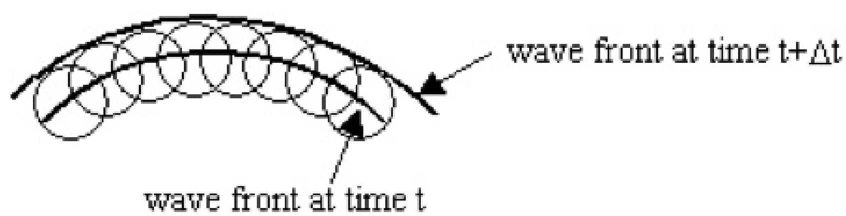
Huygens' Principle
Ponta Delgada talk
Anders Kock
University of Aarhus

1. The principle

Huygens' Principle of geometric optics describes how a wave front B proceeds: at a later time, it is an *envelope* of the spherical *wavelets* emanating from points of B.



The wave front at time $t + \Delta t$ is the *envelope* of the family of wavelets (spheres) of radius Δt, with centers on the wave front at time t. *Envelope* means that it is *touched* by each of the wavelets.

The crucial notion involved is that of *touching* of subspaces. This is here described synthetically, based on a primitive notion of when two points are (first-order) *neighbours*. The neighbour relation is reflexive and symmetric (but not transitive).

If the wave front at time t is a hypersurface B, then the wave front at time $t + \Delta t$ will be another hypersurface, $B \vdash \Delta t$. (We write s for Δt in the following.)

All spaces considered below are subspaces of a fixed space M, which we think of as Euclidean 2- or 3-space.

2. Primitive notions for an axiomatic account

We require an ambient space M; intended interpretation: Euclidean 2- or 3-space. We require two structures on it:

• A reflexive symmetric relation $\sim$ on M, the "neighbour relation" $\sim$: write $x \sim y$. It is not required to be transitive.

• A (pre-) metric dist on M.

They are used to define the two basic derived notions, entering in the statement of Huygens' principle:

- The neighbour relation to define when two subspaces of M *touch* in a point (and hence the notion of *envelope*).

- The distance function, to define the notion of *sphere* (or circle). The distance $\mathrm{dist}(x, y)$ between two distinct points takes value in a commutative semigroup $R_{>0}$ with certain properties.

2.1 Touching

Let S_1 and S_2 be subspaces of M, and assume that $b \in S_1 \cap S_2$. Then we say that S_1 and S_2 *touch in* b if

$$\mathfrak{M}(b) \cap S_1 = \mathfrak{M}(b) \cap S_2, \tag{1}$$

or, in written in pure 1st order logic:

$$b' \sim b \wedge b' \in S_1 \text{ implies } b' \in S_2$$

and

$$b' \sim b \wedge b' \in S_2 \text{ implies } b' \in S_1.$$

It is clear that for fixed b, "S_1 touches S_2 in b" is an equivalence relation on the set of subspaces containing b.

3. Neighbours and touching in SDG

To *motivate* the notions of 'neighbours' and 'touching', I shall recall some aspects of analytic geometry over a basic commutative ring R ("number line"), as it has been known now for more than half a century in SDG.

In SDG, there is a basic number line R, which has a rich supply of elements $d \in R$, meaning that they satisfy $d^k = 0$ for some integer k (with $k = 2$ being the important case).

Already then, we can define when two points x and y on the number line are (first order) neighbours:

$$x \sim y \text{ iff } (x - y)^2 = 0.$$

This is clearly a reflexive and transitive relation: $x \sim x$; and $x \sim y$ iff $y \sim x$. It is not in general a *transitive* relation. For, $(x - y)^2 = 0$ and $(y - z)^2$ does imply $(x - z)^3 = 0$, by binomial expansion, but does not imply $(x - z)^2 = 0$.

Subspaces $S \subseteq M$ of M inherit a neigbour relation from M in an obvious way.

All maps ϕ which we construct between such subspaces *preserve* the $\sim$-relation. This is a continuity property of ϕ.

(We don't put such "automatic continuity" in the axiomatics, rather, it should be a property of the ambient (un-named) category in which we work.)

In algebraic geometry, say, the categoy of affine schemes M, the subscheme $M_{(1)} \subseteq M \times M$ (first neighbourhood of the diagonal) will serve as extension of the $\sim$-relation.

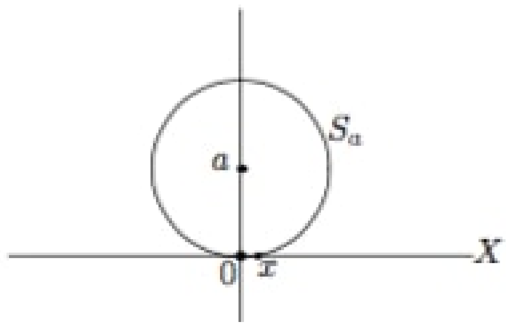
One piece of notation before we turn to "touching": for $x \in M$, we let

$$\mathfrak{M}(x) := \{y \in M \mid x \sim y\},$$

we call it the (first order) *monad* around x. By reflexivity of $\sim$, we have $x \in \mathfrak{M}(x)$.

A *necessary* condition for S_1 and S_2 to touch in a point b is that $b \in S_1 \cap S_2$. In other words, the set of points where S_1 and S_2 is a *subset* of $S_1 \cap S_2$, and this subset, we call the *touching set*. But in general, the touching set will be smaller; this will be crucial for the theory.

To understand why this is so, we turn to the model from SDG. Let S_a be the circle in R^2 with radius $a > 0$ and center $(0, a)$, and let X be the x-axis.



From Pythagoras Theorem follows that $S_a \cap X$ consists of points $(x, 0)$ with $x^2 = 0$. The picture itself is presented by the pre-socratic Greek philosopher Protagoras, who is reported to have stated that a circle and a tangent to it have more than one point in common.

Since the neighbour relation is not transitive, we cannot conclude that a neighbour $(x', 0)$ of $(x, 0) \sim (0, 0)$ has $x'^2 = 0$, and therefore such $(x, 0) \in S_a \cap X$ is not necessarily a touching point of S_a and X; *common points are not necessarily touhing points*. In fact, assuming the basic KL axiom for R, it is easy to prove that $(0, 0)$ is the *only* touching point of S_a and X.

Note that $S_a \cap X$ is independent of the radius a! (Contrary to an intuition which would perhaps expect that "larger a give larger $S_a \cap X$". So "smallness" is a qualitative determination, not a quantitative one", when we take "x is small" to mean "$x \sim 0$".)

4. (Pre-) metric dist

We now in principle forget coordinate geometry, and turn to the *axiomatic* theory for a space M. Besides the primitive *neighbour* relation $\sim$ (reflexive and symmetric), we need a weak notion of *metric* (distance). More precisely, we need a commutative semigroup to receive the values of the distance function. We write this semigroup $R_{>0}$, because what we have in mind, for the intended interpretation in analytic geometry, is e.g. the strictly positive real numbers, under addition. The distance dist(x, y) is defined whenever x and y are *distinct* (e.g. for the case where R is the basic ring in a model of SDG, x and y in R are called distinct if $x - y$ is multiplicatively invertible; thus if $x \sim y$, x and y are not distinct). We assume symmetry: dist$(x, y) =$ dist(y, x). We do not assume any kind of triangle inequality; but for some triples a, b, c of points, it may happen that we have a triangle *equality*

$$\text{dist}(a, b) + \text{dist}(b, c) = \text{dist}(a, c),$$

a weak kind of collinearity condition.

The algebraic properties of the assumed addition on $R_{>0}$ are just that the natural strict order $<$ derived from the addition should be a total order:

$$r < t := \exists! s : r + s = t \tag{2}$$

with the "dichotomy"-property: if r and t are distinct, then *either* $r < t$ *or* $t < r$. This unique s satisfying (2), we call the *difference* between t and r, written $t - r$.

Note that we do not assume any multiplication on $R_{>0}$, nor any unit. A physical model of $R_{>0}$ is realizable, using pieces of string.

Neither do we assume any 0.

With such a "pre-metric", we define for $a \in M$ and $r \in R_{>0}$ the sphere $S(a, r)$ with center a and radius r by

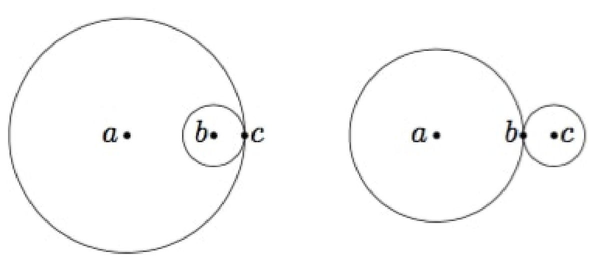
$$S(a, r) := \{b \in M \mid \text{dist}(a, b) = r\}.$$

We say that two spheres are *non-concentric* if their centers are distinct.

5. The axioms

Axiom 1: If two non-concentric spheres touch, there is a unique touching point.

Axiom 2: Two non-concentric spheres touch iff *either* the distance between their centers equals the *difference* between their radii, *or else* equals the *sum* of their radii.



Axiom 3: Given two non-concentric spheres S_1 and S_2 and given an $b \in S_1 \cap S_2$. Then if $\mathfrak{M}(b) \cap S_1 \subseteq S_2$, the two spheres touch in b; Equivalently, $\mathfrak{M}(b) \cap S_1 \subseteq S_2$ implies $\mathfrak{M}(b) \cap S_2 \subseteq S_1$.

The two geometric axioms 1. and 2. are true in any version of Euclidean geometry, provided the notion of *touching* is defined in an appropriate way; typically, for in analytic geometry, in terms of differential calculus. Axiom 3 comes, in the intended analytical interpretations, about from the fact from linear algebra, that if two (finite dimensional) linear subspaces A_1 and A_2 of a vector space have the same dimension, and if $A_1 \subseteq A_2$, then $A_1 = A_2$. - I don't know whether Axiom 3 could possibly be dispensed with.

6. Reciprocity

Given distinct points a and b with dist$(a, b) = r$, and given $s \in R_{>0}$. By Axiom 2 (first part), the spheres $S(b, s)$ and $S(a, r + s)$ touch; by Axiom 1, they touch in a unique point c.

This c we may denote $a \triangleright_s b$ (r need not be mentioned, it is known when a and b are). (Geometrically, $a \triangleright_s b$ may be seen as an extrapolation of the line segment from a to b by the amount s beyond b).

Since $c \in S(b, s) \cap S(a, r + s)$, we have dist$(a, c) = r + s$ and dist$(b, c) = s$, so together with the assuption dist$(a, b) = r$, this gives that the triangle equality holds,

$$\text{dist}(a, b) + \text{dist}(b, c) = \text{dist}(a, c); \tag{3}$$

which is thus a necessary condition for the $S(b, s)$ and $S(a, r + s)$ to touch in c; such touching of two spheres, we call *concave* or *internal* touching.

Using the second part of Axiom 2, there is similarly, a necessary condition for spheres $S(a, r)$ and $S(c, s)$ to touch in a point b, namely again the triangle equality (3); such touching of two spheres, we call *convex* or *external* touching. The touching point b, we may denote $a \triangleleft_s c$ (only meaningful for $s <$ dist(a, c)). (Geometrically, $a \triangleleft_s c$ may be seen as pushing c in the direction back to a by s units).

This far, the theory is in some sense completely classical, if one takes $x \sim y$ to mean $x = y$ (the finest possible $\sim$, i.e. as few neighbour pairs as possible). In fact, the German-American geometer Herbert Busemann developed in the 1940s to 1960s a geometric theory (which he ultimately was to call “synthetic differential geometry”). However, with a non-trivial neighbour relation, the triangle equality (3) for three points a, b, c are not sufficient for c, respectively b, to be the relevant touching point.

Let us spell out in terms of the primitive notions *what more* is required for three points a, b, c, which satisfy the triangle equality (3), in order to have c as (concave) touching point (write r for dist(a, b) and s for dist(b, c)):

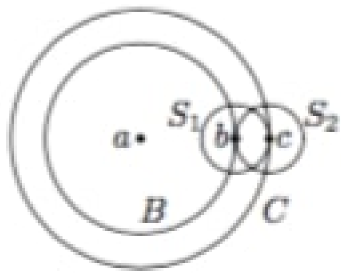
$$\forall c' \sim c : \text{dist}(a, c) = r + s \Leftrightarrow \text{dist}(b, c) = s, \tag{4}$$

Simlarly, b is convex touching point if besides the triangle equality, we have

$$\forall b' \sim b : \text{dist}(a, b') = r \Leftrightarrow \text{dist}(b', c) = s. \tag{5}$$

With a, b, c in M and r, s in $R_{>0}$, as above satisfying the triangle equality (3), we can prove that (4) is equivalent to (5):

Lemma 1 (Reciprocity) *The point b is the (convex) touching point of $S(a, r)$ and $S(c, s)$ iff c is the (concave) touching point of $S(a, r + s)$ and $S(b, s)$. Equivalently $a \lhd_s (a \rhd_s b) = b$ and similarly $a \rhd_s (a \lhd_s c) = c$.*



In view of the Dimension Axiom, Axiom 3, the $\Leftrightarrow$ may be replaced by $\Rightarrow$, or by $\Leftarrow$ in either of the formulae (4) or (5); if we replace e.g. the $\Leftrightarrow$ in (5) by $\Rightarrow$,

$$\forall b' \sim b : \text{dist}(a, b') = r \Rightarrow \text{dist}(b', c) = s, \tag{6}$$

the condition says: b is a *critical* (stationary) point of the function dist(x, c) under the constraint dist$(a, x) = r$. There are three analogous reformulations.

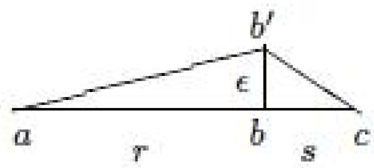
So a, b, c is strongly collinear if the weak collinearity condition (triangle equality) *plus* validity of (6) (or any other the analogous reformulations) hold.

A function on M (with values in $R_{>0}$), say), has x as a *critical point* if it is constant on $\mathfrak{M}(x)$. The *critical value* of the function is then this constant value.

If $x \in S \subseteq M$, then x is a critical point *under the constraint* $x \in S$ if the function is constant on $\mathfrak{M}(x) \cap S$; this constant value is the critical value of the function *under that constraint*.

Using this terminology, the additional condition (6) for strong linearity can be expressed: "b is a critical point for the function dist(x, c) under the constraing $x \in S(a, r)$".

The following example refers to the intended interpretation of the axiomaics in SDG. Consider in R^2 three points a, b, c on the x-axis, with $b = (0, 0)$



Then it is easy to see that a, b, c are strongly collinear. And dist$(a, b') = r$ and dist$(b', c) = s$, so a, b', c are weakly collinear. The path from a to c via b' is as short as the path via b, namely $r + s$. "Shortest path" is not enough for (strong) collinearity.

7. Contact elements

The non-triviality of the neighbour relation brings with it the notion of *contact element*; a *contact element* P *at*[1] $b \in M$ is a set of the form $\mathfrak{M}(b) \cap S$, for some sphere S, with $b \in S$. A point c distinct from b is called *orthogonal to* P if it satisfies dist$(b', c) =$ dist(b, c) for all $b' \in P$. Equivalently, if it satisfies $P \subseteq S(c, s)$, where $s :=$ dist(b, c). Write $c \perp P$ for this condition. The set of points orthogonal to P is called the *normal* to P.

Given a contact element at b. Clearly, two spheres containing b define (or represent) the same contact element iff they touch each other at b.

In the intended interpretation in analytic differential geometry, the space of contact elements in M make up the projectivized cotangent bundle of M. Contact elements were studied analytically by Lie under the name "*Linienelemente*" or "*Flächenelemente*" in his 1896 book "Berührungstransformationen" ("Contact transformations").

Sophus Lie: "It is often *practically* convenient to think of a line element as an infinitely small piece of a curve."

Significantly, he does not say "infinitely small piece of a *line*"; for, it amounts to the same thing, when "infinitely small piece" means: what you get by taking meet with $\mathfrak{M}(b)$ - as in $\mathfrak{M}(b) \cap S$.

A contact element P at b may be given a (transversal) *orientation* by selecting a sphere S representing it; any sphere touching S concavely at b then also represents this oriented contact element. The normal is a disjoint union of two parts, the positive normal and the negative one.

Given a contact element (P, b). Then we can construct points on its normal as follows. Pick a sphere $S(a, r)$ representing the contact element. For any $s \in R_{>0}$, we have the point $c := a \triangleright_s b$, the touching point of $S(a, r + s)$

[1]so P is to be considered as *pointed* set, written (P, b).

and $S(b,s)$. By the Reciprocity Lemma, b is the touching point of $S(a,r)$ and $S(c,s)$, in particular

$$b' \sim b \wedge \textbf{dist}(a,b') = r \Rightarrow \textbf{dist}(b',c) = s,$$

so c has same distance s to all points of $P = \mathfrak{M}(b) \cap S(a,r)$. This is the orthogonality condition $c \perp P$.

One may ask whether this construction of c is independent of the choice of $S(a,r)$? The answer is: almost; we shall see that one gets exactly *two* points this way. If P is given an orientation, one of them is on the positive normal.

Let c be any point on the positive normal to P with $\text{dist}(b,c) = s$ and hence $\text{dist}(b',c) = s$ for all $b' \in P = \mathfrak{M}(b) \cap S(a,r)$, for any sphere $S(a,r)$ representing P. Equivalently $\mathfrak{M}(b) \cap S(a,r) \subseteq S(c,s)$ which is to say that the spheres $S(a,r)$ and $S(c,s)$ touch in b. So by the Reciprocity Lemma, c is the touching point of $S(a,r+s)$ and $S(b,s)$, so c is $a \triangleright_s b$.

The unique point c described in this Proposition deserves a notation, we write $P \vdash s$ for it (more fully $(P,b) \vdash s$); it is thus characterized by: $c \perp P$, $\text{dist}(b,c) = s$, and c being on the positive side of P. The proof gives that it may be *constructed* by *picking* an arbitrary sphere $S(a,r)$ touching P from the inside. In terms of such sphere c is characterized by being outside $S(a,r)$ and satisfying

$$\forall b' \ [(b' \sim b \wedge \text{dist}(a,b') = r) \Rightarrow (\text{dist}(b',c) = s)]. \tag{7}$$

Note that the condition that $S(a,r)$ touches $S(c,s)$ in b implies that

$$\forall b' : [b' \in P \Rightarrow \text{dist}(b',c) = s]$$

which may be written

$$c \in \bigcap_{b' \in P} S(b',s); \tag{8}$$

The condition (8) does not involve any somewhat arbitrarily chosen sphere $S(a,r)$. It shows that the point $P \vdash s$ is a *characteristic point*, in the discriminant-sense for a discussion of this "predicative" way of constructing envelopes).

The meet on the RHS of (8) has for any P exactly two points; the orientation of P allows us to pick the right one of them for the desired c.

The construction of c in terms of a meet

$$\bigcap_{b' \in P} S(b',s)$$

in fact proves that c is a *characteristic point* in the sense of the discriminant method for constructing $B \vdash s$ as an envelope of the wavelets $S(b',s)$, see [END] for an analysis of characteristics in terms of intersections like $\bigcap_{b' \in B(b)} S(b',s)$.

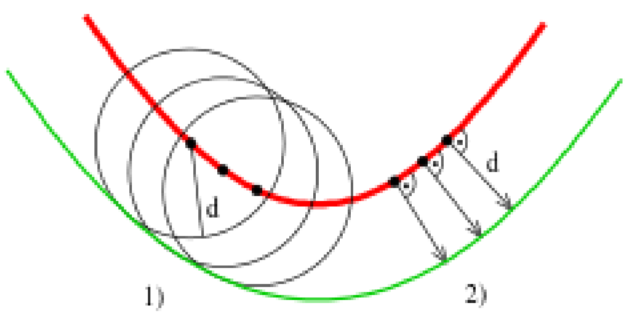
8. Hypersurfaces

We are interested in contact elements, because they are the "infinitesimal pieces" which a hypersurface is made up of:

A *hypersurface* is a subset $B \subseteq M$, such that for each $b \in B$, $\mathfrak{M}(b) \cap B$ is a contact element (which we denote $B(b)$). In particular, any sphere is a hypersurface. To give a hypersurface an orientation is to give each of its contact elements an orientation. A sphere carries a canonical orientation.

Huygens' Principle may be expressed: *given an oriented hypersurface B, and given a sufficiently small $s \in R_{>0}$. Then there exists another oriented hypersurface $B \vdash s$ "at distance s from B"*. If B is a wave front, then $B \vdash s$ is the wave front to which B develops in "time" s. Huygens constructed it as an envelope of the "wavelets", spheres of radius s wih center on B.

For s small enough, there is a bijection $B \to B \vdash s$. The point $B(b) \vdash s$ corresponding to $b \in B$ geometrically is obtained as the point one gets by "going distance s out along the normal to B at b" (in the "positive" direction). And b is the "foot" of the point $B(b) \vdash s$ on B. This makes sense in analytic/differential geometry; the point I want to make, is that it makes sense already in the present purely combinatorial context. (In the picture d denotes our s.)



One may correctly conjecture that this point may be obtained by the following procedure. Pick a sphere $S(a, r)$ such that $\mathfrak{M}(b) \cap B = \mathfrak{M}(b) \cap S(a, r)$ (with the correct orientation), and then perform the extrapolation construction $a \triangleright_s b$; i.e. put

$$B(b) \vdash s := a \triangleright_s b,$$

the point where $S(a, r + s)$ touches $S(b, s)$; this, however, leaves the problem of well-definedness: is it independent of the choice of the sphere $S(a, r)$?. In fact, it brings in data, like the point a, which are not intrinsic to B. The construction which we now give only uses the points in the contact element $P := B(b)$ itself (in particular, the construction does not involve points of B outside P, and no restriction on smallness of s is needed). Namely: $P \vdash s$ *is the unique point c on the positive normal to P which belongs to all the* *spheres* $S(b', s)$ *for* b' *in* P. (It is clear that c does belong to the normal: it is ortogonal to P, having distance s to all points of P.)

To see the uniqueness, let c belong to all $S(b', s)$ for $b' \in B(b)$. pick a sphere $S(a, r)$, as above. Then the construction gives that $S(c, s)$ touches $S(a, r)$ (convex touching) in b. By the Reciprocity Lemma, we get that $S(a, r + s)$ touches $S(b, s)$ in c (concave touching), thus $c = a \triangleright_s b$. This proves the well definedness of $P \vdash s$.

This does not yet prove that the set of points obtained as $C := B(b) \vdash s$ for $b \in B$ again constitute a hypersurface; among other things, s needs to be sufficiently small, if for instance the positive side of B is concave. Assuming that s is so small that the map $B \to C$ given by $b \mapsto B(b) \vdash s$ is a bijection, what we need to prove is that for any $c \in C$, we have

$$\mathfrak{M}(c) \cap C = \mathfrak{M}(c) \cap S(b, s), \tag{9}$$

where $B(b) \vdash s = c$. When this is proved, we that C is indeed a hypersurface: the sphere $S(b, s)$ witnesses it for c. And also it proves that C is indeed an envelope of the spheres $S(b, s)$, with $c = B(b) \vdash s$ as touching point for $S(b, s)$.

To indicate the character of the proofs of this in the present type of combinatorics, let us prove $\mathfrak{M}(c) \cap S(b, s) \subseteq C$, where $c = B(b) \vdash s$.

Let $x \in \mathfrak{M}(c) \cap S(b, s)$. Let b' be the foot of x on B. Since $x \sim c$, we have $b' \sim b$ (by automatic continuity of the foot-map).

We have $s = \text{dist}(x, b) = \text{dist}(x, b')$, the first by assumption $x \in S(b, s)$, the second since b' is foot of x on B and $b' \sim b$ in B

So $x = B(b') \vdash s$, hence b' witnesses that $x \in C$.

AK *Envelopes - notion and definiteness*, Beiträge zur Alg. und Geom. 48 (2007).
AK: *Metric spaces and SDG*, TAC 32 (2017).
AK: *Huygens' principle - a synthetic account*, arXiv:1804.05649 [mathDG] (2018).

June 28, 2018