A rigorous deductive approach to elementary Euclidean geometry

Jean-Pierre Demailly

Universit« Joseph Fourier Grenoble I

`http://www-fourier.ujf-grenoble.fr/~demailly/documents.html`

Résumé :

L'objectif de ce texte est de proposer une piste pour un enseignement logiquement rigoureux et cependant assez simple de la géométrie euclidienne au collège et au lycée. La géométrie euclidienne se trouve être un domaine très privilégié des mathématiques, à l'intérieur duquel il est possible de mettre en uvre dès le départ des raisonnements riches, tout en faisant appel de manière remarquable à la vision et à l'intuition. Notre préoccupation est d'autant plus grande que l'évolution des programmes scolaires depuis 3 ou 4 décennies révèle une diminution très marquée des contenus géométriques enseignés, en même temps qu'un affaiblissement du raisonnement mathématique auquel l'enseignement de la géométrie permettait précisément de contribuer de façon essentielle. Nous espérons que ce texte sera utile aux professeurs et aux auteurs de manuels de mathématiques qui ont la possibilité de s'affranchir des contraintes et des prescriptions trop indigentes des programmes officiels. Les premières sections devraient idéalement être maîtrisées aussi par tous les professeurs d'école, car il est à l'évidence très utile d'avoir du recul sur toutes les notions que l'on doit enseigner !

Mots-clés : géométrie euclidienne

Resumen :

El objetivo de este artículo es presentar un enfoque riguroso y aún razonablemente simples para la enseñanza de la geometría euclidiana elemental a nivel de educación secundaria. La geometría euclidiana es una área privilegiada de las matemáticas, ya que permite desde un primer nivel practicar razonamientos rigurosos y ejercitar la visión y la intuición. Nuestra preocupación es que las numerosas reformas de planes de estudio en las últimas 3 décadas en Francia, y posiblemente en otros países occidentales, han llevado a una disminución preocupante de la geometría, junto con un generalizado debilitamiento del razonamiento matemático al que la geometría contribuye específicamente de manera esencial. Esperamos que este punto de vista sea de interés para los autores de libros de texto y también para los profesores que tienen la posibilidad de no seguir exactamente las prescripciones sobre los contenidos menos relevantes, cuando están por desgracia impuestos por las autoridades educativas y por los planes de estudios. El contenido de las primeras secciones, en principio, debería también ser dominado por los profesores de la escuela primaria, ya que siempre es recomendable conocer más de lo que uno tiene que enseñar, a cualquier nivel !

Palabras clave : Geometría euclidiana

0. Introduction

The goal of this article is to explain a rigorous and still reasonably simple approach to teaching elementary Euclidean geometry at the secondary education levels. Euclidean geometry is a privileged area of mathematics, since it allows from an early stage to practice rigorous reasonings and to exercise vision and intuition. Our concern is that the successive reforms of curricula in the last 3 decades in France, and possibly in other western countries as well, have brought a worrying decline of geometry, along with a weakening of mathematical reasoning which geometry specifically contributed to in an essential way. We hope that these views will be of some interest to textbook authors and to teachers who have a possibility of not following too closely the prescriptions for weak contents, when they are unfortunately enforced by education authorities and curricula. The first sections should ideally also be mastered by primary school teachers, as it is always advisable to know more than what one has to teach at any given level !

Keywords : Euclidean geometry

1. On axiomatic approaches to geometry

As a formal discipline, geometry originates in Euclid's list of axioms and the work of his successors, even though substantial geometric knowledge existed before.

8

32. Celle qui a les coſtez egaux, mais non les angles droiƈts, ſe nomme Rhombe.
33. Mais celle qui a les coſtez oppoſez & les angles oppoſez egaux, mais n'a pas coſtez egaux enſemble ny les angles droiƈts, ſe nomme Rhomboïde.
34. Excepté celle-cy, les figures de quatre coſtez ſe nomment tablettes ou trapeſes.
35. Lignes droiƈtes paralleles ſont celles, leſquelles eſtans en vn meſme plan & menees de part & d'autre ne concourent point enſemble.
36. Parallelogramme, eſt vne figure planę contenuë de quatre lignes droiƈtes, deſquelles les oppoſees ſont paralleles entr'elles.

S'ENSVIVENT LES DEMANDES, ou poſitions ſimples.

Demande I.

D'Vn poinƈt à vn autre poinƈt mener vne ligne droiƈte.
2. Continuer vne ligne droiƈte finie tant qu'il en ſera beſoin.
3. Deſcrire vn cercle à l'entour d'vn poinƈt, & d'vne telle diſtance qu'on voudra.

9

COMMVNES SENTENCES.

Premiere.

LEs choſes egales à vne, ſont egales entr'elles.
2. Et ſi à choſes egales s'adiouſtent choſes egales, les toutes ſeront egales.
3. Et ſi de choſes egales, ſe leuent choſes egales, les reſtes ſeront egales.
4. Et ſi à choſes inegales s'adiouſtent choſes egales, les toutes ſeront inegales.
5. Et ſi des choſes inegales ſe leuent choſes egales, les reſtes ſeront inegales.
6. Les choſes qui ſont doubles à vn meſme, ſont egales entr'elles.
7. Et les choſes qui ſont moiƈtié d'vne meſme, ſont egales entr'elles.
8. Et les choſes qui conuiennent enſemble & entre elles, ſont egales entr'elles.
9. Le tout eſt plus grand que ſa partie.
10. Tous angles droiƈts ſont egaux entr'eux.
11. Si deſſus deux lignes droiƈtes tombe vne ligne droiƈte, faiſant les angles dedans d'vne meſme part, plus petits que deux droiƈts, icelles eſtans prolongees infiniment ſe rencontreront de la part en laquelle les angles ſont plus petits que deux droiƈts.
12. Deux lignes droiƈtes ne comprennent pas vn eſpace.

B

An excerpt of Euclid's book

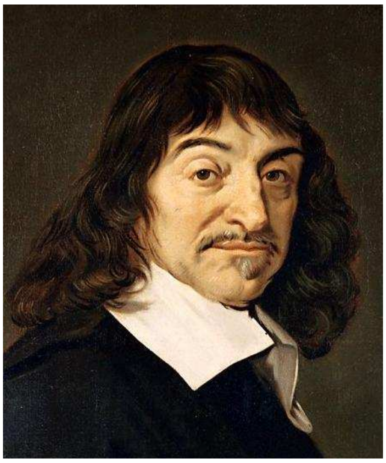
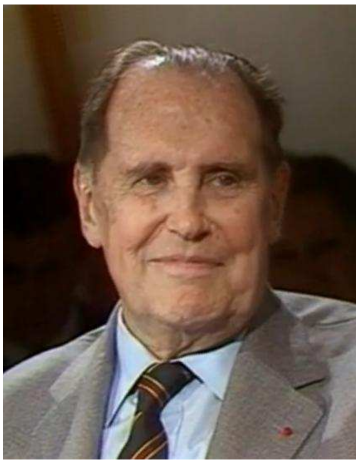
The traditional teaching of geometry that took place in France during the period 1880-1970 was directly inspired by Euclid's axioms, stating first the basic properties of geometric objects and using the "triangle isometry criteria" as the starting point of geometric reasoning. This approach had the advantage of being very effective and of quickly leading to rich contents. It also adequately reflected the intrinsic nature of geometric properties, without requiring extensive algebraic calculations. These choices echoed a mathematical tradition that was firmly rooted in the nineteenth century, aiming to develop "pure geometry", the highlight of which was the development of projective geometry by Poncelet.

Euclid's axioms, however, were neither complete nor entirely satisfactory from a logical perspective, leading mathematicians as Hilbert and Pasch to develop the system of axioms now attributed to Hilbert, that was settled in his famous memoir Grundlagen der Geometrie in 1899.

David Hilbert (1862–1943), in 1912

It should be observed, though, that the complexity of Hilbert's system of axioms makes it actually unpractical to teach geometry at an elementary level[(1)]. The result, therefore, was that only a very partial axiomatic approach was taught, leading to a situation where a large number of properties that could have been proved formally had to be stated without proof, with the mere justification that they looked intutively true. This was not necessarily a major handicap, since pupils and their teachers may not even have noticed the logical gaps. However, such an approach, even though it was in some sense quite successful, meant that a substantial shift had to be accepted with more contemporary developments in mathematics, starting already with Descartes' introduction of analytic geometry. The drastic reforms implemented in France around 1970 (with the introduction of "modern mathematics", under the direction of André Lichnerowicz) swept away all these concerns by implementing an entirely new paradigm : according to Jean Dieudonné, one of the Bourbaki founders, geometry should be taught as a corollary of linear algebra, in a completely general and formal setting. The first step of the reform implemented this approach from "classe de seconde" (grade 10) on. A major problem, of course, is that the linear algebra viewpoint completely departs from the physical intuition of Euclidean space, where the group of invariance is the group of Euclidean motions and not the group of affine transformations.

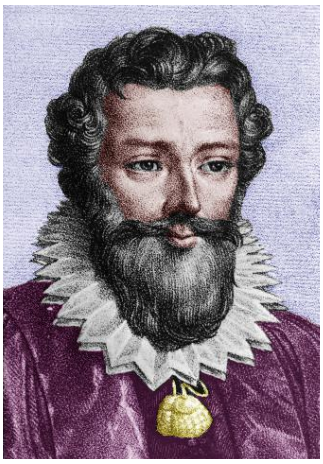
(1) Even the improved and simplified version of Hilbert's axioms presented by Emil Artin in his famous book "Geometric Algebra" can hardly be taught before the 3$^{\text{rd}}$ or 4$^{\text{th}}$ year at university.

from Descartes (1596-1650) to Dieudonné (1906-1992) and Lichnerowicz (1915-1998)

The reform could still be followed in a quite acceptable way for about one decade, as long as pupils had a solid background in elementary geometry from their earlier grades, but became more and more unpractical when primary school and junior high school curricula were themselves (quite unfortunately) downgraded. All mathematical contents of high school were then severely axed around 1986, resulting in curricula prescriptions that in fact did not allow any more the introduction of substantial deductive activity, at least in a systematic way.

As far as geometry is concerned, it is therefore essential to return to teaching practices that clearly introduce the geometric nature of objects under consideration. This means that precise definitions should be explicitly given and that curricula should be organized in a way that allows to state and prove a rich set of properties from those taken as the starting point. In a word, one should return, in a possibly not entirely explicit form, to an "axiomatic" presentation of geometry. We do not mean here that the order of presentation of concepts must necessarily follow the order implied by the internal logical constraints of "axioms" (whatever they are), but one should at least provide a clear framework that can be adopted by teachers and that serves as a firm guide for designing curricula.



François Viète (1540–1603)

Simon Stevin (1548–1620)

To justify our desire to go beyond the traditional Euclidean approach that was in use 50 or 60 years ago, our observation is that one now has a considerable advantage over the Greek, which is to have at our disposal an efficient and universally accepted algebraic notation thanks to Simon Stevin, and the possibility to investigate geometry though coordinates and analytic calculations, thanks to Descartes.

In Greek geometry, real numbers were only thought as ratios of quantities of the same kind, rather than "abstract numbers" in the modern sense. The concept of a polynomial function (or of a general function) was then much harder to tackle, although the Greeks certainly had methods for solving problems involving second and third degree equations through geometric constructions.

The approach we propose here is somehow a synthesis of the viewpoint of Pythagoras and Euclid with that of Descartes. Euclidean geometry is characterized by the expression of the distance through Pythagoras' theorem. All objects involved in Euclidean geometry can then be derived merely from the concept of length[(2)]. In this context, our reconstruction of Euclidean geometry derives Thales' theorem from Pythagoras' property. Another advantage is that all concepts can be defined using a minimal and intuitive formalism. In fact, the theory can be set to use a single axiom, directly related to Pythagoras' theorem, that can moreover be easily "justified" by simple visual considerations. However, the crucial point is certainly not the use of a single axiom – however what we call the "Pythagoras/Descartes" axiom, rather than a genuine axiom, is rather a concise description of a model of Euclidean geometry. Unlike the approach based on linear algebra, we start from points and affine concepts rather than from the much less intuitive concepts of vectors and vector spaces, and here the notion of vector will be constructed a posteriori. Another goal is to invalidate the occasional argument that elementary traditional geometry does not constitute a serious or useful part of mathematics, because is cannot be exposed in a rigorous or formalized way, in the modern sense.

However, there are certainly some disadvantages that are inherent to our exposition. One of them is that it is only a "modern reconstruction", and in spite of being somehow obvious to contemporary mathematicians (and most probably so to Klein or Hilbert), it has probably never been taught in this form on a large scale. Another feature is that giving from scratch a "rigid model" of Euclidean geometry may not be the most appropriate framework to introduce other subjects such as affine or projective geometry, which rely rather on incidence properties. Finally, real numbers are deeply embedded in the model, so, at a later stage, introducing geometries over other fields will require a different approach. Nevertheless, for a potential use at secondary school level, we have deliberately preferred simplicity to generality, and have therefore chosen to focus on the Euclidean and Archimedean model which is also the one of Newtonian mechanics . . .

2. Geometry, numbers and arithmetic operations

According to the above discussion, teaching geometry is inseparable from teaching

(2) It is well known today that a metric structure determines many other geometrical invariants, such as Hausdorff measures, curvature, tensors, etc. This viewpoint has been extensively developed by Mikhail Gromov in the last 2 or Last 3 decades. See for example M. Gromov, *Metric structures for Riemannian manifolds* by J. Lafontaine and P. Pansu, Cedic/Fernand Nathan, Paris, 1981.

numbers. This is especially true for the study of contemporary versions of the Euclidean model, which are based on real numbers. We first describe fundamental facts about numbers that should be taught accordingly.

2.1. Primary school and the 4 arithmetic operations

It is essential that handwritten calculations be taught again thoroughly at the primary school level, so as to gain a complete control of operation algorithms – calculators should be used only when students have already acquired a very good acquaintance to arithmetic operations. Safe and effective practice of handwritten calculations requires a fluid knowledge of addition and multiplication tables (as well as "subtraction and division tables" – by this, we mean being able to read addition and multiplication tables in the opposite direction). Some educators inclined to a "minimalist approach" have insisted on the fact that it is enough to mentally perform approximate calculations and estimated orders of magnitude, but the following facts are unavoidable :

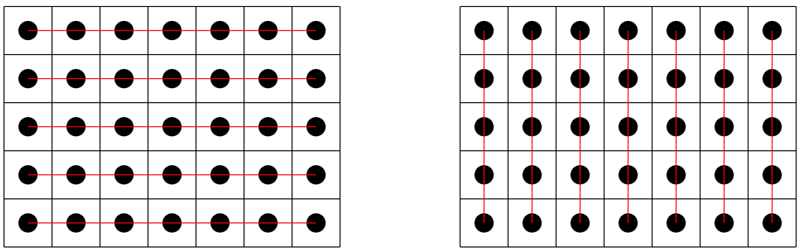
- although mental arithmetic implies a fluid knowledge of the multiplication table, just as handwritten calculations do, its procedures are different, since intermediate results have to be memorized. This is done by handling units, tens, hundreds, thousands, rather than digits taken separately, and one usually starts as well by taking into account higher weight digits rather than lower weight digits as in the handwritten algorithms. In addition to this, the smaller size of numbers involved does not usually allow to reach the level of generality that is necessary to acquire a complete understanding of the algorithms of handwritten calculations.
- even if young children may develop some sort of intuitive perception of the size of numbers before being able to perform exact calculations (this ability should of course not be neglected or negated), reaching a sufficient reliability in performing just approximate estimates is attained only by means of certain exact procedures, such as calculating powers of ten combined with a knowledge of the multiplication table.
- Finally, developing an ability to performing approximate calculations (just in case this would be the major target) is greatly enhanced when one masters e.g. the division algorithm : when the divider has two or more digits, guessing the relevant digit in the quotient requires a very effective ability to intuitively grasp the approximate value of the product of a several digit umber by a single digit number. In this circumstance, it is clear that children need precisely defined objectives to build their mental models, it is certainly not sufficient that the curricula declare the ability to perform approximate calculation a worthy target to hope that this goal will be achieved to some sort of spontaneous perception.

Of course, mastering the arithmetic algorithms is very far from being sufficient to understand numbers ; children can access the real meaning of operations only by solving numerous problems involving everyday life objects (counting apples, currency, lengths, weights...). Contrary to what has been stated in certain "modern" pedagogical theories, this intuitive capability can be built much more efficiently when all four arithmetic operations are introduced simultaneously, so that children can compare (and oppose) the use of different operations. Therefore introducing the four arithmetic operations should be done rather early, e.g. in the first grade, and certain related activities can be considered even at pre-school levels (kindergarten).

2.2. Synergy of numeracy and geometry

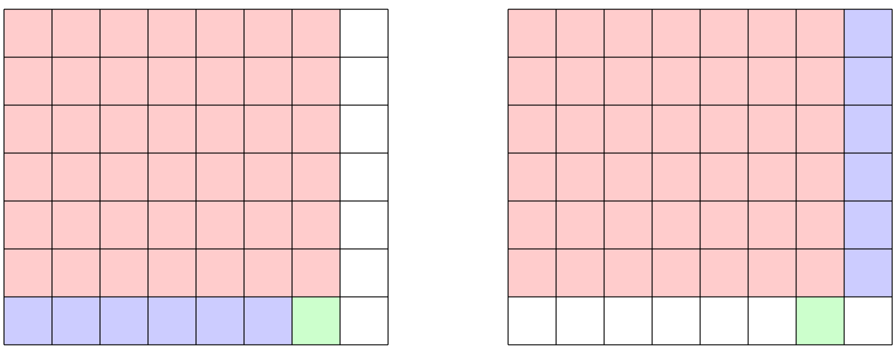
Geometric activities can already be organized at kindergarten, for example through drawing or coloring. It is certainly very useful to ask children to draw simple geometric patterns, friezes, etc.

During the first years of primary school, geometric activities can be organized in close connection with arithmetic calculations. For instance, calculating the perimeter of a rectangle involves addition, while calculating the area involves multiplication, and possibly changes of units (when dealing with the case where sides are decimal numbers). Also, the geometric representation of the rectangle immediately gives a proof of the commutativity of multiplication :

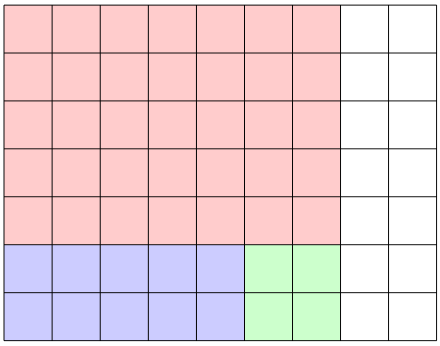


$7 \times 5 = 5 \times 7$

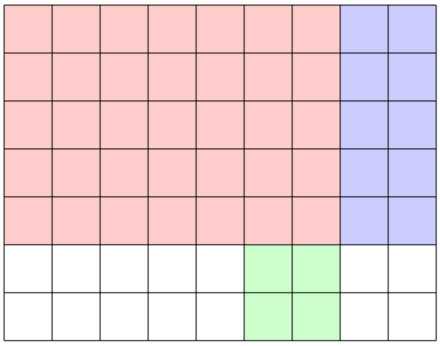
Certain numerical identities such as $6 \times 8 = 7 \times 7 - 1$ or $5 \times 9 = 7 \times 7 - 4$ can be viewed and explained through geometric considerations (in this direction, it would probably be appropriate to bring pupils to manipulate wooden pieces, so that vision can be consolidated by hand work . . .)



7×7 $\qquad$ $6 \times 8 = 7 \times 7 - 1$



7×7



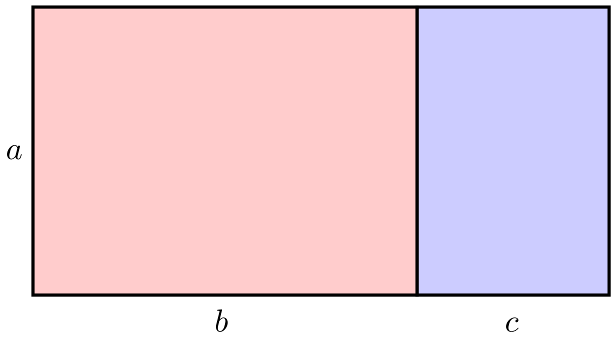
$5 \times 9 = 7 \times 7 - 2 \times 2$

These simple reasonings are actually genuine mathematical proofs, probably among the first ones that can be presented to pupils.

Let us come to the problem of calculating areas. It is natural to start by the area of a rectangle whose sides are integer multiples of the unit of length. Later on, the problem of computing the area of a rectangle whose sides are not whole numbers, such as 1.2 m by 0.7 m, is solved by converting lengths into decimeters. This yields

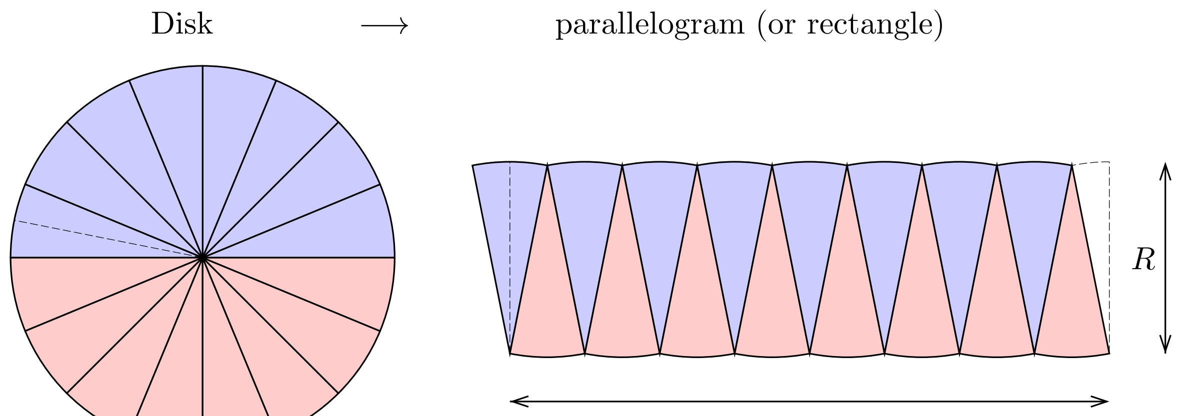
$$12\,\mathrm{dm} \times 7\,\mathrm{dm} = 84\,\mathrm{dm}^2 = 0.84\,\mathrm{m}^2,$$

once it is realized that $1\,\mathrm{m}^2 = 100\,\mathrm{dm}^2$, and thus $1\,\mathrm{dm}^2 = 0.01\,\mathrm{m}^2$. Therefore, one sees that the area of a rectangle is always the product of the lengths of its sides, even when the side lengths are decimal numbers. The distributivity of multiplication with respect to addition can also be seen geometrically :



$$a \times (b + c) = a \times b + a \times c$$

Calculating areas and volumes thus allows to firmly consolidate the understanding of the arithmetic operations, in relation e.g. with physical units and their handling in practical problems. It is possible to give, already at primary school level, genuine (i.e. non entirely trivial) mathematical proofs – for instance in grade 5, one can give a complete justification of the formula that expresses the area of a disk :



$$\text{base} \simeq \frac{P}{2} = \pi \times R$$

$$\pi = \frac{P}{D} \quad \Rightarrow \quad P = \pi \times D = 2 \times \pi \times R$$

In the limit, by increasing the number of triangular sectors, one sees that the area of a disk is given by $\pi \times R \times R = \pi R^2$. Of course, this activity presupposes that the pupils have been explained before how to compute the area of rectangles, triangles and parallelograms, again through a visualization of the classical geometric decompositions used to justify the formulas. The conceptual position of formula $P = \pi D = 2\pi R$ is different, in this case it should rather be viewed as a *definition* of number π : this is the ratio between the perimeter and the diameter, which is independent of the size of the circle under consideration (one can intuitively mention that if the diameter doubles or triple, then the perimeter changes in the same way ; of course, this follows formally from Thales theorem, by approximating circle arcs with polygons . . .). If time allows, it is of course advisable to turn this observation into a practical experiment, using a pipe with a known diameter, and putting around a few loops of string, so as to get an approximate value of π.

As a general rule, geometry should be taught through a combination of reasoning with concrete manipulations : paper and scissor decompositions, usage of instruments (ruler, compasses, protractor...), elementary constructions (midpoints, medians, bissectors, . . .). Working with graph paper helps in developing the intuitive understanding of cartesian coordinates ; it would thus be extremely useful to start such activities already from grade 2 at latest.

2.3. Negative numbers, square roots, real numbers

While developing a better understanding elementary arithmetic operations, children can be introduced simple “arithmetic and geometric” progressions

$$\begin{array}{ccccccccc} 0, & a, & 2a, & 3a, & 4a, & 5a, & 6a, & 7a, & \ldots \\ 1, & a, & a^2, & a^3, & a^4, & a^5, & a^6, & a^7, & \ldots \end{array}$$

where $n\,a = a + a + \ldots + a$ (repeated n times) and $a^n = a \times a \times \ldots \times a$ (repeated n times). The special case of squares, cubes and powers of 10 should already appear at the primary school level.

Here again, suitable practical considerations (ruler, thermometer, altitude) are the appropriate context for the introduction of negative numbers, especially when dealing with negative multiples of the unit of length :

$$\ldots\ -5u,\quad -4u,\quad -3u,\quad -2u,\quad -u,\quad 0,\quad u,\quad 2u,\quad 3u,\quad 4u,\quad 5u,\quad 6u,\quad 7u\ \ldots$$

At the beginning of junior high school, one can then pursue with negative decimal numbers and real numbers, e.g. in relation with length measurements. It is natural to extend the distributivity property of multiplication with respect to addition to numbers of arbitrary positive or negative sign, and the "sign rule" follows :

$$a \times (b + (-b)) = a \times b + a \times (-b),$$

as the left hand side is equal to $a \times 0 = 0$, one must have $a \times (-b) = -(a \times b)$.

At the end of primary chool, pupils should normally reach a rather safe and effective pratice of handwritten divisions. Such a practice should allow them to observe the periodicity of remainders, and therefore the periodicity of the decimal expansion of fractions of integers. This is especially noticeable on many fractions that have a small denominator leading to a very short periodicity cycle (e.g. fractions of denominator 3, 7, 9, 11, 21, 27, 33, 37, 41, 63, 77, 99, 101, 271 (...) and their multiples by 2 and 5, which lead to a period of length 6 at most).

Real numbers appear in a natural way as non periodic decimal expansions of certain numerical values such as square roots. However, a premature use of calculators and an insufficient practice of approximate decimal (handwritten) calculations, e.g. of divisions, may lead to a very limited and formal perception of the concept of square root. It is absolutely necessary that pupils be faced to the task of extracting square roots numerically. For instance, to start with, one can explain the numerical approximation of $\sqrt{2}$:

$$\begin{array}{llll} (1.4)^2 = 1.96, & (1.5)^2 = 2.25 & \text{thus} & 1.4 < \sqrt{2} < 1.5, \\ (1.41)^2 = 1.9881, & (1.42)^2 = 2.0164 & \text{thus} & 1.41 < \sqrt{2} < 1.42, \\ (1.414)^2 = 1.999396, & (1.415)^2 = 2.002225 & \text{thus} & 1.414 < \sqrt{2} < 1.415 \quad \ldots \end{array}$$

We recommend that such concepts be introduced at latest in grade 7, at the same time Pythagoras' theorem is explained, so that the geometric necessity of square roots clearly appeals to pupils (of course, such contents can only be introduced in grade 6 or 7 if one assumes that the severe deficiencies of the current French primary curricula have been alleviated...).

When these concepts are understood, it becomes possible to give a precise general definition of the concept of real number – at the same time, this is a good opportunity to approach implicitly the important concept of limit :

(2.3.1) Definition. *Real numbers are numbers that are used to measure physical quantities with an unlimited accuracy. A real number is therefore expressed by an arbitrary (non necessarily periodic) decimal expansion, in other words a sequence* $\pm\,\square\square\ldots\square\square\,\textbf{.}\,\square\square\square\square\square\ldots$ *of digits* $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ *that is finite on the left side of the dot, and infinite on the right side, with a + or − initial sign (the absence of any sign implicitly means that we have a +sign, except for the null value which has no*

definite sign, actually one simply writes $+0.000\ldots = -\ 0.000\ldots = 0$*).*
From a geometric viewpoint, a real number corresponds to a point on an oriented axis, that would be set by means of a "metric ruler with infinite accuracy".

We firmly recommend that the manual algorithm to extract square roots be taught at school at the same period in time; in fact, mastering such an algorithm "enpowers" children by giving them a lot of control on their numerical environment – and at the same time it provides a strong hint that a square root is going to lead to a non periodic expansion, since one sees no reason that the algorithm would exhibit periodicity, as was the case with the division of integers. This would be an excellent way of consolidating the practice of handwritten claculations. Experiments have shown that well trained pupils can learn the square root algorithm in only one or two hours, assuming that they already have a good practice of handwritten division. Unfortunately, it appears to be almost impossible to test these issues with the immense majority of French pupils, at a time when the mathematical menu confines to extreme deficiency and the algorithms are not sufficiently mastered (of course we do not mean that being able to perform the formal algorithms suffices in any way to grasp primary mathematics; understanding the meaning of operations and their use in solving problems is even more important.)

To turn Definition (2.3.1) into a precise and mathematically adequate statement, one must also explain what are proper and improper decimal expansions [(3)]. First one should make pupils realize that $0,999999\ldots = 1$, in fact, if one sets $x = 0.999999\ldots$, then $10\,x = 9.999999\ldots$, hence $10\,x - x = 9$, and therefore one is led to admit that necessarily $x = 1$, at least under the assumption that the usual rules applied to decimal numbers still apply to infinite decimal expansions. More generally, one has to take e.g.

$$0.34999999\ldots = 0.35 = 0.35000000\ldots$$

These observations appear as a complementary specification to add in Definition (2.3.1).

(2.3.2) Complementary specification in the definition of real numbers. *Decimal numbers have two distinct expansions, one that is finite* (*or, eqivalently, one that contains an infinite sequence of consecutive* 0 *digits*)*, called "proper expansion", the other one containing an infinite sequence of consecutive* 9 *digits* (*and the previous digit decreased by* 1)*, called the "improper expansion". Real numbers that are not decimal numbers have only one infinite decimal expansion.*

Arithmetic operations on decimal numbers can be extended to find the expansion of the sum and product of two real numbers with any accuracy given in advance – and therefore to calculate the sum and product of two real numbers, at least in principle[(4)]. The order relation is obtained by comparing digit one by one in "lexicographic" order.

At this point, say in the 7th grade, one should be able to reach the following important characterizations (provided that curricula of all previous levels have been upgraded to provide a sufficient background!)

(3) Once this is done, Definition (2.3.1) can be considered as a perfectly acceptable formal definition of real numbers – even though it has the draw-back, in reality more a lack of elegance than a genuine inconvenience, to seemingly depend on the numeration system (base 10 here).

(4) For a complete theoretical justification of this claim, we need the theorem that asserts the convergence of bounded increasing sequences of real numbers, a theorem that is at best seen at high school level, cf. our manuscript *Puissances, exponentielles, logarithmes, . . .* for details.

(2.3.3) Characterization of rational and decimal numbers.

(a) *An infinite decimal expansion represents a rational number (i.e. a fraction of integers) if and only if the development is periodic from a certain index.*

(b) *Among the rational numbers, decimal numbers are those for which the expansion has an infinite sequence of consecutive* 0 *digits ("proper decimal expansion") or an infinite sequence of consecutive* 9 *digits ("improper decimal expansion").*

(c) *Real numbers that are not decimal are characterized by the fact that their expansion does not possess an infinite sequence of consecutive digits which are all equal to* 0 *or all equal to* 9 *; in other words, either they have an infinite number of digits that are neither* 0 *or* 9*, or an infinite alternating sequence (possibly irregular) of digits* 0 *and* 9*, starting from a certain index.*

Proof. (a) Indeed, given a simplified fraction p/q which is not a decimal number (i.e., q has at least one prime factor distinct from 2 and 5), the division algorithm of p by q never terminates and leads to remainders that fall in the finite sequence $1, 2, \ldots, q-1$. After at most $q-1$ steps beyond the dot, the remainder necessarily reproduces a value that has been already found. This implies that the expansion is periodic and that the period length is at most $q-1$. Conversely, if the expansion is periodic, say of length 5, and if we have e.g.

$$x = 0.10723114231142311423114\ldots,$$

we observe that the quotient $1 : 99999$ yields

$$\frac{1}{99999} = 0.00001000010000100001\ldots,$$

therefore

$$\frac{23114}{99999} = 23114 \times \frac{1}{99999} = 0.23114231142311423114\ldots$$

$$\frac{23114}{99999000} = 0.00023114231142311423114\ldots$$

Finally, since $0.107 = \dfrac{107}{1000}$, we get

$$x = \frac{107}{1000} + \frac{23114}{99999000} = \frac{107 \times 99999 + 23114}{99999000} = \frac{10723007}{99999000}$$

which is actually a rational number. This process is easily generalized to convert any periodic decimal expansion into a fraction. Statement (b) is just a reformulation of Definition (2.3.2), and (c) is also an equivalent statement. The final situation described in (c) occurs for instance when one considers the rational number

$$1/11 = 0.09090909\ldots$$

or an irrational number like $0.909009000900009\ldots$ (a clearly aperiodic sequence). □

All these considerations can be strengthened by introducing calculations with polynomials, by manipulating inequalities, approximations and algebraic identities. It is important to visualize geometrically $(a+b)^2$, $(a+b)(a-b)$. In fact, all primary schools and junior high schools should have wooden pieces designed to visualize quantities such as $(a+b)^2$, $(a+b)^3$ (such matters can even be approached at the end of primary school, in a non necessarily formalized manner, on the occasion of the introduction of areas and volumes). The identity $(10a+b)^2 - 100a^2 = (20a+b)b$ occurs in the justification of the square root algorithm. At a more basic level, say in grade 5, and possibly with a justification relying only on a use of squares drawn on graph paper (take a side which is e.g. 75 mm), the formula

$$(10a+5)^2 = 100\,a(a+1) + 25$$

can be used to perform efficient mental calculations of squares of numbers ending by the digit 5 in decimal base : for instance $(75)^2 = 5625$, where the number 56 is obtained by computing $a(a+1) = 7 \times 8$.

3. First steps of the introduction of Euclidean geometry

3.1. Fundamental concepts

The primitive concepts we are going to use freely are :

- real numbers, with their properties already discussed above ;
- points and geometric objects as sets of points : a point should be thought of as a geometric object with no extension, as can be represented with a sharp pencil ; a line or a curve are infinite sets of points (at this point, this is given only for intuition, but will not be needed formally) ;
- distances between points.

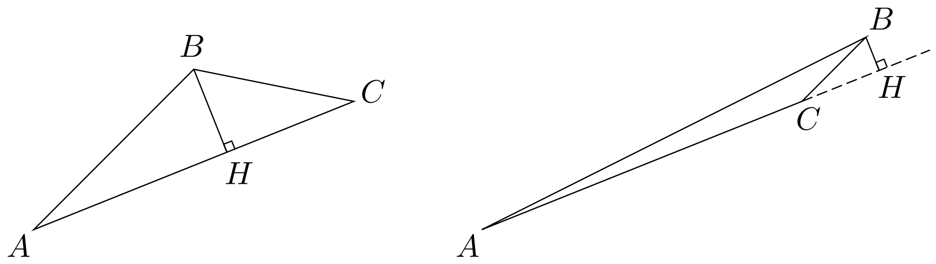
Let us mention that the language of set theory has been for more than one century the universal language of mathematicians. Although excessive abstraction should be avoided at early stages, we feel that it is appropriate to introduce at the beginning of junior high school the useful concepts of sets, of inclusion, the notation $x \in E$, operations on sets such as union, intersection and difference ; geometry and numbers already provide rich and concrete illustrations.

A geometric figure is simply an ordered finite collection of points A_j and sets S_k (vertices, segments, circles, arcs, ...)

Given two points A, B of the plane or of space, we denote by $d(A,B)$ (or simply by AB) their distance, which is in general a positive number, equal to zero when the points A and B coincide – concretely, this distance can be mesured with a ruler. A fundamental property of distances is :

3.1.1. Triangular inequality. *For any triple of points A, B, C, their mutual distances always satisfy the inequality $AC \leqslant AB + BC$, in other words the length of any side of a triangle is always at most equal to the sum of the lengths of the two other sides.*

Intuitive justification.



Let us draw the height of the triangle joining vertex B to point H on the opposite side (AC).

If H is located between A and C, we get $AC = AH + HC$; on the other hand, if the triangle is not flat (i.e. if $H \neq B$), we have $AH < AB$ and $HC < BC$ (since the hypotenuse is longer than the right-angle sides in a right-angle triangle – this will be checked formally thanks to Pythagoras' theorem). If H is located outside of the segment $[A, C]$, for instance beyond C, we already have $AC < AH \leqslant AB$, therefore $AC < AB \leqslant AB + BC$.

This justification(5) shows that the equality $AC = AB + BC$ holds if and only if the points A, B, C are aligned with B located between A and C (in this case, we have $H = B$ on the left part of the above figure). This leads to the following intrinsic definitions that rely on the concept of distance, and nothing more(6).

3.1.2. Definitions (segments, lines, half-lines).

(a) *Given two points A, B in a plane or in space, the segment $[A, B]$ of extremities A, B is the set of points M such that $AM + MB = AB$.*

(b) *We say that three points A, B, C are aligned with B located between A and C if $B \in [A, C]$, and we say that they are aligned (without further specification) if one of the three points belongs to the segment determined by the two other points.*

(c) *Given two distinct points A, B, the line (AB) is the set of points M that are aligned with A and B ; the half-line $[A, B)$ of origin A containing point B is the set of points M aligned with A and B such that either M is located between A and B, or B between A and M. Two half-lines with the same origin are said to be opposite if their union is a line.*

In the definition, part (a) admits the following physical interpretration : a line segment can be realized by stretching a thin and light wire between two points A and B : when the wire is stretched, the points M located between A and B cannot "deviate", otherwise the distance AB would be shorter than the length of the wire, and the latter could still be stretched further ...

We next discuss the notion of an axis : this is a line $\mathcal{D}$ equipped with an origin O and a direction, which one can choose by specifying one of the two points located at unit

(5) This is not a real proof since one relies on undefined concepts and on facts that have not yet been proved, for example, the concept of line, of perpendicularity, the existence of a point of intersection of a line with its perpendicular, etc ... This will actually come later (without any vicious circle, the justifications just serve to bring us to the appropriate definitions!)

(6) As far as they are concerned, these definitions are perfectly legitimate and rigorous, starting from our primitive concepts of points and their mutual distances. They would still work for other geometries such as hyperbolic geometry or general Riemannian geometry, at least when geodesic arcs are uniquely defined globally.

instance from O, with the abscissas $+1$ and -1 ; let us denote them respectively by I and I'. A point $M \in [O, I)$ is represented by the real value $x_M = +OM$ and a point M on the opposite half-line $[O, I')$ by the real value $x_M = -OM$. The algebraic measure of a bipoint (A, B) of the axis is defined by $\overline{AB} = x_B - x_A$, which is equal to $+AB$ or $-AB$ according to whether the ordering of A, B corresponds to the orientation or to its opposite. For any three points A, B, C of $\mathcal{D}$, we have the Chasles relation

$$\overline{AB} + \overline{BC} = \overline{AC}.$$

This relation can be derived from the equality $(x_B - x_A) + (x_C - x_B) = (x_C - x_A)$ after a simplification of the algebraic expression.

Building on the above concepts of distance, segments, lines and half-lines, we can now define rigorously what are planes, half-planes, circles, circle arcs, angles ...(7)

3.1.3. Definitions.

(a) *Two lines $\mathcal{D}$, $\mathcal{D}'$ are said to be concurrent if their intersection consists of exactly one point.*

(b) *A plane $\mathcal{P}$ is a set of points that can be realized as the union of a family of lines (UV) such that U describes a line $\mathcal{D}$ and V a line $\mathcal{D}'$, for some concurrent lines $\mathcal{D}$ and $\mathcal{D}'$ in space. If A, B, C are 3 non aligned points, we denote by (ABC) the plane defined by the lines $\mathcal{D} = (AB)$ and $\mathcal{D}' = (AC)$ (say)(8).*

(c) *Two lines $\mathcal{D}$ and $\mathcal{D}'$ are said to be parallel if they coincide, or if they are both contained in a certain plane $\mathcal{P}$ and do not intersect.*

(d) *A salient angle $\widehat{BAC}$ (or a salient angular sector) defined by two non opposite half-lines $[A, B)$, $[A, C)$ with the same origin is the set obtained as the union of the family of segments $[U, V]$ with $\in [A, B)$ and $V \in [A, C)$.*

(e) *A reflex angle (or a reflex angular sector) $\widetilde{BAC}$ is the complement of the corresponding salient angle $\widehat{BAC}$ in the plane (ABC), in which we agree to include the half-lines $[A, B)$ and $[A, C)$ in the boundary.*

(f) *Given a line $\mathcal{D}$ and a point M outside $\mathcal{D}$, the half-plane bounded by $\mathcal{D}$ containing M is the union of the two angular sectors $\widehat{BAM}$ and $\widehat{CAM}$ obtained by expressing $\mathcal{D}$ as the union of two opposite half-lines $[A, B)$ and $[A, C)$; this is the union of all segments $[U, V]$ such that $U \in \mathcal{D}$ and $V \in [A, M)$. The opposite half-plane is the one associated with the half-line $[A, M')$ opposite to $[A, M)$. In that situation, we also say that we have flat angles of vertex A.*

(g) *In a given plane $\mathcal{P}$, a circle of center A and radius $R > 0$ is the set of points M in the plane $\mathcal{P}$ such that $d(A, M) = AM = R$.*

---

(7) Of course, this long series of definitions is merely intended to explain the sequence of concepts in a logical order. When teaching to pupils, it would be necessary to approach the concepts progressively, to give examples and illustrations, to let the pupils solve exercises and produce related constructions with instruments (ru2ler, compasses ...).

(8) In a general manner, one could define by induction on n the concept of an affine subspace $\mathcal{S}_n$ of dimension n : this is the set obtained as the union of a family of lines (UV), where U describes a line $\mathcal{D}$ and V describes an affine subspace $\mathcal{S}_{n-1}$ of dimension $n-1$ intersecting $\mathcal{D}$ in exactly one point. Our definitions are valid in any dimension (even in an infinite dimensional ambient space), without taking special care !

(h) *A circular arc is the intersection of a circle with an angular sector, the vertex of which is the center of the circle.*

(i) *The measure of an angle (in degrees) is proportional to the length of the circular arc that it intercepts on a circle whose center coincides with the vertex of the angle, in such a way that the full circle corresponds to* $360°$. *A flat angle (cut by a half-plane bounded by a diameter of the circle) corresponds to an arc formed by a half-circle and has measure* $180°$. *A right angle is one half of a flat angle, that is, an angle corresponding to the quarter of a circle, in other words, an angle of measure equal to* $90°$.

(j) *Two half-lines with the same origin are said to be perpendicular if they form a right angle.*[(8)]

The usual properties of parallel lines and of angles intercepted by such lines ("corresponding angles" vs "alternate angles") easily leads to establishing the value of the sum of angles in a triangle (and, from there, in a quadrilateral).

Definition (i) requires of course a few comments. The first and most obvious comment is that one needs to define what is the length of a circular arc, or more generally of a curvilign arc : this is the limit (or the upper bound) of the lengths of polygonal line inscribed in the curve, when the curve is divided into smaller and smaller portions (cf. 2.2)[(8)]. The second one is that the measure of an angle is independent of the radius R of the circle used to evaluate arc lengths; this follows from the fact that arc lengths are proportional to the radius R, which itself follows from Thales' theorem (see below).

Moreover, a proportionality argument yields the formula for the length of a circular arc located on a circle of radius R : a full arc ($360°$) has length $2\pi R$, hence the length of an arc of $1°$ is 360 times smaller, that is $2\pi R/360 = \pi R/180$, and an arc of measure a (in degrees) has length

$$\ell = (\pi R/180) \times a = R \times a \times \pi/180.$$

3.2. Construction with instruments and isometry criteria for triangles

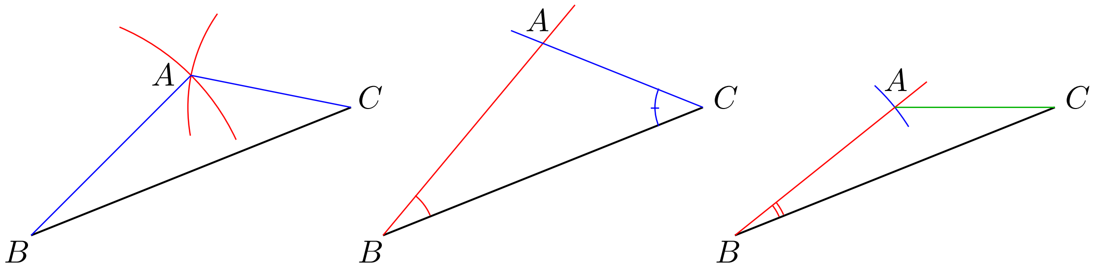
As soon as they are introduced, it is extremely important to illustrate geometric concepts with figures and construction activities with instruments. Basic constructions with ruler and compasses, such as midpoints, medians, bissectors, are of an elementary level and should be already taught at primary school. The step that follows immediately next consists of constructing perpendiculars and parallel lines passing through a given point.

At the beginning of junior high school, it becomes possible to consider conceptually more advanced matters, e.g. the problem of constructing a triangle ABC with a given base BC and two other elements, for instance :

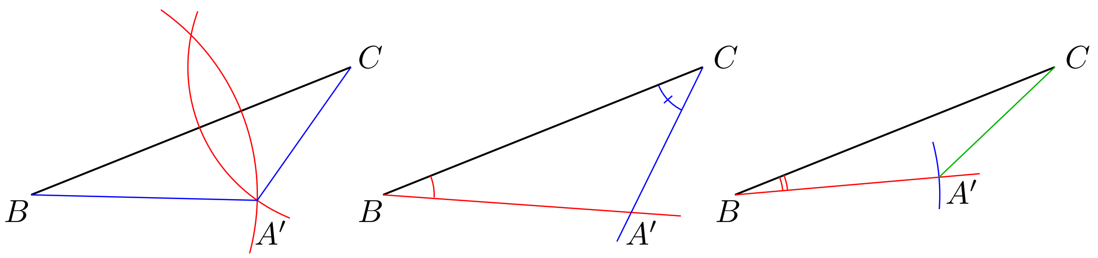
(8) The concepts of right and flat angles, as well as the notion of half angle are already primary school concerns. At this level, the best way to address these issues is probably to let pupils practice paper folding (the notion of horizontality and verticality are relative concepts, it is better to avoid them when introducing perpendicularity, so as to avoid any potential confusion).

(8) The definition and existence of limits are difficult issues that cannot be addressed before high school, but it seems appropriate to introduce this idea at least intuitively.

(3.2.1) the lengths of sides AB and AC,
(3.2.2) the measures of angles $\widehat{ABC}$ and $\widehat{ACB}$,
(3.2.3) the length of AB and the measure of angle $\widehat{ABC}$.



In the first case, the solution is obtained by constructing circles of centers B, C and radii equal to the given lengths AB and AC, in the second case a protractor is used to draw two angular sectors with respective vertices B and C, in the third case one draws an angular sector of vertex B and a circle of center B. In each case it can be seen that there are exactly two solutions, the second solution being obtained as a triangle $A'BC$ that is symmetric of ABC with respect to line (BC) :



One sees that the triangles ABC and $A'BC$ have in each case sides with the same lengths. This leads to the important concept of *isometric figures*.

3.2.4. Definition.

(a) *One says that two triangles are isometric if the sides that are in correspondence have the same lengths, in such a way that if the first triangle has vertices* A, B, C *and the corresponding vertices of the second one are* A', B', C', *then* $A'B' = AB$, $B'C' = BC$, $C'A' = CA$.

(b) *More generally, one says that two figures in a plane or in space are isometric, the first one being defined by points* A_1, A_2, A_3, A_4 ... *and the second one by corresponding points* A'_1, A'_2, A'_3, A'_4 ... *if all mutual distances coincide.*

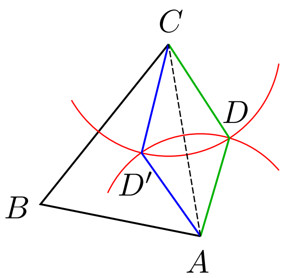
The concept of isometric figures is related to the physical concept of *solid body* : a body is said to be a solid if the mutual distances of its constituents (molecules, atoms) do not vary while the object is moved; after such a mo

ve, atoms which occupied certain positions A_i occupy new positions A'_i and we have $A'_i A'_j = A_i A_j$. This leads to a rigorous definition of solid displacements, that have a meaning from the viewpoints of mathematics and physics as well.

3.2.5. Definition. *Given a geometric figure* (*or a solid body in space*) *defined by* *characteristic points A_1, A_2, A_3, A_4 ..., a solid move is a continuous succession of positions $A_i(t)$ of these points with respect to the time t, in such a way that all distances $A_i(t)A_j(t)$ are constant. If the points A_i were the initial positions and the points A'_i are the final positions, we say that the figure $(A'_1A'_2A'_3A'_4\ \ldots)$ is obtained by a displacement of figure $(A_1A_2A_3A_4\ \ldots)$.*(9)

Beyond displacements, another way of producing isometric figures is to use a reflection (with respect to a line in a plane, or with respect to a plane in space, as obtained by taking the image of an object through reflection in a mirror)(9). This fact is already observed with triangles, the use of transparent graph paper is then a good way of visualizing isometric triangles that cannot be superimposed by a displacement without "getting things out of the plane" ; in a similar way, it can be useful to construct elementary solid shapes (e.g. non regular tetrahedra) that cannot be superimposed by a solid move.

3.2.6. Exercise. In order to ensure that two quadrilaterals $ABCD$ and $A'B'C'D'$ are isometric, it is not sufficient to check that the four sides $A'B' = AB$, $B'C' = BC$, $C'D' = CD$, $D'A' = DA$ possess equal lengths, one must also check that the two diagonals $A'C' = AC$ and $B'D' = BD$ be equal; equaling only one diagonal is not enough as shown by the following construction :



The construction problems considered above for triangles lead us to state the following fundamental isometry criteria.

3.2.7. Isometry criteria for triangles(10)**.** *In order that two triangles be isometric, it is necessary and sufficient to check one of the following cases*

(a) *that the three sides be respectively equal (this is just the definition), or*

(b) *that they possess one angle with the same value and its adjacent sides equal, or*

(c) *that they possess one side with the same length and its adjacent angles of equal values.*

(9) The concept of continuity that we use is the standard continuity property for functions of one real variable - one can of course introuduce this only intuitively at the junior high school level. One can further show that an isometry between two figures or solids extends an affine isometry of the whole space, and that a solid move is represented by a positive affine isometry, see Section 10. The formal proof is not very hard, but certainly cannot be given before the end of high school (this would have been possible with the rather strong French curricula as they were 50 years ago in the grade 12 science class, but doing so would be nowadays completely impossible).

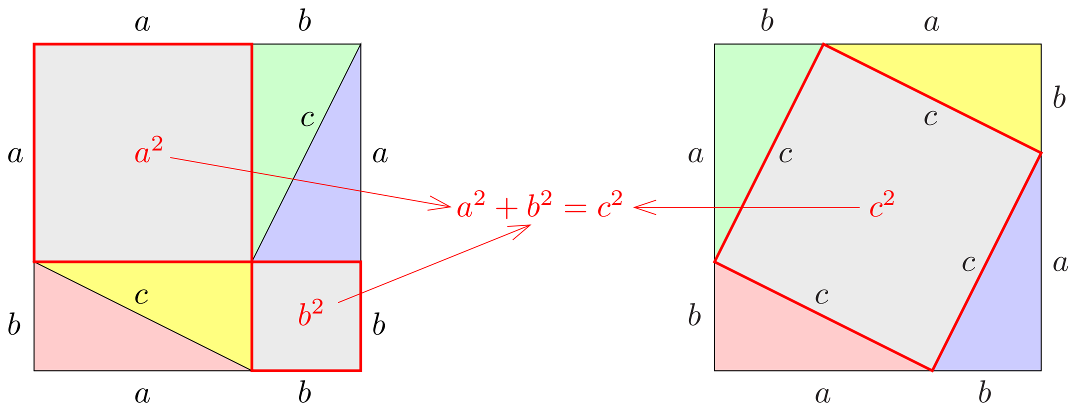
(9) Conversely, an important theorem - which we will show later (see section 10) says that isometric figures can be deduced from each other either by a solid move or by a solid move preceded (or followed) by a reflection.

(10) A rigorous formal proof of of these 3 isometry criteria will be given later, cf. Section 8.

One should observe that conditions (b) and (c) are not sufficient if the adjacency specification is omitted - and it would be good to introduce (or to let pupils perform) constructions demonstrating this fact. A use of isometry criteria in conjunction with properties of alternate or corresponding angles leads to the various usual characterizations of quadrilaterals - parallelograms, lozenges, rectangles, squares . . .

3.3. Pythagoras' theorem

We first give the classical "Chinese" proof of Pythagoras' theorem, which is derived by a simple area argument based on moving four triangles (represented here in green, blue, yellow and light red). Its main advantage is to be visual and convincing[(10)].



The point is to compare, in the left hand and right hand figures, the remaining grey area, which is the difference of the area of the square of side a + b with the area of the four rectangle triangles of sides a, b, c. The equality of the grey areas implies $a^2 + b^2 = c^2$.

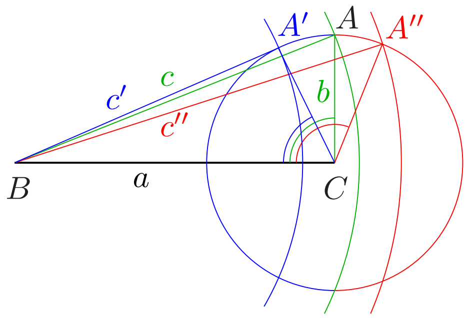
Complement. *Let (ABC) be a triangle and a, b, c the lengths of the sides that are opposite to vertices A, B, C.*

(i) *If the angle $\widehat{C}$ is smaller than a right angle, we have $c^2 < a^2 + b^2$,*

(ii) *If the angle $\widehat{C}$ is larger than a right angle, we have $c^2 > a^2 + b^2$.*

Proof. First consider the case where (ABC) is rectangle : we have $c^2 = a^2 + b^2$ and the angle is equal to 90°.

(10) Again, in our context, the argument that will be described here is a justification rather than a formal proof. In fact, it would be needed to prove that the quadrilateral central figure on the right hand side is a square - this could certainly be checked with isometry properties of triangles - but one should not forget that they are not yet really proven at this stage. More seriously, the argument uses the concept of area, and it would be needed tp prove the existence of an area measure in the plane with all the desired propertiesÂă: additivity by disjoint unions, translation invariance . . .



If angle $\widehat{C}$ is $< 90°$, we have $c' < c$.

If angle $\widehat{C}$ is $> 90°$, we have $c'' > c$.

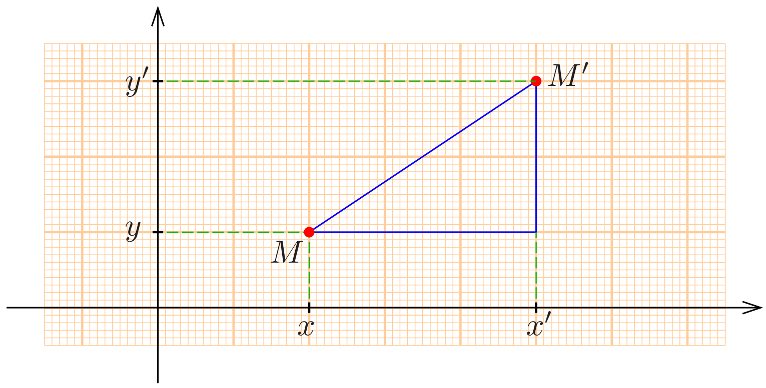
We argue by either increasing or decreasing the angle: if angle $\widehat{C}$ is $< 90°$, we have $c' < c$; if angle $\widehat{C}$ is $> 90°$, we have $c'' > c$. By this reasoning, we conclude :

Converse of Pythagoras' theorem. *With the above notation, if* $c^2 = a^2 + b^2$, *then angle* $\widehat{C}$ *must be a right angle, hence the given triangle is rectangle in* C.

4. Cartesian coordinates in the plane

The next fundamental step of our approach is the introduction of cartesian coordinates and their use to give *formal proofs of properties that had previously been taken for granted* (or given with a partial justification only). This is done by working in orthonormal frames.

4.1. Expression of Euclidean distance



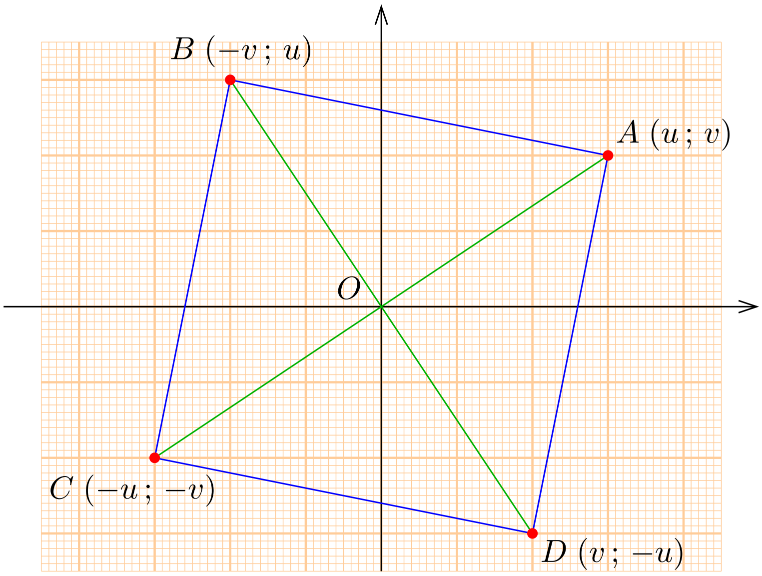
Pythagoras' theorem shows that the length MM' of the hypotenuse is given by the formula $MM'^2 = (x'-x)^2 + (y'-y)^2$, as the two sides of the right angle are $x'-x$ and $y'-y$ (up to sign). The distance from M to M' is therefore equal to

$$d(M, M') = MM' = \sqrt{(x'-x)^2 + (y'-y)^2}. \tag{4.1.1}$$

(It is of course advisable to first present the argument with simple numerical values).

4.2. Squares

Let us consider the figure formed by points $A\,(u\,;\,v)$, $B\,(-v\,;\,u)$, $C\,(-u\,;\,-v)$, $D\,(v\,;\,-u)$.



Formula (4.1.1) yields

$$AB^2 = BC^2 = CD^2 = DA^2 = (u+v)^2 + (u-v)^2 = 2(u^2+v^2),$$

hence the four sides have the same length, equal to $\sqrt{2}\sqrt{u^2+v^2}$. Similarly, we find

$$OA = OB = OC = OD = \sqrt{u^2+v^2},$$

therefore the 4 isoceles triangles OAB, OBC, OCD and ODA are isometric, and as a consequence we have $\widehat{OAB} = \widehat{OBC} = \widehat{OCD} = \widehat{ODA} = 90°$ and the other angles are equal to 45°. Hence $\widehat{DAB} = \widehat{ABC} = \widehat{BCD} = \widehat{CDA} = 90°$, and we have proved that our figure is a square.

4.3. "Horizontal and vertical" lines

The set $\mathcal{D}$ of points $M(x\,;\,y)$ such that $y = c$ (where c is a given numerical value) is a "horizontal" line. In fact, given any three points M, M', M'' of abscissas $x < x' < x''$ we have

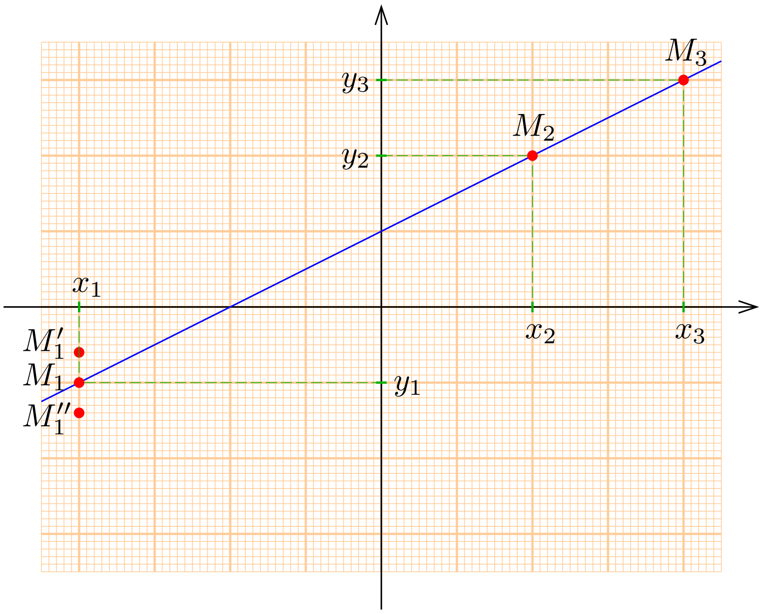
$$MM' = x' - x, \qquad M'M'' = x'' - x', \qquad MM'' = x'' - x$$

and therefore $MM' + M'M'' = MM''$. This implies by definition that our points M, M', M''. If we consider the line $\mathcal{D}_1$ given by the equation $y = c_1$ with $c_1 \neq c$, this is another horizontal line, and we have clearly $\mathcal{D} \cap \mathcal{D}_1 = \emptyset$, therefore our lines $\mathcal{D}$ and $\mathcal{D}_1$ are parallel.

Similarly, the set $\mathcal{D}$ of points $M(x\,;\,y)$ such that $x = c$ is a "vertical line" and the lines $\mathcal{D} : x = c$, $\mathcal{D}_1 : x = c_1$ are parallel.

4.4. Line defined by an equation $y = ax + b$

We start right away with the general case $y = ax + b$ to avoid any unnecessary repetitions, but with pupils it would be of course more appropriate to treat first the linear case $y = ax$.



Consider three points $M_1\ (x_1\,;\,y_1)$, $M_2\ (x_2\,;\,y_2)$, $M_3\ (x_3\,;\,y_3)$ satisfying the relations $y_1 = ax_1 + b$, $y_2 = ax_2 + b$ and $y_3 = ax_3 + b$, with $x_1 < x_2 < x_3$, say. As $y_2 - y_1 = a(x_2 - x_1)$, we find

$$M_1M_2 = \sqrt{(x_2 - x_1)^2 + a^2(x_2 - x_1)^2} = \sqrt{(x_2 - x_1)^2(1 + a^2)} = (x_2 - x_1)\sqrt{1 + a^2},$$

and likewise $M_2M_3 = (x_3 - x_2)\sqrt{1 + a^2}$, $M_1M_3 = (x_3 - x_1)\sqrt{1 + a^2}$. This shows that $M_1M_2 + M_2M_3 = M_1M_3$, hence our points M_1, M_2, M_3 are aligned. Moreover[11], we see that for any point $M'_1\ (x, y'_1)$ with $y'_1 > ax_1 + b$, then this point is not aligned with M_2 and M_3, and similarly for $M''_1\ (x, y''_1)$ such that $y''_1 < ax_1 + b$.

Consequence. *The set $\mathcal{D}$ of points $M\ (x\,;\,y)$ such that $y = ax + b$ is a line.*

The slope of line $\mathcal{D}$ is the ratio between the "vertical variation" and the "horizontal variation", that is, for two points $M_1\ (x_1\,;\,y_1)$, $M_2\ (x_2\,;\,y_2)$ of $\mathcal{D}$ the ratio

$$\frac{y_2 - y_1}{x_2 - x_1} = a.$$

A horizontal line is a line of slope $a = 0$. When the slope a becomes very large, the inclination of the line $\mathcal{D}$ becomess intuitively close to being vertical. We therefore agree that a vertical line has infinite slope. Such an infinite value will be denoted by the symbol ∞ (without sign).

Consider two distinct points $M_1\ (x_1, y_1)$, $M_2\ (x_2, y_2)$. If $x_1 \neq x_2$, we see that there exists a unique line $\mathcal{D} : y = ax + b$ passing through M_1 and M_2 : its slope is given by $a = \frac{y_2 - y_1}{x_2 - x_1}$ and we infer $b = y_1 - ax_1 = y_2 - ax_2$. If $x_1 = x_2$, the unique line $\mathcal{D}$ passing through M_1, M_2 is the vertical line of equation $x = x_1$.

(11) A rigorous formal proof would of course be possible by using a distance calculation, but this is much less obvious thanwhat we have done until now. One could however argue as in § 5.2 and use a new coordinate frame to reduce the situation to the case of the horizontal line $Y = 0$, in which case the proof is much easier.

4.5. Intersection of two lines defined by their equations

Consider two lines $\mathcal{D} : y = ax+b$ and $\mathcal{D}' : y = a'x+b'$. In order to find the intersection $\mathcal{D} \cap \mathcal{D}'$ we write $y = ax + b = a'x + b'$, and get in this way $(a' - a)x = -(b' - b)$. Therefore, if $a \neq a'$, there is a unique intersection point $M(x\,;\,y)$ such that

$$x = -\frac{b'-b}{a'-a}, \qquad y = ax + b = \frac{-a(b'-b) + b(a'-a)}{a'-a} = \frac{ba' - ab'}{a'-a}.$$

The intersection of $\mathcal{D}$ with a vertical line $\mathcal{D}' : x = c$ is still unique, as we immediately find the solution $x = c$, $y = ac + b$. From this discussion, we can conclude:

Theorem. *Two lines $\mathcal{D}$ and $\mathcal{D}'$ possessing distinct slopes a, a' have a unique intersection point: we say that they are concurrent lines.*

On the contrary, if $a = a'$ and moreover $b \neq b'$, there is no possible solution, hence $\mathcal{D} \cap \mathcal{D}' = \emptyset$, our lines are distinct parallel lines. If $a = a'$ and $b = b'$, the lines $\mathcal{D}$ and $\mathcal{D}'$ are equal, and they are still considered as being parallel.

Consequence 1. *Consequence 1. Two lines $\mathcal{D}$ and $\mathcal{D}'$ of slopes a, a' are parallel if and only if their slopes are equal (finite or infinite).*

Consequence 2. *If $\mathcal{D}$ is parallel to $\mathcal{D}'$ and if $\mathcal{D}'$ is parallel to $\mathcal{D}''$, then $\mathcal{D}$ is parallel to $\mathcal{D}''$.*

Proof. In fact, if $a = a'$ and $a' = a''$, then $a = a''$. □

We can finally prove "Euclid's parallel postulate" (in our approach, this is indeed a rather obvious theorem, and not a postulate !).

Consequence 3. *Given a line $\mathcal{D}$ and a point M_0, there is a unique line $\mathcal{D}'$ parallel to $\mathcal{D}$ that passes through M_0.*

Proof. In fact, if $\mathcal{D}$ has a slope a and if $M_0(x_0\,;\,y_0)$, we see that

- for $a = \infty$, the unique possible line is the line $\mathcal{D}'$ of equation $x = x_0$;
- for $a \neq \infty$, the line $\mathcal{D}'$ has an equation $y = ax + b$ with $b = y_0 - ax_0$, therefore $\mathcal{D}'$ is the line that is uniquely defined by the equation $\mathcal{D}' : y - y_0 = a(x - x_0)$.

4.6. Orthogonality condition for two lines

Let us consider a line passing through the origin $\mathcal{D} : y = ax$. Select a point $M(u\,;\,v)$ located on $\mathcal{D}$, $M \neq O$, that is $u \neq 0$. Then $a = \frac{v}{u}$. We know that the point $M'\,(u'\,;\,v') = (-v\,;\,u)$ is such that the lines $\mathcal{D} = (OM)$ and (OM') are perpendicular, thanks to the construction of squares presented in section 4.2. Therefore, the slope of the line $\mathcal{D}' = (OM')$ perpendicular to $\mathcal{D}$ is given by

$$a' = \frac{v'}{u'} = \frac{u}{-v} = -\frac{u}{v} = -\frac{1}{a}$$

if $a \neq 0$. If $a = 0$, the line $\mathcal{D}$ coincides with the horizontal axis, its perpencular through O is the vertical axis of infinite slope. The formula $a' = -\frac{1}{a}$ is still true in that case if we agree that $\frac{1}{0} = \infty$ (let us repeat again that here ∞ means an infinite non signed value).

Consequence 1. *Two lines $\mathcal{D}$ and $\mathcal{D}'$ of slopes a, a' are perpendicular if and only if their slopes satisfy the condition $a' = -\frac{1}{a} \quad\Leftrightarrow\quad a = -\frac{1}{a'}$ (agreeing that $\frac{1}{\infty} = 0$ and $\frac{1}{0} = \infty$)).*

Consequence 2. *If $\mathcal{D} \perp \mathcal{D}'$ and $\mathcal{D}' \perp \mathcal{D}''$ then $\mathcal{D}$ and $\mathcal{D}''$ are parallel.*

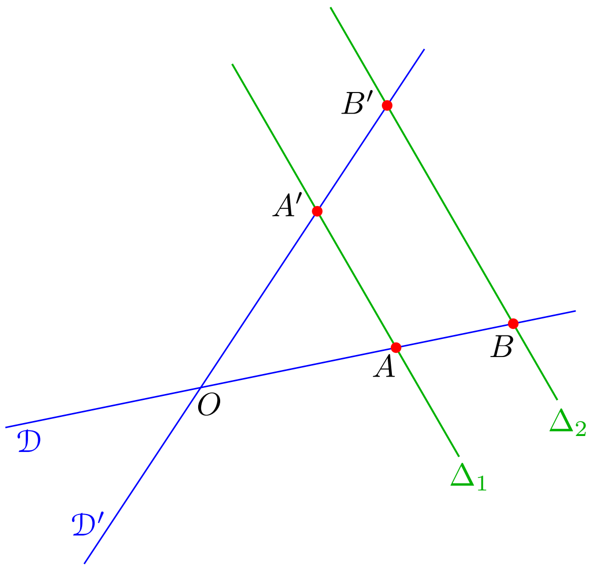
Proof. In fact, the slopes satisfy $a = -\frac{1}{a'}$ and $a'' = -\frac{1}{a'}$, hence $a'' = a$. □

4.7. Thales' theorem

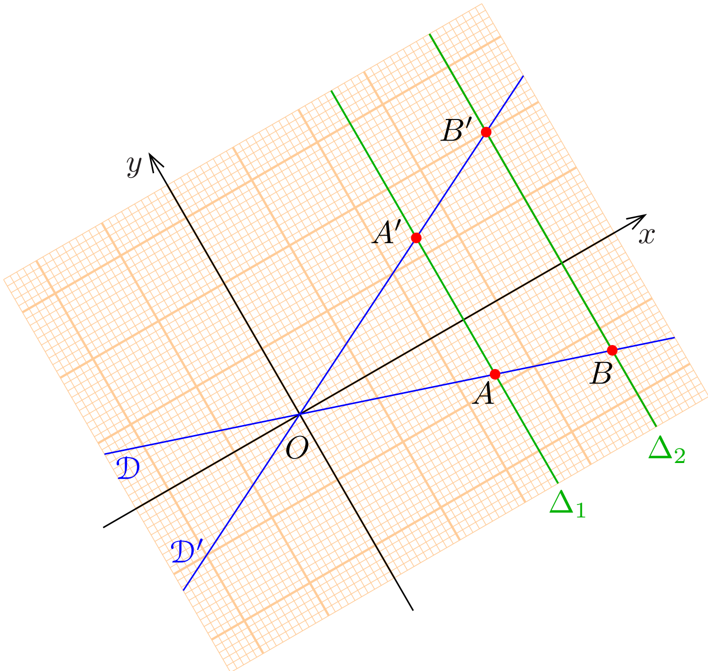
We start by stating a "Euclidean version" of the theorem, involving ratios of distances rather than ratios of algebraic measures.

Thales' theorem. *Consider two concurrent lines $\mathcal{D}$, $\mathcal{D}'$ intersecting in a point O, and two parallel lines Δ_1, Δ_2 that intersect $\mathcal{D}$ in points A, B, and $\mathcal{D}'$ in points A', B' ; we assume that A, B, A', B' are different from O. Then the length ratios satisfy*

$$\frac{OB}{OA} = \frac{OB'}{OA'} = \frac{BB'}{AA'}.$$



Proof. We argue by means of a coordinate calculation, in an orthonormal frame Oxy such that Ox is perpendicular to lines Δ_1, Δ_2, and Oy is parallel to lines Δ_1, Δ_2.



In these coordinates, lines Δ_1, Δ_2 are "vertical" lines of respective equations

$$\Delta_1 : x = c_1, \qquad \Delta_2 : x = c_2$$

with c_1, $c_2 \neq 0$, and our lines $\mathcal{D}$, $\mathcal{D}'$ admit respective equations $\mathcal{D} : y = ax$, $\mathcal{D}' : y = a'x$. Therefore

$$A\ (c_1, ac_1), \quad B\ (c_2, ac_2), \quad A'\ (c_1, a'c_1), \quad B'\ (c_2, a'c_2).$$

By Pythagoras' theorem we infer (after taking absolute values) :

$$OA = |c_1|\sqrt{1+a^2}, \quad OB = |c_2|\sqrt{1+a^2}, \quad OA' = |c_1|\sqrt{1+a'^2}, \quad OB' = |c_2|\sqrt{1+a'^2},$$
$$AA' = |(a'-a)c_1|, \quad BB' = |(a'-a)c_2|.$$

We have $a' \neq a$ since $\mathcal{D}$ and cD' are concurrent by our assumption, hence $a' - a \neq 0$, and we then conclude easily that

$$\frac{OB}{OA} = \frac{OB'}{OA'} = \frac{BB'}{AA'} = \frac{|c_2|}{|c_1|}. \qquad \square$$

In a more precise manner, if we choose orientations on $\mathcal{D}$, $\mathcal{D}'$ so as to turn them into axes, and also an orientation on Δ_1 and Δ_2, we see that in fact we have an equality of algebraic measures

$$\frac{\overline{OB}}{\overline{OA}} = \frac{\overline{OB'}}{\overline{OA'}} = \frac{\overline{BB'}}{\overline{AA'}}.$$

Converse of Thales' theorem. *Let $\mathcal{D}$, $\mathcal{D}'$ be concurrent lines intersecting in O. If Δ_1 intersects $\mathcal{D}$, $\mathcal{D}'$ in distinct points A, A', and Δ_2 intersects $\mathcal{D}$, $\mathcal{D}'$ in distinct points B, B' and if*

$$\frac{\overline{OB}}{\overline{OA}} = \frac{\overline{OB'}}{\overline{OA'}}$$

then Δ_1 *and* Δ_2 *are parallel.*

Proof. It is easily obtained by considering the line δ_2 parallel to Δ_1 that passes through B, and its intersection point β' with $\mathcal{D}'$. We then see that $\overline{O\beta'} = \overline{OB'}$, hence $\beta' = B'$ and $\delta_2 = \Delta_2$, and as a consequence $\Delta_2 = \delta_2 \mathbin{//} \Delta_1$. □

4.8. Consequences of Thales and Pythagoras theorems

The conjunction of isometry criteria for triangles and Thales and Pythagoras theorems already allows (in a very classical way !) to establish many basic theorems of elementary geometry. An important concept in this respect is the concept of similitude.

Definition. *Two figures* $(A_1A_2A_3A_4\ \ldots)$ *and* $(A'_1A'_2A'_3A'_4\ \ldots)$ *are said to be similar in the ratio* k $(k > 0)$ *if we have* $A'_iA'_j/A_iA_j = k$ *for all segments* $[A_i, A_j]$ *and* $[A'_i, A'_j]$ *that are in correspondance.*

An important case where similar figures are obtained is by applying a homothety with a given center, say point O : if O is chosen as the origin of coordinates and if to each point $M(x\,;\,y)$ we associate the point $M'(x'\,;\,y')$ such that $x' = kx$, $y' = ky$, then formula (4.1.1) shows that we indeed have $A'B' = |k|\,AB$, hence by assigning to each point A_i the corresponding point A'_i we obtain similar figures in the ratio $|k|$; this situation is described by saying that we have homothetic figures in the ratio k ; this ratio can be positive or negative (for instance, if $k = -1$, this is a central symmetry with respect to O). The isometry criteria for triangles immediately extend into criteria for similarity.

Similarity criteria for triangles. *In order to conclude that two triangles are similar, ii is necessary and sufficient that one of the following conditions is met*:

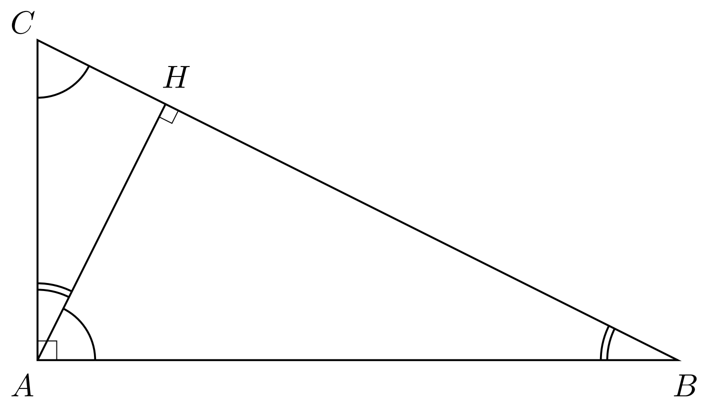
(a) *the corresponding three sides are proportional in a certain ratio* $k > 0$ (*this is the definition*);

(b) *the triangles have a corresponding equal angle and the adjacent sides are proportional*;

(c) *the triangles have two equal angles in correspondence.*

An interesting application of the similarity criteria consists in stating and proving the basic metric relations in rectangle triangles: if the triangle ABC is rectangle in A and if H is the foot of the altitude drawn from vertex A, we have the basic relations

$$AB^2 = BH \cdot BC, \quad AC^2 = CH \cdot CB, \quad AH^2 = BH \cdot CH, \quad AB \cdot AC = AH \cdot BC.$$



In fact (for example) the similarity of rectangle triangles ABH and ABC leads to the equality of ratios

$$\frac{AB}{BC} = \frac{BH}{AB} \implies AB^2 = BH \cdot BC.$$

One is also led in a natural way to the definition of sine, cosine and tangent of an acute angle in a rectangle triangle.

Definition. *Consider a triangle ABC that is rectangle in A. One defines*

$$\cos\widehat{ABC} = \frac{AB}{BC}, \qquad \sin\widehat{ABC} = \frac{AC}{BC}, \qquad \tan\widehat{ABC} = \frac{AC}{AB}.$$

In fact, the ratios only depend on the angle $\widehat{ABC}$ (which also determines uniquely the complementary angle $\widehat{ACB} = 90° - \widehat{ABC}$), since rectangle triangles that share a common angle else than their right angle are always similar by criterion (c). Pythagoras' theorem then quickly leads to computing the values of cos, sin, tan for angles with "remarkable values" 0°, 30°, 45°, 60°, 90°.

4.9. Computing areas and volumes

It is possible – and therefore probably desirable – to justify many basic formulas concerning areas and volumes of usual shapes and solid bodies (cylinders, pyramids, cones, spheres), just by using Thales and Pythagoras theorems, combined with elementary geometric arguments[(12)]. We give here some indication on such techniques, in the case of cones and spheres. The arguments are close to those developed by Archimedes more than two centuries BC (except that we take here the liberty of reformulating them in modern algebraic notations).

Volume of a cone with an arbitrary planar base

Our goal is to evaluate the volume of a cone whose base B is an arbitrary bounded measurable planar domain. Let $\mathcal{P}$ be the plane containing the base domain B and let O be a point that does not belong to $\mathcal{P}$.

Definition. *The cone of vertex O and of base B is the union of all segments $[O, M]$ staring at O and ending at a point $M \in B$. The altitude h of the cone is the between O and the plane $\mathcal{P}$ that contains B.*

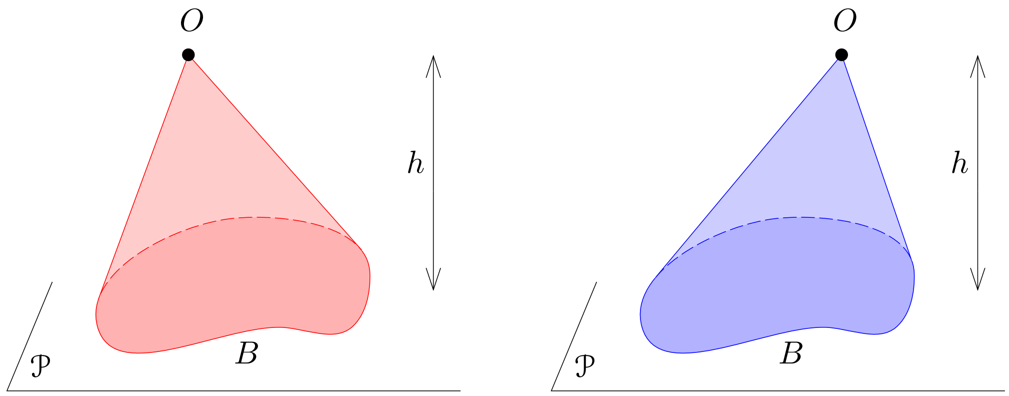
In the case where the base B is a disk, we say that this is a *circular or elliptic cone* (straight or oblique). When the base is a triangle, the cone is actually a *tetrahedron*, and when the base is a square or a rectangle, the cone is a *pyramid* (straight or oblique).

Here we make the hypothesis that the area A of base B is measurable, and we mean by this that by using a sufficiently fine grid, the approximate area obtained by multiplying

(12) We are using here the word "justify" rather than "prove" because the necessary theoretical foundations (e.g. measure theory) are missing – and will probably be missing for 5-6 years or more. But in reality, one can see that these justifications can be made perfectly rigorous once the foundations considered here as intuitive are rigorously established. The concept of Hausdorff measure, as briefly explained in (11.3), can be used e.g. to give a rigorous definition of the p-dimensional measure of any object in a metric space, even when p is not an integer.

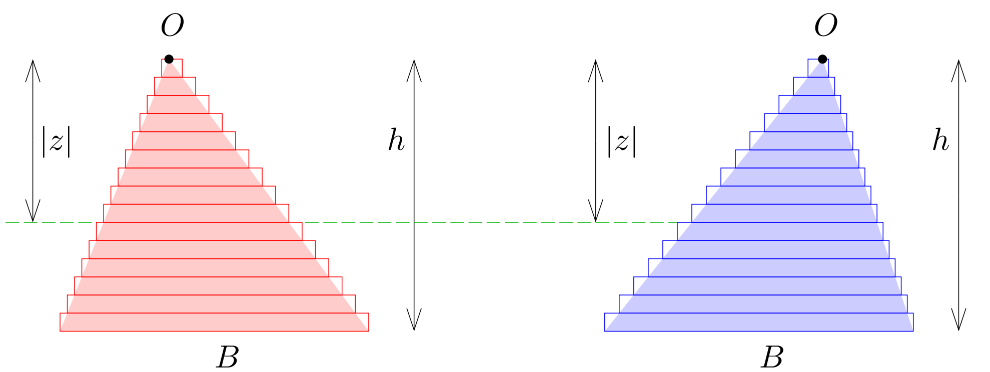
the area of a square by the number of squares that are contained in B tends to a limit when the side of squares tends to zero (the same limit being also attained when one takes all squares intersecting B, so that their union contains B).

If we choose the vertex O as the origin and coordinates (x, y, z) such that the plane $\mathcal{P}$ is horizontal and located "below" O, then $\mathcal{P}$ can be expressed as a plane of equation $z = -h$ where h is the altitude. Our first important observation is :



(4.9.1) *The volume of the cone V depends only on the base B and on the altitude h, but not the relative position of O and B (i.e., if B is displaced horizontally with respect to O, the volume does not change).*

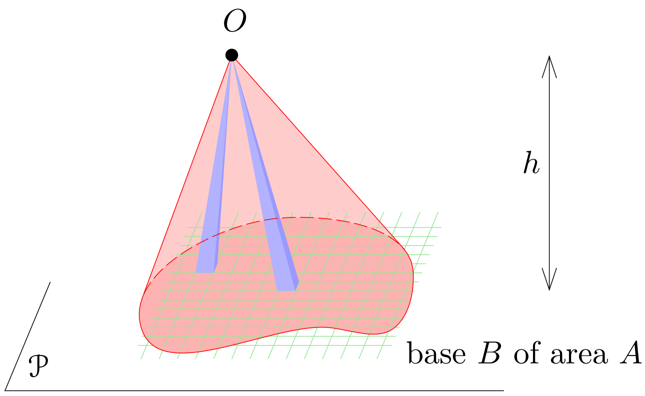
To prove (4.9.1), we compare our cone and the deformed one by slicing both horizontally in very thin slices, as illustrated in the figure below :



If we replace the slices by cylinders with vertical lateral surface (and whose bases are homothetic to the base B in the ratio $|z|/h$ as shown above), one gets a small error, since the calculated volumes become a little larger than those of the corresponding truncated cones , however the error becomes smaller and smaller when the number of slices increases. However, it is clear that the volumes of the two stacks of cylindrical slices are identical, as they have just been "dragged" horizontally against each other.

The second observation is :

(4.9.2) *When the altitude h is fixed, the volume V of the cone is proportional to the area A of base B.*

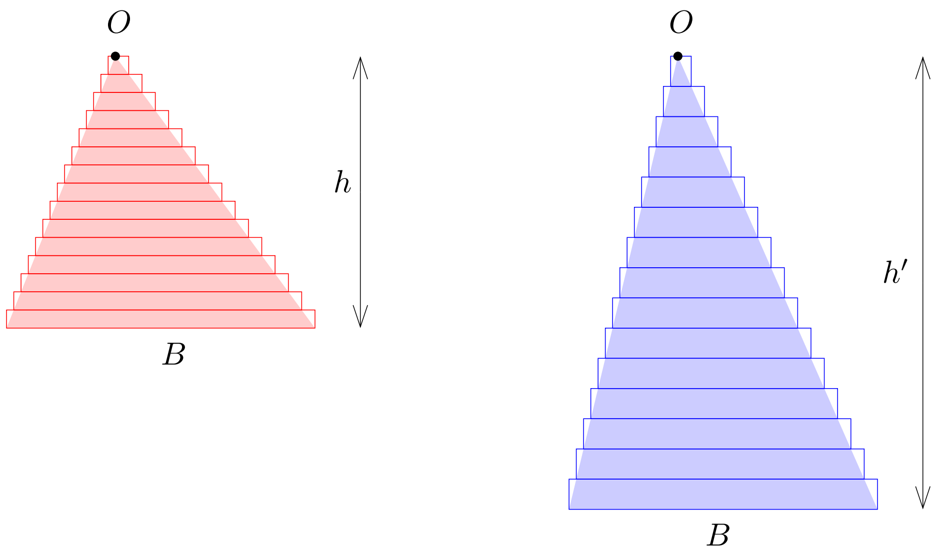


Indeed, we can calculate an approximate value of the area A of the base by using a grid, and then it follows from (4.9.1) that the volumes of all oblique pyramids based on different grid squares are all identical. This shows that the total volume of these pyramids is proportional to the number of n square grids included in B, and thus, in the limit, the volume of the cone V is proportional to the area A of the base.

Our third observation is :

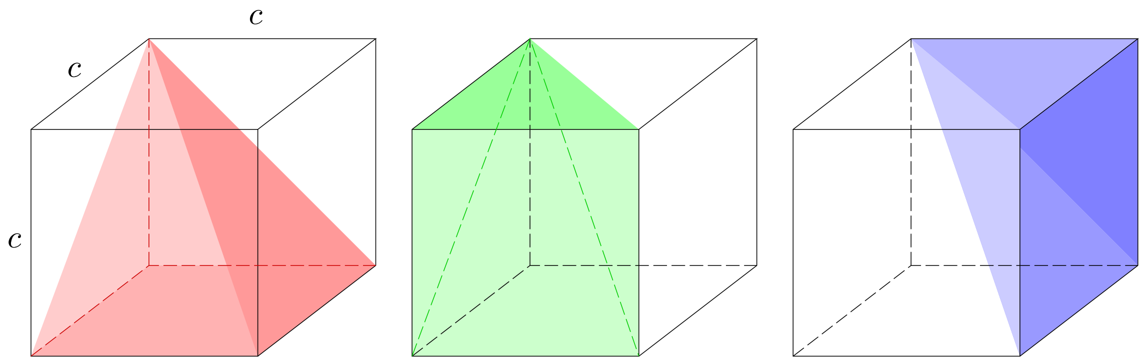
(4.9.3) *When the base B is fixed, the volume V of the cone is proportional to the altitude h.*

To prove this property, we consider two cones with the same base B and different altitudes h, h', and we compute their volumes through an approximation by a union of thin cylindrical slices :



To obtain the second cone, each slice is scaled vertically in the ratio h'/h. As the volume of a cylinder is the product of the area of the base by the height, we see that the volume of the second stack is proportional to volume of the first, multiplied by the ratio h'/h. In the limit, when the slices become thinner and thinner, this clearly shows that the ratio of the volumes of our cones is $V'/V = h'/h$.

The fourth and final step is to calculate the volume of a pyramid : we start with the case of a pyramid whose base is a square of side c and whose altitude is also $h = c$. The picture below shows that one can fill the cube with side c by 3 isometric pyramids of this kind (it would be useful to actually build them with paper and scissors and verify that they can be assembled into a cube!).



The volume of each of the three pyramids represented above is therefore one third of the volume of the cube, that is $V = \frac{1}{3}c^3$. If we consider a pyramid with a square base $A = c^2$ and an arbitrary altitude h, the volume must be multiplied by h/c by (4.9.3), hence its volume is equal to

$$V = \frac{h}{c} \times \frac{1}{3}c^3 = \frac{1}{3}c^2 h.$$

For an arbitrary base B, we can use a grid in the plane $\mathcal{P}$ containing B, exactly as we did in the proof of (4.9.2). If n is the number of squares of side c contained in the base B, the area A of the base is approximately $A \approx n\,c^2$ and therefore we get

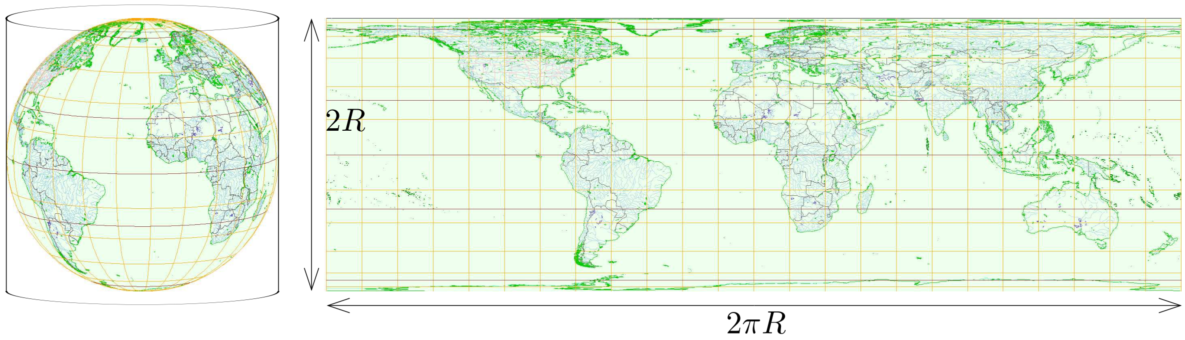
$$V \approx n \times \frac{1}{3}c^2 h = \frac{1}{3}(n\,c^2)h \approx \frac{1}{3}Ah.$$

The approximation becomes better and better when the side c tends to 0, and in the limit, we see that the volume of an arbitrary cone is given by

$$V = \frac{1}{3}Ah. \tag{4.9.4}$$

Archimedes formula for the area of a sphere of radius R

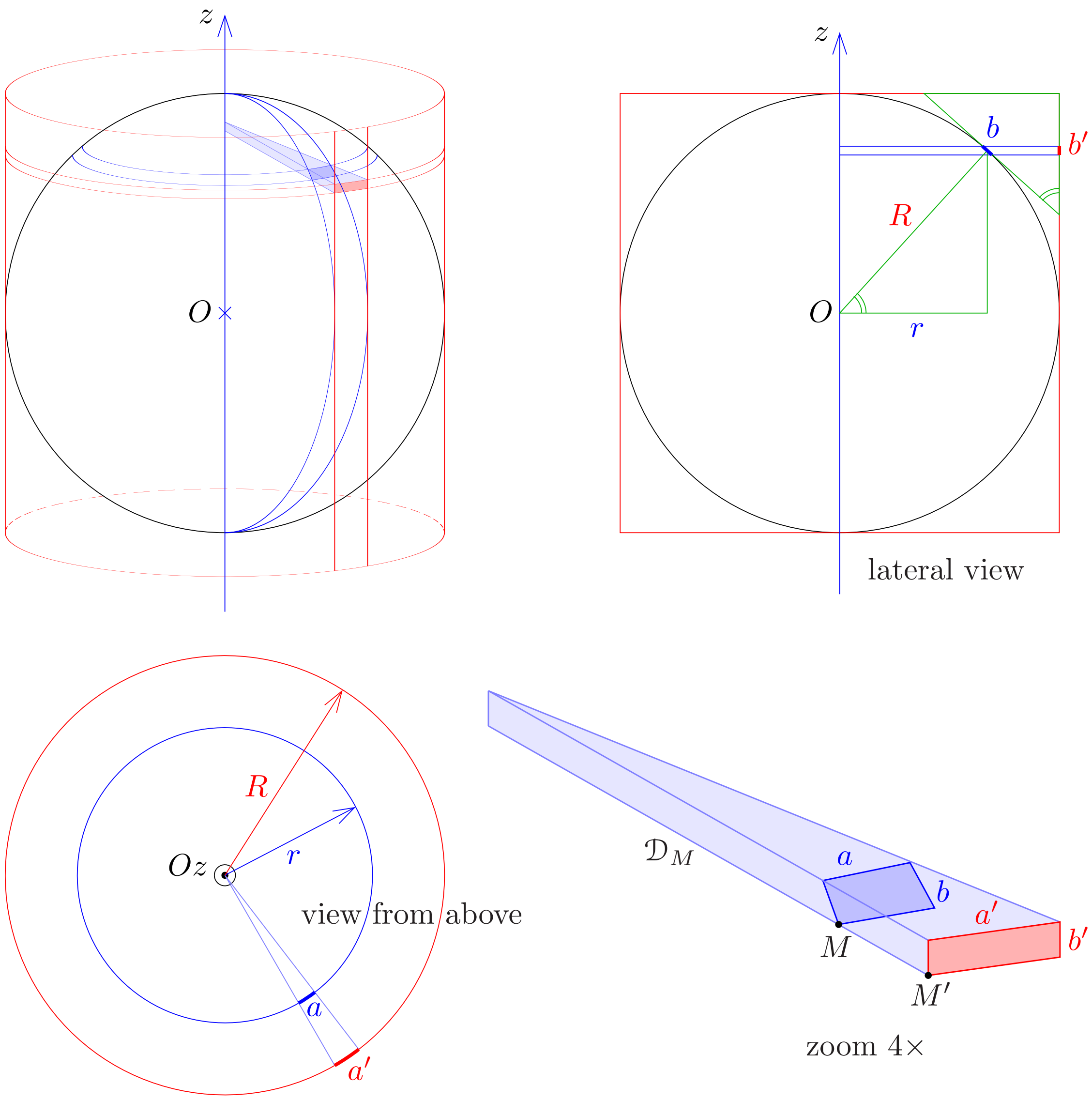
Since any two spheres of the same radius are isometric, their area depends only on the radius R. Let us take the center O of the sphere as the origin, and consider the "vertical" cylinder of radius R tangent to the sphere along the equator, and more precisely, the portion of cylinder located between the "horizontal" planes $z = -R$ and $z = R$. We use a "projection" of the sphere to the cylinder: for each point M of the sphere, we consider the point M' on the cylinder which is the intersection of the cylinder with the horizontal line $\mathcal{D}_M$ passing by M and intersecting the Oz axis. This projection is actually one of the simplest possible cartographic representations of the Earth. After cutting the cylinder along a meridian (say the meridian of longitude 180°), and unrolling the cylinder into a rectangle, we obtain the following cartographic map.



We are going to check that the cylindrical projection preserves areas, hence that the area of the sphere is equal to that of the corresponding rectangular map of sides $2R$ and $2\pi R$:

(4.9.5) $$A = 2R \times 2\pi R = 4\pi R^2.$$

In order to check that the areas are equal, we consider a "rectangular field" delimited by parallel and meridian lines, of very small size with respect to the sphere, in such a way that it can be seen as a planar surface, i.e. to a rectangle (for instance, on Earth, one certainly does not realize the rotundity of the globe when the size of the field does not exceed a few hundred meters).



Let a, b be the side lengths of our "rectangular field", respectively along parallel lines direction and meridian lines direction, and a', b' the side lengths of the corresponding rectangle projected on the tangent cylinder.

In the view from above, Thales' theorem immediately implies

$$\frac{a'}{a} = \frac{R}{r}.$$

In the lateral view, the two triangles represented in green are homothetic (they share a common angle, as the adjacent sides are perpendicular to each other). If we apply again Thales' theorem to the tangent triangle and more specifically to the sides adjacent to the common angle, we get

$$\frac{b'}{b} = \frac{\text{adjacent small side}}{\text{hypotenuse}} = \frac{r}{R}.$$

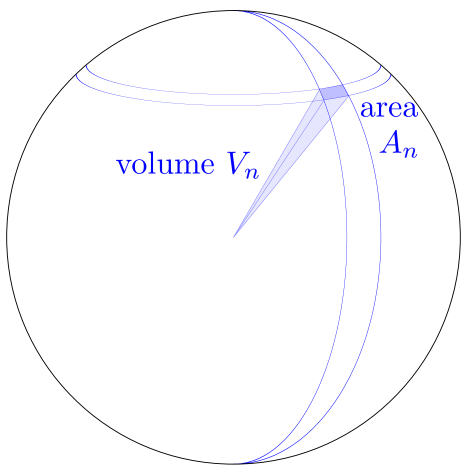
The product of these equalities yields

$$\frac{a' \times b'}{a \times b} = \frac{a'}{a} \times \frac{b'}{b} = \frac{R}{r} \times \frac{r}{R} = 1.$$

We conclude from there that the rectangle areas $a \times b$ and $a' \times b'$ are equal. This implies that the cylindrical projection preserves areas, and formula (4.9.5) follows.

Volume of a ball of radius R

To obtain the volume of a ball(13) of radius R, we use a gridding of the sphere by meridians and parallels, and calculate the volume of pyramids based on the "rectangular fields" delimited in this way. When the grid is sufficiently fine, we can consider that the fields are actually almost planar and rectangular (note that this spatial argument is the exact analogue of the planar argument used to derive the area of a disc through angular sectors, see section 2.2). Let A_1, A_2, A_3, ... be the areas of the rectangular fields and let V_1, V_2, V_3, ... be the volumes of the corresponding pyramids (clearly, the altitude of these pyramids is $h = R$). We denote here $n = 1$ or 2 or 3 ...



Volume of pyramids:

$$V_n = \frac{1}{3} A_n \times R$$

If $A = 4\pi R^2$ denotes the total area of the sphere, distributivity of multiplication withrespect to addition shows that the volume V of the ball is given by

$$V = V_1 + V_2 + V_3 + \ldots = \frac{1}{3}(A_1 + A_2 + A_3 + \ldots) \times R = \frac{1}{3} AR.$$

This gives the formula for the volume of the ball that we were aiming at:

$$V = \frac{1}{3} AR = \frac{1}{3} \times 4\pi R^2 \times R = \frac{4}{3}\pi R^3. \qquad (4.9.6)$$

5. An axiomatic approach to Euclidean geometry

Although we have been able to follow a deductive presentation when it is compared to some of the more traditional approaches – almost all of the statements were "proven" from the definitions – it should nevertheless be observed that some proofs relied merely on intuitive facts – this was for instance the case of the "proof" of Pythagoras' Theorem. The only way to break the vicious circle is to take some of the facts that we feel

(13) In mathematics, a ball is the portion of space delimited by a sphere, just as a disk is the plane domain bounded by a circle.

necessary to use as "axioms", that is to say, to consider them as assumptions from which we first deduct all other properties by logical deduction ; a choice of other assumptions as our initial premises leads to non-Euclidean geometries (see section 11).

As we shall see, the notion of a Euclidean plane can be defined using a single axiom, essentially equivalent to the conjunction of Pythagoras' Theorem - which was only partially justified - and the existence of Cartesian coordinates - which we had not discussed either. In case the idea of using an axiomatic approach would look frightening, we want to stress that this section may be omitted altogether – provided pupils are in some way brought to the idea that the coordinate systems can be changed (translated, rotated, etc.) according to the needs.

5.1. The "Pythagoras/Descartes" model

In our vision, plane Euclidean geometry is based on the following "axiomatic definition".

Definition. *What we will call a Euclidean plane is a set of points denoted $\mathcal{P}$, for which mutual distances of points are supposed to be known, i.e. there is a predefined function*

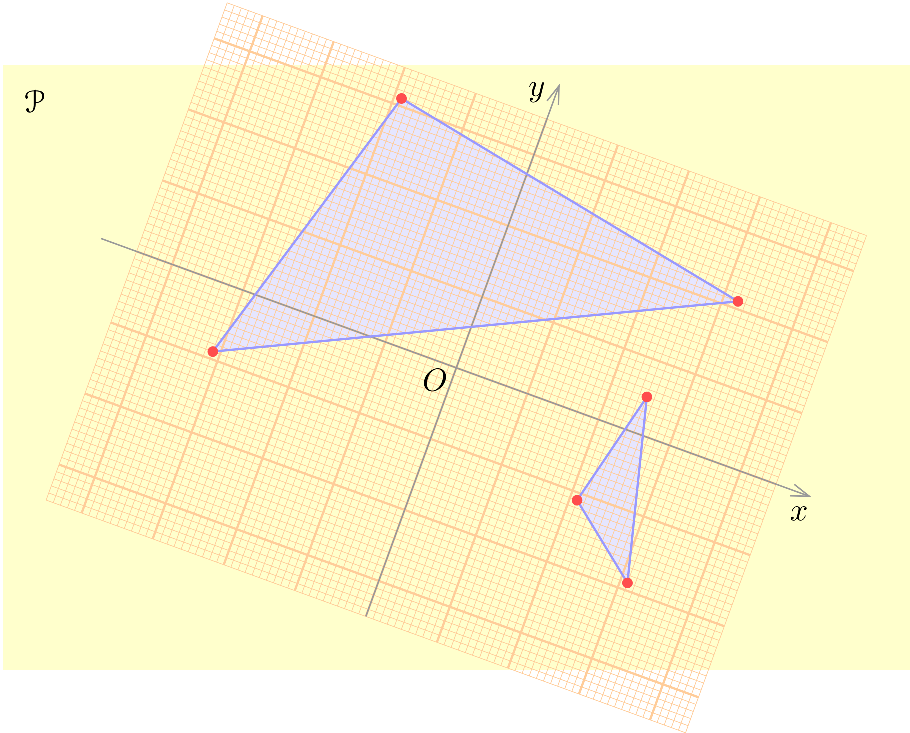
$$d : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}_+, \qquad (M, M') \longmapsto d(M, M') = MM' \geqslant 0,$$

and we assume that there exist "orthonormal coordinate systems": to each point one can assign a pair of coordinates, by means of a one-to-one correspondence $M \mapsto (x\,;\,y)$ satisfying the axiom[14]

(Pythagoras/Descartes) $$d(M, M') = \sqrt{(x' - x)^2 + (y' - y)^2}$$

for all points $M\ (x\,;y)$ and $M'\ (x'\,;\,y')$.

It is certainly a good practice to represent the choice of an orthonomal coordinate system by using a transparent sheet of graph paper and placing it over the paper sheet that contains the working area of the Euclidean plane (here that area contains two triangles depicted in blue, above which the transparent sheet of graph paper has been placed).



This already shows (at an intuitive level only at this point) that there is an infinite number of possible choices for the coordinate systems. We now investigate this in more detail.

5.1.1. Rotating the sheet of graph paper around O by $180°$

A rotation of $180°$ of the graph paper around O has the effect of just changing the orientation of axes. The new coordinates $(X\,;Y)$ are given with respect to the old ones by

$$X = -x, \qquad Y = -y.$$

Since $(-u)^2 = u^2$ for every real number u, we see that the formula

$$(*) \qquad d(M, M') = \sqrt{(X' - X)^2 + (Y' - Y)^2}$$

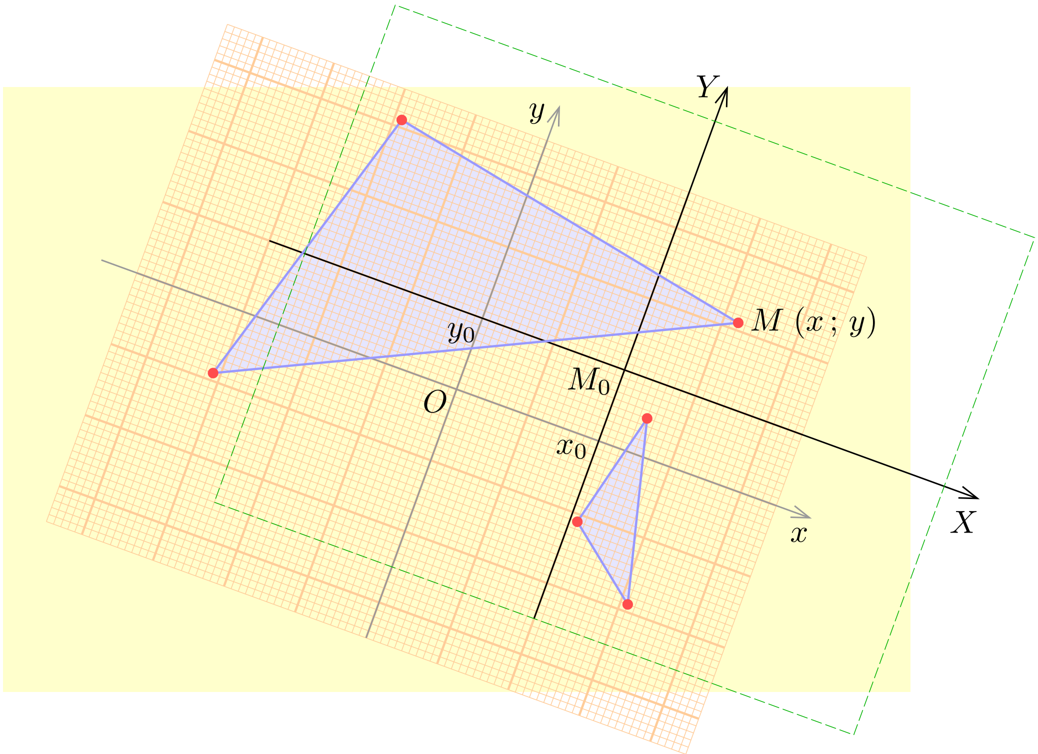
is still valid in the new coordinates, assuming it was valid in the original coordinates $(x\,;y)$.

5.1.2. Reversing the sheet of graph paper along one axis

If we reverse along Ox, we get $X = -x$, $Y = y$ and formula $(*)$ is still true. The argument is similar when reversing the sheet along Oy, we get the change of coordinates $X = x$, $Y = -y$ in that case.

5.1.3. Change of origin

Here we replace the origin O by an arbitrary point $M_0\ (x_0\,;y_0)$.



The new coordinates of point $M\ (x\,;\,y)$ are given by

$$X = x - x_0, \qquad Y = y - y_0.$$

For any two points M, M', we get in this situation

$$X' - X = (x' - x_0) - (x - x_0) = x' - x, \qquad Y' - Y = (y' - y_0) - (y - y_0) = y' - y$$

and we see that formula $(*)$ is still unchanged.

5.1.4. Rotation of axes

We will show that when the origin O is chosen, one can get the half-line Ox to pass through an aribrary point $M_1\ (x_1\,;\,y_1)$ distinct from O. This is intuitively obvious by “rotating” the sheet of graph paper around point O, but requires a formal proof relying on our “Pythagoras/Descartes” axiom. This proof is substantially more involved than what we have done yet, and can probably be jumped over at first – we give it here to show that there is no logical flaw in our approach. We start from the algebraic equality called *Lagrange's identity*

$$(au + bv)^2 + (-bu + av)^2 = a^2u^2 + b^2v^2 + b^2u^2 + a^2v^2 = (a^2 + b^2)(u^2 + v^2),$$

which is valid for all real numbers a, b, u, v. It can be obtained by developping the squares on the left and observing that the double products annihilate. As a consequence, if a and b satisfy $a^2 + b^2 = 1$ (such an example is $a = 3/5$, $b = 4/5$) and if we perform the change of coordinates

$$X = ax + by, \qquad Y = -bx + ay$$

we get, for any two points M, M' in the plane

$$X' - X = a(x' - x) + b(y' - y), \qquad Y' - Y = -b(x' - x) + a(y' - y),$$
$$(X' - X)^2 + (Y' - Y)^2 = (x' - x)^2 + (y' - y)^2$$

by Lagrange's identity with $u = x' - x$, $v = y' - y$. On the other hand, it is easy to check that

$$aX - bY = x, \qquad bX + aY = y,$$

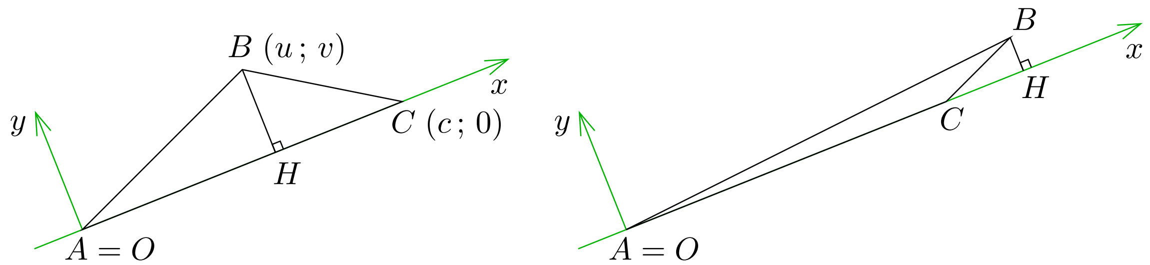
hence the assignment $(x\,;\,y) \mapsto (X\,;\,Y)$ is one-to-one. We infer from there that in the sense of our definition, $(X\,;\,Y)$ is indeed an orthonormal coordinate system. If we now choose $a = kx_1$, $b = ky_1$, the coordinates of point M_1 $(x_1\,;\,y_1)$ are transformed into

$$X_1 = ax_1 + by_1 = k(x_1^2 + y_1^2), \qquad Y_1 = -bx_1 + ay_1 = k(-y_1x_1 + x_1y_1) = 0,$$

and the condition $a^2 + b^2 = k^2(x_1^2 + y_1^2) = 1$ is satisfied by taking $k = 1/\sqrt{x_1^2 + y_1^2}$. Since $X_1 = \sqrt{x_1^2 + y_1^2} > 0$ and $Y_1 = 0$, the point M_1 is actually located on the half-line OX in the new coordinate system.

5.2. Revisiting the triangular inequality

The proof given in 3.1.1, which relied on facts that were not entirely settled, can now be made completely rigorous.



Given three distinct points A, B, C distincts, we select $O = A$ as the origin and the half line $[A, C)$ as the Ox axis. Our three points then have coordinates

$$A\ (0\,;\,0), \qquad B\ (u\,;\,v), \qquad C\ (c\,;\,0), \qquad c > 0,$$

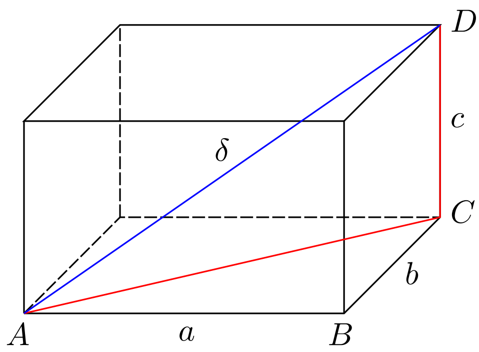
and the foot H of the altitude starting at B is $H\ (u\,;\,0)$. We find $AC = c$ and

$$AB = \sqrt{u^2 + v^2} \geqslant AH = |u| \geqslant u, \quad BC = \sqrt{(c-u)^2 + v^2} \geqslant HC = |c - u| \geqslant c - u.$$

Therefore $AC = c = u + (c - u) \leqslant AB + BC$ in all cases. The equality only holds when we have at the same time $v = 0$, $u \geqslant 0$ and $c - u \geqslant 0$, i.e. $u \in [0, c]$ and $v = 0$, in other words when B is located on the segment $[A, C]$ of the Ox axis.

5.3. Axioms of higher dimensional affine spaces

The approach that we have described is also appropriate for the introduction of Euclidean geometry in any dimension, especially in dimension 3. The starting point is the calculation of the diagonal δ of a rectangular parallelepiped with sides a, b, c :



As the triangles ACD and ABC are rectangle in C and B respectively, we have

$$AD^2 = AC^2 + CD^2 \quad \text{and} \quad AC^2 = AB^2 + BC^2$$

hence the "great diagonal" of our rectangle parallelepiped is given by

$$\delta^2 = AD^2 = AB^2 + BC^2 + CD^2 = a^2 + b^2 + c^2 \quad \Rightarrow \quad \delta = \sqrt{a^2 + b^2 + c^2}.$$

This leads to the following definition

Definition. *A Euclidean space of dimension* 3 *is a set of points denoted* $\mathcal{E}$*, equipped with a distance* d*, namely a mapping*

$$d : \mathcal{E} \times \mathcal{E} \longrightarrow \mathbb{R}_+, \qquad (M, M') \longmapsto d(M, M') = MM' \geqslant 0,$$

such that one can find "orthonormal coordinate systems" in the form of a one-to-one correspondence $\mathcal{E} \ni M \longmapsto (x\,;\,y\,;\,z) \in \mathbb{R}^3$ *satisfying the axiom*

(Pythagore/Descartes) $$d(M, M') = \sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2}$$

for all points $M\ (x\,;\,y\,;\,z)$ *and* $M'\ (x'\,;\,y'\,;\,z')$ *in* $\mathcal{E}$*.*

One could of course give a similar definition in any dimension n by taking coordinate systems $(x_1\,,\,\ldots\,,\,,x_n) \in \mathbb{R}^n$. The triangle inequality can be proved with almost no changed: given three points A, B, C, we choose A as the origin, the half-line $[A, C)$ as the axis Ox. This reduces the calculation to the case where $B = (u\,;\,v\,;\,w)$ and $C = (c\,;\,0\,;\,0)$. For this, one must first verify that it is possible to find an orthonormal coordinate system pointing the Ox axis in the direction of an arbitrary point $M_1\ (x_1\,;\,y_1\,;\,z_1)$. One begins by getting z_1 to vanish by a change of variables $Y = ay + bz$, $Z = by - az$ in the last two coordinates; this brings M_1 in the "horizontal" plane $Z = 0$; then one makes y_1 vanish with the same method, using a change of variables in x, y only. Idem in higher dimensions.

6. Foundations of vector calculus

We will work here in the plane to simplify the exposition, but the only change in higher dimension would be the appearance of additional coordinates.

6.1. Median formula

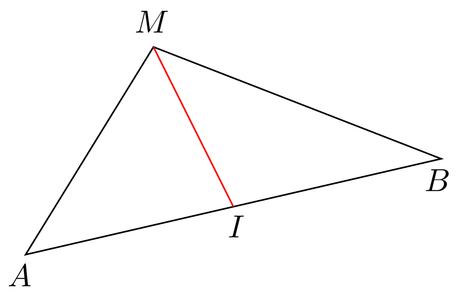
Consider points A, B with coordinates $(x_A\,;y_A)$, $(x_B\,;y_B)$ in an orthonormal frame Oxy.The point I of coordinates

$$x_I = \frac{x_A + x_B}{2}, \qquad y_I = \frac{y_A + y_B}{2}$$

satisfies $IA = IB = \frac{1}{2}AB$: this is the midpoint of segment $[A, B]$.

Median formula. *For every point M $(x\,;y)$, one has*

$$MA^2 + MB^2 = 2\,MI^2 + \frac{1}{2}AB^2 = 2\,MI^2 + 2\,IA^2.$$



Proof. In fact, by expanding the squares, we get

$$(x - x_A)^2 + (x - x_B)^2 = 2x^2 - 2(x_A + x_B)x + x_A^2 + x_B^2,$$

while

$$\begin{aligned} 2(x - x_I)^2 + \frac{1}{2}(x_B - x_A)^2 &= 2(x^2 - 2x_I x + x_I^2) + \frac{1}{2}(x_B - x_A)^2 \\ &= 2\Big(x^2 - (x_A + x_B)x + \frac{1}{4}(x_A + x_B)^2\Big) + \frac{1}{2}(x_B - x_A)^2 \\ &= 2x^2 - 2(x_A + x_B)x + x_A^2 + x_B^2. \end{aligned}$$

Therefor we get

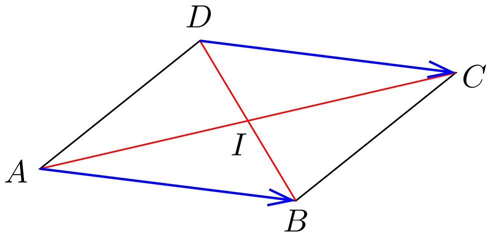
$$(x - x_A)^2 + (x - x_B)^2 = 2(x - x_I)^2 + \frac{1}{2}(x_B - x_A)^2.$$

The median formula is obtained by adding the analogous equality for coordinates y and applying Pythagoras' theorem. □

It follows from the median formula that there is a unique point M such that $MA = MB = \frac{1}{2}\,AB$, in fact we then find $MI^2 = 0$, hence $M = I$. The coordinate formulas that we initially gave to define midpoints are therefore independent of the choice of coordinates.

6.2. Parallelograms

A quadrilateral $ABCD$ is a parallelogram if and only if its diagonals $[A, C]$ and $[B, D]$ intersect at their midpoint :



In this way, we find the necessary and sufficient condition

$$x_I = \frac{1}{2}(x_B + x_D) = \frac{1}{2}(x_A + x_C), \qquad y_I = \frac{1}{2}(y_B + y_D) = \frac{1}{2}(y_A + y_C),$$

which is equivalent to

$$x_B + x_D = x_A + x_C, \qquad y_B + y_D = y_A + y_C$$

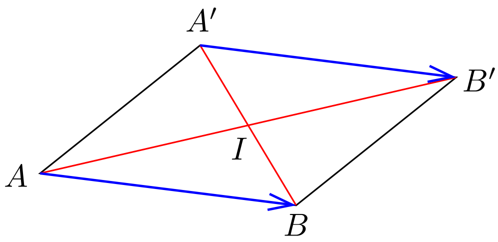
or, alternatively, t

$$x_B - x_A = x_C - x_D, \qquad y_B - y_A = y_C - y_D,$$

in other words, the variation of coordinates involved in getting from A to B is the same as the one involved in getting from D to C.

6.3. Vectors

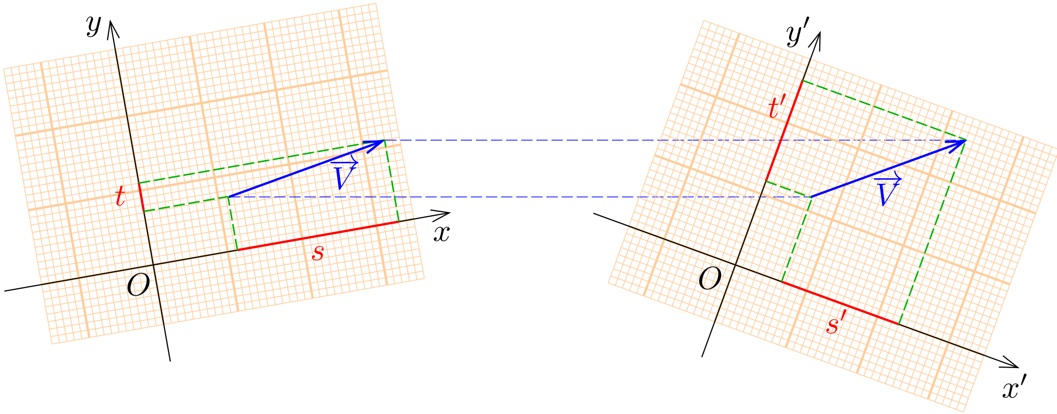
A *bipoint* is an ordered pair (A, B) of points ; we say that A is the *origin* and that B is the *extremity* of the bipoint. The bipoints (A, B) and (A', B') are said to be *equipollent* if the quadrilateral $ABB'A'$ is a parallelogram (which can possibly be a “flat” parallelogram in case the four points are aligned).



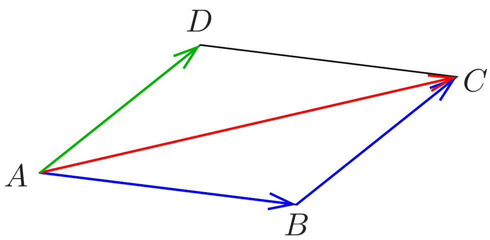
Definition. *Given two points A, B, the vector $\overrightarrow{AB}$ is the “variation of position” needed to get from A to B. Given a coordinate frame Oxy, this “variation of position” is expressed along the Ox axis by $x_B - x_A$ and along the Oy axis by $y_B - y_A$. If the bipoints (A, B) and (A', B') are equipollent, the vectors $\overrightarrow{AB}$ and $\overrightarrow{A'B'}$ are equal since the variations $x_{B'} - x_{A'} = x_B - x_A$ and $y_{B'} - y_{A'} = y_B - y_A$ are the same (this is true in any coordinate system).*

The “component” of vector $\overrightarrow{AB}$ in the coordinate system Oxy are the numbers denoted in the form of an ordered pair $(x_B - x_A\,; y_B - y_A)$. The components $(s\,; t)$ of a vector

$\overrightarrow{V}$ depend of course on the choice of the coordinate frame Oxy : to a given vector $\overrightarrow{V}$ one assigns different components $(s\,;\,t)$, $(s'\,;\,t')$ in different coordinate frames Oxy, $Ox'y'$.



6.4. Addition of vectors



The addition of vectors is defined by means of *Chasles' relation*

(6.4.1) $$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$$

for any three points A, B, C : when one takes the sum of the variation of position required to get from A to B, and then from B to C, one finds the variation of position to get from A to C ; actually, we have for instance

$$(x_B - x_A) + (x_C - x_B) = x_C - x_A$$

for the component along the Ox axis. Equivalently, if $ABCD$ is a parallelogram, one can also put

(6.4.2) $$\overrightarrow{AB} + \overrightarrow{AD} = \overrightarrow{AC}.$$

That (6.4.1) and (6.4.2) are equivalent follows from the fact that $\overrightarrow{AD} = \overrightarrow{BC}$ in parallelogram $ABCD$. For any choice of coordinae frame Oxy, the sum of vectors of components $(s\,;\,t)$, $(s'\,;\,t')$ has components $(s+s'\,;\,t+t')$.

For every point A, the vector $\overrightarrow{AA}$ has zero componaents : it will be denoted simply $\overrightarrow{0}$. Obviously, we have $\overrightarrow{V} + \overrightarrow{0} = \overrightarrow{0} + \overrightarrow{V} = \overrightarrow{V}$ for every vector $\overrightarrow{V}$. On the other hand, Chasles' relation yields

$$\overrightarrow{AB} + \overrightarrow{BA} = \overrightarrow{AA} = \overrightarrow{0}$$

for all points A, B. Therefore we define

$$-\overrightarrow{AB} = \overrightarrow{BA},$$

in other words, the opposite of a vector is obtained by exchanging the origin and extremity of any corresponding bipoint.

6.5. Multiplication of a vector by a real number

Given a vector $\overrightarrow{V}$ of components $(s\,;\,t)$ in a coordinate frame Oxy and an arbitrary real number λ, we define $\lambda\overrightarrow{V}$ as the vector of components $(\lambda s\,;\,\lambda t)$.

This definition is actually independent of the coordinate frame Oxy. In fact if $\overrightarrow{V} = \overrightarrow{AB} \neq \overrightarrow{0}$ and $\lambda \geqslant 0$, we have $\lambda\overrightarrow{AB} = \overrightarrow{AC}$ where C is the unique point located on the half-line $[A, B)$ such that $AC = \lambda\, AB$. On the other hand, if $\lambda \leqslant 0$, we have $-\lambda \geqslant 0$ and

$$\lambda\overrightarrow{AB} = (-\lambda)(-\overrightarrow{AB}) = (-\lambda)\overrightarrow{BA}.$$

Finally, it is clear that $\lambda\,\overrightarrow{0} = \overrightarrow{0}$. Multiplication of vectors by a number is distributive with respect to the addition of vectors (this is a consequence of the distributivity of multiplication with respect to addition in the set of real numbers).

7. Cartesian equation of circles and trigonometric functions

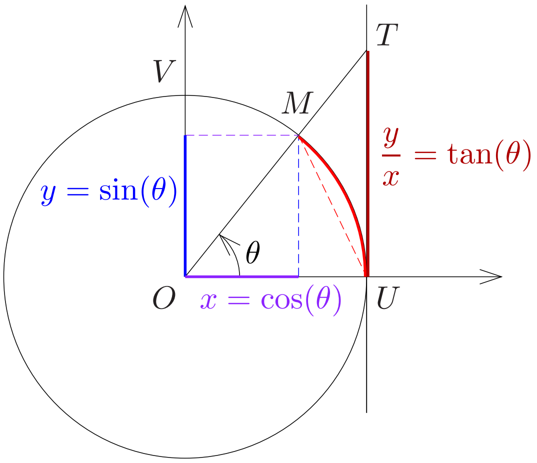
By Pythagoras' thorem, the circle of center $A\ (a, b)$ and radius R in the plane is the set of points M satisfying the equation

$$AM = R \quad\Leftrightarrow\quad AM^2 = R^2 \quad\Leftrightarrow\quad (x-a)^2 + (y-b) = R^2,$$

which can also be put in the form $x^2 + y^2 - 2ax - 2by + c = 0$ with $c = a^2 + b^2 - R^2$. Conversely, the set of solutions of such an equation defines a circle of center $A\ (a\,;\,b)$ and of radius $R = \sqrt{a^2 + b^2 - c}$ if $c < a^2 + b^2$, is reduced to point A if $c = a^2 + b^2$, and is empty if $c > a^2 + b^2$.

The *trigonometric circle* $\mathcal{C}$ is defined to be the unit circle centered at the origin in an orthormal coordinate system Oxy, that is, the of points $M\ (x\,;\,y)$ such that $x^2 + y^2 = 1$. Let U be the point of coordinates $(1\,;\,0)$ and V the point of coordinates $(0\,;\,1)$. The usual trigonometric functions cos, sin and tan are then defined for arbitrary angle arguments as shown on the above figure[14] :

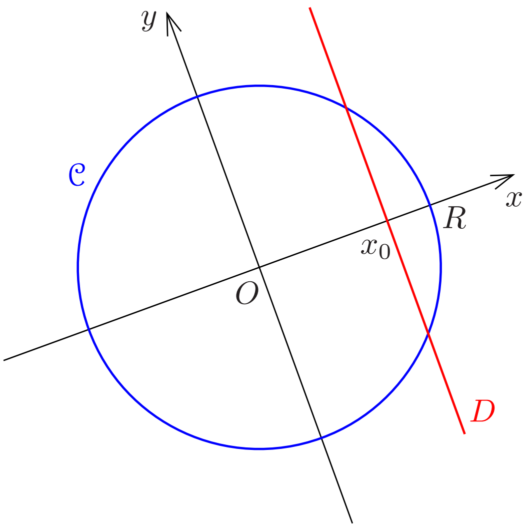
[14] It seems essential at this stage that functions cos, sin, tan have already been introduced as the ad hoc ratios of sides in a right triangle, i.e. at least for the case of acute angles, and that their values for the remarkable angle values 0°, 30°, 45°, 60°, 90° are known.



The equation of the circle implies the relation $(\cos\theta)^2 + (\sin\theta)^2 = 1$ for every θ.

8. Intersection of lines and circles

Let us begin by intersecting a circle $\mathcal{C}$ of center A and radius R with an arbitrary line $\mathcal{D}$. In order to simplify the calculation, we take $A = O$ as the origin and we take the axis Ox to be perpendicular to the line $\mathcal{D}$. The line $\mathcal{D}$ is then “vertical” in the coordinate frame Oxy. (We start here right away with the most general case, but, once again, it would be desirable to approach the question by treating first simple numerical examples ...).



This leads to equation

$$\mathcal{C} : x^2 + y^2 = R^2, \qquad D : x = x_0,$$

hence

$$y^2 = R^2 - x_0^2.$$

As a consequence, if $|x_0| < R$, we have $R^2 - x_0^2 > 0$ and there are two solutions $y = \sqrt{R^2 - x_0^2}$ and $y = -\sqrt{R^2 - x_0^2}$, corresponding to two intersection points$(x_0, \sqrt{R^2 - x_0^2})$ and $(x_0, -\sqrt{R^2 - x_0^2})$ that are symmetric with respect to the Ox axis. If $|x_0| = R$, we find a single solution $y = 0$: the line $\mathcal{D} : x = x_0$ is tangent to circle $\mathcal{C}$ at point $(x_0\,;\,0)$. If $|x_0| > R$, the equation $y^2 = R^2 - x_0^2 < 0$ has no solution ; the line $\mathcal{D}$ does not intersect the circle.

Consider now the intersection of a circle $\mathcal{C}$ of center A and radius R with a circle $\mathcal{C}'$ of center A' and radius R'. Let $d = AA'$ be the distance between their centers. If $d = 0$ the circles are concentric and the discussion is easy (the circles coincide if $R = R'$, and are disjoint if $R \neq R'$). We will therefore assume that $A \neq A'$, i.e. $d > 0$. By selecting $O = A$ as the origin and $Ox = [A, A')$ as the positive x axis, we are reduces to the case where $A\,(0\,;\,0)$ and $A'\,(d\,;\,0)$. We then get equations

$$\mathcal{C} : x^2 + y^2 = R^2, \qquad \mathcal{C}' : (x-d)^2 + y^2 = R'^2 \iff x^2 + y^2 = 2dx + R'^2 - d^2.$$

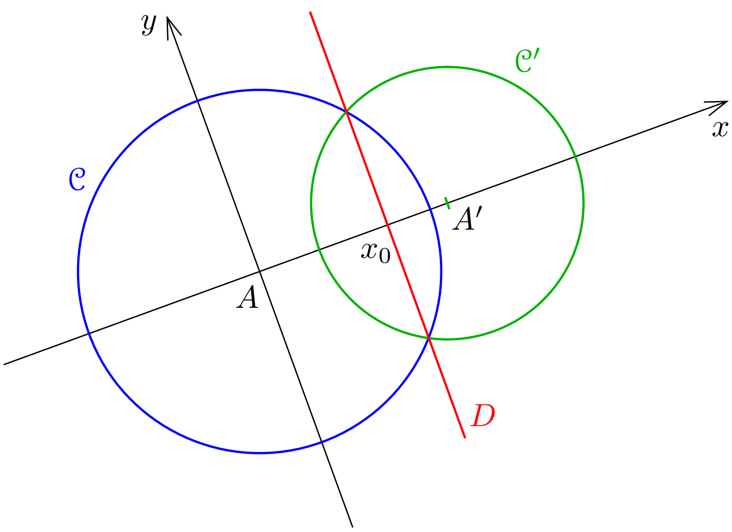
For any point M in the intersection $\mathcal{C} \cap \mathcal{C}'$, we thus get $2dx + R'^2 - d^2 = R^2$, hence

$$x = x_0 = \frac{1}{2d}(d^2 + R^2 - R'^2).$$

This shows that the intersection $\mathcal{C} \cap \mathcal{C}'$ is contained in the intersection $\mathcal{C} \cap D$ of $\mathcal{C}$ with the line $\mathcal{D} : x = x_0$. Conversely, one sees that if $x^2 + y^2 = R^2$ and $x = x_0$, then $(x\,;\,y)$ also satisfies the equation

$$x^2 + y^2 - 2dx = R^2 - 2dx_0 = R^2 - (d^2 + R^2 - R'^2) = R'^2 - d^2$$

which is the equation of $\mathcal{C}'$, hence $\mathcal{C} \cap D \subset \mathcal{C} \cap \mathcal{C}'$ and finally $\mathcal{C} \cap \mathcal{C}' = \mathcal{C} \cap D$.



The intersection points are thus given by $y = \pm\sqrt{R^2 - x_0^2}$. As a consequene, we have exactly two solutions that are symmetric with respect to the line (AA') as soon as $-R < x_0 < R$, or equivalently

$$\begin{aligned} -2dR < d^2 + R^2 - R'^2 < 2dR \quad &\iff \quad (d+R)^2 > R'^2 \ \text{ et } \ (d-R)^2 < R'^2 \\ &\iff \quad d + R > R', \ \ d - R < R', \ \ d - R > -R', \end{aligned}$$

i.e. $|R - R'| < d < R + R'$. If one of the inequalities is an equality, we get $x_0 = \pm R$ and we thus find a single solution $y = 0$. The circles are tangent internally if $d = |R - R'|$ and tangent externally if $d = R + R'$.

Note that these results lead to a complete and rigorous proof of the isometry criteria for triangles : up to an orthonormal change of coordinates, each of the three cases entirely determines the coordinates of the triangles modulo a reflection with respect to Ox (in this argument, the origin O is chosen as one of the vertices and the axis Ox is taken to be the direction of a side of known length). The triangles specified in that way are thus isometric.

9. Scalar product

The *norm* $\|\overrightarrow{V}\|$ of a vector $\overrightarrow{V} = \overrightarrow{AB}$ is the length $AB = d(A, B)$ of an arbitrary bipoint that defines $\overrightarrow{V}$. From there, we put

$$\overrightarrow{U} \cdot \overrightarrow{V} = \frac{1}{2}\big(\|\overrightarrow{U} + \overrightarrow{V}\|^2 - \|\overrightarrow{U}\|^2 - \|\overrightarrow{V}\|^2\big) \tag{9.1}$$

in particular $\overrightarrow{U} \cdot \overrightarrow{U} = \|\overrightarrow{U}\|^2$. The real number $\overrightarrow{U} \cdot \overrightarrow{V}$ is called the *inner product* of $\overrightarrow{U}$ and $\overrightarrow{V}$, and $\overrightarrow{U} \cdot \overrightarrow{U}$ is also defined to be the inner square of $\overrightarrow{U}$, denoted $\overrightarrow{U}^2$. Consequently we obtain

$$\overrightarrow{U}^2 = \overrightarrow{U} \cdot \overrightarrow{U} = \|\overrightarrow{U}\|^2.$$

By definition (9.1), we have

$$\|\overrightarrow{U} + \overrightarrow{V}\|^2 = \|\overrightarrow{U}\|^2 + \|\overrightarrow{V}\|^2 + 2\,\overrightarrow{U} \cdot \overrightarrow{V}, \tag{9.2}$$

and this formula can also be rewritten

$$(\overrightarrow{U} + \overrightarrow{V})^2 = \overrightarrow{U}^2 + \overrightarrow{V}^2 + 2\,\overrightarrow{U} \cdot \overrightarrow{V}. \tag{9.2$'$}$$

This was the main motivation of the definition : that the usual identity for the square of a sum be valid for inner products. In dimension 2 and in an orthonormal frame Oxy, we find $\overrightarrow{U}^2 = x^2 + y^2$; if $\overrightarrow{V}$ has components $(x' ; y')$, Definition (9.1) implies

$$\overrightarrow{U} \cdot \overrightarrow{V} = \frac{1}{2}\big((x + x')^2 + (y + y')^2 - (x^2 + y^2) - (x'^2 + y'^2)\big) = xx' + yy'. \tag{9.3}$$

In dimension n, we would find similarly

$$\overrightarrow{U} \cdot \overrightarrow{V} = x_1 x'_1 + x_2 x'_2 + \ldots + x_n x'_n.$$

From there, we derive that the inner product is “bilinear”, namely that

$$\begin{aligned}
(k\overrightarrow{U}) \cdot \overrightarrow{V} &= \overrightarrow{U} \cdot (k\overrightarrow{V}) = k\,\overrightarrow{U} \cdot \overrightarrow{V},\\
(\overrightarrow{U_1} + \overrightarrow{U_2}) \cdot \overrightarrow{V} &= \overrightarrow{U_1} \cdot \overrightarrow{V} + \overrightarrow{U_2} \cdot \overrightarrow{V},\\
\overrightarrow{U} \cdot (\overrightarrow{V_1} + \overrightarrow{V_1}) &= \overrightarrow{U} \cdot \overrightarrow{V_1} + \overrightarrow{U} \cdot \overrightarrow{V_2}.
\end{aligned}$$

if $\overrightarrow{U}$, $\overrightarrow{V}$ are two vectors, we can pick a point A and write $\overrightarrow{U} = \overrightarrow{AB}$, then $\overrightarrow{V} = \overrightarrow{BC}$, so that $\overrightarrow{U} + \overrightarrow{V} = \overrightarrow{AC}$. The triangle ABC is rectangle if and only if we have Pythagoras' relation $AC^2 = AB^2 + BC^2$, i.e.

$$\|\overrightarrow{U} + \overrightarrow{V}\|^2 = \|\overrightarrow{U}\|^2 + \|\overrightarrow{V}\|^2,$$

in other words, by (9.2), if and only if $\overrightarrow{U} \cdot \overrightarrow{V} = 0$.

Consequence. *Tw vectors $\overrightarrow{U}$ and $\overrightarrow{V}$ are perpendicular if and only if $\overrightarrow{U} \cdot \overrightarrow{V} = 0$.*

More generally, if we fix an origin O and a point A such that $\overrightarrow{U} = \overrightarrow{OA}$, one can also pick a coordinate system such that A belongs to the Ox axis, that is, $A = (u\,;\,0)$. For every vector$\overrightarrow{V} = \overrightarrow{OB}$ $(v\,;\,w)$ in Oxy, we then get

$$\overrightarrow{U} \cdot \overrightarrow{V} = uv$$

whereas

$$\|\overrightarrow{U}\| = u, \qquad \|\overrightarrow{V}\| = \sqrt{v^2 + w^2}.$$

As the half-line $[O, B)$ intersects the trigonometric circle at point $(kv\,;\,kw)$ with $k = 1/\sqrt{v^2 + w^2}$, we get by definition

$$\cos(\widehat{\overrightarrow{U}, \overrightarrow{V}}) = \cos(\widehat{AOB}) = kv = \frac{v}{\sqrt{v^2 + w^2}}.$$

This leads to the very useful formulas

$$\text{(9.4)} \qquad \overrightarrow{U} \cdot \overrightarrow{V} = \|\overrightarrow{U}\|\,\|\overrightarrow{V}\|\,\cos(\widehat{\overrightarrow{U}, \overrightarrow{V}}), \qquad \cos(\widehat{\overrightarrow{U}, \overrightarrow{V}}) = \frac{\overrightarrow{U} \cdot \overrightarrow{V}}{\|\overrightarrow{U}\|\,\|\overrightarrow{V}\|}.$$

10. Vector spaces, linear and affine linear maps

At this point, we have all the necessary foundations, and the succession of concepts to be introduced becomes much more flexible – much of what we discuss below only concerns high school level and beyond.

One can for example study further properties of triangles and circles, and gradually introduce the main geometric transformations (in the plane to start with) : translations, homotheties, affinities, axial symmetries, projections, rotations with respect to a point ; and in space, symmetries with respect to a point, a line or a plane, orthogonal projections on a plane or on a line, rotation around an axis. Available tools allow making either intrinsic geometric reasonings (with angles, distances, similarity ratios, ...), or calculations in Cartesian coordinates. It is actually desirable that these techniques remain intimately connected, as this is common practice in contemporary mathematics (the period that we describe as "contemporary" actually going back to several centuries for mathematicians, engineers, physicists ...)

It is then time to investigate the phenomenon of linearity, independently of any distance consideration. This leads to the concepts of linear combinations of vectors, linear dependence and independence, non orthonormal frames, etc, in relation with the resolution of systems of linear equations. One is quickly led to determinants 2×2, 3×3, to equations of lines, planes, etc.

Given an affine Euclidean space $\mathcal{E}$ of dimension n, what has been done in Section 6 works without change. We denote by $\overrightarrow{\mathcal{E}}$ the set of vectors $\overrightarrow{V} = \overrightarrow{AB}$ associated with the points of $\mathcal{E}$. This set is provided with two laws, namely $+$, the addition of vectors, and $\cdot$, the scalar multiplication, satisfying the usual properties:

(A 0) Addition is a "law of composition" $\overrightarrow{\mathcal{E}} \times \overrightarrow{\mathcal{E}} \to \overrightarrow{\mathcal{E}}$;

(A 1) addition of vectors is associative;

(A 2) addition of vectors is commutative;

(A 3) addition possesses a neutral element denoted $\overrightarrow{0}$;

(A 4) every vector $\overrightarrow{V}$ has an opposite $-\overrightarrow{V}$ such that $\overrightarrow{V} + (-\overrightarrow{V}) = (-\overrightarrow{V}) + \overrightarrow{V} = \overrightarrow{0}$.

(M 0) Scalar multiplication $(\lambda, \overrightarrow{V}) \mapsto \lambda\overrightarrow{V}$ is an operation (or external law of composition) $\mathbb{R} \times \overrightarrow{\mathcal{E}} \to \overrightarrow{\mathcal{E}}$;

(M 1) The element $1 \in \mathbb{R}$ operates trivially: $1 \cdot \overrightarrow{V} = \overrightarrow{V}$;

(M 2) scalar multiplication satisfies the "pseudo-associativity" rule $\lambda \cdot (\mu \cdot \overrightarrow{V}) = (\lambda\mu) \cdot \overrightarrow{V}$;

(M 3) scalar multiplication is left distributive: $(\lambda + \mu) \cdot \overrightarrow{V} = \lambda \cdot \overrightarrow{V} + \mu \cdot \overrightarrow{V}$;

(M 4) scalar multiplication is right distributive: $\lambda \cdot (\overrightarrow{V} + \overrightarrow{W}) = \lambda \cdot \overrightarrow{V} + \lambda \cdot \overrightarrow{W}$.

Whenever we have a set $(\overrightarrow{\mathcal{E}}, +, \cdot)$ equipped with two laws $+$ and $\cdot$ satisfying properties (A 0,1,2,3,4) and (M 0,1,2,3,4) as above for all scalar elements λ, $\mu \in \mathbb{R}$ and all vectors $\overrightarrow{V}$, $\overrightarrow{W}$, we say that $\overrightarrow{\mathcal{E}}$ has a vector space structure over the field of real numbers (this terminology comes from the fact that one can also consider vector spaces over other fields, for instance the field $\mathbb{Q}$ of rational numbers).

For any of the previously mentioned transformations $s : M \mapsto s(M)$, we see that the transformation is given in coordinates by formulas of the form: for M $(x_i)_{1 \leqslant i \leqslant n}$ and $s(M)$ $(y_i)_{1 \leqslant i \leqslant n}$, we have

$$y_i = \sum_{i=1}^{n} a_{ij} x_j + b_i, \tag{10.1}$$

i.e., the coordinates (y_i) of $s(M)$ are affine linear functions of the coordinates (x_i) of M. More generally, one can consider a transformation $s : \mathcal{E} \to \mathcal{F}$ of a space $\mathcal{E}$ of dimension n to a space $\mathcal{F}$ of dimension p not necessarily equal to n (e.g., a projection of the space $\mathcal{E}$ of dimension 3 on a plane $\mathcal{F} = \mathcal{P} \subset \mathcal{E}$). Such a mapping can be expressed in the similar way. In terms of matrices (in our opinion, matrix formalism should be introduced already at high school level – at least in dimensions 1, 2 and 3), one can always write such mappings in the form $Y = AX + B$. We then say that s is an affine linear transformation. If N is another point of coordinates $N : X' = (x'_i)_{1 \leqslant i \leqslant n}$ and $s(N) : Y = (y'_i)_{1 \leqslant i \leqslant p}$ is its image, we see that $Y' = AX' + B$, thus

$$\overrightarrow{s(M)s(N)} : Y' - Y = A(X' - X).$$

Therefore, if we denote by $\sigma : \overrightarrow{\mathcal{E}} \to \overrightarrow{\mathcal{F}}$, $\overrightarrow{V} \mapsto \sigma(\overrightarrow{V})$, the vector transformation defined by $Y = AX$, we get the formula

$$\overrightarrow{s(M)s(N)} = \sigma(\overrightarrow{MN}), \tag{10.2}$$

and σ possesses the following essential property referred to as "linearity"

$$\sigma(\overrightarrow{V} + \overrightarrow{W}) = \sigma(\overrightarrow{V}) + \sigma(\overrightarrow{W}), \qquad \sigma(\lambda\overrightarrow{V}) = \lambda\sigma(\overrightarrow{V}). \tag{10.3}$$

Conversely, if $\sigma : \overrightarrow{\mathcal{E}} \to \overrightarrow{\mathcal{F}}$ satisfies (10.3), it is easily shown by using bases that σ is given in coordinates by a formula of the form $Y = AX$; moreover, if $s : \mathcal{E} \to \mathcal{F}$ satisfies (10.2), by applying the formula the the pair (O, M) and by denoting B the colum vector of coordinates of $s(O)$, we see that $Y - B = AX$, i.e. $Y = AX + B$, thus s is an affine linear transformation. We now state a very important "rigidity theorem" of Euclidean geometry.

Definition. *Let $\mathcal{E}$ and $\mathcal{F}$ be two Euclidean spaces and let $s : \mathcal{E} \to \mathcal{F}$ be an arbitrary map between these. We say that s is an isometry from $\mathcal{E}$ to $\mathcal{F}$ if for every pair of points (M, N) of $\mathcal{E}$, we have $d(s(M), s(N)) = d(M, N)$.*

Theorem. *If $s : \mathcal{E} \to \mathcal{F}$ is an isometry, then s is an affine transformation, and its associated linear map $\sigma : \overrightarrow{\mathcal{E}} \to \overrightarrow{\mathcal{F}}$ is an orthogonal transform of Euclidean vector spaces, namely a linear map preserving orthogonality and inner products :*

$$\sigma(\overrightarrow{V}) \cdot \sigma(\overrightarrow{W}) = \overrightarrow{V} \cdot \overrightarrow{W} \tag{10.4}$$

for all vectors $\overrightarrow{V}, \overrightarrow{W} \in \overrightarrow{E}$.

Proof. Fix an origin O, and define $\sigma : \overrightarrow{\mathcal{E}} \to \overrightarrow{\mathcal{F}}$ by

$$\sigma(\overrightarrow{OM}) = \overrightarrow{s(O)s(M)}.$$

The properties of inner products yield

$$\begin{aligned} \overrightarrow{OM} \cdot \overrightarrow{ON} &= \frac{1}{2}\big(\|\overrightarrow{OM}\|^2 + \|\overrightarrow{ON}\|^2 - \|\overrightarrow{OM} - \overrightarrow{ON}\|^2\big) \\ &= \frac{1}{2}\big(d(O,M)^2 + d(O,N)^2 - d(M,N)^2\big). \end{aligned} \tag{10.5}$$

Therefore we get

$$\begin{aligned} \sigma(\overrightarrow{OM}) \cdot \sigma(\overrightarrow{ON}) &= \overrightarrow{s(O)s(M)} \cdot \overrightarrow{s(O)s(N)} \\ &= \frac{1}{2}\big(d(s(O), s(M))^2 + d(s(O), s(N))^2 - d(s(M), s(N))^2\big) \\ &= \frac{1}{2}\big(d(O,M)^2 + d(O,N)^2 - d(M,N)^2\big) = \overrightarrow{OM} \cdot \overrightarrow{ON}. \end{aligned}$$

This shows that σ satisfies (10.4). We now expand by bilinearity the inner square

$$\big(\sigma(\lambda\overrightarrow{V} + \mu\overrightarrow{W}) - \lambda\sigma(\overrightarrow{V}) - \mu\sigma(\overrightarrow{W})\big)^2.$$

By using (10.4), one then sees that the result is

$$\big((\lambda\overrightarrow{V}+\mu\overrightarrow{W})-\lambda\overrightarrow{V}-\mu\overrightarrow{W}\big)^2=0,$$

hence $\sigma(\lambda\overrightarrow{V}+\mu\overrightarrow{W})-\lambda\sigma(\overrightarrow{V})-\mu\sigma(\overrightarrow{W})=\overrightarrow{0}$ and σ is linear. By taking differences, the formula $\sigma(\overrightarrow{OM})=\overrightarrow{s(O)s(M)}$ yields $\sigma(\overrightarrow{MN})=\overrightarrow{s(M)s(N)}$ for any two points M, N, thus s is an affine linear map associated with the linear map σ, and the above theorem is proved. □

In the same vein, one can prove the "theorem of isometric figures", which provides a rigorous mathematical justification to all definitions and physical considerations presented in section 3.2.

Theorem. *Let $(A_1A_2A_3A_4\,\ldots)$ and $(A'_1A'_2A'_3A'_4\,\ldots)$ be two isometric figures formed by points A_i, A'_i of a Euclidean space $\mathcal{E}$. Then there exists an isometry s of the entire Euclidean space $\mathcal{E}$ such that $A'_i=s(A_i)$ for all i.*

Proof. Consider the vector subspaces $\overrightarrow{\mathcal{V}}$ and $\overrightarrow{\mathcal{V}'}$ of $\overrightarrow{\mathcal{E}}$ generated by the linear combinations of vectors $\overrightarrow{A_iA_j}$ (resp. $\overrightarrow{A'_iA'_j}$). Since

$$\overrightarrow{A_iA_j}=\overrightarrow{A_1A_j}-\overrightarrow{A_1A_i},$$

it suffices to take the linear combinations of all vectors $\overrightarrow{V_i}=\overrightarrow{A_1A_i}$ (resp. $\overrightarrow{V'_i}=\overrightarrow{A'_1A'_i}$). By replacing O with A_1, M with A_i and N with A_j, formula (10.5) yields

$$\overrightarrow{V_i}\cdot\overrightarrow{V_j}=\overrightarrow{A_1A_i}\cdot\overrightarrow{A_1A_j}=\frac{1}{2}\big((A_1A_i)^2+(A_1A_j)^2-(A_iA_j)^2\big).$$

This implies that $\overrightarrow{V'_i}\cdot\overrightarrow{V'_j}=\overrightarrow{V_i}\cdot\overrightarrow{V_j}$, therefore the vector transformation

$$\sigma_0:\sum\lambda_i\overrightarrow{V_i}\mapsto\sum\lambda_i\overrightarrow{V'_i}$$

defines an orthogonal linear map from $\overrightarrow{\mathcal{V}}$ onto $\overrightarrow{\mathcal{V}'}$ (σ_0 is well defined; indeed, if a vector admits several representations $\sum\lambda_i\overrightarrow{V_i}$, the corresponding images $\sum\lambda_i\overrightarrow{V'_i}$ are the same, as one can see by observing that the inner square of the difference vanishes). In particular, the spaces $\overrightarrow{\mathcal{V}}$ and $\overrightarrow{\mathcal{V}'}$ have the same dimension, and one can extend σ_0 into an isometry σ from $\overrightarrow{\mathcal{E}}$ onto $\overrightarrow{\mathcal{E}}$ by taking the orthogonal direct sum $\sigma=\sigma_0\oplus\tau$ with an arbitrary orthogonal transformation τ from $\overrightarrow{\mathcal{V}}^{\perp}$ onto $\overrightarrow{\mathcal{V}'}^{\perp}$ (the latter exists since $\overrightarrow{\mathcal{V}}^{\perp}$ and $\overrightarrow{\mathcal{V}'}^{\perp}$ have the same dimension $=\dim\overrightarrow{\mathcal{E}}-\dim\overrightarrow{\mathcal{V}}$). We now define an affine isometry s by considering the unique affine linear map $s:\mathcal{E}\to\mathcal{E}$ that sends A_1 to A'_1, and is associated with the linear orthogonal transformation σ. As $s(A_1)=A'_1$ and $\sigma(\overrightarrow{V_i})=\overrightarrow{V'_i}$, that is, $\sigma(\overrightarrow{A_1A_i})=\overrightarrow{A'_1A'_i}$, we get $s(A_i)=A'_i$ for all i, as expected.
□

To go a little further, we need to know the precise structure of orthogonal transformations: a first easy characterization is that a vector transformation given by a matrix

A in an orthonormal basis is orthogonal if and only if the matrix A satisfies $A^t A = \mathrm{Id}$, or equivalently, if and only if the column vectors form an orthonormal basis. In dimension 2, one obtains matrices of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \qquad \begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \qquad a^2+b^2=1,$$

which can be also written

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \qquad \begin{pmatrix} \sin\alpha & \cos\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix},$$

and correspond respectively to a vector rotation of angle α and to a reflection through the line $\mathcal{D}_{\alpha/2}$ of polar angle $\alpha/2$ with respect to the Ox axis. In the latter case, the choice of a suitable new orthonormal basis leads to the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In general, in any dimension, one shows that for any orthogonal vector transformation, there exists an orthonormal basis wherein the matrix is of the form

$$A = \begin{pmatrix}
\cos\alpha_1 & -\sin\alpha_1 & 0 & 0 & & & \dots & & & & 0 \\
\sin\alpha_1 & \cos\alpha_1 & 0 & 0 & & & \dots & & & & 0 \\
0 & 0 & \cos\alpha_2 & -\sin\alpha_2 & & & \dots & & & & 0 \\
0 & 0 & \sin\alpha_2 & \cos\alpha_2 & & & \dots & & & & 0 \\
\vdots & & & 0 & \ddots & & \dots & & & & \vdots \\
0 & \dots & & 0 & & \cos\alpha_k & -\sin\alpha_k & 0 & \dots & & 0 \\
0 & \dots & & 0 & & \sin\alpha_k & \cos\alpha_k & 0 & \dots & & 0 \\
0 & \dots & & & & & 0 & 1 & 0 & \dots & 0 \\
0 & \dots & & \vdots & & & 0 & 0 & 1 & \dots & 0 \\
0 & \dots & & & & & & & & \ddots & 0 \\
0 & \dots & & 0 & & & 0 & & & \dots & \varepsilon
\end{pmatrix},$$

with suitable angles α_i and $\varepsilon = \pm 1$.

The proof is obtained by induction on the dimension, observing that either there exists a real eigenvalue (necessarily equal to ± 1) or there exists a pair of conjugate nonn-real complex eigenvalues, corresponding to a stable real plane in which the transformation operates as a rotation of angle α_i ; one relies on the fact that the orthogonal of a stable subspace by an orthogonal transformation is stable, and that the -1 eigenvalues can be grouped by 2×2 blocks corresponding to rotations of angle $\alpha_i = \pi$. This leaves the possibility of at most one isolated -1 eigenvalue (which appears here in the last line of the matrix in the form of the eigenvalue $\varepsilon = \pm 1$ (but this line may be absent when the dimension is even)[15].

---

(15) In dimension 3, the proof is easily reduced to the case of dimension 2, thanks to the fact that the characteristic polynomial has degree 3 and therefore possesses necessarily a real eigenvalue. This result was part of the math curriculum of grade 12 ("Terminale C") if the French science classes in the years 1970-1985 ; the 2-dimensional case was introduced even before, in grade 10 or 11. Unfortunately, this is a good measure of the incredible and appalling degradation of the level of education in France – it is nowadays harder and harder to give a complete proof of this result with students specializing in math/physics at the second year university level ...

Let us choose a parameter $t \in [0,1]$ (thought of as a time parameter), and define a matrix A_t as above, with angles α_i replaced by $t\alpha_i$ and the parameter $\varepsilon = \pm 1$ kept constant. We obtain a continuous variation of matrices such that $A_1 = A$ and either $A_0 = \mathrm{Id}$ or $A_0 =$ matrix of the orthogonal reflection through the hyperplane $x_n = 0$. This shows that for any isometric transformation $s : X \mapsto Y = AX + B$, there is a continuous time dependent variation $s_t : X \mapsto A_t X + tB$ such that $s_1 = s$, while s_0 is either the identity or the reflection through the affine hyperplane $x_n = 0$. Therefore, given two isometric figures $\mathcal{F}$ and $\mathcal{F}' = s(\mathcal{F})$, $\mathcal{F}'$ can be deduced from $\mathcal{F}$ by a continuous motion $\mathcal{F}_t = s_t(\mathcal{F})$ such that $\mathcal{F}_1 = s_1(\mathcal{F}) = s(\mathcal{F}) = \mathcal{F}'$, where $\mathcal{F}_0 = s_0(\mathcal{F})$ coincides with either $\mathcal{F}$ or with the figure obtained by a mirror symmetry through a hyperplane.

Notice that with the above notation we have $\det(A) = \varepsilon = \pm 1$. One says that the corresponding isometry s is a *positive isometry* if $\varepsilon = 1$ and a *negative isometry* if $\varepsilon = -1$. In the case of a continuous matrix function $t \mapsto A(t)$, the determinant $\det(A(t))$ is also a continuous function. Therefore it cannot "jump" from value 1 to value -1, in other words, the determinant is constant when the time t varies. The motion of a solid body $t \mapsto s_t(\mathcal{F})$ from an initial position (corresponding to $s_0 = \mathrm{Id}$, thus to $\det A(0) = 1$) is thus realized only through positive isometries. Conversely, the above discussion shows that any positive isometry can indeed be achieved by a continuous displacement in the sense of definition (3.2.5) : mathematically, there is a perfect coincidence between the concepts of displacement and of positive isometry.

11. Non euclidean geometries

The approach to geometry based on investigating properties of metric structures is not a mere curiosity that is specific to Euclidean geometry. In fact, non-Euclidean geometries find a very natural place within the general framework of Riemannian geometry, so named in reference to Bernhard Riemann, one of the main founders of complex analysis and modern differential geometry.

Bernhard Riemann (1826–1866)

A *Riemannian manifold* is by definition a differential manifold M, that is, a space M which possesses local coordinate systems $x = (x_1\,; x_2\,; \ldots\,; x_n)$ of real variables, equipped with an infinitesimal metric g of the form

$$ds^2 = g(x) = \sum_{1 \leqslant i,j \leqslant n} a_{ij}(x)\, dx_i dx_j.$$

This metric represents the square of the length of a small displacement $dx = (dx_i)$ on the manifold. We assume here that $(a_{ij}(x))$ is a positive definite symmetric matrix and that the coefficients $a_{ij}(x)$ are infinitely differentiable functions of x. Given a path $\gamma : [\alpha, \beta] \to M$ that is continuous and piecewise differentiable, the length of γ is computed by putting

$$x = \gamma(t) = (\gamma_1(t)\,;\, \ldots\,;\, \gamma_n(t)), \qquad dx = \gamma'(t)\, dt.$$

This gives

$$ds = \|dx\|_{g(\gamma(t))} = \sqrt{\sum_{1 \leqslant i,j \leqslant n} a_{ij}(\gamma(t))\, \gamma_i'(t)\gamma_j'(t)}\, dt, \tag{11.1$'$}$$

$$\mathrm{long}(\gamma) = \int_\alpha^\beta ds = \int_\alpha^\beta \sqrt{\sum_{1 \leqslant i,j \leqslant n} a_{ij}(\gamma(t))\, \gamma_i'(t)\gamma_j'(t)}\, dt. \tag{11.1$''$}$$

Given two points a, $b \in M$, the "geodesic distance" $d_g(a,b)$ of a to b is defined by $d_g(a,b) = \inf_\gamma \mathrm{long}(\gamma)$ where the inf is extended to all paths $\gamma : [\alpha, \beta] \to M$ of extremities $\gamma(\alpha) = a$ and $\gamma(\beta) = b$. Euclidean geometry as described in the previous sections corresponds to the particular case of a constant Riemannian metric (also referred to as a "flat metric" – by this, one means that the curvature is zero) :

$$ds^2 = dx_1^2 + \ldots + dx_n^2. \tag{11.2$'$}$$

In this case, one gets the simpler expression

$$\mathrm{long}(\gamma) = \int_\alpha^\beta ds = \int_\alpha^\beta \sqrt{\sum_{1 \leqslant i \leqslant n} \gamma_i'(t)^2}\, dt. \tag{11.2$''$}$$

Einstein's theory of special relativity relies on a metric of a somewhat different kind, called a Lorentzian metric, which is a metric of signature $(3,1)$ with respect to spacetime

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 - c^2\, dt^2 \tag{11.3}$$

where c denotes the speed of light. The theory of general relativity corresponds to the case of Lorentzian metrics with variable coefficients, and its study leads to curvature phenomena describing the effects of gravitation in a geometric framework.

Coming back to geometry in dimension 2, we recall briefly the description of the non-Euclidean geometry of Lobachevski, through the model of the "Poincaré disk" equipped with the invariant hyperbolic metric.

Nicolai Lobatchevski (1793–1856)

Henri Poincaré (1854–1912)

Definition. *On the unit disk* $\mathbb{D} = \{z \in \mathbb{C}\,;\, |z| < 1\}$ *of the complex plane, denoting* $z = x + iy$*, one considers the "Poincaré metric"*

$$(11.4) \qquad ds = \frac{|dz|}{1 - |z|^2} \quad \Leftrightarrow \quad ds^2 = \frac{dx^2 + dy^2}{(1 - (x^2 + y^2))^2}.$$

A calculation shows that this metric is kept invariant by the group $\mathrm{Aut}(\mathbb{D})$ *of holomorphic automorphisms of* $\mathbb{D}$*, where* $\mathrm{Aut}(\mathbb{D})$ *consists of homographic transformations*

$$h(z) = \lambda \frac{z - a}{1 - \overline{a}z}, \qquad |\lambda| = 1, \quad a \in \mathbb{D}$$

that preserve $\mathbb{D}$*. With respect to the above metric, the length of a continous piecewise differentiable path* $\gamma : [\alpha, \beta] \to \mathbb{D}$ *is computed by putting* $z = \gamma(t)$*,* $dz = \gamma'(t)\,dt$*, whence*

$$\mathrm{long}(\gamma) = \int_\alpha^\beta ds = \int_\alpha^\beta \frac{|\gamma'(t)|}{1 - |\gamma(t)|^2}\,dt.$$

According to the general definition, the Poincaré distance $d_{\mathrm{P}}(a, b)$ from a to b is the geodesic distance defined by $d_{\mathrm{P}}(a, b) = \inf_\gamma \mathrm{long}(\gamma)$ where the inf is extended to all paths $\gamma : [\alpha, \beta] \to \mathbb{D}$ of extremitiess $\gamma(\alpha) = a$ and $\gamma(\beta) = b$. Since the Poincaré metric ds is invariant under $\mathrm{Aut}(\mathbb{D})$, we immediately infer that the distance d_{P} is also invariant, i.e. $d_{\mathrm{P}}(h(a), h(b)) = d_{\mathrm{P}}(a, b)$ for all $h \in \mathrm{Aut}(\mathbb{D})$ and all points $a, b \in \mathbb{D}$.

We claim that the length minimzing path joining the center 0 of the disk to any point $w \in \mathbb{D}$ is the segment $[0, w]$. In fact, if we write $z = \gamma(t) = r(t)\,e^{i\theta(t)}$ in polar coordinates, we have

$$dz = (dr + ir\,d\theta)\,e^{i\theta}, \qquad |dz| = \sqrt{dr^2 + r^2 d\theta^2} \geqslant dr = r'(t)dt.$$

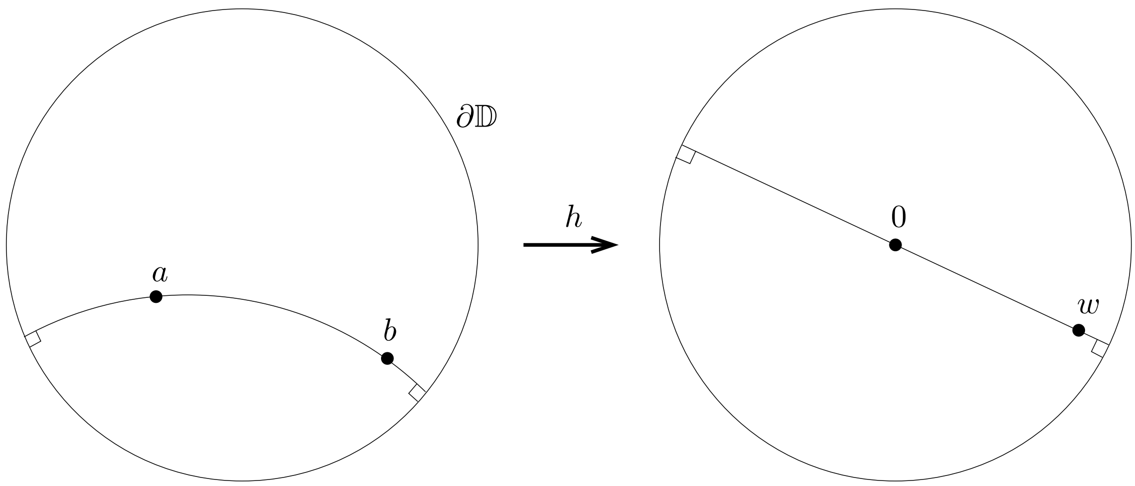
Therefore

$$\mathrm{long}(\gamma) = \int_\gamma \frac{|dz|}{1 - |z|^2} \geqslant \int_0^{|w|} \frac{dr}{1 - r^2} = \frac{1}{2} \ln \frac{1 + |w|}{1 - |w|},$$

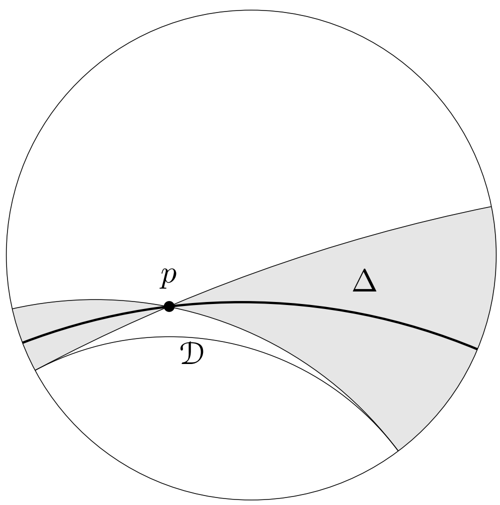
and the right hand side is precisely the hyperbolic length of segment $[0,w]$. To get the distance of two arbitrary points a, b, one can apply the automorphism $h(z) = \frac{z-a}{1-\overline{a}z}$ that maps a to $h(a) = 0$ and b to $w = h(b) = \frac{b-a}{1-\overline{a}b}$. In this way, we find

$$d_{\mathrm{P}}(a,b) = d_{\mathrm{P}}(h(a),h(b)) = d_{\mathrm{P}}(0,w) = \frac{1}{2}\ln\frac{1+\frac{|b-a|}{|1-\overline{a}b|}}{1-\frac{|b-a|}{|1-\overline{a}b|}}. \tag{11.5}$$

This distance "governs" hyperbolic geometry just in the same way Pythagoras' theorem governs Euclidean geometry.



The geodesic (length minimizing path) joining a to b is the inverse image by h of segment $[0,w]$. This is an arc of extremities a and b, which is part of the circle orthogonal to the boundary $\partial\mathbb{D}$ (this is so because the diameter $\mathbb{R}w$ is orthogonal to the boundary $\partial\mathbb{D}$, so their inverse images by h are also orthogonal, thanks to the fact that h is a conformal transform preserving $\partial\mathbb{D}$). The only exception to this is the case where 0, a, b are aligned, in which case the geodesic is the line segment $[a,b]$ of extremities a, b, included in a diameter of $\mathbb{D}$. The "hyperbolic lines" of $\mathbb{D}$ are therefore diameters and arcs of circles that are orthogonal to the boundary $\partial\mathbb{D}$.



In hyperbolic geometry, it is easy to see that Euclid's incidence axioms are satisfied, ex cept precisely the $5^{\rm th}$ postulate (according to which there is a single parallel to a given line passing through a given point) : here, for any point p outside a given "line" $\mathcal{D}$, there are uncountably many lines Δ that do not intersect the initially given line. This is the surprising discovery made by Lobachevski in 1826, thereby destroying the millenary hope to derive the $5^{\rm th}$ postulate from the other axioms.

12. A few important ideas due to Felix Hausdorff

There are many other circumstances where metric structures play a crucial role. We describe here a few important ideas due to Felix Hausdorff (1868 - 1942), one of the founders of modern topology [Hau].

Felix Hausdorff (1868–1942)

The first one is that the Lebesgue measure of $\mathbb{R}^n$ can be generalized without any reference to the vector space structure, but just by using the metric. If $(\mathcal{E}, d)$ is an arbitrary metric space, one defines the p-dimensional Hausdorff measure of a subset A of $\mathcal{E}$ as

$$(12.1)\qquad \mathcal{H}_p(A) = \lim_{\varepsilon\to 0} \mathcal{H}_{p,\varepsilon}(A), \qquad \mathcal{H}_{p,\varepsilon}(A) = \inf_{\operatorname{diam} A_i \leqslant \varepsilon} \sum_i (\operatorname{diam} A_i)^p$$

where $\mathcal{H}_{p,\varepsilon}(A)$ is the least upper bound of sums $\sum_i (\operatorname{diam} A_i)^p$ running over all countable partitions $A = \bigcup A_i$ with $\dim A_i \leqslant \varepsilon$. In $\mathcal{E} = \mathbb{R}^n$, for $p = 1$ (resp. $p = 2$, $p = 3$), one recovers the usual concepts of length, area, volume, and the definition even works when p is not an integer ; it is then extremely useful to define the dimension of fractal sets. Definition (12.1) works equally well in an arbitrary metric space, for instance in any Riemannian manifold.

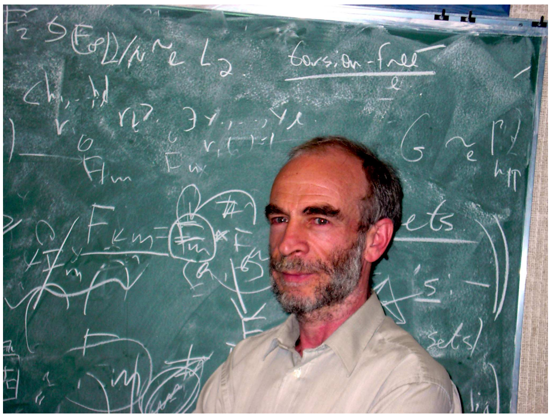
Another important idea of Hausdorff is the existence of a natural metric structure on the set of compact subsets of a given metric space $(\mathcal{E}, d)$. If K, L are two compact subsets of $\mathcal{E}$, the *Hausdorff distance* of K and L is defined to be

$$(12.2)\qquad d_{\rm H}(K, L) = \max\Big\{ \max_{x\in K} \min_{y\in L} d(x,y), \max_{y\in L} \min_{x\in K} d(x,y) \Big\}.$$

One can check that $d_{\rm H}$ is indeed a distance function ; in this way the set $\mathcal{K}(\mathcal{E})$ of compact subsets of $\mathcal{E}$ is provided with the structure of a metric space. When $(\mathcal{E}, d)$ is compact, one can show that $(\mathcal{K}(E), d_{\rm H})$ is itself compact.

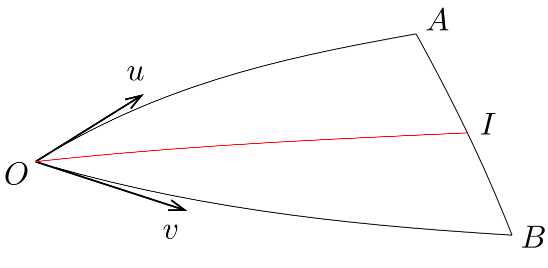
13. On the work of Mikhail Gromov

The study of metric structures has still been an extremely active research subject in the recent period ; one can cite in particular the work of Mikhail Gromov on length spaces and the related definition of "moduli spaces" of Riemannian manifolds and their compactifications.



Mikhail Gromov (1943–), Abel prize 2009

A length space is by definition a metric space $(\mathcal{E}, d)$ such that for all points A, B of $\mathcal{E}$ there exists a "midpoint" I such that $d(A, I) = d(I, B) = \frac{1}{2} d(A, B)$. If the space $\mathcal{E}$ is complete, one can then build by successive dichotomies a path γ with end points A, B such that $d(A, \gamma(t)) = t\, d(A, B)$ and $d(\gamma(t), B) = (1 - t)\, d(A, B)$ for all $t \in [0, 1]$, which can be seen as a geodesic joining A and B. This is a fruitful generalization of Riemannian manifolds – and thus in particular of Euclidean and non-Euclidean geometries. A remarkable fact is that one can define for example the curvature tensor of a Riemannian manifold (M, g) by using only the properties of the infinitesimal distance of the associated length space : let O be a point chosen as the origin, $A = \exp_O(\varepsilon u)$ and $B = \exp_O(\varepsilon v)$ where u, v are tangent vectors to M and $\varepsilon > 0$ is a small real number. Finally, let I be the midpoint of A, B with respect to the geodesic distance.



Then infinitesimally small geodesic triangles OAB satisfy in the limit

$$\lim_{\varepsilon \to 0} \frac{OA^2 + OB^2 - 2OI^2 - 2AI^2}{OA^2\, OB^2} = -\frac{1}{6} \frac{\langle R(u, v)u, v\rangle_g}{\|u\|_g^2\, \|v\|_g^2}, \tag{13.1}$$

$$\lim_{\varepsilon \to 0} \frac{OA^2 + OB^2 - 2OI^2 - 2AI^2}{OA^2\, OB^2 - (OI^2 - AI^2)^2} = -\frac{1}{6} \frac{\langle R(u, v)u, v\rangle_g}{\|u \wedge v\|_g^2}, \tag{13.2}$$

where R is the Riemannian curvature tensor. The numerator $OA^2+OB^2-2OI^2-2AI^2$ vanishes for Euclidean geometry (median theorem !), and the above formulas tell us that the deviation with respect to the Euclidean situation is essentially described by the sectional curvature.
Proof : this is left as a very interesting – and non trivial – exercise to readers !

If X and Y are two compact metric spaces, one defines their *Gromov-Hausdorff distance* $d_{\mathrm{GH}}(X,Y)$ to be the infimum of all Hausdorff distances $d_{\mathrm{H}}(f(X),f(Y))$ for all possible isometric embeddings $f:X\to\mathcal{E}$, $g:Y\to\mathcal{E}$ of X and Y in another compact metric space $\mathcal{E}$. This provides a crucial tool to study deformations and degenerations of Riemannian manifolds.

References

[Art] Emil Artin, *Geometric Algebra*, Interscience publishers Inc., New York, 1957, download at `http://ia700300.us.archive.org/24/items/geometricalgebra033556mbp/geometricalgebra033556mbp.pdf`

[Des] René Descartes, *La Géométrie,* appendix of *Discours de la méthode*, available from Project Gutenberg `http://www.gutenberg.org/files/26400/26400-pdf.pdf`

[Die] Jean Dieudonné, *Algèbre linéaire et géométrie élémentaire*, Hermann, 1968, 240 pages

[Euc] Euclid, *Euclid's elements of geometry*, English translation from Greek by Richard Fitzpatrick, `http://farside.ph.utexas.edu/euclid/Elements.pdf`

[Gro] Mikhail Gromov, *Structures métriques pour les variétés riemanniennes*, edited by J. Lafontaine and P. Pansu, Textes Mathématiques, no. 1, CEDIC, Paris, 1981, translated by S.M. Bates, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhauser, 1999

[Hau] Felix Hausdorff, *Dimension und äusseres Mass*, Mathematische Annalen 79, 1-2 (1919) 157-179, and *Grundzüge der Mengenlehre*, Leipzig, 1914. viii + 476 pages, English translation *Set Theory*, American Mathematical Soc., 1957, 352 pages

[Hil] David Hilbert, *Grundlagen der Geometrie*, Teubner, Leipzig, 1899, English translation *The Foundations of Geometry*, by E.J. Townsend (2005), available from project Gutenberg, `http://www.gutenberg.org/files/17384/17384-pdf.pdf`

[Lob] Nikolai Lobachevski, *Geometrical researches on the theory of parallels*, `http://quod.lib.umich.edu/u/umhistmath/AAN2339.0001.001/1?rgn=full+text;view=pdf`

[Pas] Moritz Pasch, *Vorlesungen über neuere Geometrie*, Teubner, Leipzig, 1882, download at `https://ia700308.us.archive.org/32/items/vorlesungenbern00pascgoog/vorlesungenbern00pascgoog.pdf`

[Poi] Henri Poincaré, *ÅŠuvres*, tome 2, 1916 - Fonctions fuchsiennes, et *Les géométries non euclidiennes* `http://www.univ-nancy2.fr/poincare/bhp/`

[Pon] Jean-Victor Poncelet, *Traité des propriétés projectives des figures*, Gauthier-Villars, Paris, 1866, download at `http://gallica.bnf.fr/ark:/12148/bpt6k5484980j`

[Rie] Bernhard Riemann, *Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen complexen Grösse*, Inauguraldissertation, Göttingen, 1851, and *Ueber die Hypothesen, welche der Geometrie zu Grunde liegen*, Habilitationsschrift, 1854, Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen, 13 (1868), English translation *On the Hypotheses which lie at the Bases of Geometry*, `http://www.emis.de/classics/Riemann/`