Foundations of Constructive Geometry

Michael Beeson

San José State University

March 13, 2012
Stanford

Euclidean Constructive Geometry **ECG**

Does *constructive* refer to the use of intuitionistic logic?

Or does it refer to geometrical constructions with ruler and compass?

What is the relation between these two?

In our constructive geometry, they are closely related: things proved to exist can be constructed with ruler and compass.

Continuous geometry

Constructive proofs yield continuity in parameters.

In practice, constructive proofs *require* continuity in parameters.

There are thus *three* ways of looking at this subject: geometry with intuitionistic logic, geometry of ruler-and-compass constructions, geometry with continuous dependence on parameters.

Is Euclid's reasoning constructive?

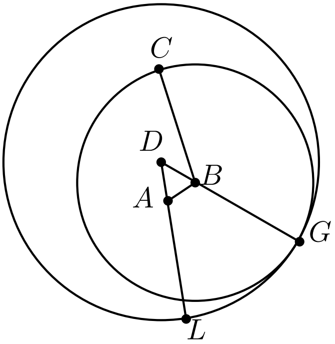
Yes, Euclid's reasoning is generally constructive; indeed the only irreparably non-constructive proposition is Book I, Prop. 2, which shows that a rigid compass can be simulated by a collapsible compass. We just take Euclid I.2 as an axiom, thus requiring a rigid compass in $\mathbf{ECG}$. Only one other repair is needed, in the formulation of the parallel axiom, as we shall see below.

Book I, Proposition 2

To place at a given point a straight line equal to a given straight line.

Let A be the given point, and BC the given straight line. It is required to place at A a straight line AL equal to BC.

Euclid requires [without explicit mention] a case distinction, whether $A = B$ or not.



The form of Euclid's theorems

Euclid did not deal with disjunctions explicitly, and all his theorems are of the form: Given certain points related in certain ways, we can construct other points related to the given points and each other in certain ways. Euclid has been criticized (as far back as Geminus and Proclus) for ignoring case distinctions in a proof, giving a diagram and proof for only one case. Since case distinctions (on whether $ab = cd$ or not) are non-constructive, these omissions are *prima facie* non-constructive. However, these non-constructive proof steps are eliminable, as we will explain.

An example of such an argument in Euclid is Prop. I.6, whose proof begins

> *Let ABC be a triangle having the angle ABC equal to the angle ACB. I say that the side AB is also equal to the side AC. For, if AB is unequal to AC, one of them is greater. Let AB be greater, ...*

The same proof also uses an argument by contradiction in the form $\neg x \neq y \ \rightarrow \ x = y$. This principle, the "stability of equality", is an axiom of **ECG**, and is universally regarded as constructively acceptable. The conclusion of I.6, however, is negative (has no $\exists$ or $\vee$), so we can simply put double negations in front of every step, and apply the stability of equality once at the end.

Prop. I.26 is another example of the use of the stability of equality: "… DE is not unequal to AB, and is therefore equal to it." To put the matter more technically, in constructive logic we have $P \rightarrow \neg\neg P$, and although generally we do not have $\neg\neg P \rightarrow P$, we do have it for quantifier-free, disjunction-free P. We can double-negate $A \vee \neg A \rightarrow B$, obtaining $\neg\neg(A \vee \neg A) \rightarrow \neg\neg B$, and then the hypothesis is provable, so we have $\neg\neg B$, and hence B since B is quantifier-free and disjunction-free. The reason why this works throughout Euclid is that the *conclusions* of Euclid's theorems are all quantifier-free and disjunction-free. Euclid never even *thought* of stating a theorem with an "or" in it.

The bottom line is that Euclid is constructive as it stands, except for Book I, Prop. 2, and the exact formulation of the parallel postulate.

To remedy these problems in **ECG**:

Take Book I, Prop. 2 as an axiom.

Strengthen the parallel postulate as discussed below.

We also take as an axiom $\neg\neg\mathbf{B}(x,y,z) \rightarrow \mathbf{B}(x,y,z)$, or "Markov's principle for betweenness", enabling us to drop double negations on atomic sentences. Here betweenness is strict.

The Elementary Constructions

- *Line*(A,B)
- *Circle*(A,B) (center A, passes through B) collapsible compass
- *Circle*(A,B,C) (center A, radius BC) rigid compass

The Elementary Constructions

- *Line*(A,B)
- *Circle*(A,B) (center A, passes through B) collapsible compass
- *Circle*(A,B,C) (center A, radius BC) rigid compass
- *IntersectLines*(A,B,C,D) (AB meets CD)

The Elementary Constructions

- *Line* (A,B)
- *Circle* (A,B) (center A, passes through B) collapsible compass
- *Circle* (A,B,C) (center A, radius BC) rigid compass
- *IntersectLines* (A,B,C,D) (AB meets CD)
- *IntersectLineCircle1* (A,B,C,D)
  (Line AB meets circle with center C through D)
- *IntersectLineCircle2* (A,B,C,D)

The Elementary Constructions

- *Line* (A,B)
- *Circle* (A,B) (center A, passes through B) collapsible compass
- *Circle* (A,B,C) (center A, radius BC) rigid compass
- *IntersectLines* (A,B,C,D) (AB meets CD)
- *IntersectLineCircle1* (A,B,C,D)
  (Line AB meets circle with center C through D)
- *IntersectLineCircle2* (A,B,C,D)
- *IntersectCircles1* (c_1,c_2)
- *IntersectCircles2* (c_1,c_2)

How the sorts mix

You can also write

$$\mathit{IntersectLines}\,(L, K)$$

if L and K are lines. Then

$$\mathit{IntersectLines}\,(A, B, C, D) = \mathit{IntersectLines}\,(\mathit{Line}\,(A, B), \mathit{Line}\,(C, D))$$

and so on.

Models of the Elementary Constructions

- The "standard plane" $\mathbb{R}^2$
- The "recursive plane". Points are given by recursive functions giving rational approximations to within $1/n$.
- The minimal model, the points constructible by ruler and compass
- The algebraic plane, points with algebraic coordinates
- The Poincaré model. These constructions work in non-Euclidean geometry too.

First order theories of geometry

- Angles can be treated as ordered triples of points.
- Rays and segments are needed only for visual effect; for theory we need only points, lines, and circles.
- We don't even need lines and circles; every theorem comes down to constructing some points from given points, so that the constructed points bear certain relations to the original points.
- The relations in question can be expressed in terms of *betweenness* and *equidistance*.
- **ECG** uses a three-sorted language, for points, lines, and circles.

Tarski geometry and Hilbert geometry

Just to avoid confusion: today we are concerned with "elementary" geometry in the sense that only line-circle and circle-circle continuity are used. Hilbert's geometry included a second-order continuity axiom; we may compare it to requiring that Dedekind cuts be filled, although Hilbert formulated it differently.

"Tarski geometry" is a first-order theory with a continuity schema, essentially requiring that first-order definable Dedekind cuts be filled. Sometimes "elementary" means first-order, and Tarski wrote a famous paper, *What is Elementary Geometry*, in which "elementary geometry" meant Tarski geometry. But "elementary" can also refer to the *Elements* of Euclid, which is a weaker theory.

Remarks on axiomatizations of geometry

There are numerous issues concerning the axiomatization of geometry. Here are a few:

- What are the primitive sorts of the theory?
- What are the primitive relations?
- What (if any) are the function symbols?
- What are the continuity axioms?
- How is congruence of angles defined?
- How is the SAS principle built into the axioms?
- How close are the axioms to Euclid?
- Are the axioms few and elegant, or numerous and powerful?
- Are the axioms strictly first-order?

Remarks on axiomatizations of geometry

Famous axiomatizations have been given by Veblen, Pieri, Hilbert, Tarski, Borsuk and Szmielew, and Szmielew, and that list is by no means comprehensive. Nearly every possible combination of answers to the "issues" has something to recommend it. For example, Hilbert has several sorts, and his axioms are not strictly first-order; Tarski has only one sort (points) and ten axioms. Our results require a quantifier-free and disjunction-free system, but otherwise do not depend on the exact choice of formalization. Our first presentation was Hilbert-style and our final presentation will be Tarski-style. It is a lot of work to develop geometry constructively from ten or so axioms about points, but very elegant. However, we are not going there today, as few of the issues one encounters are specifically constructive.

Three important issues

- When there are two intersection points, which one is denoted by which term?
- In degenerate situations, such as *Line* (P, P), what do we do?
- When the indicated lines and/or circles do not intersect, what do we do about the term(s) for their intersection point(s)?

Undefined terms

When the indicated lines or circles do not intersect, then the term for their intersection is "undefined". This can best be handled formally using the logic of partial terms, which we do in **ECG**; it can also be handled in other more cumbersome ways without modifying first-order logic.

Degenerate circles

We take *Circle* (P, P) to be defined, i.e., we allow circles of zero radius; that technicality makes the formal development smoother and seems philosophically unobjectionable–we just allow the two points of the compass to coincide. The point here is not so much that circles of zero radius are of interest, but that we do not want to force a case distinction as to whether the two points of the compass are, or are not, coincident.

Order of points on a line meeting a circle

We take the two points of intersection of a line *Line*(A, B) and a circle to occur in the same order as A and B occur on L. That means that lines are treated as having direction. Not only do they have direction, they "come equipped" with two points from which they were constructed. There are function symbols to recover those points from a line. *Line*(P, P) is undefined, since having it defined would destroy continuous dependence of *Line*(P, Q) on P and Q.

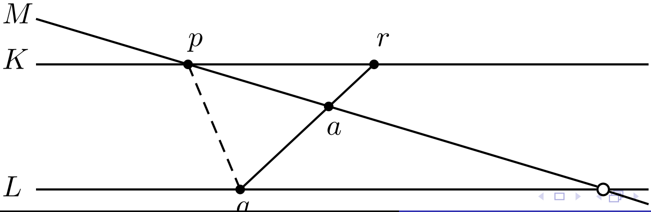
Right and left turns

The two intersection points $p = \mathit{IntersectCircles1}(C, K)$ and $q = \mathit{IntersectCircles2}(C, K)$ are to be distinguished as follows: With a the center of C and b the center of K we should have abp a right turn, and abq a left turn. But can "right turn" and "left turn" be defined? What we do is to *define Right* and *Left* using equations involving *IntersectCircles1* and *IntersectCircles2*; then we give axioms about *Right* and *Left*, namely that if abc is a left turn, then c and d are on the same side of $\mathit{Line}(a, b)$ if and only if abd is a left turn, and c and d are on opposite sides of $\mathit{Line}(a, b)$ if and only if abd is a right turn. Note that neither this issue nor its solution has to do with constructivity, but simply with the introduction of function symbols corresponding to the elementary constructions.

Euclid's Parallel Postulate

If a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

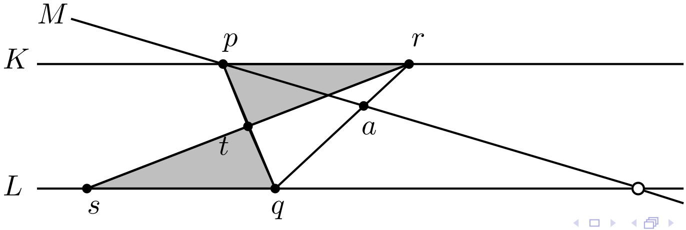
That is, M and L must meet on the right side, provided $\mathbf{B}(q, a, r)$ and pq makes alternate interior angles equal with K and L. The point at the open circle is asserted to exist.



Euclid 5 without mentioning angles

We need to eliminate mention of "alternate interior angles", because angles are not directly treated in **ECG**, but instead are treated as triples of points.

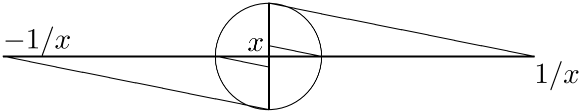
M and L must meet on the right side, provided $\mathbf{B}(q, a, r)$ and $pt = qt$ and $rt = st$.



Inadequacy of Euclid 5

Although we have finally arrived at a satisfactory formulation of Euclid 5, that formulation is satisfactory only in the sense that it accurately expresses what Euclid said. It turns out that this axiom is not satisfactory as a parallel postulate for **ECG**. The most obvious reason is that it is inadequate to define division geometrically. But it turns out to also be necessary for addition and multiplication!

Division and parallels



The circle has radius 1. The slanted lines are parallel. $1/x$ is defined if and only if the horizontal line intersects the long slanted line. If we know the sign of x then Euclid 5 suffices; the vertical line is a transversal and on one side the interior angles are less than two right angles.

Division and parallels

Without knowing the sign of x, we will not know on which side of the transversal pq the two adjacent interior angles will make less than two right angles. In other words, with Euclid 5, we will only be able to divide by a number whose sign we know; and the principle $x \neq 0 \;\rightarrow\; x < 0 \vee x > 0$ is not an axiom (or theorem) of **ECG**. The conclusion is that if we want to divide by nonzero numbers, we need to strengthen Euclid's parallel axiom.

Strengthening Euclid 5

We make three changes in Euclid 5 to get the "strong parallel postulate":

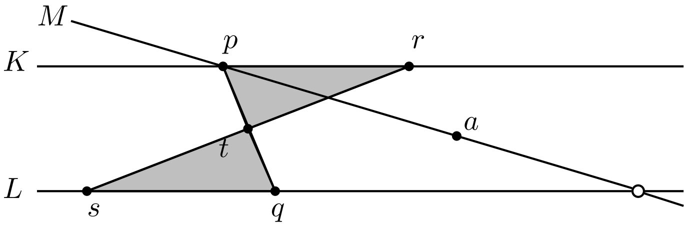
(i) We change the hypothesis $\mathbf{B}(q,a,r)$ to $\neg on(a,K)$. In other words, we require that the two adjacent interior angles do not make exactly two right angles, instead of requiring that they make less than two right angles.

(ii) We change the conclusion to state only that M meets L, without specifying on which side of the transversal pq the intersection lies.

(iii) We drop the hypothesis $\neg$ *on* (p, L).

The Strong Parallel Postulate of **ECG**

Figure: Strong Parallel Postulate: M and L must meet (somewhere) provided a is not on K and $pt = qt$ and $rt = st$.



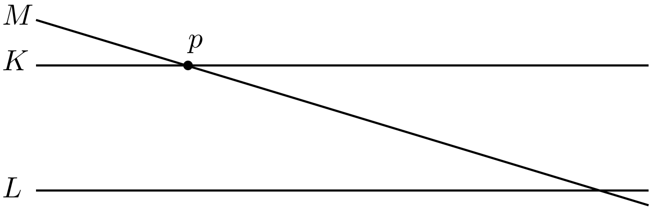
Is the "strong" parallel axiom stronger than Euclid 5?

The strong parallel axiom differs from Euclid's version in that we are not required to know in what *direction* M passes through P; but also the conclusion is weaker, in that it does not specify *where* M must meet L. In other words, the betweenness hypothesis of Euclid 5 is removed, and so is the betweenness conclusion. Since both the hypothesis and conclusion have been changed, it is not immediate whether this new postulate is stronger the Euclid 5, or equivalent, or possibly even weaker, but it turns out to be stronger–hence the name.

Playfair's Axiom

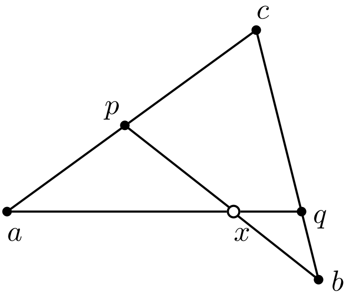
Let P be a point not on line L. We consider lines through P that do not meet L (i.e., are parallel to L). Playfair's version of the parallel postulate says that two parallels to L through P are equal. That is, in the picture not both M and K are parallel to L.

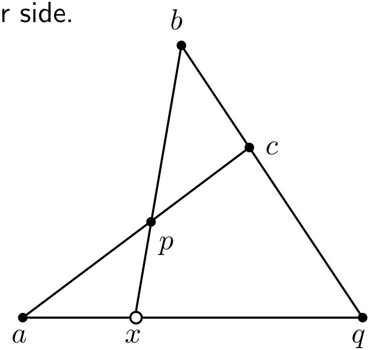
But no point is asserted to exist.



Pasch's Axiom Inner Pasch (left) and Outer Pasch (right)

Line pb meets triangle acq in one side. The open circles show the points asserted to exist on the other side.





Same Side and Opposite Side

Two points a and b not on line L are on opposite sides of L if $a \neq b$ and there is a point of L between a and b, i.e., the segment ab meets L. Two points a and b are on the same side of L if they are both on the opposite side of L from the same point. That turns out to be equivalent to "no point on L is between a and b", but the proof is not easy. Then we have both a $\forall$ version and an $\exists$ version of "same side", which enables us to get the axioms involving *Right* and *Left* quantifier-free.

Euclidean rings

Classical Euclidean geometry has models $K^2 = K \times K$ where K is a Euclidean field, i.e. an ordered field in which nonnegative elements have square roots.

We define a Euclidean ring to be an ordered ring in which nonnegative elements have square roots. We use a language with symbols $+$ for addition and $\cdot$ for multiplication, and a unary predicate $P(x)$ for "x is positive".

Euclidean fields

Euclidean fields: nonzero elements have reciprocals

Weakly Euclidean fields: positive elements have reciprocals

Playfair rings: elements without reciprocals are zero, and if x is greater than a positive invertible element, then x is invertible.

Axioms of Euclidean field theory

Ring axioms plus

$0 \neq 1$	EF0
$x \neq 0 \ \rightarrow\ \exists y\,(x \cdot y = 1)$	EF1
$P(x) \wedge P(y) \ \rightarrow\ P(x+y) \wedge P(x \cdot y)$	EF2
$x + y = 0 \ \rightarrow\ \neg(P(x) \wedge P(y))$	EF3
$x + y = 0 \wedge \neg P(x) \wedge \neg P(y) \ \rightarrow\ x = 0$	EF4
$x + y = 0 \wedge \neg P(y) \ \rightarrow\ \exists\, z(z \cdot z = x)$	EF5
$\neg\neg P(x) \ \rightarrow\ P(x)$	EF6 (Markov's principle)

Weakly Euclidean fields

Replace

$$x \neq 0 \;\rightarrow\; \exists y\,(x \cdot y = 1) \qquad \text{EF1}$$

by

$$P(x) \;\rightarrow\; \exists y\,(x \cdot y = 1) \qquad \text{EF7}$$

Playfair fields

Replace EF1 by

$$(\forall y(x \cdot y \neq 1)) \rightarrow x = 0 \qquad \text{EF9}$$

$$P(y) \wedge P(z) \wedge y + z = x \wedge (y \cdot v = 1) \rightarrow \exists\, w\, (w \cdot x = 1) \qquad \text{EF10}$$

EF10 says "elements greater than a positive invertible are invertible."

EF9 enables us to verify the Playfair axiom.
EF10 enables us to verify Pasch's axiom.

Addition and Multiplication

In order to show that the models of some geometrical theory T have the form F^2, one has to define addition and multiplication (of segments or points on a line) within T. This was first done by Descartes, and again (in a different way) by Hilbert in his 1899 book, *Foundations of Geometry*. These constructions, however, involve a non-constructive case distinction on the sign of the numbers being added or multiplied. It is not trivial to define addition continuously in parameters, i.e., with no case distinction. Hilbert's second definition of multiplication is OK as it stands.

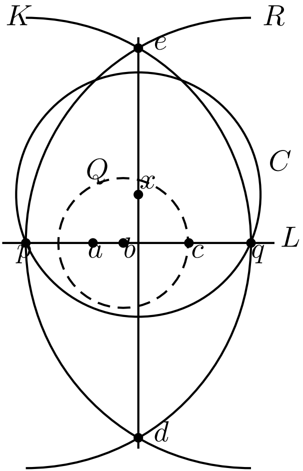
Euclidean geometry without case distinctions

To constructivize Euclidean geometry turns out to involve replacing familiar constructions that involve case distinctions by more elaborate constructions that work without case distinctions.

Example: to construct a line through a point P perpendicular to a line L, without a case distinction as to whether P is or is not on L.

Uniform perpendicular

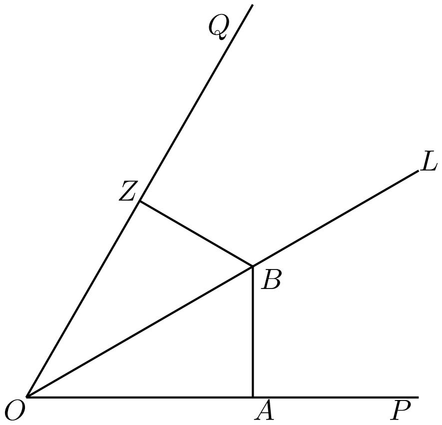
$M = \textit{Perp}(x, L))$ is constructed perpendicular to L without a case distinction whether x is on L or not. Note $bc = xa$ so the radius ac of C is long enough to meet L twice.



Rotation

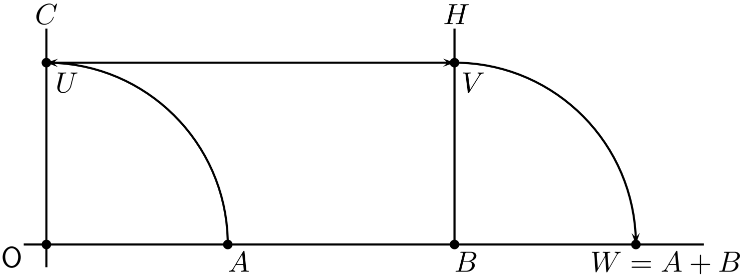
Rotate a given point through a given angle (without a case distinction whether the point lies on the vertex of the angle or not).

Z is defined even when $A = O$ (in which case it is just O, of course), and if A moves along *Line* (O, P) through O, then Z moves along *Line* (O, Q), passing through O as A does.

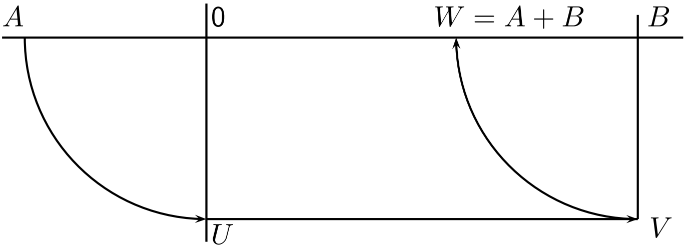


Signed addition without test-for-equality

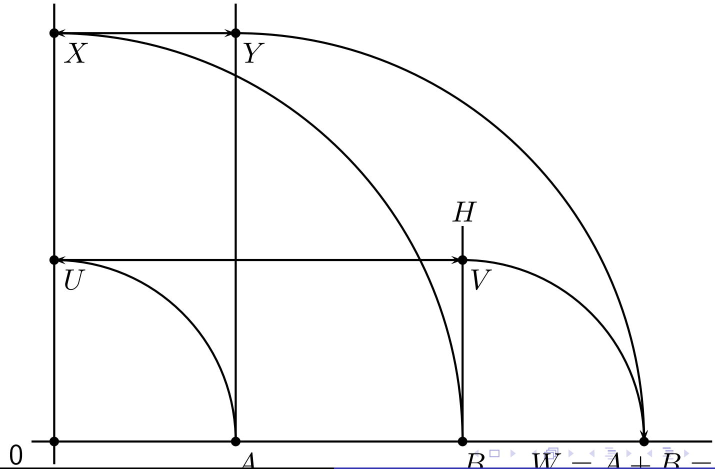
A is rotated to U, then projected to V, then rotated to W.



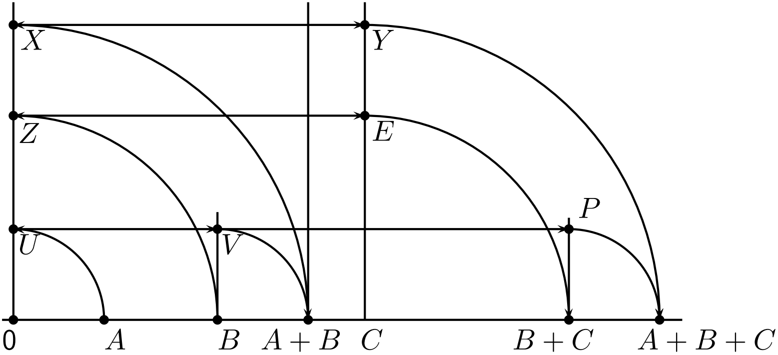
Signed addition when A is negative



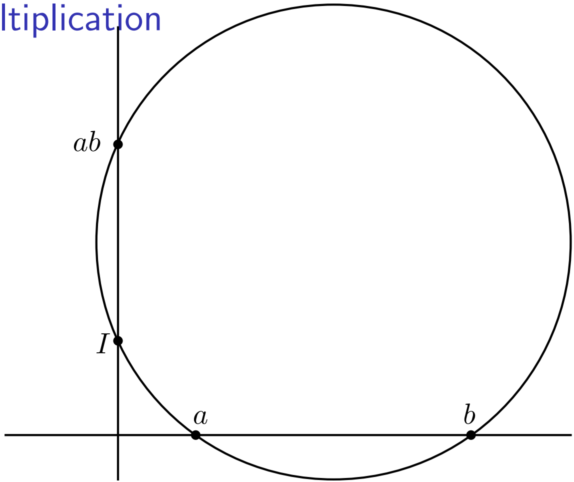
Commutativity of Addition



Associativity of Addition



Hilbert multiplication



ECG corresponds to Euclidean fields

The models of $\mathbf{ECG}$ are of the form F^2, where F is a Euclidean field. More specifically, given such a field, we can define betweenness, incidence, and equidistance by analytic geometry and verify the axioms of **ECG**. Conversely, and this is the hard part, we can define multiplication, addition, and division of points on a line (having chosen one point as zero), in **ECG**. It turns out that we need the strong parallel axiom to do that.

Rings corresponding to Euclid 5

If we replace the parallel axiom of $\mathbf{ECG}$ by Euclid's parallel postulate, we get instead models of the form F^2, where F is a weakly Euclidean field (that is, a Euclidean ring in which positive elements have reciprocals).

We cannot go the other way by defining multiplication and addition geometrically without the strong parallel axiom. (That is, if we only had Euclid 5, we would need case distinctions, as Hilbert and Descartes did.)

Rings corresonding to Playfair's axiom

We now work out the field-theoretic version of Playfair's axiom. Playfair says, if P is not on L and K is parallel to L through P, that if line M through P does not meet L then $M = K$. Since $\neg\neg M = K \rightarrow M = K$, Playfair is just the contrapositive of the parallel axiom of **ECG**, which says that if $M \neq K$ then M meets L. Hence it corresponds to the contrapositive of $x \neq 0 \rightarrow 1/x \downarrow$; that contrapositive says that if x has no multiplicative inverse, then $x = 0$. Thus Playfair geometries have models F^2 where F is a Playfair ring (as defined above). (We cannot prove the converse because we need the strong parallel axiom to verify multiplication.) We also need EF10 to verify Pasch's axiom.

Interpretations versus Models

Classically one shows that the models of some geometrical theory T have the form F^2 for a certain kind of field F. That doesn't work for intuitionistic theories. One needs *interpretations* both ways.

An interpretation can be thought of as a formalized model. It is a map from formulas to formulas, preserving provability. To show a geometrical theory is equivalent to a field theory we need two interpretations, each of which is sound (preserves provability) and such that they are *inverses* in a suitable sense.

This requires *many pages* of detailed work. Even in the classical case there is a payoff: proofs in one theory are not much longer than proofs in the other theory, e.g. the geometries and field theories are polynomial-time equivalent.

What ECG proves to exist, can be constructed with ruler and compass

In earlier work I proved that if **ECG** proves an existential statement $\exists y A(x,y)$, with A negative, then there is a term t of **ECG** such that **ECG** proves $A(x, t(x))$. In words: things that **ECG** can prove to exist, can be constructed with ruler and compass. Of course, the converse is immediate: things that can be constructed with ruler and compass can be proved to exist in **ECG**. Hence the two meanings of "constructive" coincide for **ECG**: it could mean "proved to exist with intuitionistic logic" or it could mean "constructed with ruler and compass."

What makes that metatheorem work

The technique of the proof is to apply Gentzen's cut-elimination theorem. What makes it applicable is that the axiomatization of **ECG** has two important properties: it is *quantifier-free*, and it is *disjunction-free*. What is particularly interesting about this is that it was not difficult to axiomatize **ECG** in this way–we just followed Euclid.

Strong Parallel Axiom implies Euclid 5

The reduction of geometry to field theory described above shows that (relative to a base theory), the strong parallel axiom implies Euclid's postulate 5 (since if reciprocals of non-zero elements exist, then of course reciprocals of positive elements exist). (A direct proof has also been given.)
And Euclid 5 easily implies Playfair's postulate.

Euclid 5 does not imply the strong parallel axiom

That is, relative to the base theory **ECG** minus its parallel axiom. Since non-constructively, the implications *are* reversible, we cannot hope to give counterexamples. In terms of field theory, we won't be able to construct a Euclidean ring in which positive elements have reciprocals but nonzero elements do not. The proof proceeds by constructing appropriate Kripke models.

Reduction to ordered field theory independence results

To show that Euclid 5 does not prove the strong parallel axiom, it suffices to prove the corresponding result in ordered field theory: the axiom that positive elements have reciprocals does not imply that all nonzero elements have reciprocals. That does suffice, in spite of the fact that we have full equivalence between geometry and field theory only for **ECG** and Euclidean fields, for if the weaker geometry proved the strong parallel axiom SP, then the interpretation of SP in field theory would be provable, as *that* direction does work, and the interpretation of SP implies that nonzero elements have reciprocals.

A ring of real-valued functions

We need a Kripke model in which positive elements have reciprocals, but nonzero elements do not necessarily have reciprocals. We construct such a Kripke model whose "points" are functions from $\mathbb{R}$ to $\mathbb{R}$. The function f is *positive semidefinite* if $f(x) \geq 0$ for all real x. Let $\mathbb{K}$ be the least subfield of the reals closed under square roots of positive elements. Let $\mathcal{A}$ be the least ring of real-valued functions containing polynomials with coefficients in $\mathbb{K}$, and closed under reciprocals and square roots of positive semidefinite functions. For example

$$\sqrt{\sqrt{1+t^2}+\sqrt{1+t^4}}+\frac{1}{1+t^2}$$

is in $\mathcal{A}$, but $1/t$ is not in $\mathcal{A}$.

Zeroes of functions in $\mathcal{A}$

We show (using Pusieux series) that each member of $\mathcal{A}$ has finitely many zeroes and singularities and that there is a countable set Ω including all zeroes and singularities, whose complement is dense in $\mathbb{R}$.

Kripke models based on rings of functions

We take $\mathcal{A}$ as the root of a Kripke model, interpreting the positivity predicate $P(x)$ to mean x is positive definite. For $\alpha \notin \Omega$, we define $\mathcal{A}_\alpha$ by interpreting $P(x)$ to hold if and only if $x(\alpha) > 0$. In our Kripke model, A_α lies immediately above the root. Now t is a nonzero element without a reciprocal. But if x is positive, then $x(\alpha) > 0$ for all $\alpha \notin \Omega$, and since the complement of Ω is dense and x is continuous, x is positive semidefinite, so $1/x$ exists in $\mathcal{A}$.

Playfair does not imply Euclid 5

To prove that Playfair does not imply Euclid 5, we use a similar Kripke model, starting with a different ring $\mathcal{A}$. This time $\mathcal{A}$ is constructed in stages, each time adding square roots of positive definite functions and reciprocals (not of every positive definite function already added but) of positive definite functions bounded below by functions that already have reciprocals.

Summary

Euclid needs only two modifications to be completely constructive: we have to postulate a rigid compass, rather than relying on Prop. I.2 to simulate it, and we have to take the strong parallel axiom instead of Euclid 5. With those changes Euclid is entirely constructive, and **ECG** formalizes Euclid nicely.

The classical constructions used to define addition and multiplication involve non-constructive case distinctions, but these can be replaced by more elaborate constructions that are continuous (and constructive), so geometry can still be shown equivalent to the theory of Euclidean fields, and different versions of the parallel axiom correspond to weakenings of the field axiom about reciprocals.

Summary

ECG has the nice property that things it can prove to exist can be constructed with ruler and compass, and hence depend continuously on parameters.

ECG permits us to distinguish between versions of the parallel axiom with different constructive content, even though non-constructively they are equivalent, and using Kripke models whose "points" are real-valued functions, we proved formal independence results to make those distinctions sharp.

To read more

Google *Michael Beeson Cambridge.pdf*
for the 10-page version.

Google *Michael Beeson ConstructiveGeometryLong.pdf*
for the unfinished 260 page version (it is unfinished, but it has complete proofs).