset of $\overline{\overline{ST}}$, this point of intersection is $\Phi(Y)$. Therefore $\Phi(X)$–$\Phi(Y)$–T. Combining this with S–$\Phi(X)$–T we have S–$\Phi(X)$–$\Phi(Y)$–T. Since $S < T$, by Theorem ORD.6 $\Phi(X) < \Phi(Y)$.

Claim 8: The mapping Φ is a bijection of $\{X \mid X \leq A$ or $X \geq B\}$ onto $\overline{\overline{ST}}$.

By Claim 4 every point X for which $\Phi(X)$ is defined must either satisfy $X \leq A$ or $X \geq B$. Suppose $X \neq X'$ are such points; if both $X \leq A$ and $X' \leq A$, by Claims 2 and 5 $\Phi(X) \neq \Phi(X')$; if both $X \geq B$ and $X' \geq B$, by Claims 3 and 6 $\Phi(X) \neq \Phi(X')$; if $X \leq A$ and $X' \geq B$, by Claim 7 $\Phi(X) \neq \Phi(X')$. Therefore Φ is a one-to-one mapping defined on $\{X \mid X \leq A$ or $X \geq B\}$.

Now let Z be any point of $\overline{\overline{ST}}$; then the line $\overleftrightarrow{ZQ}$ intersects $\mathcal{L}$ at some point, because, by the negation of property PE, it cannot be parallel to $\mathcal{L}$. By Claim 4 there can be no point $X \in \overline{AB}$ such that $\Phi(X)$ is defined. Therefore every point of $\overline{\overline{ST}}$ is an image point of some X such that either $X \leq A$ or $X \geq B$.

Claim 9: If Axiom LUB holds, the assumption that there is no line through Q which is parallel to $\mathcal{L}$ yields a contradiction. By Claim 7, $\{\Phi(Y) \mid Y \geq B\}$ is the set of upper bounds for $\mathcal{E} = \{\Phi(X) \mid X \leq A\}$. Let U be the least upper bound for $\mathcal{E}$. Since there are upper bounds for $\mathcal{E}$ which belong to $\overline{\overline{ST}}$, $U \in \overline{\overline{ST}}$. Moreover, $U \neq S$ since $\mathcal{L}$ contains a point $X < A$, for which $\Phi(X) > S$; also $U \neq T$, because $\mathcal{L}$ contains a point $Y > B$ for which $\Phi(Y) < T$, so that $\Phi(Y)$ is an upper bound for $\mathcal{E}$.

It follows that either $\Phi^{-1}(U) < A$ or $\Phi^{-1}(U) > B$.

If the former, by property B.3 of Definition IB.1 there exists a point Y of $\mathcal{L}$ such that Y–$\Phi^{-1}(U)$–A, that is, $Y < \Phi^{-1}(U) < A$. Then by Claim 5, $\Phi(Y) > U > \Phi(A) = S$. But $\Phi(Y) \in \{\Phi(X) \mid X \leq A\}$ so that U is not an upper bound for this set, contradicting Axiom LUB.

If $B < \Phi^{-1}(U)$, by property B.3 of Definition IB.1 there exists a point Y of $\mathcal{L}$ such that B–$\Phi^{-1}(U)$–Y, that is, $B < \Phi^{-1}(U) < Y$. Then by Claim 6, $T = \Phi(B) > U > \Phi(Y)$. But $\Phi(Y) \in \{\Phi(X) \mid X \geq B\}$ and is therefore an upper bound for $\{\Phi(X) \mid X \leq A\}$, and U is not the least upper bound for this set. Again, this is a contradiction. □

Chapter 8
Existence of Midpoints in the Presence of a Parallel Axiom (NEUTM)

Dependencies: *Chapters 1, 4, 5, 6, 7, and 8 from Euclidean Geometry and its Subgeometries (Specht); Axiom PW*
Acronym: *NEUTM*

Property R.6 of *Specht* Ch.8 Definition NEUT.2 says that every segment in a neutral plane has a midpoint. In this short chapter we prove this as a theorem from the other properties (R.1 through R.5) of this definition, in the case that one of the parallel axioms holds, either Axiom PW or PS.

If we had been able to prove this without invoking a parallel axiom, we could have eliminated Property R.6 from Definition NEUT.2.

All references in this chapter of the Supplement are to *Specht*; many of them are to Chapter 8 (neutral geometry), but none of them cite anything in that chapter after Theorem NEUT.48. There are no references to other chapters of this Supplement.

For the record, we restate Definition NEUT.85 from *Specht* Chapter 8: A triangle $\mathcal{T}$ is *right* iff an angle of $\mathcal{T}$ is right. In the first theorem below we provide an alternate characterization of an *acute angle*, and prove some preliminary lemmas, some of which duplicate theorems and exercises from Chapter 8 of *Specht*.

Theorem NEUTM.1 *An angle $\angle BAC$ is acute according to Definition NEUT.81 iff there exists a point Q such that $\angle BAQ$ is right and $C \in \operatorname{ins} \angle BAQ$.*

Proof. If $\angle BAC$ is acute according to Definition NEUT.81, there exist noncollinear points E, D, and F such that $\angle BAC < \angle EDF$ and $\angle EDF$ is right. By Definition NEUT.70 there exists a point $P \in \operatorname{ins} \angle EDF$ such that $\angle BAC \cong \angle EDP$. By Theorem NEUT.38 let α be the isometry such that $\alpha(\angle EDP) = \angle BAC$ and $\alpha(\overrightarrow{DE}) = \overrightarrow{AB}$ and $\alpha(\overrightarrow{DP}) = \overrightarrow{AC}$. Then let $Q = \alpha(F)$. By Corollary NEUT.44.2 $\alpha(\angle EDF) = \angle BAQ$ is right, and by Theorem NEUT.15(11), $C \in \operatorname{ins} \angle BAQ$.

Conversely, suppose there exists a point Q such that $\angle BAQ$ is right and $C \in \operatorname{ins} \angle BAQ$. Since $\angle BAQ \cong \angle BAQ$ and $\angle BAC \cong \angle BAC$, by Definition NEUT.70 $\angle BAC < \angle BAQ$. Since $\angle BAQ$ is right, by Definition NEUT.81, $\angle BAC$ is acute. $\square$

Lemma NEUTM.2 (*cf* Theorem NEUT.84) *If A, B, and C are noncollinear points on a neutral plane, and if $\angle BAC$ is right, then $\angle ABC$ and $\angle ACB$ are both acute.*

Proof. Let P be a point on the C-side of $\overleftrightarrow{AB}$ such that $\angle ABP$ is right, and let A' be such that A'–B–A. Then by Theorem NEUT.47(A) $\overleftrightarrow{BP} \parallel \overleftrightarrow{AC}$. Since C and P are on the same side of $\overleftrightarrow{AB}$, by Exercise PSH.32 either $C \in \overrightarrow{BP}$ or $C \in \operatorname{ins} \angle ABP$ or $C \in \operatorname{ins} \angle A'BP$.

If $C \in \operatorname{ins} \angle A'BP$ then by Definition PSH.36 $C \in A'$-side of $\overleftrightarrow{BP}$ and by Theorem PSH.12 (Plane Separation Theorem), $\overline{AC}$ must intersect $\overleftrightarrow{BP}$ which is impossible since $\overleftrightarrow{BP} \parallel \overleftrightarrow{AC}$. If $C \in \overrightarrow{BP}$ then both $\overleftrightarrow{AC}$ and $\overleftrightarrow{BP}$ are lines perpendicular to $\overleftrightarrow{AB}$ and containing C, so that by Theorem NEUT.47(B) these are the same line; then $A = B$ which contradicts the hypothesis that A, B, and C are noncollinear (and hence distinct). Therefore $C \in \operatorname{ins} \angle ABP$ and $\angle ABC$ is acute. Similar reasoning, interchanging B and C, shows that $\angle ACB$ is acute. $\square$

Lemma NEUTM.3 (*cf* Exercise NEUT.20) *Let $\triangle ABC$ be a right triangle where $\angle BAC$ is the right angle; then if $E \in \overleftrightarrow{BC}$ and $\overleftrightarrow{EA} \perp \overleftrightarrow{BC}$, B–E–C. That is, E lies on the hypotenuse, but is not one of its endpoints.*

Proof. If the conclusion is false, either $E = B$, $E = C$, B–C–E or E–B–C.

If $E = B$ then $\overleftrightarrow{AB} \perp \overleftrightarrow{BC}$ so that by Theorem NEUT.47(A), $\overleftrightarrow{BC} \parallel \overleftrightarrow{AC}$; this is impossible because a triangle has three corners, so $E \neq B$. A similar argument shows that $E \neq C$.

If C–B–E let D be any point on the A-side of $\overleftrightarrow{BC}$ such that $\angle CBD$ is right. Then since $\angle CBA$ is acute (by Lemma NEUTM.2), there exists a point Q such that $\angle CBQ$ is right and $A \in \text{ins}\,\angle CBQ$. Then by Definition PSH.36 A lies on the Q-side of $\overleftrightarrow{BC}$; by Theorem NEUT.48(A) there is only one line through B which is perpendicular to $\overleftrightarrow{BC}$ and both Q and D belong to this line; moreover, both these points are on the A-side of $\overleftrightarrow{BC}$. Therefore $\overrightarrow{BD} = \overrightarrow{BQ}$ and $\angle CBQ = \angle CBD$, so by Theorem NEUTM.1 $A \in \text{ins}\,\angle CBD$, and $A \in \overrightarrow{BDC}$.

Now $\overleftrightarrow{BD} \perp \overleftrightarrow{BC}$ and $\overleftrightarrow{EA} \perp \overleftrightarrow{BC}$ so that by Theorem NEUT.47(A) $\overleftrightarrow{BD} \parallel \overleftrightarrow{EA}$. Since C–B–E, E is on the side of $\overleftrightarrow{BD}$ opposite C so by Theorem PSH.12 (Plane Separation), $\overline{EA}$ and $\overleftrightarrow{BD}$ must intersect, contradicting their parallelism. Thus C–B–E is false. A similar proof shows that B–C–E is false. $\square$

Lemma NEUTM.4 *Let $\angle ACB$ and $\angle ADB$ be two right angles on a neutral plane, where C and D are on the same side of $\overleftrightarrow{AB}$. Then neither $C \in \text{ins}\,\angle ADB$ or $D \in \text{ins}\,\angle ACB$.*

$\angle ADB$ cannot be a right angle
if $\angle ACB$ is right.

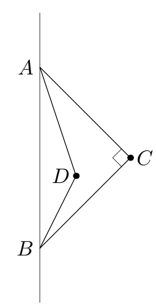


Fig. 8.1 For Lemma NEUTM.4.

Proof. See Figure 8.1. If $D \in \text{ins}\,\angle ACB$, by Definition PSH.36, $D \in A$-side of $\overleftrightarrow{BC}$ and the B-side of $\overleftrightarrow{AC}$. Since $D \in C$-side of $\overleftrightarrow{AB}$, $D \in \text{ins}\,\angle ABC$. By Theorem PSH.40, $\overline{AC} \subseteq \text{ins}\,\angle ABC$; by Theorem PSH.39, there exists a point F of intersection of $\overline{AC}$ and $\overleftrightarrow{BD}$, so that A–F–C. By Corollary 2 of Theorem PSH.39, C is on the side of $\overleftrightarrow{BD}$ opposite A. A similar argument will show that C is on the side of $\overleftrightarrow{AD}$ opposite B.

Let $E = \operatorname{ftpr}(D, \overleftrightarrow{AC})$. Apply Lemma NEUTM.3 to $\triangle ADF$; since $\angle ADF$ is right, A–E–F and hence A–E–F–C, so $E \in A$-side of $\overleftrightarrow{BD}$. Let E'–D–E; then $E' \in C$-side of $\overleftrightarrow{BD}$. Since B–D–F and F is on the opposite side of $\overleftrightarrow{DE}$ to A, $B \in A$-side of $\overleftrightarrow{DE}$.

Now $\overleftrightarrow{CB} \perp \overleftrightarrow{AC}$ and $\overleftrightarrow{DE} \perp \overleftrightarrow{AC}$ so by Theorem NEUT.47(A) $\overleftrightarrow{CB} \parallel \overleftrightarrow{DE}$. But B and C are on opposite sides of $\overleftrightarrow{DE}$, so that $\overline{CB}$ intersects $\overleftrightarrow{DE}$, contradicting their parallelism. Therefore $D \notin \operatorname{ins}\angle ACB$; interchanging C and D in the above argument shows that $C \notin \operatorname{ins}\angle ADB$. $\square$

Theorem NEUTM.5 *Let C, D, E, F be points on the neutral plane such that*

(1) *$\angle CED$ and $\angle CFD$ are right angles, and $\overleftrightarrow{EF}$ is the line of symmetry for both; and*

(2) *E and F are on opposite sides of $\overleftrightarrow{CD}$.*

Then

(A) *$\angle CED \cong \angle CFD$ and $\overleftrightarrow{CD} \perp \overleftrightarrow{EF}$;*

(B) *if $\{H\} = \overleftrightarrow{CD} \cap \overleftrightarrow{EF}$, $\overline{EH} \cong \overline{FH}$, $\overline{CH} \cong \overline{DH}$, and $\overline{EC} \cong \overline{FC} \cong \overline{ED} \cong \overline{FD}$.*

Proof. First note that the first conclusion of (A) follows from Euclid's fourth postulate, Theorem NEUT.69, but in this chapter we are not assuming anything from our development after Theorem NEUT.48, so this needs to be proved.

Since $\overleftrightarrow{EF}$ is the line of symmetry for both $\angle CED$ and $\angle CFD$, $\mathcal{R}_{\overleftrightarrow{EF}}(\overrightarrow{FC}) = \overrightarrow{FD}$ and $\mathcal{R}_{\overleftrightarrow{EF}}(\overrightarrow{EC}) = \overrightarrow{ED}$; then by elementary mapping theory

$$\begin{aligned}\{\mathcal{R}_{\overleftrightarrow{EF}}(C)\} &= \mathcal{R}_{\overleftrightarrow{EF}}(\overrightarrow{FC} \cap \overrightarrow{EC}) = \mathcal{R}_{\overleftrightarrow{EF}}(\overrightarrow{FC}) \cap \mathcal{R}_{\overleftrightarrow{EF}}(\overrightarrow{EC}) \\ &= (\overrightarrow{FD}) \cap (\overrightarrow{ED}) = \{D\}.\end{aligned}$$

By Theorem NEUT.22(A), $\overleftrightarrow{CD}$ is a fixed line for $\mathcal{R}_{\overleftrightarrow{EF}}$ and $\overleftrightarrow{CD} \perp \overleftrightarrow{EF}$. Now H, E, and F all belong to $\overleftrightarrow{EF}$, so are fixed points for $\mathcal{R}_{\overleftrightarrow{EF}}$; since $\mathcal{R}_{\overleftrightarrow{EF}}(C) = D$, by Theorem NEUT.15(5) $\overline{EC} \cong \overline{ED}$, $\overline{FC} \cong \overline{FD}$, and $\overline{HC} \cong \overline{HD}$.

By Theorem NEUT.44, $\overleftrightarrow{CD}$ is a line of symmetry for $\overleftrightarrow{EF}$. Let $B = \mathcal{R}_{\overleftrightarrow{CD}}(E)$; then B is a point on $\overleftrightarrow{EF}$ because $\overleftrightarrow{EF}$ is a fixed line for $\mathcal{R}_{\overleftrightarrow{CD}}$. Then F and B are on the same side of $\overleftrightarrow{CD}$ because they are both in the side opposite E, and are both on the same side of H. By Corollary NEUT.44.1 $\angle CBD$ is a right angle since it is congruent to $\angle CED$. We prove that $B = F$.

Suppose otherwise, that $B \neq F$; by Theorem PSH.38 $\overrightarrow{HF} = \overrightarrow{HB}$. By Definition IB.4 exactly one of H–F–B, H–B–F, or $B = F$ is true.

If $H{-}F{-}B$, then $F \in \overrightarrow{BH}$; since $H \in \overrightarrow{CD}$, by Theorem PSH.40 $H \in \text{ins}\,\angle CBD$ and hence $F \in \text{ins}\,\angle CBD$. By Lemma NEUTM.4, this is impossible. Likewise, $H{-}B{-}F$ is impossible. Therefore $F = B = \mathcal{R}_{\overleftrightarrow{CD}}(E)$; since H, C, and D are all fixed points for $\mathcal{R}_{\overleftrightarrow{CD}}$, it follows from Theorem NEUT.15(5) that $\overline{EC} \cong \overline{FC}$, $\overline{ED} \cong \overline{FD}$, and $\overline{HE} \cong \overline{HF}$. $\square$

Theorem NEUTM.6 *If either Axiom PW or PS holds, every segment $\overline{AB}$ on a neutral plane has a midpoint; this result is independent of Property R.6 of Definition NEUT.2.*

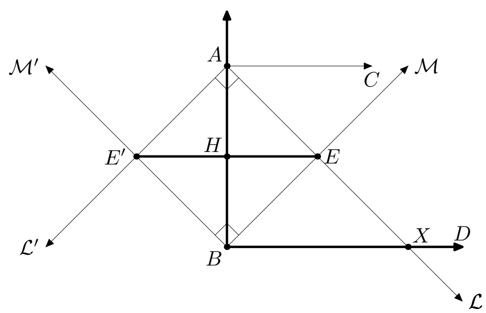


Fig. 8.2 For Theorem NEUTM.6.

Proof. See Figure 8.2. Let C and D be points on the same side of $\overleftrightarrow{AB}$ such that $\overleftrightarrow{AC} \perp \overleftrightarrow{AB}$ and $\overleftrightarrow{BD} \perp \overleftrightarrow{AB}$. Let $\mathcal{L}$ and $\mathcal{M}$ be the lines of symmetry of $\angle BAC$ and $\angle ABD$ respectively.

By Theorem NEUT.47(A), $\overleftrightarrow{AC} \parallel \overleftrightarrow{BD}$. If $\mathcal{L}$ does not intersect $\overleftrightarrow{BD}$, then the two lines are parallel; both $\overleftrightarrow{AC}$ and $\mathcal{L}$ contain A and are parallel to $\overleftrightarrow{BD}$. This is impossible by either Axiom PS or PW. Therefore $\mathcal{L}$ intersects $\overleftrightarrow{BD}$ at some point X. By Theorem NEUT.20(E) $\mathcal{L}$ contains a point $P \in \text{ins}\,\angle CAB$ and $\overrightarrow{AP} \subseteq \text{ins}\,\angle BAC$. Since $\mathcal{L}$ intersects $\overleftrightarrow{AC}$ at A, all points of the opposing ray are on the side of $\overleftrightarrow{AC}$ opposite B so that $X \in \overrightarrow{AP} \subseteq \text{ins}\,\angle BAC$. Therefore X belongs to the C-side, that is, the D-side of $\overleftrightarrow{AB}$, and thus $X \in \overrightarrow{BD}$.

By Theorem PSH.37 $\overline{XA} \subseteq \text{ins}\,\angle ABD$; by Corollary PSH.39.2 A and X are on opposite sides of $\mathcal{M}$; thus by Theorem PSH.12 (Plane Separation) $\overline{AX} \subseteq \overrightarrow{AX}$ intersects $\mathcal{M}$ at some point E.

Reflect the lines $\mathcal{L}$ and $\mathcal{M}$, and the points C, D, and E in $\overleftrightarrow{AB}$. Let $\mathcal{L}' = \mathcal{R}_{\overleftrightarrow{AB}}(\mathcal{L})$, $\mathcal{M}' = \mathcal{R}_{\overleftrightarrow{AB}}(\mathcal{M})$, and $E' = \mathcal{R}_{\overleftrightarrow{AB}}(E)$. Then $\angle ABE' \cong \angle ABE$, and by Theorem NEUT.39, $\angle ABE \cong \angle EBD$; by Theorem NEUT.14, $\angle EBD \cong \angle ABE \cong \angle ABE'$. By Exercise NEUT.40(A), $\angle EBE' \cong \angle ABD$, and therefore $\angle EBE'$ is a right angle. Similarly, $\angle EAE'$ is a right angle, and both these angles have $\overleftrightarrow{AB}$ as their common line of symmetry. Let H be the intersection of $\overleftrightarrow{EE'}$ with $\overline{AB}$. Then by Theorem NEUTM.5, $\overline{AH} \cong \overline{BH}$, so that H is a midpoint of $\overline{AB}$. $\square$

Remark NEUTM.7 We would be most grateful if a reader with more perspicacity than we should come up a proof of the existence of midpoints using only properties R.1 through R.5 of Definition NEUT.2 without invoking parallelism; this would make it possible, in the main development of *Specht* to dispense with Property R.6 of this definition.

References

1. Jordan, C.: Cours d'analyse de l'École Polytechnique, Cambridge University Press (Reprint in 3 vols.) (2013) ISBN 978-1108064699, 978-1108064705, 978-1108064712
2. Kelley, J. L.: Algebra: A Modern Introduction, Van Nostrand Reinhold (1965) ISBN 978-0442043001
3. Martin, G. E.: Transformation Geometry, An Introduction to Symmetry, Springer (1982) ISBN 978-1461256823
4. Whyburn, G. T.: Analytic Topology, AMS Colloquium Publications, Chapter VI Reprint by Literary Licensing, LLC (2012) ISBN 978-1258240264

Index

http://www.springer.com/978-3-319-23774-9

Euclidean Geometry and its Subgeometries
Specht, E.J.; Jones, H.T.; Calkins, K.G.; Rhoads, D.H.
2015, XIX, 527 p. 59 illus., Hardcover
ISBN: 978-3-319-23774-9
A product of Birkhäuser Basel