# II : Vector algebra and Euclidean geometry

> As long as algebra and geometry proceeded along separate paths their advance was slow and their applications limited. But when these sciences joined company, they drew from each other fresh vitality and thenceforward marched on at a rapid pace toward perfection.
>
> J. – L. Lagrange (1736 – 1813)

We have already given some indications of how one can study geometry using vectors, or more generally linear algebra. In this unit we shall give a more systematic description of the framework for using linear algebra to study problems from classical Euclidean geometry in a comprehensive manner.

One major goal of this unit is to give a modern and logically complete list of axioms for Euclidean geometry which is more or less in the spirit of Euclid's ***Elements***. Generally we shall view these axioms as facts about the approach to geometry through linear algebra, which we began in the first unit. The axioms split naturally into several groups which are discussed separately; namely, **incidence**, **betweenness**, **separation**, **linear measurement**, **angular measurement** and **parallelism.**

The classical idea of ***congruence*** is closely related to the idea of moving an object without changing its size or shape. Operations of this sort are special cases of ***geometric transformations,*** and we shall also cover this topic, partly for its own sake but mainly for its use as a mathematical model for the physical concept of ***rigid motion.***

In the course of discussing the various groups of axioms, we shall also prove some of their logical consequences, include a few remarks about the logical independence of certain axioms with respect to others, and present a few nonstandard examples of systems which satisfy some of the axioms but not others. The main discussion of geometrical theorems will be given in the next unit.

*Historical background*

The following edited passages from Chapter **0** of Ryan's book give some historical perspectives on the material in the next two units. The comments in brackets have been added to amplify and clarify certain points and to avoid making statements that might be misleading, inaccurate or impossible to verify.

> In the beginning, geometry was a collection of rules for computing lengths, areas and volumes. Many were crude approximations arrived at by trial and error. This body of knowledge, developed and used in [*numerous areas including*] construction, navigation and surveying by the Babylonians and Egyptians, was passed along to ... [*the Grecian culture*] ... the Greeks transformed geometry into a [*systematically*] deductive science. Around 300 B. C. E., Euclid of Alexandria organized ... [*the most basic mathematical*] knowledge of his day in such an effective fashion that [*virtually*] all geometers for the next 2000 years used his ... ***Elements*** as their starting point. ...

> Although a great breakthrough at the time, the methods of Euclid are imperfect by [*the much stricter*] modern standards [*which have been forced on the subject as it made enormous advances, particularly over the past two centuries*]. ...
>
> Because progress in geometry had been frequently hampered by lack of computational facility, the invention of analytic geometry ... [*mainly in the* 17th *century*] made simpler approaches to more problems possible. For example, it allowed an easy treatment of the theory of conics, a subject which had previously been very complicated [*and whose importance in several areas of physics was increasing rapidly at the time*] ... analytic methods have continued to be fruitful because they have allowed geometers to make use of new developments in algebra and calculus [*and also the dramatic breakthroughs in computer technology over the past few decades*]. ...
>
> Although Euclid [*presumably*] believed that his geometry contained true facts about the physical world, he realized that he was dealing with an idealization of reality. [*For example*,] he [*presumably*] did not mean that there was such a thing physically as a breadthless length. But he was relying on many of the intuitive properties of real objects.

The latter is closely related to the logical gaps in the ***Elements*** that were mentioned earlier in the quotation. In Ryan's words, one very striking example is that "Euclid ... did not enunciate the following proposition, even though he used it in his very first theorem: *Two circles, the sum of whose radii is greater than the distance between their centers, and the difference of whose radii is less than that distance, must have a point of intersection*." We shall discuss this result in Section **III.6** of the notes. There were also many other such issues; near the end of the 19th century several mathematicians brought the mathematical content of the ***Elements*** up to modern standards for logical completeness, and the 1900 publication of ***Foundations of Geometry*** by D. Hilbert (1862 – 1943) is often taken to mark the completion of this work.

Further information about the history of analytic geometry is contained in the following standard reference:

> C. B. Boyer. ***History of Analytic Geometry.*** Dover Books, New York, NY, 2004. ISBN: 0–486–43832–5.

## II.1 : Approaches to geometry

> In geometry there is no royal road.
>
> Euclid (*c*. 325 B.C.E. – *c*. 265 B.C.E.) OR
> Menaechmus (*c*. 380 B.C.E. – *c*. 320 B.C.E.)

It is elusive — and perhaps hopelessly naïve — to reduce a major part of mathematics to a single definition, but in any case one can informally describe ***geometry*** as ***the study of spatial configurations, relationships and measurements.***

Like nearly all branches of the sciences, geometry has ***theoretical*** and ***experimental*** components. The latter corresponds to the "empirical approach" mentioned in Ryan. Current scientific thought is that Relativity Theory provides the best known model for physical space, and there is experimental evidence to support the relativistic viewpoint. This means that the large – scale geometry of physical space (or space – time) is not given by classical Euclidean geometry, but the latter is a perfectly good approximation for small – scale purposes. The situation is comparable to the geometry of the surface of the earth; it is not really flat, but if we only look at small pieces Euclidean geometry is completely adequate for many purposes. A more substantive discussion of the geometry of physical space would require a background in physics well beyond the course prerequisites, so we shall not try to cover the experimental side of geometry here.

On the theoretical side, there are two main approaches to the geometry, and both are mentioned in Ryan; these are the ***synthetic*** and ***analytic*** approaches. The names arose from basic philosophical considerations that are described in the online reference

http://plato.stanford.edu/entries/analytic-synthetic/

but for our purposes the following rough descriptions will suffice:

- The ***synthetic approach*** deals with abstract geometric objects that are assumed to satisfy certain geometrical properties given by abstract axioms or postulates (in current usage, these words are synonymous). Starting with this foundation, the approach uses deductive logic to draw further conclusions regarding points, lines, angles, triangles, circles, and other such plane and solid figures. This is the kind of geometry that appears in Euclid's ***Elements*** and has been the standard approach in high school geometry classes for generations. One major advantage of such an approach is that one can begin very quickly, with a minimum of background or preparation.
- The ***analytic approach*** models points by ordered pairs or triples of real numbers, and views objects like lines and planes as sets of such ordered pairs or triples. Starting with this foundation, the approach combines deductive logic with the full power of algebra and calculus to discover results about geometric objects such as systems of straight lines, conics, or more complicated curves and surfaces. This is the approach to geometry that is taught in advanced high school and introductory college courses. One major advantage of such an approach is that the systematic use of algebra streamlines the later development of the subject, replacing some complicated arguments by straighforward calculations.

We shall take a ***combined approach*** to Euclidean geometry, in which we set things up analytically and take most basic axioms of synthetic Euclidean geometry for granted. The main advantage is that this will allow us to develop the subject far more quickly than we could if we limited ourselves to one approach. However, there is also a theoretical disadvantage that should at least be mentioned.

> ***In mathematics, logical consistency is a fundamentally important issue.*** Logically inconsistent systems always lead to conclusions which undermine the value of the work. Unfortunately, there are ***no absolute tests for logical consistency,*** but there is a very useful criterion called ***relative consistency,*** which means that if there is a logical problem with some given mathematical system then there is also a logical problem with our standard assumptions about the nonnegative integers (and no such problems have been discovered in the 75 years since relative consistency became a standard criterion, despite enormous mathematical progress during that time). Of course, it is easier to test a system for relative consistency if it is based upon fewer rather than more assumptions. The combined approach to geometry requires ***all*** the assumptions in ***both*** the synthetic and analytic approaches to the subject, and with so many assumptions there are reasons for concern about consistency questions. Fortunately, it turns out that the combined approach does satisfy the relative consistency test; a proof

requires a very large amount of work, much of which is well beyond the scope of this course, so for our purposes it will suffice to note this relative consistency and proceed without worrying further about such issues.

More specific comments on the logical issues discussed above will appear in the online document http://math.ucr.edu/~res/math144/coursenotes8.pdf.

*Setting up the combined approach*

Our geometry is an abstract geometry. The reasoning could be followed by a disembodied spirit with no concept of a physical point, just as a man blind from birth could understand the electromagnetic theory of light.

H. G. Forder (1889 – 1981)

Mathematicians are like Frenchmen; whatever you say to them they translate into their own language and forthwith it is something entirely different.

J. W. von Goethe (1749 – 1832)

Before proceeding, we shall include some explanatory comments. These are adapted from the following online document:

http://www.math.uh.edu/~dog/Math3305/Axiomatic%20Development.doc

In all deductive systems it is necessary to view some concepts as **undefined**. Any attempt to define everything ends up circling around the terms and using one to define the other. This can be illustrated very well by looking up a simple word like "point" in a dictionary, then looking up the words used in the definition, and so on; eventually one of the definitions is going to contain the original word or some other word whose definition has already been checked.

Since much of the early material below is probably covers topics that are extremely familiar, the reasons for doing so should also be clarified. It is assumed that the reader has at least some familiarity with Euclidean geometry. Our goal here is ***to deepen and widen an already established body of knowledge.***

***The synthetic setting.*** There are $2-$dimensional and $3-$dimensional versions, each of which begins with a nonempty set, which is called ***the plane*** or ***the space.*** The elements of this set are generally called ***points.*** The "undefined concepts" of ***lines*** and (in the $3-$dimensional case) ***planes*** are families of proper subsets of the plane or the space, and a point is said to *lie on a line or a plane* if and only if it is a member of the appropriate subset. There are several equivalent ways to formulate the other "undefined concepts" in Euclidean geometry, and our choices will be *a priori* notions of **(1)** ***distance between two points*** and **(2)** ***angle measurement.*** These data are assumed to satisfy certain *rules* or *geometric axioms*. These rules split naturally into several groups. We shall discuss the first of these (the Axioms of Incidence) below, and the remaining groups will be covered in the following three sections.

***The analytic setting.*** Once again, there are $\mathbf{2}$ – dimensional and $\mathbf{3}$ – dimensional versions, and the points in these respective cases are elements of $\mathbf{R}^2$ and $\mathbf{R}^3$. The lines and planes are the subsets with those names that we described in Unit **I:** Lines are translates of $\mathbf{1}$ – dimensional vector subspaces, and similarly the planes are translates of $\mathbf{2}$ – dimensional subspaces. Equivalently, lines in $\mathbf{R}^2$ and planes in $\mathbf{R}^3$ can also be described as subsets defined by nontrivial linear equations of the form $\mathbf{a} \cdot \mathbf{x} = \mathbf{b}$. The distance between two points is merely the usual distance given by $||\mathbf{Y} - \mathbf{X}||$, and we ***define*** angle measurement $\boldsymbol{\theta}$ of angle $\angle \mathbf{PXQ}$ by setting $\mathbf{a} = \mathbf{P} - \mathbf{X}$ and $\mathbf{b} = \mathbf{Q} - \mathbf{X}$, and finally taking

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

This definition might appear circular since trigonometric functions are usually defined in elementary courses by means of geometry. However, it is also possible to define the sine and cosine functions (and hence their inverses) with no formal use of geometry; in particular, this is described in Appendix **F** of Ryan.

In a completely analytic treatment of Euclidean geometry, it would be necessary to ***verify explicitly that the interpretations for "undefined concepts" satisfy the synthetic axioms.*** We shall simply ***assume*** that the analytic interpretations satisfy all of the geometric axioms to be described in this unit. Although the formal verifications of these geometric axioms are theoretically indispensable, the details of the proofs are difficult to follow in places and not necessarily enlightening for the purposes of this course.

In all approaches to Euclidean geometry, drawings and figures to illustrate proofs are important, both as indispensable motivation and as aids to understanding the arguments. Without visual intuition, it is difficult to imagine how or why classical Euclidean geometry might ever have been developed. However, in keeping with the remark by Forder, the arguments must be constructed so that the figures play no formal role in the sequence of logical deductions (see also the quotations at the end of this section).

Here are some additional references in which the synthetic and analytic approaches to the subject are developed in considerable detail:


G. D. Birkhoff, "A set of postulates for plane geometry (based on scale and protractors)," *Annals of Mathematics* (2) **33** (1932), pp. 329 – 345.

G. D. Birkhoff and R. Beatley, ***Basic Geometry*** (3rd Edition). Chelsea – American Mathematical Society, Providence, RI, 1999. ISBN: 0–821–82101–6.

H. G. Forder, ***The foundations of Euclidean geometry*** (Reprint of the original 1927 edition). Dover Books, New York, NY, 1958. ASIN: B0007F8NLG.

E. E. Moïse, ***Elementary Geometry from an Advanced Standpoint*** (3rd Edition). Addison – Wesley, Reading, MA, 1990. ISBN: 0–201–50867–2.

We shall also mention two online references. The first one goes quite far into the subject, but at the undergraduate level, and comments on axiom systems may be found in the link Euclid's Mathematical System listed there:

http://www.math.uncc.edu/~droyster/math3181/notes/hyprgeom/hyprgeom.html

The following reference contains important additional material related to G. D. Birkhoff's paper in the *Annals of Mathematics*:

http://www.math.uiuc.edu/~gfrancis/M302/handouts/postulates.pdf

*The Incidence Axioms*

> As a basis for our study we assume an arbitrary collection of entities of an arbitrary nature, entities which for brevity, we shall call points, and this quite independently of their nature.
>
> G. Fano (1871 – 1952)
>
> One must be able at any time to replace ***points, lines and planes*** with ***tables, chairs and beer mugs***.
>
> D. Hilbert (1862 – 1943)
>
> What matters in mathematics ... is not the intrinsic nature of our terms but the logical nature of their interrelations.
>
> B. Russell (1872 – 1970)
>
> The unproved postulates with which we start are purely arbitrary. They must be consistent, but they had better lead to something interesting.
>
> J. L. Coolidge (1873 – 1954)

This group of axioms describes basic properties of points, lines and planes and how they are related to each other. Sometimes these are also called the *axioms of connection* (a suitable translation of Hilbert's usage **Verknüpfung** into a single English word is elusive). The **2 –** dimensional version is relatively simple, so we shall consider it first.

**Axiom I – 1 :** *Given any two distinct points there is exactly one line that contains them.*

**Axiom I – 2 :** *Every line contains at least two points.*

It is not possible to do much with these two assumptions, but at least we can show they imply another basic property of points and lines.

**Proposition 1.** *Two distinct lines have at most one point in common.*

***Proof.*** Suppose that **L** and **M** are lines, and both contain two distinct points, say **A** and **B.** By the first axiom, there is a unique line **N** containing these points. Since **L** and **M** are lines containing these points, it follows that **L = N** and **M = N,** so that **L = M.** Therefore **L** and **M** are not distinct, and we have proved the contrapositive of the proposition. By standard rules of logic, this means that we have also proved the original statement.■

The significance of this proof is not that it tells us something previously unknown (we **expect** lines to have the given property !), but rather that it illustrates the logical relationship between basic geometric facts.

We now proceed to describe the $3$ – dimensional incidence axioms. These start with the preceding two axioms and also include the following:

**Axiom I – 3 :** *Given any three distinct points that are not contained in a line, there is exactly one plane that contains them.*

**Axiom I – 4 :** *Every plane contains at least three distinct points.*

**Axiom I – 5 :** *If* **L** *is a line and two distinct points* **A, B** *on* **L** *also lie on the plane* **P,** *then all points of* **L** *are contained in* **P.**

**Axiom I – 6 :** *If two distinct planes have one point in common, then their intersection is a line.*

We shall next describe some simple but important consequences of the $3$ – dimensional incidence axioms.

**Proposition 2.** *If* **L** *is a line and* **X** *is a point not on* **L,** *then there is a unique plane* **P** *which contains* **X** *and (all the points of)* **L.**

***Proof.*** We know that the line **L** contains at least two points, say **A** and **B.** There is no line containing the three points **A, B** and **X,** for the only line containing the first two points is the line **L** and **X** does not lie on **L.** Let **P** be the unique plane containing the points **A, B, X.** We claim that **P** contains **X** and **L,** and in fact **P** is the only such plane. To see the first part, note that since **A** and **B** lie on **P** we also know that **L** is contained in **P** by the fifth axiom. Therefore **P** is a plane which contains **L** and **X.** To see that **P** is the only such plane, note that every plane **Q** which contains **L** and **X** automatically contains **A, B** and **X;** since **P** is the only such plane, it follows that **Q** must be equal to **P.**■

**Proposition 3.** *If* **L** *and* **M** *are distinct lines that have one point in common, then there is a unique plane* **P** *which contains both* **L** *and* **M.**

In other words, ***two intersecting lines in space determine a unique plane.***

***Proof.*** Let **X** be the point where **L** and **N** meet. Since lines have at least two points, we know there is a second point **A** on **L** and a second point **B** on **M.**

We claim that there is no line containing all of the points **A, B, X;** if so, this means there is a unique plane **P** containing them. If there is a line **N** containing the given three points, then by the uniqueness of lines containing two points we have **L = N** and **M = N,** contradicting our assumption that **L** and **M** are distinct.

By the conclusions of the preceding paragraph and the fourth axiom, we know that the plane **P** contains both **L** and **M.** To complete the proof we need to show that **P** is the only such plane. But this follows directly, for if **Q** is a plane containing both lines then it will also contain **A, B** and **X;** since **P** is the only such plane, it follows that **Q** must be equal to **P.**■

In many situations it is important to have explicit analytic information about the lines or planes containing a given configuration. It is generally easy to do this for the axioms described above.

***<u>Examples,</u>*** **1.** If **a, b, c** are noncollinear points (not contained on any line), then the unique plane containing them is equal to **a + W,** where **W** is the $2-$ dimensional vector subspace spanned by **b – a** and **c – a.**

**2.** If two distinct planes **P, Q** are defined by nontrivial equations of the forms $\mathbf{a}\cdot\mathbf{x} = b$ and $\mathbf{c}\cdot\mathbf{x} = d$, and there is a point $\mathbf{y_0}$ which lies on both planes, then the line of intersection consists of all vectors expressible as $\mathbf{y_0} + t(\mathbf{a}\times\mathbf{c})$ for some scalar $t$.

## *Finite incidence planes*

> There is no branch of mathematics, however abstract, which may not some day be applied to phenomena of the real world.
>
> N. Lobachevsky (1793 – 1856)

The fictional writings of Charles Dickens (1812 – 1870) and Alexandre Dumas the elder (1802 – 1870) are filled with incredible coincidences in which characters turn out to have more ties to each other than a reader would initially expect. Similar things happen repeatedly in the relatively real world of mathematics. One particularly striking and easily stated example is the appearance of the geometrically important number $\pi$ in the theory of mathematical probability. Our interest here lies with some novel examples of mathematical systems satisfying the incidence axioms given above.

There are many different ways of constructing mathematical systems satisfying the $2-$ and $3-$ dimensional incidence axioms given above. In particular, here is one class of examples that are fairly trivial.

***<u>Planes with two – point lines.</u>*** Let **S** be any set which contains at least three elements, and define lines to be subsets with exactly two elements, Then lines are nonempty proper subsets and they satisfy incidence axioms $\mathbf{I-1}$ and $\mathbf{I-2}$. In particular, if we take **S** to be a set with exactly three elements, then we have a ***finite geometry.***

There are also many other examples of such finite geometries, but before presenting any of them we should address a fundamental question:

***<u>Are finite geometries anything more than abstract intellectual curiosities?</u>***

Perhaps the best way to do so is to consider a question that is possibly interesting in its own right: Suppose that we have a long (but finite) list of golfers and we wish to arrange a tournament consisting of several matches so that the following requirements will be satisfied:

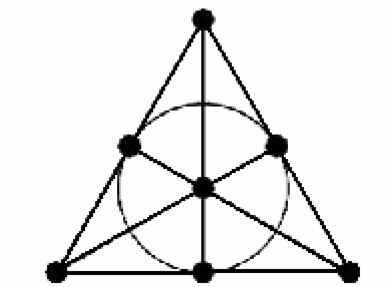
1. Every pair of players will play against each other exactly once.
2. The number of players in every match will be at least some lower limit $L$, but no greater than some upper limit $U$.
3. The number of matches will not exceed some upper limit $N$.

Finding a solution to this problem can be viewed as searching for a model of the first two incidence axioms with certain additional properties. Specifically, the "plane" is the list of golfers, the "points" are the golfers themselves, and the "lines" are the lists of golfers who play in the various matches. The second and third conditions amount to assuming that the number of points on a line must be between $L$ and $U$, and the number of points in the plane is at most $N$.

Geometrical and combinatorial ideas allow one to answer such questions in numerous cases. For example, suppose we want to assume that the number of players in a given match is always $3$ or $4$, and suppose we also want to add the following condition, perhaps in order to regulate the number of matches:

4. For each pair of matches, there is one golfer who plays in both of them.

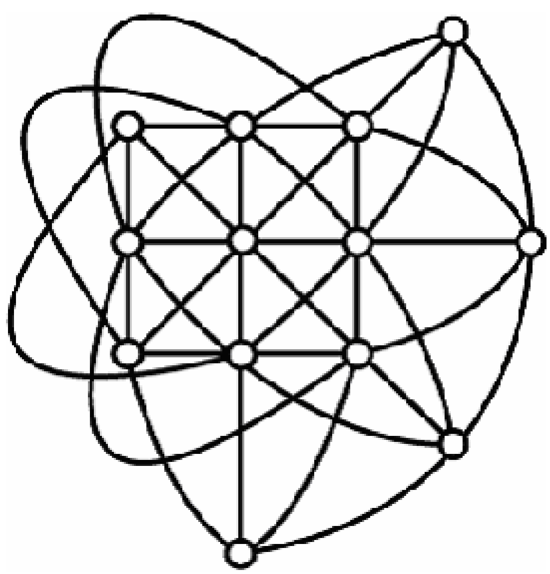
If we insist that each match will have a threesome or foursome of golfers, then there is a tournament with the specified properties. In particular, the following picture (showing the ***Fano plane***) suggests a tournament of seven games for seven golfers in threesomes:



(Sources: http://en.wikipedia.org/wiki/Finite_geometry , http://en.wikipedia.org/wiki/Projective_plane )

As in the preceding discussion, the points correspond to the seven golfers, and the seven curves (six straight lines plus a circle) correspond to the seven matches.

Here is an example involving foursomes, with $13$ golfers and $13$ matches:



(Source: http://home.wlu.edu/~mcraea/Finite_Geometry/NoneuclideanGeometry/Prob14ProjPlane/problem14.html)

Further information on similar objects appears in the online reference cited below (however, it is written at a somewhat higher level than this course):

http://planetmath.org/encyclopedia/FiniteProjectivePlane4.html

More serious and systematic applications of finite geometries arise in subjects such as coding theory and experimental design. The following online reference discusses some of these:

http://www.mnstate.edu/peil/geometry/C1AxiomSystem/4summary.htm

## II.2 : Synthetic axioms of order and separation

Man is going to err so long as he is striving.

J. W. von Goethe, ***Faust***

Sufficient to the day is the rigor thereof [***cf.*** Matthew 6:34 in the King James Bible].

E. H. Moore (1862 – 1932)

Very few books have ever been as widely circulated or as influential as Euclid's ***Elements***, and in many respects it serves as an excellent model for organizing and presenting a subject by means of deductive logic. More details on this can be found in the following online document:

http://math.ucr.edu/~res/math153/history03.pdf

For many centuries the ***Elements*** was the definitive work for the mathematics it covered, but during the 19th century several mathematicians noticed that the ***Elements*** did not treat certain geometrical points in a logically complete (or ***rigorous***) manner. The first difficulty is the attempt to define basic terms line point, line and plane at the beginning of the first book. As noted in Section **I.1**, a deductive treatment of geometry must start with some things that are undefined and subsequently define other concepts in terms of these primitive objects. One can overcome this problem in the ***Elements*** quite easily by simply agreeing that the definitions at the beginning are only meant to provide the crucial physical motivation and are not part of the formal deductive structure. More serious problems arise with other necessary terms that are not defined and results that are used without either proving them earlier or assuming them explicitly.

> In fact, the first deficiency of this kind arises in the very first proposition of the Elements, which aims to prove that one can construct an equilateral triangle whose base is a given line segment, say **[AB]**. The idea is simple: One draws the circle with center **A** passing through **B** and the circle with center **A** passing through **B.** If we do this, we "see" that the circles intersect in two points, one on each side of the line **AB,** and either of these points together with **A** and **B** will give the vertices of an equilateral triangle which has **[AB]** as one of its sides. If we carry this out by drawing on a sheet of paper with an ordinary compass, everything works just fine. However, the conclusion does not follow logically from the material introduced before the statement and proof of the proposition; in particular, this was noted by G. W. von Leibniz (1646 – 1716), who is better known in mathematics for his role in the development of calculus. We need

some additional assumption which yields information about the intersections of two circles in order to make the argument rigorous. We shall discuss this further in the final section of Unit **III.**

Another problem involving a concept called ***superposition*** will be treated in Section **II.4** (since Euclid used this idea only twice when it could have been used at many other points to simplify proofs, he may have suspected it was problematic). In this section we shall consider the following notions that generally received inadequate attention in the ***Elements***:

1. The concept of **betweenness** for three points on the same line.
2. The concepts of ***two sides of a line*** in a plane, and ***two sides of a plane*** in space.
3. The concepts of ***interiors*** and ***exteriors*** for angles, triangles and other such figures.

The need for more specific treatment of such matters (generally known as ***order and separation properties***) was noted explicitly by C. F. Gauss (1777 – 1855), and during the 19th century mathematicians developed logically rigorous methods for bridging the gaps. Fortunately, these issues do not lead to problems with the ultimate correctness of any conclusions or results in the ***Elements.*** However, they do have practical as well as theoretical significance, and there are examples to show that a lack of careful attention to the concepts listed above can lead to outrageous mistakes. At the end of this section we shall describe a standard looking "proof" due to W. W. Rouse Ball (1850 – 1925), which at first may seem very reasonable and similar to arguments in elementary geometry textbooks but "proves" the ridiculous conclusion that ***every triangle is isosceles.* —** As we shall note in our discussion, the mistake in the proof involves a lack of adequate attention to order and separation properties. In other contexts it might not be so obvious that carelessness with such points leads to a false conclusion, and *the only way to eliminate such unwelcome mistakes is to set things up as carefully and completely as possible*.

*Balancing logical correctness, motivation and intelligibility*

> Mathematics [sometimes] consists of proving the most obvious thing in the least obvious way.
>
> When introduced at the wrong time or place, good logic may be the worst enemy of good teaching.
>
> G. Polyá (1887–1985)
>
> If [most of] the best mathematicians did not recognize the need for these axioms and theorems for over two thousand years, how can we expect young people to see the need for them?
>
> M. Kline (1908 – 1992)
>
> Sir, I have found you an argument. I am not obliged to find you an understanding.
>
> Samuel Johnson (1709 – 1784)

This [a theorem entirely unrelated to order and separation] … contains something to displease everybody.

J. F. Adams (1930 – 1989)

Unfortunately, although the modern mathematical treatment of order and separation properties resolves some nontrivial logical deficiencies in the ***<u>Elements</u>***, it also generates some major pedagogical difficulties that can be difficult to resolve. Most of the main points involving order and separation are intuitively fairly clear, but giving an organized and logically sound account of these matters turns out to be more complicated than one might expect at first glance. No matter what framework one selects, the basic assumptions, definitions and theorems generally tend to be obvious looking statements with dull and unmotivated proofs. The use of linear algebra does simplify things somewhat, but only to a limited extent. Kline discussed the problem at considerable length in the following book:

> M. Kline, ***Why Johnny Can't Add: The Failure of the New Math.*** Random House, New York, 1974. ISBN: 0–394–71981–6.

This book takes some highly controversial positions on many issues which are far beyond our scope, but there is widespread agreement that the following highly edited passage raises a point that must be taken into account:

> Another consequence … is that a host of ~~trivial~~ theorems must be proved before the significant ones [*from classical geometry*] are reached. The number of ~~minor~~ theorems is so large that the major features of the subject [*can very easily*] fail to stand out.

A similar opinion appears in the following passage from the State of California standards for teacher certification in mathematics:

> One should de–emphasize the proofs of simple theorems that come near the beginning of the axiomatic development. The proofs of such theorems are harder to learn than those of theorems that follow, and this is true not only for beginners but also for professional mathematicians as well. These proofs also tend to be tedious and uninspiring.

Yet another way of expressing the problem is to say the ***logical*** and ***psychological*** orders of the subject matter are quite different.

So what does this mean for geometry courses? At the high school level it seems clear that one should restrict the rigorous discussion or order and separation properties to a bare minimum, but some intuitive discussion of the underlying concepts and facts seems unavoidable. It would probably be appropriate to mention the ultimate need for a more rigorous approach to order and separation properties and that this can be worked out if one approaches everything from a more advanced standpoint. In any case, all these considerations raise questions about presenting the usual proofs from high school geometry, and the following passage from the California standards makes the following specific suggestion:

> One way to [introduce] … the proofs of more substantive theorems as soon as possible is to adopt the method of "local axiomatics," which is to list the facts one needs for a particular proof, and then proceed to construct the proof on the basis of these facts. This approach mirrors the axiomatic method because, in effect, these facts are the "axioms" in this particular setting.

Kline summarizes the situation very effectively in one sentence:

> Students can be much more readily attracted to the fruits rather than to the roots of mathematics.

The implications for an upper division undergraduate course are related but differ in some key respects. As noted in the California standards, a complete and rigorous account of elementary geometry is also challenging at the graduate level or perhaps even higher. On the other hand, the discussion of order and separation needs to be more detailed, if for no other reason that the instructor of a high school course in geometry should know the subject matter than he or she is required to teach, and in particular should have some understanding of the hidden issues in the subject such as the ultimate need for proofs involving order and separation. We have already noted the usefulness of linear algebra for analyzing and understanding such issues. Some further general suggestions for studying the material of this section are given below.

### *The betweenness relation*

> All human knowledge begins with intuition, thence proceeds to concepts and ends with [abstract] ideas.
>
> I. Kant (1724 – 1804), ***Critique of Pure Reason***, Elementarlehre, Pt. 2, Sec. 2

Before we begin presenting definitions and logical consequences, we shall make a few general remarks related to the previous discussion. We have noted that much if not all of the content will seem intuitively clear to many if not all readers, and this intuitive transparency clashes with the logical need for a slow, deliberate, perhaps even boring presentation of the material. A reader who finds the treatment too tedious or dense may find it better to pick and choose material using the following priorities:

1. The statements of the axioms, the definitions (and other terminology) and the statements of the results are the most important points; *it is more important to recognize the basic contents and the correctness of the theorems than it is to follow the proofs*.
2. A solid understanding of what the preceding items mean in terms of coordinates and linear algebra is nearly as important as the first priority.
3. The understanding of the proofs themselves is a distant third in importance.

These guidelines are designed to be consistent with the previously quoted passage from the California standards for teaching certification in mathematics.

***Managing conflicts between logical correctness and clarity.*** The preceding discussion reflects an important fact: *An argument which is logically complete is not necessarily easier to understand than an argument that is logically incomplete*. In particular, the proof of an "obvious" statement may require a great deal of work and some lengthy digressions from the mainstream of the argument. Such interruptions can cause a reader to lose track of the main ideas in a proof. In such cases it may be good to view such potential distractions as logical subroutines; a basic fact is needed to take the argument to the next step, but it may require some work to justify, so a reader might simply assume the crucial point has been justified, proceed with the rest of the argument, and come back to the justification later when it will not disrupt the chain of thought. Most professional mathematicians use this approach frequently, and there are obvious analogs in writing computer programs when the programmer inserts subroutines at appropriate steps. This may also be viewed as a form of "working backwards" to complete a proof, a rearrangement of the material in psychological rather than logical order, or as a splitting of the proof into pieces that can be handled separately by the notion of "local axiomatics" described earlier.

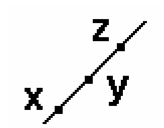
Considerably more could be said about these issues, but we have already said a great deal, and ***we shall now proceed to the mathematics itself.***

<u>**Definition.**</u> Let **x, y, z** be three distinct collinear points (in our plane or $3-$dimensional space). We shall say **y** <u>*is between*</u> **x** <u>*and*</u> **z**, and write **x*y*z,** if we have the distance equation: $d(\mathbf{x},\mathbf{z}) = d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z})$. By the symmetry of this equation we see that

**y** <u>*is between*</u> **x** <u>*and*</u> **z** **if and only if** **y** <u>*is between*</u> **z** <u>*and*</u> **x.**

Frequently one says that the points **x, y, z** <u>*are in the*</u> <u>***order***</u> **x*y*z,** and for this reason the basic properties of betweenness are frequently called ***axioms of order.*** Results from Section **I.1** and some simple algebraic rewriting yield the following alternate characterizations of betweenness.

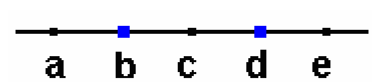
<u>**Proposition 1.**</u> *Given three distinct collinear points* **x, y, z** *the relation* **x*y*z** *holds if and only if* $\mathbf{y} = \mathbf{x} + t(\mathbf{z}-\mathbf{x})$ *where* $0 < t < 1$, *or equivalently if and only if we have* $\mathbf{y} = s\mathbf{x} + t\mathbf{z}$ *where* $s$ *and* $t$ *are positive and* $s + t = 1$.




<u>***Proof.***</u> The first equivalence follows from the result in Section **I.1** on the conditions under which the Triangle Inequality becomes an equation, and the second follows by simply expressing **y** as a linear combination of **x** and **z** on one hand and as a linear combination of **x** and **z – x** on the other.■

We shall now state several basic properties of betweenness that may be taken as axioms in the synthetic approach.

<u>**Axiom B – 1 :**</u> *Given any two distinct points* **b** *and* **d,** *there exist points* **a, c,** *and* **e** *lying on the line* **bd** *such that* **a*b*d, b*c*d,** *and* **b*d*e.**




<u>**Axiom B – 2 :**</u> *If* **a, b** *and* **c** *are three distinct points lying on the same line, then one and only one of the points is between the other two.*

We need the concept of betweenness in order to define several frequently used subsets of lines.

<u>**Definition.**</u> Let **A** and **B** be distinct points.

- The ***closed segment*** **[AB]** consists of all **X** on line **AB** such that one of **X** = **A, X** = **B,** or **A*X*B** is true.

- The ***open segment*** **(AB)** consists of all **X** on **AB** such that **A∗X∗B** is true.
- The ***closed ray*** **[AB** consists of all **X** on line **AB** such that one of **X** = **A, X** = **B**, **A∗X∗B,** or **A∗B∗X** is true.
- The ***open ray*** **(AB** consists of all **X** on **AB** such that one of **X** = **B**, **A∗X∗B,** or **A∗B∗X** is true.
- The ***opposite closed ray*** $\mathbf{[AB}^{\mathbf{OP}}$ consists of all **X** on line **AB** such that **X** = **A** or **X∗A∗B** is true.
- The ***opposite open ray*** $\mathbf{(AB}^{\mathbf{OP}}$ consists of all **X** on **AB** such that **X∗A∗B** is true.

The algebraic characterizations of these sets will be fundamentally important for our purposes. Before stating and proving these characterizations, we must descrive the meanings of various betweenness possibilities algebraically.

**<u>Theorem 2.</u>** *Suppose that* **a, b** *and* **x** *are distinct collinear points, and express* **x** *in the form* $\mathbf{a} + t(\mathbf{b} - \mathbf{a})$ *for some (uniquely determined) scalar* $t$. *Then the following hold:*

> **(1)** **x** *is between* **a** *and* **b** *if and only if* $0 < t < 1$.
> **(2)** **a** *is between* **x** *and* **b** *if and only if* $t < 0$.
> **(3)** **b** *is between* **x** *and* **a** *if and only if* $t > 1$.

***<u>Proof.</u>*** There are exactly five mutually exclusive possibilities for $t$ and five mutually exclusive possibilities for the relation of **x** to **a** and **b.** We claim it will suffice to have the following conclusions.

> **(1)** If $t = 0$ then **x = a.**
> **(2)** If $t = 1$ then **x = b.**
> **(3)** If $0 < t < 1$ then **x** is between **a** and **b.**
> **(4)** If $t < 0$ then **a** is between **x** and **b.**
> **(5)** If $t > 1$ then **b** is between **x** and **a.**

These statements immediately imply the "if" implications in the theorem. To see that they also imply the converses, proceed as follows: If **x** is between **a** and **b,** we may use the previous result on betweenness to conclude that $0 < t < 1$. If **a** is between **x** and **b,** then by **(1) – (5)** any condition other than $t < 0$ will yield a condition that contradicts the known betweenness relation. Similarly, if **b** is between **x** and **a,** then by **(1) – (5)** any condition other than $t > 1$ will yield a condition that contradicts the known betweenness relation. Therefore the "only if" implications also hold.

Thus we are reduced to verifying **(1) – (5).** The first two are trivial, and the third follows from the eariler characterization of betweenness. Suppose that we have $\mathbf{a} + t(\mathbf{b} - \mathbf{a})$, where $t < 0$. Then we have $(\mathbf{a} - \mathbf{x}) = |t|(\mathbf{b} - \mathbf{a})$ so that

$$(\mathbf{b} - \mathbf{x}) = (\mathbf{b} - \mathbf{a}) + (\mathbf{a} - \mathbf{x}) = (\mathbf{b} - \mathbf{a}) + |t|(\mathbf{b} - \mathbf{a})$$

which in turn implies $\|\mathbf{b} - \mathbf{x}\| = \|\mathbf{b} - \mathbf{a}\| + \|\mathbf{a} - \mathbf{x}\|$, so that **a** is between **b** and **x** by the previous proposition. Now suppose that $\mathbf{a} + t(\mathbf{b} - \mathbf{a})$, where $t < 1$. Then we have $(\mathbf{x} - \mathbf{a}) = t(\mathbf{b} - \mathbf{a})$, so that

$$(\mathbf{x} - \mathbf{b}) = (\mathbf{x} - \mathbf{a}) - (\mathbf{b} - \mathbf{a}) = (t - 1)(\mathbf{b} - \mathbf{a})$$

and since $t - 1 > 0$ the latter implies that $\|\mathbf{x} - \mathbf{a}\| = \|\mathbf{b} - \mathbf{x}\| + \|\mathbf{b} - \mathbf{a}\|$, so that **b** is between **a** and **x** by the previous proposition.■

It is now an easy exercise to translate the preceding into information about the six subsets defined above.

**Theorem 3.** *Suppose that* **a, b** *and* **x** *are distinct collinear points, and express* **x** *in the form* $\mathbf{a} + t(\mathbf{b} - \mathbf{a})$ *for some (uniquely determined) scalar* $t$. *Then the following hold*:

1. *The point* **x** *lies on* **[ab]** *if and only if* $0 \le t \le 1$.
2. *The point* **x** *lies on* **(ab)** *if and only if* $0 < t < 1$.
3. *The point* **x** *lies on* **[ab** *if and only if* $t \ge 0$.
4. *The point* **x** *lies on* **(ab** *if and only if* $t > 0$.
5. *The point* **x** *lies on* $\mathbf{[ab}^{\mathbf{OP}}$ *if and only if* $t \le 0$.
6. *The point* **x** *lies on* $\mathbf{(ab}^{\mathbf{OP}}$ *if and only if* $t < 0$.

These follow immediately from **(1) – (5)** in the proof of the previous theorem.■

Here is a typical elementary consequence of the preceding material. Once again, the conclusion is what one would expect:

**Proposition 4.** *Suppose that* **A, B, C, D** *are four distinct collinear points satisfying the conditions* **A∗B∗D** *and* **B∗C∗D**. *Then* **A∗B∗C** *and* **A∗C∗D** *also hold*.

***Proof.*** The simplest characterization of betweenness is the additivity of distances, and this is what we shall use. The assumptions imply that $d(\mathrm{A}, \mathrm{D}) = d(\mathrm{A}, \mathrm{B}) + d(\mathrm{B}, \mathrm{D})$ and $d(\mathrm{A}, \mathrm{C}) = d(\mathrm{A}, \mathrm{B}) + d(\mathrm{B}, \mathrm{C})$, and if we combine these we obtain the equation

$$d(\mathrm{A}, \mathrm{D}) = d(\mathrm{A}, \mathrm{B}) + d(\mathrm{B}, \mathrm{C}) + d(\mathrm{C}, \mathrm{D}).$$

Since $d(\mathrm{A}, \mathrm{C}) \le d(\mathrm{A}, \mathrm{B}) + d(\mathrm{B}, \mathrm{C})$ and $d(\mathrm{A}, \mathrm{D}) \le d(\mathrm{A}, \mathrm{C}) + d(\mathrm{C}, \mathrm{D})$ we have

$$d(\mathrm{A}, \mathrm{C}) + d(\mathrm{C}, \mathrm{D}) \le d(\mathrm{A}, \mathrm{B}) + d(\mathrm{B}, \mathrm{C}) + d(\mathrm{C}, \mathrm{D}) =$$

$$d(\mathrm{A}, \mathrm{D}) \le d(\mathrm{A}, \mathrm{C}) + d(\mathrm{C}, \mathrm{D}).$$

Since the left and right hand expressions are identical, the inequalities in the preceding expression must be equalities; therefore we have $d(\mathrm{A}, \mathrm{D}) \le d(\mathrm{A}, \mathrm{C}) + d(\mathrm{C}, \mathrm{D})$ and hence **A∗C∗D** holds. Furthermore, if we subtract $d(\mathrm{C}, \mathrm{D})$ from the first and second expressions in the display above we obtain

$$d(\mathrm{A}, \mathrm{C}) = d(\mathrm{A}, \mathrm{B}) + d(\mathrm{B}, \mathrm{C})$$

and hence **A∗B∗C** holds.■

Given three collinear points **a, b** and **c** it is frequently important to understand the relationships among the closed rays **[ab, [ac,** $\mathbf{[ab}^{\mathbf{OP}}$, and $\mathbf{[ac}^{\mathbf{OP}}$, and likewise for the relationships among the corresponding open rays **(ab, (ac,** $\mathbf{(ab}^{\mathbf{OP}}$, and $\mathbf{(ac}^{\mathbf{OP}}$. The following result addresses these questions.■

**Theorem 5.** *Let* **a, b, c** *be three distinct collinear points. Then we have either* $\mathbf{c} \in \mathbf{(ab}$ *or else* $\mathbf{c} \in \mathbf{(ab}^{\mathbf{OP}}$.. *In the first case we have*

$$[\mathbf{ab} = [\mathbf{ac},\ [\mathbf{ab}^{\mathbf{OP}} = [\mathbf{ac}^{\mathbf{OP}},\ (\mathbf{ab} = (\mathbf{ac},\ \textit{and}\ \ (\mathbf{ab}^{\mathbf{OP}} = (\mathbf{ac}^{\mathbf{OP}}.$$

*In the second case we have*

$$[\mathbf{ab} = [\mathbf{ac}^{\mathbf{OP}},\ [\mathbf{ab}^{\mathbf{OP}} = [\mathbf{ac},\ (\mathbf{ab} = (\mathbf{ac}^{\mathbf{OP}},\ \textit{and}\ \ (\mathbf{ab}^{\mathbf{OP}} = (\mathbf{ac}.$$

***Proof.*** We have $\mathbf{c} - \mathbf{a} = k(\mathbf{b} - \mathbf{a})$ where $k$ is a constant not equal to $0$ or $1$. By the previous result on rays we know that $k$ is positive if and only if **c** lies on **(ab** and $k$ is negative if and only if **c** lies on $(\mathbf{ab}^{\mathbf{OP}}$. This proves the first part. For the second, note that a point **x** on the line **ab** may be written in the form

$$\mathbf{x} = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = \mathbf{a} + t\,k^{-1}(\mathbf{c} - \mathbf{a}).$$

If $k$ is positive then the coefficients of $(\mathbf{b} - \mathbf{a})$ and $(\mathbf{c} - \mathbf{a})$ have the same signs, and if $k$ is negative then these coefficients have opposite signs. If we combine this with the previous result on rays, we obtain the remaining conclusions in the result.■

*Separation axioms*

There are two more axioms related to $\mathbf{B-1}$ and $\mathbf{B-2}$, and they are stronger than the latter. Given a plane **P**, a line **L** contained in **P**, and a point **X** which is on **P** but not **L**, then experience suggests that every other point **Y** on **P** satisfies exactly one of the following three conditions:

- **X** and **Y** lie on the same side of **L.**
- **Y** lies on **L.**
- **X** and **Y** lie on opposite sides of **L.**

Similarly, if we are given a plane **P** in space and a point **X** not on **P**, then experience suggests that every other point **Y** satisfies exactly one of the following three conditions:

- **X** and **Y** lie on the same side of **P.**
- **Y** lies on **P.**
- **X** and **Y** lie on opposite sides of **P.**

In order to state the axioms it is helpful to introduce a standard definition:

**Definition.** A subset of **K** of the plane or space is said to be ***convex*** if $\mathbf{a}, \mathbf{b} \in \mathbf{K}$ implies that the entire segment **[ab]** is contained in **K.** Alternatively, **K** is convex if and only if $\mathbf{a}, \mathbf{b} \in \mathbf{K}$ and $\mathbf{a}*\mathbf{x}*\mathbf{b}$ implies $\mathbf{x} \in \mathbf{K}$.

***Examples.*** **1.** Lines are convex by the definition of betweenness, and by incidence axiom $\mathbf{I-5}$ we know that planes are convex.

**2.** Open and closed rays are convex. **PROOF:** Suppose that we are given a closed ray **[ab** with $\mathbf{x}, \mathbf{y} \in [\mathbf{ab}$. Then we have $\mathbf{x} = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ and $\mathbf{y} = \mathbf{a} + s(\mathbf{b} - \mathbf{a})$ where $s$ and $t$ are nonnegative scalars. Suppose now that **z** is between **x** and **y**, so that we have $\mathbf{z} = \mathbf{x} + u(\mathbf{y} - \mathbf{x})$ where $0 < u < 1$. Straightforward algebraic computation shows that $\mathbf{z} = \mathbf{a} + v(\mathbf{b} - \mathbf{a})$, where $v = t + u(t - s) = (1 - u)t + us$. Now all the numbers $s, t, u, (1 - u)$ are nonnegative, and therefore $v$ is also nonnegative, which implies $\mathbf{z} \in$ **[ab.** A similar argument works for **(ab,** the main differences being that $s$ and $t$ are now positive and this is enough to imply that $v$ is also positive.■

The following result is often useful for showing that sets are convex.

**Proposition 6.** *The intersection of two convex sets is convex.*

***Proof.*** Let $\mathbf{K_1}$ and $\mathbf{K_2}$ be convex sets, and let **a** and **b** belong to their intersection. Then we have $\mathbf{a}, \mathbf{b} \in \mathbf{K_1}$ and $\mathbf{a}, \mathbf{b} \in \mathbf{K_2}$. Since $\mathbf{K_1}$ and $\mathbf{K_2}$ are convex, it follows that the segment **[ab]** is contained in both $\mathbf{K_1}$ and $\mathbf{K_2}$, so that **[ab]** is also contained in the intersection $\mathbf{K_1} \cap \mathbf{K_2}$. Therefore $\mathbf{K_1} \cap \mathbf{K_2}$ is convex.■

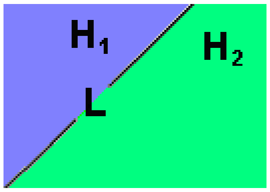
**Corollary 7.** . *Open and closed segments are convex.*

***Proof.*** Let **A** and **B** be distinct points. Then by definition the open segment **(AB)** is equal to the intersection of **(AB** and **(BA,** and the closed segment **[AB]** is equal to the intersection of **[AB** and **[BA.**■

We shall now state the two axioms.

**Axiom B – 3** (*Plane Separation Postulate*) **:** *Given a plane* **P,** *a line* **L** *contained in* **P,** *and a point* **X** *which is on* **P** *but not* **L,** *the set of all points on* **P** *but not on* **L** *is a union of two disjoint, nonempty convex subsets* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *such that if* $\mathbf{Y_1} \in \mathbf{H_1}$ *and* $\mathbf{Y_2} \in \mathbf{H_2}$ *then the open segment* $(\mathbf{Y_1\,Y_2})$ *and the line* **L** *have a point in common.*

Here is a picture illustrating the Plane Separation Postulate:




The subsets $\mathbf{H_1}$ and $\mathbf{H_2}$ are called the ***half – planes*** determined by the line **L,** or the ***sides*** of **L** in **P.** If **A** is a point which lies on, say, $\mathbf{H_1}$ then we shall say that $\mathbf{H_1}$ is the *side of* **L** *containing* **A** and $\mathbf{H_2}$ is the *side of* **L** *opposite* **A.** Similarly, if **A** and **B** are points of **P** that do not lie on **L,** we shall say that **A** *and* **B** *lie on the same side of* **L** if they both lie in either $\mathbf{H_1}$ and $\mathbf{H_2}$ and that **A** *and* **B** *lie on opposite sides of* **L** if one lies in $\mathbf{H_1}$ and the other lies in $\mathbf{H_2}$.

For **3 –** dimensional space there is also a corresponding assumption involving planes in space.

**Axiom B – 4** (*Space Separation Postulate*) **:** *Given a plane* **P** *in space and a point* **X** *which is not on* **P,** *the set of all points in space that are not on* **L** *is a union of two disjoint, nonempty convex subsets* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *such that if* $\mathbf{Y_1} \in \mathbf{H_1}$ *and* $\mathbf{Y_2} \in \mathbf{H_2}$ *then the open segment* $(\mathbf{Y_1\,Y_2})$ *and the plane* **P** *have a point in common.*

There are analogs of the notational conventions following **B – 3:** The subsets $\mathbf{H_1}$ and $\mathbf{H_2}$ are called the ***half – spaces*** determined by the line **P,** or the sides of **P** in space. If **A** is a point which lies in, say, $\mathbf{H_1}$ then we shall say that $\mathbf{H_1}$ is the *side of* **P** *containing* **A** and $\mathbf{H_2}$ is the *side of* **P** *opposite* **A.** Similarly, if **A** and **B** are points of space that do not lie on **P,** we shall say that **A** *and* **B** *lie on the same side of* **P** if they both lie in either $\mathbf{H_1}$ and $\mathbf{H_2}$ and that **A** *and* **B** *lie on opposite sides of* **P** if one of the points lies in $\mathbf{H_1}$ and the other point lies in $\mathbf{H_2}$**.**

The subsets $\mathbf{H_1}$ and $\mathbf{H_2}$ in the separation postulates have simple analytic descriptions in the most basic cases. A line in $\mathbf{R}^2$ or a plane in $\mathbf{R}^3$ is always given by a nontrivial equation of the form $\mathbf{a} \cdot \mathbf{x} = b$, and the half – planes or half – spaces are simply the sets of points where $\mathbf{a} \cdot \mathbf{x} > b$ and $\mathbf{a} \cdot \mathbf{x} < b$. In order to justify this statement, one should check that the sets defined by the given inequalities actually satisfy the properties stated in the postulates; one of the exercises for this section provides such a verifiction.

Most of the discussion below will be restricted to the $2-$ dimensional case, but the first result applies to both the $2-$ and $3-$ dimensional cases.

**Proposition 8.** *Let* **M** *denote either a line* **L** *in a plane* **P** *or a plane* **Q** *in space. Then the following hold:*

1. *If* **A** *and* **B** *are on the same side of* **M** *and* **B** *and* **C** *are on the same side of* **M,** *then* **A** *and* **C** *are on the same side of* **M.**
2. *If* **A** *and* **B** *are on opposite sides of* **M** *and* **B** *and* **C** *are on opposite sides of* **M,** *then* **A** *and* **C** *are on the same side of* **M.**
3. *If* **A** *and* **B** *are on opposite sides of* **M** *and* **B** *and* **C** *are on the same side of* **M,** *then* **A** *and* **C** *are on opposite sides of* **M.**

***Proof.*** Let $\mathbf{H_1}$ and $\mathbf{H_2}$ be the subsets given by the appropriate separation postulate. The point **A** belongs to exactly one of them, so relabel these sets as $\mathbf{K_1}$ and $\mathbf{K_2}$ such that $\mathbf{A} \in \mathbf{K_1}$. In part **(1)** the first hypothesis implies that $\mathbf{B} \in \mathbf{K_1}$, and therefore the second hypothesis implies that $\mathbf{C} \in \mathbf{K_1}$, which yields the stated conclusion for this part of the result. In part **(2)** the first hypothesis implies that $\mathbf{B} \in \mathbf{K_2}$, and therefore the second hypothesis implies that $\mathbf{C} \in \mathbf{K_1}$, which yields the stated conclusion for this part of the result. Finally, in part **(3)** the first hypothesis implies that $\mathbf{B} \in \mathbf{K_2}$, and therefore the second hypothesis implies that $\mathbf{C} \in \mathbf{K_2}$ which yields the stated conclusion for this part of the result.■

Needless to say, it is useful to have a simple analytic criterion for two points in a plane to lie on the same or opposite sides of a line. The next result provides one.

**Proposition 9.** *Suppose that* **A, B, C, D** *are points in* $\mathbf{R}^2$ *such that* **C** *and* **D** *do not lie on the line* **AB.** *Let* $\mathbf{D} = x\mathbf{A} + y\mathbf{B} + z\mathbf{C}$ *be the unique expression for* **D** *using barycentric coordinates, so that* $x + y + z = 1$. *Then* **C** *and* **D** *lie on the same side of* **AB** *if* $z$ *is positive, and they lie on opposite sides of* **AB** *if* $z$ *is negative.*

***Proof.*** Before beginning, we note that $z$ must be nonzero, for otherwise **D** would lie on the line **AB.** Suppose that **AB** is defined by the equation $\mathbf{m} \cdot \mathbf{P} = k$ for some nonzero vector **m** and some constant $k$, and let $q = \mathbf{m} \cdot \mathbf{C}$, so that $q - k$ is nonzero. We then have

$$\mathbf{m} \cdot \mathbf{D} \;=\; (x\mathbf{A} + y\mathbf{B} + z\mathbf{C}) \cdot \mathbf{D} \;=\; xk + yk + zq \;=\; (1-z)k + q \;=\; k + z(q-k).$$

Suppose that $z > 0$. Then the right hand side is greater than $k$ if $q > k$ and less than $k$ if $q < k$, so it follows that both of $\mathbf{m} \cdot \mathbf{C}$ and $\mathbf{m} \cdot \mathbf{D}$ are either greater than $k$ or less than $k$, which means that they lie on the same side of **AB**. Now suppose that $z < 0$. Then the right hand side is less than $k$ if $q > k$ and greater than $k$ if $q < k$, so it follows that one of $\mathbf{m} \cdot \mathbf{C}$ and $\mathbf{m} \cdot \mathbf{D}$ is greater than $k$ and the other is less than $k$, which means that they lie on opposite sides of **AB**.■

The next obsevation is simple but important.

**Lemma 10.** *Let* **L** *be a line in the plane, and let* **M** *be a line in the plane which meets* **L** *in exactly one point. Then* **M** *contains points on both sides of* **L.**

***Proof.*** Let **A** be the point where the lines meet, and let **B** be a second point of **M.** If **C** is a second point of **L,** then **A, B** and **C** are noncollinear (otherwise **L** = **AC** = **AB** = **M**). Let **D** = **A** – (**B** – **A**) = 2**A** – **B,** so that **D** lies on **M.** It follows that the expression for **D** as a linear combination of **A** and **B** also gives the expression for **D** in terms of **A, B** and **C** by barycentric coordinates (since the latter are uniquely determined), and therefore by the previous result we know that **B** and **D** lie on opposite sides of the line **L** = **AC**.■

The preceding result is helpful in proving a result about the intersection of a line and a half – plane.

**Proposition 11.** *Let* **L** *be a line in the plane, let* $\mathbf{H_1}$ *and* $\mathbf{H_2}$ *be the two half – planes determined by* **L,** *and let* **M** *be a line in the plane which meets* **L** *in exactly one point. Then each of the intersections* $\mathbf{H_1} \cap \mathbf{M}$ *and* $\mathbf{H_2} \cap \mathbf{M}$ *is an open ray.*

***Proof.*** By the symmetry of the hypotheses, it suffices to prove the result for $\mathbf{H_1} \cap \mathbf{L}$. Let **A** be the point where the lines meet and let **B** be a point on $\mathbf{H_1} \cap \mathbf{M}$ whose existence is guaranteed by the preceding lemma. Once again let **C** be a second pont of **L.** Given an arbitrary point **X** on **M,** express it as **A** + $k$(**B** – **A**) for a suitable scalar $k$. This linear combination can be rewritten as **X** = $(1-k)$**A** + $k$**B,** and by the uniqueness of barycentric coordinates it follows that the latter is the expression for **X** in terms of **A, B, C** by barycentric coordinates. Thus **X** lies on $\mathbf{H_1} \cap \mathbf{M}$ if and only if $k$ is positive. Since this is the same condition for **X** to lie on **(AB** it follows that $\mathbf{H_1} \cap \mathbf{M}$ = **(AB** as required.■

The following properties of betweenness and separation are often needed as steps in other geometric proofs.

**Proposition 12.** *Let* **L** *be a line in the plane, let* **M** *be a line in the plane which meets* **L** *in exactly one point* **A,** *and let* **B** *and* **C** *be two other points on* **M.** *Then* **B** *and* **C** *lie on the same side of the line* **L** *if either* **A*****C*****B** *or* **A*****B*****C** *is true, and they lie on opposite sides of the line* **L** *if* **B*****A*****C** *is true.*

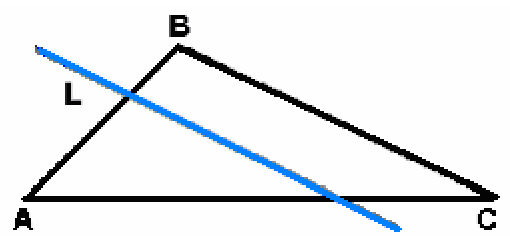
***Proof.*** Once again write **C** = $(1-k)$**A** + $k$**B,** and assume that **C** is distinct from **A** and **B** so that $k$ is not equal to $0$ or $1$. If **A*****C*****B** is true then we have $0 < k < 1$, while if **A*****B*****C** is true then $k > 1$, so in both cases **B** and **C** lie on the same side of **L**

by previous results. On the other hand, if **B*A*C** is true then $k < 0$, so that **B** and **C** lie on opposite sides of **L** by previous results.■

We shall conclude this discussion with a result that might look a bit more interesting. It was used implicitly in the ***Elements***, and the formal statement of it is due to M. Pasch (1843 – 1930). Before stating the result we need to formalize a standard concept.

**Definition.** Let **A, B** and **C** be noncollinear points. Then ***triangle* ABC,** often written △**ABC,** is given by the union of the three closed segments **[AB], [BC]** and **[AC].** Each of these segments is called a ***side*** of the triangle, and the original three points are called its ***vertices*** (*sing.* ***vertex***).

**Theorem 13.** (Pasch's "Postulate") *Suppose we are given* △**ABC** *and a line* **L** *in the same plane as the triangle such that* **L** *meets the open side* **(AB)** *in esactly one point. Then either* **L** *passes through* **C** *or else* **L** *has a point in common with* **(AC)** *or* **(BC).**




***Proof.*** This turns out to be fairly simple. By the previous result we know that **A** and **B** lie on opposite sides of **L.** What can we say about **C?** If **C** does not lie on **L,** then it lies on one of the two sides of **L.** If **C** lies on the same side of **L** as **A,** then **C** and **B** lie on opposite sides of **L,** and therefore the line **L** and the segment **(BC)** have a point in common. On the other hand, if **C** lies on the same side of **L** as **B,** then **C** and **A** lie on opposite sides of **L,** and therefore the line **L** and the segment **(AC)** have a point in common. In every case one of the options in the conclusion is satisfied.■

We have tried to limit our discussion to results that will be needed at several points later in the course. A few additional results of a similar nature will be covered after we introduce angles and their interiors in the next section.

*Numerical examples*

Here are two problems which illustrate the concepts in described above; they are similar to some of the homework exercises.

**PROBLEM 14.** *Let* **L** *be the line in* $\mathbf{R}^2$ *defined by* $y = 2x + 1$. *Determine which of the three points* $(4, 10)$, $(5, 8)$ *and* $(2, 3)$ *lie on one side of* **L** *and which lie on the other.*

**SOLUTION.** The sides of **L** are given by the inequalities $y < 2x + 1$ and $y > 2x + 1$. Since we have $10 > 2\cdot4 + 1$, $8 < 2\cdot5 + 1$, and $3 < 2\cdot2 + 1$, it follows that the second and third points lie on one side of the line, while the first lies on the opposite side.■

**PROBLEM 15.** *If* $\mathbf{A} = (0, 2)$, $\mathbf{B} = (1, 3)$ *and* $\mathbf{C} = (6, 10)$, *determine whether the* *points* $(10, 6)$ *lie on the same side of* **AB** *as* **C**, *and do the same for* $(5, 9)$.

**SOLUTION.** If **D** is either of the two points, we need to find the barycentric coordinates of **D** with respect to **A**, **B**, and **C**. In other words, we need to express **D** as a linear combination $x$ **A** + $y$ **B** + $z$ **C**, where $x + y + z = 1$. Then **D** will lie on the same side of **AB** as **C** if $z > 0$, and **D** will lie on the opposite side of **AB** if $z < 0$.

The standard method to find the barycentric coordinates is to write **D – A** = $y$(**B – A**) + $z$(**C – A**) by means of the equation $x = 1 - y - z$; since **B – A** and **C – A** are linearly independent, it follows that there are unique choices for $y$ and $z$. Now **B – A** = $(1, 1)$, and **C – A** = $(6, 8)$. Therefore if **D** = $(10, 6)$ we obtain the vector equation

$$(10, 4) = y(1, 1) + z(6, 8)$$

and if we solve the associated system of scalar equations we find that $z = -3$, so that ***the points*** $(10, 6)$ ***and*** $(6, 10)$ ***lie on opposite sides of*** **AB**. Next, if we take the point **D** = $(5, 9)$, then we get the equation

$$(5, 7) = y(1, 1) + z(6, 8)$$

and if we solve the associated system of scalar equations we find that $z = 1$, so that ***the points*** $(5, 9)$ ***and*** $(6, 10)$ ***lie on the same side of*** **AB**.■

**GENERAL SUGGESTION.** In each exercise, it may be helpful to confirm the results of the calculations by plotting the lines and points on graph paper.■

*Comments on the Appendix*

In our discussion of the need to be careful about issues of betweenness and separation, we mentioned that too much reliance on drawings can lead to inaccurate conclusions, and we mentioned a classic example in which reasonable but inaccurate drawings can lead to the obviously false conclusion that every triangle is isosceles. Most of the appendix is devoted to presenting this example, but there are also some general comments at the end that are important.

A reader who does not wish to go through the details of the example may skip to the final subheading (*The role of drawings in geometrical proofs*) without loss of continuity.

## Appendix – The isosceles triangle fallacy

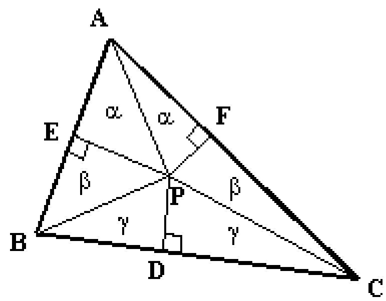
> The truth of a theory is in your mind, not in your eyes.
>
> Albert Einstein (1879 – 1955)

The treatment below is adapted from the following source:

http://www.mathpages.com/home/kmath392.htm

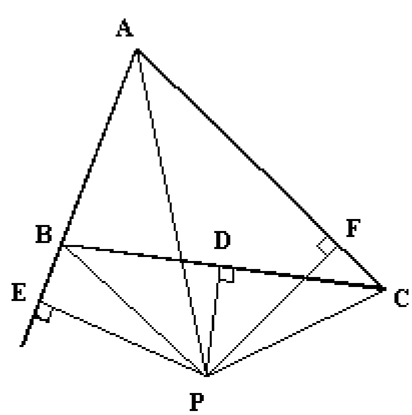
One well-known illustration of the logical fallacies to which Euclid's methods are vulnerable (or at least would be vulnerable if we didn't "cheat" by allowing ourselves to be guided by accurately drawn figures) is the "proof" that all triangles are isosceles. *The discussion below explicitly assumes that the reader is familiar with some basic ideas from elementary (high school) geometry*.

Given an arbitrary triangle **ABC,** draw the angle bisector of the interior angle at **A,** and draw the perpendicular bisector of the closed segment **[BC]** with midpoint **D,** as shown below:




If the angle bisector at **A** and the perpendicular bisector of **[BC]** are parallel or identical, then **ABC** is isosceles (this is a valid result in Euclidean geometry and can be shown directly by standard methods; we shall omit the details because they do not involve the fallacy). On the other hand, if the lines in questions are not parallel, then they intersect at a point, which we shall call **P**, and we can drop the perpendiculars from **P** to **AB** and **AC,** which will meet these lines at **E** and **F** respectively. Now the two triangles △**APE** and △**APF** have equal angles and share a common side, so they are congruent. Therefore, ***d*(P, E) = *d*(P, F).** Also, since **D** is the midpoint of **[BC],** the triangles △**PDB** and △**PDC** are congruent right triangles, and hence ***d*(P, B) = *d*(P, C).** From this it follows that the triangles labeled △**PEB** and △**PFC** are similar and equal to each other, so we have ***d*(B, E) + *d*(E, A) = *d*(C, F) + *d*(F, A),** and therefore △**ABC** is isosceles **(???).**

Of course, if we attempt to accurately construct the points and lines described in this proof accurately, we see that the actual configuration doesn't look like the picture above. The point **P** necessarily falls outside the triangle △**ABC.** However, if we carry out the proof on this basis, and if we now assume the points **E** and **F** also fall outside the triangle, we still conclude that the triangle is isosceles. This too is an incorrect configuration. The ***actual configuration*** of points given by the stated construction is for the point **P** to be outside the triangle **ABC,** and for exactly one of the points **E, F** to be between the vertices of the triangle, as shown below:



We still have $d$**(A, E) = $d$(A, F),** $d$**(P, E) = $d$(P, F),** and $d$**(P, B) = $d$(P, C),** and it still follows that $d$**(B, E) = $d$(F, C),** but now we see that even though $d$**(A, E) = $d$(A, F)** and $d$**(B, E) = $d$(F, C)** it does ***not*** follow that $d$**(A, B) = $d$(A, C),** for even though **F** is between **A** and **C,** the point **E** is ***not*** between **A** and **B.**

Actually, the argument above does yield valid (but far less intriguing) conclusions. For example, if sides **[AB]** and **[AC]** have unequal lengths, then either **E** does not lie on **[AB]** or else **F** does not lie on **[AC].** For if both were true the argument above would show that △**ABC** is isosceles, and we have assumed this is false.

***Additional examples*.** There are several other well – known examples of fallacious, but reasonable looking, geometrical arguments which purportedly yield conclusions that are patently false. Many such arguments are contained in the second part of the following book:

A. I. Fetisov and Ya. S. Dubnov, ***Proof in Geometry: With "Mistakes in Geometric Proofs."*** Dover Books, New York, 2006.
ISBN: 0–486–45354–5.

Also, here is an online reference with a (fallacious) geometrical "proof" that $99 = 100$:

http://www.cut-the-knot.com/pythagoras/tricky.html

*The role of drawings in geometrical proofs*

Geometry is the science of correct reasoning on incorrect figures.

G. Polyá

You can see a lot just by looking.

L. "Yogi" Berra (1925 – )

For me, following a geometrical argument purely logically, without a picture for it constantly in front of me, is impossible.

F. Klein (1849 – 1925)

To paraphrase the online reference given above, one may view the logical proofs in the ***Elements*** as fairly well constructed arguments which are ***sometimes intuitive*** and ***based upon accurately drawn figures.*** The modern standard for logical correctness is to give completely rigorous proofs of abstract concepts in which ***the reasoning may be frequently suggested by roughly drawn figures*** but ***does not formally depend upon them.***