Special Relativity and Geometry

C. E. Harle Ricardo Bianconi

December 6, 1999

Foreword

This is a preliminary version of the book. There will be a chapter on the four dimensional model, but the main ideas are already in the two and three dimensional model. Some notes to the chapters will come later. There are still many typos and the index is not in final form.

São Paulo, November 30, 1999.

Carlos Edgar Harle
Ricardo Bianconi (bianconi@ime.usp.br, http://www.ime.usp.br/~bianconi)

Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 66281
05315-970 São Paulo, SP, Brazil.

Contents

List of Figures

Chapter 1

Introduction

Usually special relativity is introduced via the use of certain frames of reference, called inertial frames, in which one supposes that processes of measure of time and length are defined (rulers and clocks). With these processes one introduces coordinates on each of those frames and the coordinates of the various frames are related in such a way that the speed of light remains independent of them.

Such presentation of the subject is not free from conceptual difficulties though. For instance, lengths of the same object may be distinct when measured in distinct frames. The same happens with measure of time.

This work presents a completely different approach to the subject. Firstly, we do not use coordinate systems to introduce the theory and secondly we strongly stress its geometrical aspect (even its graphical aspect).

Chapter 2

The 2-dimensional model

2.1 The Space-Time

Let us start this section considering the following simple example from classical kinematics to motivate the definition of space-time.

We know that the movement of a material point on a straight line when we know the position of this point at each instant t (measured in some clock). We also know that such position is represented by a point in that line. Taking a second straight line coplanar to the first one and marking the time instants in it we obtain a geometrical representation of the movement by a plane curve. Thus each point of this curve represents the position of the material point at some given instant of time. So, each point of the curve embodies two bits of information, namely the position of the moving point and the instant of time in which such position is occupied. We indicate this fact calling events the points of the curve.

In order to formulate the relativistic version of this situation we begin with a plane whose points are called *events*, without any notion of distance as yet. We stress this fact by calling such a plane an *affine plane* or 2-dimensional space-time.

2.2 Light

In order to construct the model we make use of communication between events through light signals.

In the 2-dimensional space-time we arbitrarily choose two distinct classes of parallel straight lines and we orient them so that two lines of the same class have the same direction (see Figure 2.1).

Once this is stablished we can describe the processes of communication among events. We say that two distinct events P and Q can communicate if both belong to the same light ray. If the

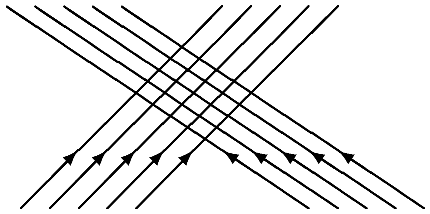

Figure 2.1: Light rays.

vector $\mathbf{PQ}$ has the same direction of this ray we say that in the event P one can emit a light signal to Q and in the event Q one can recieve a light signal emitted in P. (See Figure 2.2.)

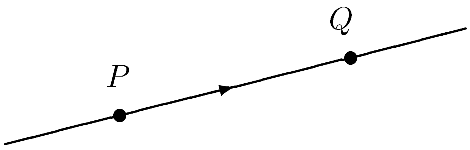

Figure 2.2: P emits a light ray to Q.

It is easy to see that in any event two distinct light signals can be emitted and two distinct light signals can be recieved. (See Figure 2.3.) Notice also that if Q can recieve a light signal from P then P cannot recieve a light signal from Q.

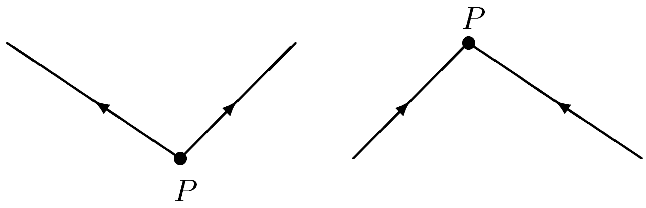

Figure 2.3: P emits/recieves two light rays.

The non null vectors of the affine plane which are parallel to a light ray and having the same direction are called *lightlike vectors*.

2.3 Time

One of the remarkable features of relativity is that the notion of time can be introduced via processes of emission and reception of light signals.

Let P and Q be two events which do not communicate with each other. We say that the pair PQ is a *time interval* if there is a third event S which can recieve a light signal from P and emit another (distinct) light signal to Q. (See Figure 2.4.) In this case we say that P is the beginning and Q is the end of this interval.

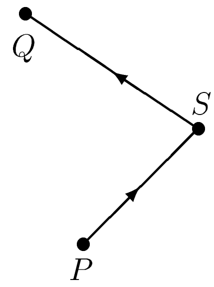

Figure 2.4: The time interval PQ.

We also say that an event Q is posterior to P (or P is anterior to Q) if PQ is a time interval.

Notice that the definition excludes events which are in the same light ray. This is usually related to the fact that at the speed of light any clock fails to mark the time.

Notice also that if A and B are two distinct events which cannot communicate with each other and C is a third event which can recieve light signals from both A and B, then neither of the pairs AB or BA can be a time interval. (See Figure 2.5.) Therefore between these events there is no chronological relationship, that is neither is anterior to the other.

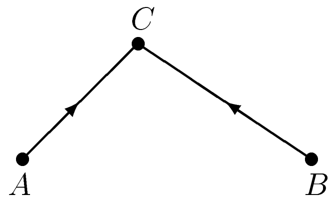

Figure 2.5: Neither AB nor BA is a time interval.

A vector $\mathbf{PQ}$ is called a *timelike vector*, where PQ is a time interval.

2.4 Observers

According to the Principle of Inertia from Classical Mechanics the movement of a free material point (that is, without any force applied to it) is in a straight line with constant speed. As we have mentioned in section 2.1, this movement can be represented by a straight line in a plane.

In Special Relativity we also formulate the *Principle of Inertia* as follows. Movements of free material points are represented in the space-time as straight lines. These cannot be arbitrarily chosen because we impose that between any two events of this line there must be a relation of chronological precedence, that is, one of the events must be anterior to the other. Such lines are called timelike. We can easily see that these lines are characterized by the fact that they must contain a timelike interval. (See Figure 2.6.)

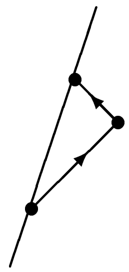

Figure 2.6: Timelike line.

The timelike lines are also called *observers* because of the following property:

Let r be an observer and X an event not in r. Then in X one can emit a light signal which is recieved by some event in r. This means that any event outside r can be observed in r. (See Figure 2.7.)

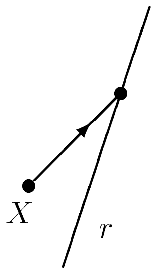

Figure 2.7: Observer.

The fact that a light signal can reach any observer is usually mentioned by saying that the speed of the observer is lesser than that of the light.

Given an event X outside an observer r, two events X' and X'' in r are determined by the fact that X' can emit a light signal to X and X'' can recieve a light signal from X. We say that $X'X''$ is the time interval of r determined by X. (See Figure 2.8.)

Two events X and Y outside an observer r are *symmetric with respect to* r if the time intervals determined by both X and Y are the same. (See Figure 2.9.)

Let r and r' be two distinct observers and X an event in r not in r'. The echo of the event X produced by r' is defined as the event Y in r which can recieve a light signal from the event W in r' such that W can recieve a light signal from X. (See Figure 2.10.)

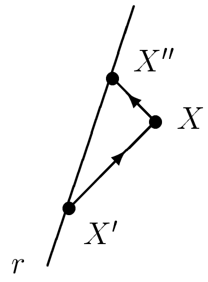

Figure 2.8: Definition of X' and X''.

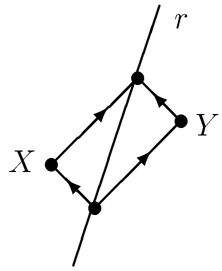

Figure 2.9: Symmetric events X and Y.

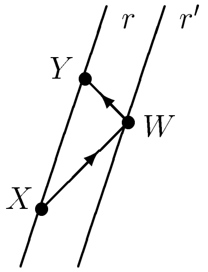

Figure 2.10: Echo of X produced by r'.

Given two distinct observers r and r' we say that an observer h is *equidistant* from r and r' if for each event X in h the echoes of X by r and r' coincide. (See Figure 2.11.)

Given an event R in an observer r and two events X and Y not in r and symmetrical with respect to r, we have that r is the observer equidistant from the observers determined by the time intervals RX and RY.

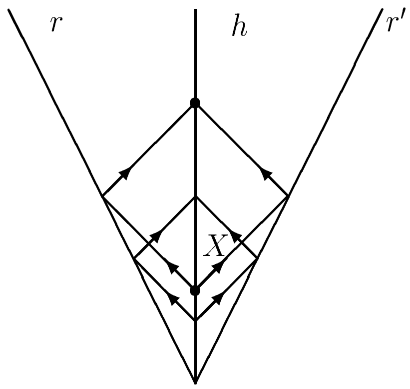

Figure 2.11: Observer h is equidistant from r and r'.

2.5 Rest

In the 2-dimensional space-time we say that two observers are *at rest with respect to each other* if they are the same one or if they are distinct they never meet. So the concept of rest coincides with the concept of parallel lines from affine geometry.

Two observers at rest with respect to each other are represented by two parallel timelike straight lines. (See Figure 2.12.)

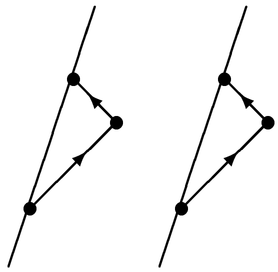

Figure 2.12: Two observers at rest.

We use this concept to define a criterion to compare time intervals in the same observer.

Let PQ and RS two time intervals in the same observer r, where we assume that P is anterior to R. We say that these intervals have the same duration if there exists an observer s at rest with respect to r such that the events R and S are the echos of P and Q, respectively, produced by s. In the case where P and Q coincide we say that the intervals have the same duration if R and S also coincide. (See Figure 2.13.) Notice that two time intervals in the same observer have the same duration if, and only if, the segments which represent them are equipolents in the sense of affine geometry.

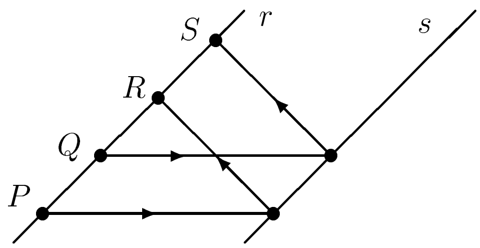

Figure 2.13: Comparing time intervals in the same observer.

Related to the notion of rest we have the notion of simultaneous events, which can be introduced as follows.

Let h be an observer, M an event in h and X an event outside h. Let $X'X''$ be the time interval determined by X in h. We say that X is *simultaneous* to M with respect to h if M is anterior to X'' and posterior to X' and the time intervals $X'M$ and MX'' have the same duration. In terms of geometry we can say that M is the middle point of X' and X''. (See Figure 2.14.)

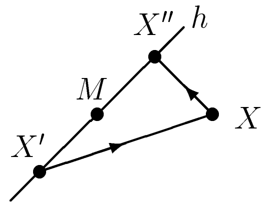

Figure 2.14: The event X is simultaneous to M with respect to h.

Let A and B be two events simultaneous to M in h and let h_A be the observer containing A and at rest with respect to h. Then we can see that A is simultaneous to B with respect to h_A. (See Figure 2.15.)

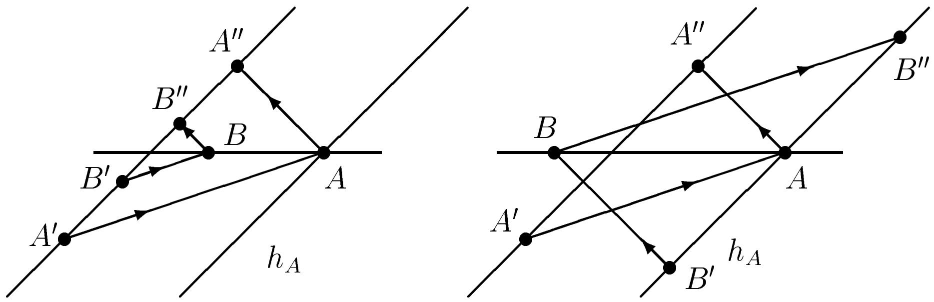

Figure 2.15: A is simultaneous to B with respect to h_A.

2.6 Comparing time intervals

In section 2.5 we treated the comparison of intervals of time in the same observer. Here we study this problem in the case where the time intervals are in distinct observers.

In order to make the exposition more suggestive, let us introduce some auxiliary concepts.

Let XY be a time interval and h an observer. We now define another time interval which we call the *observation* of XY in h.

We distinguish three cases. If neither X nor Y are in h, we define the observation as the interval $X''Y''$ (X'' and Y'' as defined in section 2.4). If X is in h and Y not in h, the observation is XY''. If X is not in h and Y is in h, the observation is $X''Y$. Finally, if both X and Y are in h, the observation is XY. (See Figure 2.16.)

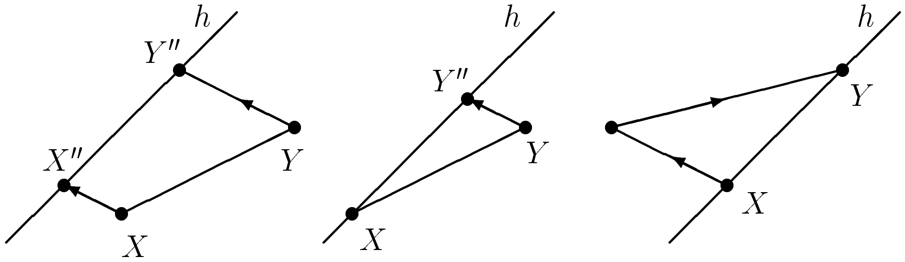

Figure 2.16: Observation of a time interval.

Now we can formulate the principle of comparison of time intervals.

We say that two time intervals AA' and BB' in the distinct observers a and b, respectively, have the same duration if their observations in the observer equidistant from a and b (see section 2.4) have the same duration. (See Figure 2.17.)

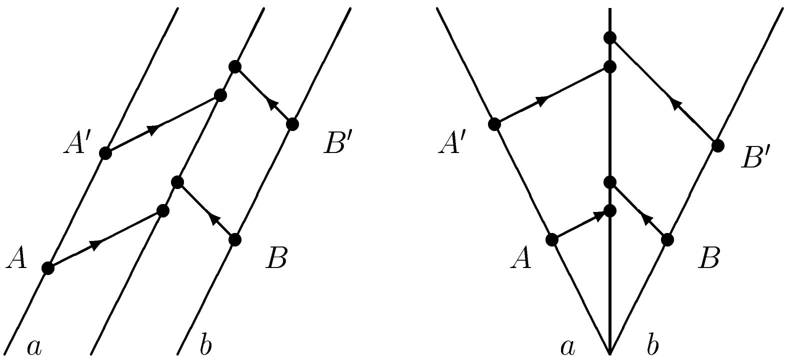

Figure 2.17: Comparing time intervals.

It is easy to see that if PX and PY are time intervals and h is an observer through P such that

X and Y are symmetric with respect to h then those time intervals have the same duration. (See Figure 2.18.)

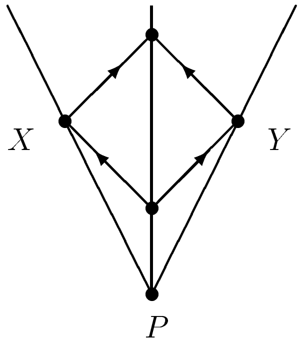

Figure 2.18: The time intervals PX and PY have the same duration.

The most important property of this notion is that it defines a coherent criterion for the comparison of time intervals, that is, if two time intervals have the same duration as a third interval then the first two intervals also have the same duration.

Starting with a time interval PQ and considering the several observers through P we can construct a bunch of time intervals PX wiht the same duration as PQ. We observe that in section 2.14 we will show that the set of points X thus obtained lies in a hyperbola. (See Figure 2.19.)

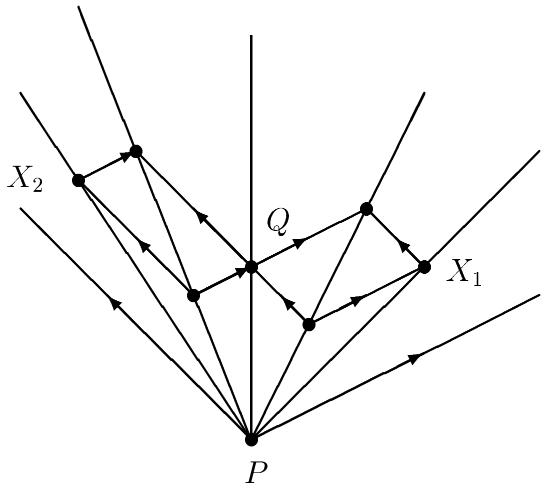

Figure 2.19: Constructing time intervals PX_1 and PX_2 with the same duration as PQ.

2.7 Clocks

Informally, a clock can be understood as any device that can register a series of events in a given observer which constitute time intervals of same duration.

More precisely, the *clock of an observer* is characterized by a sequence of events T_n (for n an integer, positive or negative or zero), such that all the time intervals T_nT_{n+1} have the same duration.

Once such sequence is determined by the events T_0 and T_1 in the observer h, we indicate the clock by the triplet (h, T_0, T_1).

The integer n associate to the event T_n is defined as the time of this event. By definition, an event X posterior or equal to T_n and anterior to T_{n+1} has time n. (See Figure 2.20.)

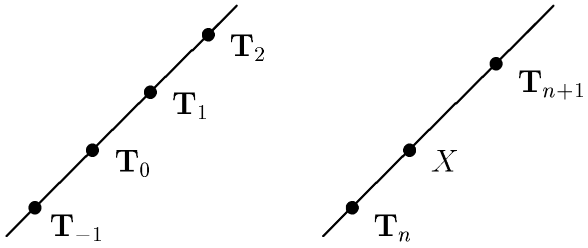

Figure 2.20: Definition of clock and time of X.

These clocks are too rough because they indicate only integer times. So there is the need to refine them.

In order to do this we firstly introduce the *light clocks* also known as *Einstein clocks*.

Let h and h' be two distinct observers at rest with respect to each other and T an event in h. We consider the sequence of events T_n in h defined by the recurrence rule: $T_0 = T$ and T_{n+1} is the echo of T_n by h'. The clock thus obtained is called *light clock* and is denoted by the triplet (h, h', T_0). (See Figure 2.21.)

Now we describe the refinement process of a clock (h, T_0, T_1), which we assume that coincides with the light clock (h, h', T_0).

Consider the sequence of observers r_k obtained by recurrence as follows: $r_0 = h'$ and r_{k+1} is the observer equidistant from h and r_k. Thus we obtain a sequence of light clocks (h, r_k, T_0). The event of index 1 of the k^{th} clock is denoted by $T_1^{[k]}$. (See Figure 2.22.)

Clearly, the clocks we just defined are $(h, T_0, T_1^{[k]})$. We denote vectors determined by two events by using the same symbols in boldface. We have the following relation:

$$\mathbf{T_0T_1^{[k]}} = \frac{1}{2^k}\mathbf{T_0T_1}.$$

Given an event X in h, let n_k be its time in the k^{th} clock, that is, X is equal or posterior to T_{n_k} and anterior to T_{n_k+1}.

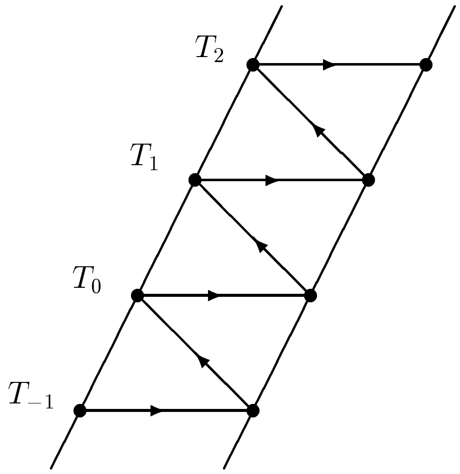

Figure 2.21: Einstein's light clock.

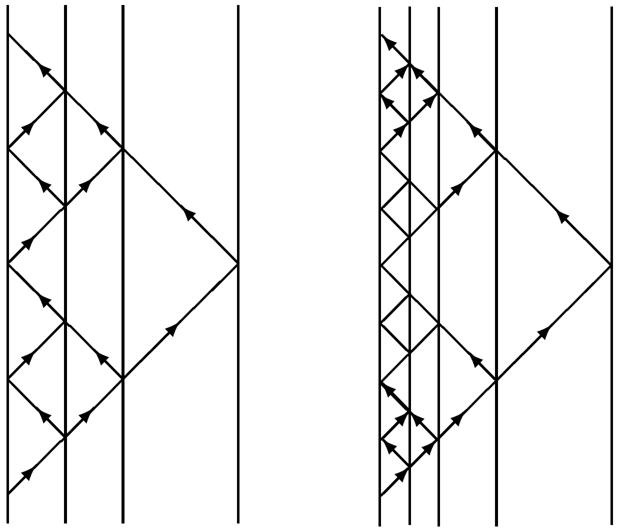

Figure 2.22: Process of refinement of clocks.

In terms of ratio of segments we have:

$$n_k \leq \frac{T_0X}{T_0T_1^{[k]}} < n_k + 1.$$

Considering the previous expression, we have

$$\frac{n_k}{2^k} \leq \frac{T_0X}{T_0T_1} < \frac{n_k}{2^k} + \frac{1}{2^k}.$$

Therefore we see that the number sequence $n_k/2^k$ has as limit the real number T_0X/T_0T_1, as k goes to infinity.

Thus we have an interpretation of the geometrical concept of ratio of segments in terms of concepts from special relativity. Once the number T_0X/T_0T_1 can be determined (with the desired approximation) from the reading of time in clocks, we (re)define the time of the event X in the clock h, T_0, T_1 as this number.

2.8 Distance

One of the most remarkable characterisitics of relativity is that distances can be measured with clocks. Therefore we can define a unit of length through the use of a unit of time.

We fix a unit of time by taking a clock as a standard clock and considering only standardized clocks, that is, clocks whose time intervals have the same duration as the standard one. Henceforward we assume that all observers use standardized clocks.

Given two observers h and r at rest with respect to each other and an event X in h, we can emit a light signal from X to r and consider its echo Y produced by r. (See Figure 2.23.)

We define the distance of h and r as the real number

$$\operatorname{dist}(h, r) = \frac{t(Y) - t(X)}{2},$$

where $T(X)$ and $T(Y)$ are the times of the events X and Y respectively, measured in a standardized clock of h.

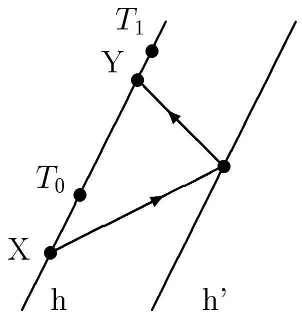

Figure 2.23: Echo Y of X in h.

The problem of defining the distance of two events is a bit more delicate. Let A and B be two events. We could consider two observers at rest h_A and h_B through A and B respectively, and define the distance of A and B as the distance of h_A and h_B. (See Figure 2.24.)

But this definition presents some difficulties. Firstly, we do not have a unique way to choose the observers h_A and h_B, so we could obtain more than one value for the distance fo A and B depending on the choice of those observers.

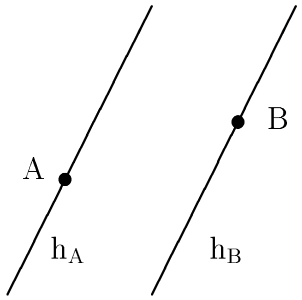

Figure 2.24: Events A and B in observers h_A and h_B.

Another difficulty surges when we try to measure the distance of two distinct events C and D which determine a time interval CD. If we choose observers h_C and h_D as being the observer which contain the interval CD then the distance of these events would be zero. This is undesirable because we want distinct events to have nonzero distance.

A way out of these difficulties is to restrict the events for which we can define distance and describe how to measure such distance without ambiguities.

In order to do this we start defining the notion of spatial interval. We say that two (distinct) events A and B constitute a *spatial interval* if there is an event C which can send light signals to both A and B. Such event C is unique because is determined by the intersection of two light rays, one going through A and other through B. The other point D with such property is the event which recieves light rays from A and B. Of course CD is a time interval, thus determining an observer h_{CD} which contain it.

We define the distance of A and B as the distance of the observers h_A and h_B through A and B respectively and at rest with respect to h_{CD}. (See Figure 2.25.)

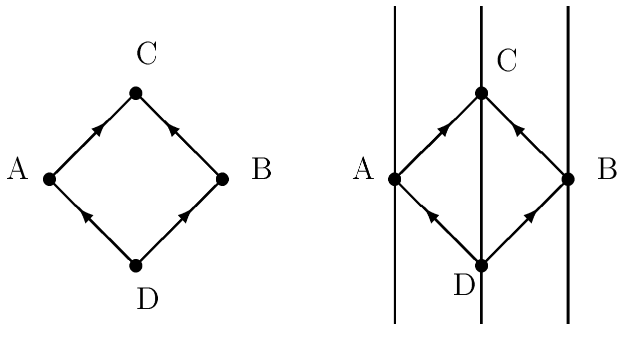

Figure 2.25: Events A and B constitute spatial interval.

Notice that if the clock of the observer h_A is set in such a way that $T_0 = A$ then the distance of A and B coincides with the time of the event in h_A which recieves a light signal emitted in B.

2.9 Spaces at rest

The *space at rest* of an event M in a given observer h is the set of the events X simultaneous to M with respect to h.

In the 2-dimensional space-time, spaces at rest (with respect to an observer h) are represented by parallel lines. This follows easily from the fact that light rays are represented by parallel lines and using Thales Theorem from affine geometry. Notice that the space at rest of M with respect to h meets h in M. (See Figure 2.26.)

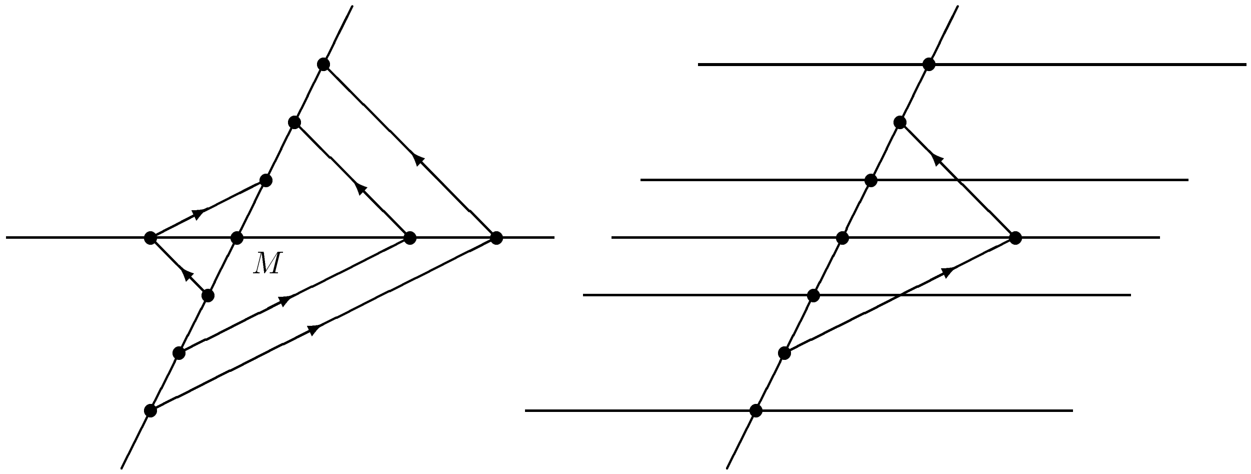

Figure 2.26: Space at rest of the event M.

Let A and B be two events in the space at rest of M (with respect to h), with $A \neq M$ and such that the light rays emitted in A and B have the same direction. (See Figure 2.27.) Then we have the following relation:

$$\frac{MB}{MA} = \frac{\text{dist}(M, B)}{\text{dist}(M, B)}$$

So we relate two concepts, one from affine geometry (segment ratio) and other from special relativity (distance of two events which constitute a spatial interval). The proof of such equation is easy. Just notice that distances are segment ratios in h (see section 2.7) and use the Thales Theorem.

2.10 Coordinates

In its most general sense a coordinate system is any process which gives a bijective correspondence between events in the 2-dimensional space-time and ordered pairs of real numbers.

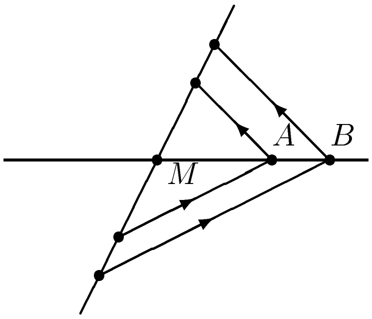

Figure 2.27: Distance and segment ratio.

An *affine coordinate system* in the plane is given by a sequence of noncollinear points (P_0, P_1, P_2) in that plane.

Let X be any point of the affine plane. We define the ordered pair of real numbers $(x_1(X), x_2(X))$ by the vectorial equation

$$\mathbf{P_0X} = x_1(X)\mathbf{P_0P_1} + x_2(X)\mathbf{P_0P_2}$$

(see Figure 2.28).

These numbers are called the coordinates of X in the system (P_0, P_1, P_2).

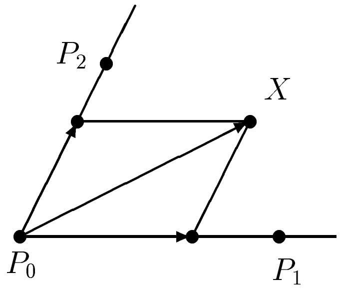

Figure 2.28: Vectorial relation connecting the points X, P_0, P_1 and P_2.

Saying with a more geometrical flavour, the coordinates $x_1(X)$ and $x_2(X)$ are given by the segment ratios

$$x_1(X) = \frac{P_0X_1}{P_0P_1} \quad \text{and} \quad x_2(X) = \frac{P_0X_2}{P_0P_2}$$

where X_1 and X_2 are the points in the lines P_0P_1 and P_0P_2 respectively which meet the lines parallel to P_0P_2 and P_0P_1 (respectively) through X.

The points P_0, P_1 and P_2 are called the origin, the unit point of the abscissa and the unit point of the ordinates, respectively.

We often use the notation $X : (\mu, \nu)$ to indicate that the point X has abscissa μ and ordinate ν.

Another often used fact is that, given $A : (\alpha, \alpha')$ and $B : (\beta, \beta')$, then the vector $\mathbf{AB}$ is given by

$$\mathbf{AB} = (\beta - \alpha)\mathbf{P}_0\mathbf{P}_1 + (\beta' - \alpha')\mathbf{P}_0\mathbf{P}_2.$$

2.11 Inertial frames

One of the most interesting features of special relativity is that an observer with a clock can naturally define an affine coordinate system in the 2-dimensional space-time. Such a system is called inertial system.

Let h be an observer with a standardized clock (h, T_0, T_1). We know that there are two events which can send light signals to T_1 and recieve signals from T_{-1}. We arbitrarily choose one of these events and call it E_1. (See Figure 2.29.)

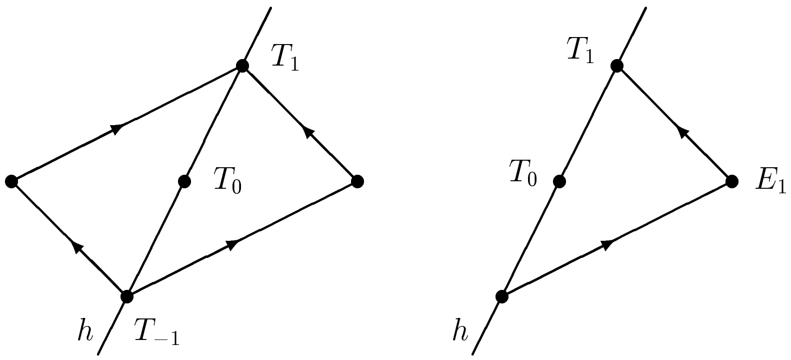

Figure 2.29: Affine coordinate system.

Once E_1 is chosen, we have the affine coordinate system determined by the sequence (T_0, E_1, T_1). This will be called a *inertial coordinate system* (or also inertial frame) of the clock (h, T_0, T_1). The other event with the same properties as the chosen E_1 will be denoted E_{-1}. Clearly $\mathbf{T_0E_{-1}} = -\mathbf{T_0E_1}$.

Let us denote ε_0 the space at rest of h in T_0. Given an event X there are unique events X_T in h and X_E in ε_0 such that X_T is simultaneous to X with respect to h and X_E is also in the observer h_X at rest with respect to h and containing X. (See Figure 2.30.)

Let $x(X)$ be the abscissa of X and $t(X)$ be the ordinate of X. In terms of vectors we have

$$\mathbf{T_0X} = x(X)\mathbf{T_0E_1} + t(X)\mathbf{T_0T_1}.$$

In section 2.10 we have seen that, in terms of segment ratios,

$$t(X) = \frac{T_0X_T}{T_0T_1}, \text{ and } x(X) = \frac{T_0X_E}{T_0E_1}$$

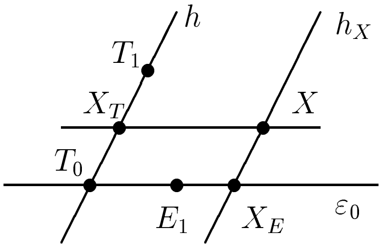

Figure 2.30: The events X_T and X_E determined by X.

The first of the above relations says that $t(X)$ is the time of X_T in the clock (h, T_0, T_1).

In order to interpret the number $x(X)$ we consider two possibilities.

Firstly, the events X_E and E_1 can emit light signals in the same direction to h. (See Figure 2.31.)

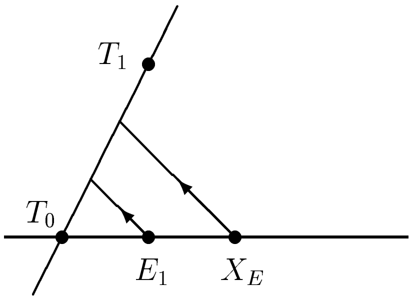

Figure 2.31: X_E and E_1 can emit light signals in the same direction to h.

In this case $x(X)$ is the distance of T_0 and X_E as we have seen in section 2.8.

Secondly, the events X_E and E_1 emit light signals from distinct classes to h. (See Figure 2.32.)

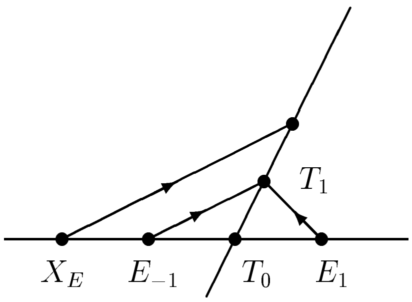

Figure 2.32: X_E and E_1 emit light signals from distinct classes to h.

In this case X_E and E_{-1} can emit light signals in the same direction to h.

From what we have seen in section 2.8 we have

$$\frac{T_0 X_E}{T_0 E_1} = \mathrm{dist}\,(X_E, E_1)$$

and because $T_0 X_E / T_0 E_1 = -T_0 X_E T_0 E_{-1}$ we have

$$x(X) = -\mathrm{dist}\,(X_E, E_1)$$

giving thus the interpretation of the coordinates $x(X)$ and $t(X)$.

In order to express properties of relativity in coordinates it is necessary to know the expression of lightlike vectors, what we do now.

We know that from the event T_{-1} it is possible to emit a light signal to E_1. In other words, the vector $\mathbf{T_{-1}E_1}$ is a lightlike vector. Notice that $T_{-1} : (0, -1)$ and $E_1 : (1, 0)$, so the vector $\mathbf{T_{-1}E_1}$ is given by

$$\mathbf{T_{-1}E_1} = \mathbf{T_0E_1} + \mathbf{T_0T_1}$$

On the other hand, E_1 can emit a light signal to T_1, so

$$\mathbf{E_1T_1} = -\mathbf{T_0E_1} + \mathbf{T_0T_1}$$

is also a lightlike vector.

From this we obtain the remarkable fact that the events $S_1 : (1, 1)$ and $S_{-1} : (-1, 1)$ can recieve light signals from the event $T_0 : (0, 0)$. (See Figure 2.33.)

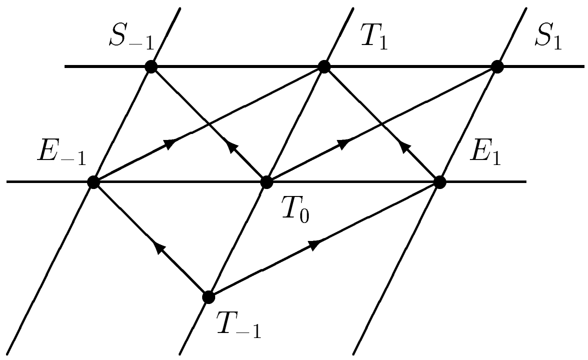

Figure 2.33: S_1 and S_{-1} recieve light signals from T_0.

Notice that this fact occurs in any inertial coordinate system with origin at T_0 and defined by standardized clocks.

2.12 Velocity

In this section we introduce the notion of velocity of an observer with respect to an inertial frame.

Let $(T_0,\, E_1,\, T_1)$ be an inertial frame defined by a clock $(h,\, T_0,\, T_1)$.

Firstly we characterize time intervals of the form T_0X in terms of coordinates. We will show that T_0X is a time interval if, and only if, $|x(X)| < t(X)$.

Indeed, saying that T_0X is a time interval is equivalent to say that the vector $\mathbf{T_0X}$ is a sum

$$\mathbf{T_0X} = \alpha\mathbf{T_0S_1} + \beta\mathbf{T_0S_{-1}}$$

with $\alpha, \beta > 0$. (This is the coordinate equivalent of the definition of time intervals.)

From this we obtain (see section 2.11)

$$\mathbf{T_0X} = (\alpha - \beta)\mathbf{T_0E_1} + (\alpha + \beta)\mathbf{T_0T_1}.$$

That is, the event X has coordinates

$$x(X) = \alpha - \beta, \text{ and } t(X) = \alpha + \beta.$$

From these equations we conclude that $\alpha, \beta > 0$ if, and only if, $|x(X)| < t(X)$.

We can also easily see that T_0X is a lightlike interval if, and only if, $|x(X)| = t(X)$.

Now consider an observer k containing the event T_0, and let X be an event in k posterior to T_0. We have that T_0X is a time interval and $|x(X)| < t(X)$.

The number $v = x(X)/t(X)$ is called the *velocity* of k is the frame $(T_0,\, E_1,\, T_1)$.

It is clear that v does not depend on the choice of X in k (any other Y in k posterior to T_0 would give the same result) and $|v| < 1$. Notice that because $X : (x(X), t(X))$ is in k and k goes through T_0, the event $V : (v, 1)$ is in k (see Figure 2.34).

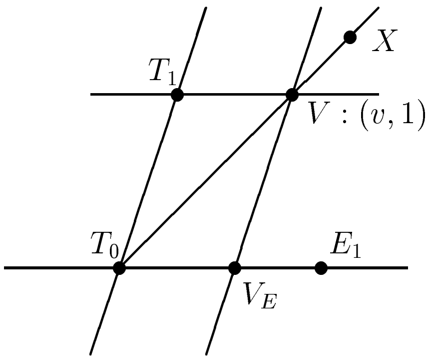

Figure 2.34: The coordinate $x(V) = v$ is the velocity of k.

The velocity of an observer r which not necessarily contains T_0 is defined as the velocity of the observer r_{t_0} which is at rest with respect to r and contains T_0.

If we apply this definition to a light ray we obtain that the velocity of light is 1 in any inertial frame (defined by a standardized clock). This fact was experimentally verified by Michelson and Morley and was one of the motivations to the creation of relativity theory.

We finally observe that the inequality $|v| < 1$ means that no observer can move with velocity equal or greater than the velocity of light.

2.13 Symmetry

In section 2.4 we have introduced the notion of symmetry with respect to an observer. Recall that an event X is symmetric to Y with respect to an observer r if X and Y are distinct and there are events A which can emit light signals to X and Y and B which can recieve light signal from X and Y.

Geometrically this means that the lines XY and AB are the diagonals of a parallelogram whose sides are lines parallel to light rays (see Figure 2.35).

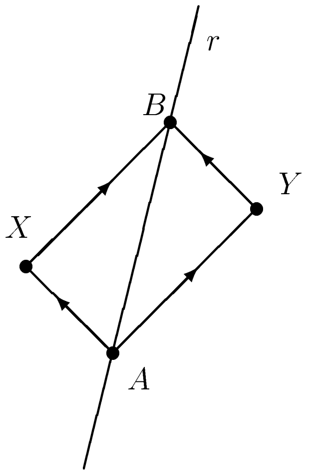

Figure 2.35: X and Y are symmetric with respect to r.

Our goal in this section is to describe symmetry in terms of coordinates.

Let us start fixing an inertial frame (T_0, E_1, T_1) (see section 2.11).

We use repeatedly the usual procedure from Plane Analytic Geometry of determining the intersection points of straight lines by solving their cartesian equations.

Let P be an event. The straight line through P and parallel to the vector $\mathbf{T_0S_1}$ (see section 2.11) has the equation

$$x - t = x(P) - t(P).$$

Analogously, the line through P and parallel to the vector $\mathbf{T_0 S_{-1}}$ has the equation

$$x + t = x(P) + t(P).$$

Also, the equation of a line which representes an observer (through T_0) with velocity v is

$$x = vt.$$

Let W be the event in r which meets the light ray through X and parallel to the vector $\mathbf{T_0 S_1}$. (See Figure 2.36.) The coordinates $x(W)$ and $t(W)$ are the solution to the system of equations

$$\begin{aligned} x - t &= x(X) - t(X) \\ x - vt &= 0 \end{aligned}$$

that is,

$$x(W) = \frac{v}{1-v}(t(X) - x(X)), \text{ and } t(W) = \frac{1}{1-v}(t(X) - x(X))$$

On the other hand, the event U in r which meets the light ray through X and parallel to the vector $\mathbf{T_0 S_{-1}}$ (see Figure 2.36) has coordinates

$$x(U) = \frac{v}{1+v}(t(X) + x(X)), \text{ and } t(U) = \frac{1}{1+v}(t(X) + x(X))$$

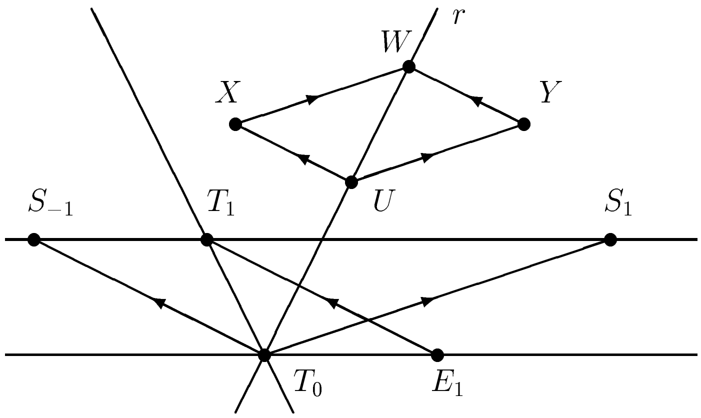

Figure 2.36: Relating coordinates to symmetry.

It is clear that, possibly except by their order, the events U and W coincide with A and B respectively (compare with Figure 2.35).

Now consider the event Y meeting the light ray through U and parallel to the vector $\mathbf{T_0S_1}$ and the ray through W and parallel to $\mathbf{T_0S_{-1}}$. (See Figure 2.36.)

The coordinates of Y are the solutions to the system

$$\begin{aligned} x - t &= x(U) - t(U) \\ x + t &= x(U) + t(U) \end{aligned}$$

A simple calculation gives

$$x(U) - t(U) = \frac{v-1}{1+v}(t(X) + x(X))$$

$$x(W) + t(W) = \frac{1+v}{1-v}(t(X) + x(X))$$

From these we obtain the coordinates of Y

$$x(Y) = -\frac{1+v^2}{1-v^2}x(X) + \frac{2v}{1-v^2}t(X)$$

$$t(Y) = -\frac{2v}{1-v^2}x(X) + \frac{1+v^2}{1-v^2}t(X)$$

Finally we obtain the remarkable relation

$$x(Y)^2 - t(Y)^2 = x(X)^2 - t(X)^2$$

In other words, if T_0X and T_0Y are two time intervals such that Y is symmetric to X with respect to an observer r through T_0, then the coordinates of these events satisfy the relation above.

The converse can be easily proven. It is enough to show that if X and Y are two events satisfying the above relation then X and Y are symmetric with respect to the observer through T_0 and through the middle point of those events.

2.14 Synchronization hyperbola

Here we look again at the subject of section 2.6, that is, the question of comparing time intervals. The following fact is probably the most important of this chapter.

Let T_0T_1 be a given time interval. The set of events X such that the time intervals T_0T_1 and T_0X have the same duration is called *synchronization hyperbola* of T_0T_1. We will show that this set is a branch of the hyperbola with center T_0, containing the event T_1 and having as assintotes the light rays through T_0. (See Figure 2.37.)

Fix an inertial frame (T_0, E_1, T_1). From section 2.13 we have

$$x(X)^2 - t(X)^2 = x(T_1)^2 - t(T_1)^2$$

that is

$$x(X)^2 - t(X)^2 = -1$$

for any event X in the synchronization hyperbola. This is the equation of a hyperbola in those conditions.

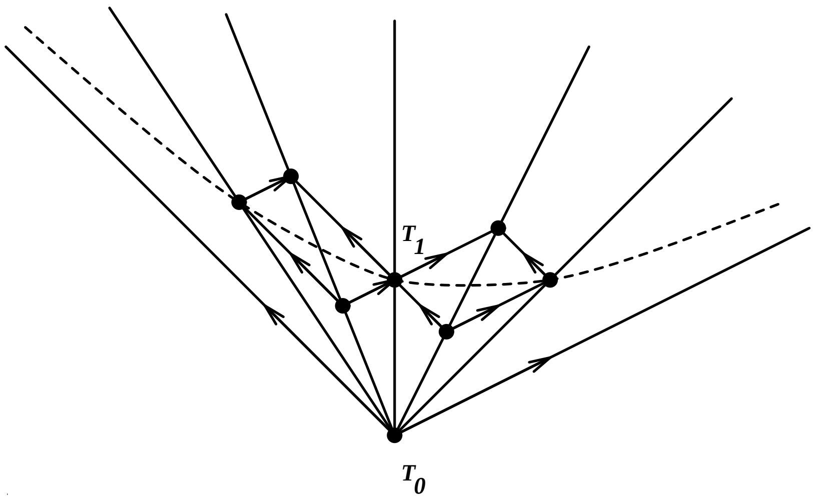

Figure 2.37: Synchronization hyperbola.

2.15 Lorentz transformations

We now study the relations which connect coordinates in distinct inertial frames. In other treatments of relativity these transformations are taken as the starting point of the theory.

We consider inertial frames with the same origin T_0 and we assume that the clocks are synchronized.

Two frames (T_0, E_1, T_1) and (T_0, E_1', T_1') are said to be *coherent* if the light rays which can be emited from E_1 and E_1' to T_1 and T_1', respectively, have the same direction (see Figure 2.38).

Let h and h' be the observers containing the time intervlas T_0T_1 and T_0T_1', respectively, and let v be the velocity of h' in (T_0, E_1, T_1). The coordinates of T_1' in (T_0, E_1, T_1) are

$$T_1' : \left(\frac{v}{\sqrt{1-v^2}}, \frac{1}{\sqrt{1-v^2}}\right)$$

(see the equation in the end of section 2.13).

We will show that in (T_0, E_1, T_1) the coordinates of E_1' are given by:

$$E_1' : \left(\frac{1}{\sqrt{1-v^2}}, \frac{v}{\sqrt{1-v^2}}\right)$$

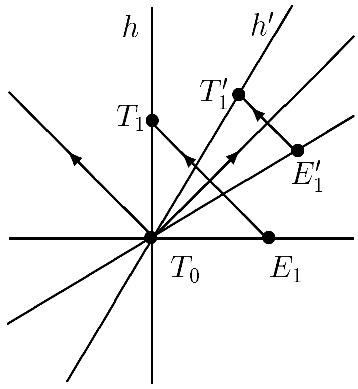

Figure 2.38: Coherent inertial frames.

In the frame $(T_0,\, E_1,\, T_1)$ T'_{-1} has coordinates:

$$T'_{-1} : \left(-\frac{v}{\sqrt{1-v^2}}, -\frac{1}{\sqrt{1-v^2}}\right)$$

From these we obtain the following vector relations:

$$\begin{aligned}
\mathbf{E'_1T'_1} &= (1-v)(1-v^2)^{-1/2}\,\mathbf{T_0S_{-1}}\\
\mathbf{T'_{-1}E'_1} &= (1+v)(1-v^2)^{-1/2}\,\mathbf{T_0S_1}\\
\mathbf{E_1T_1} &= \mathbf{T_0S_{-1}}
\end{aligned}$$

From this we see that E'_1 is simultaneous to T_0 with respect to h and that the light signal emitted from E'_1 to T'_1 has the same direction as the one emitted from E_1 to T_1. We also see that the distance of E'_1 and T_0 equals 1.

From the coordinates of E'_1 and T'_1 we obtain the vector relations:

$$\begin{aligned}
\mathbf{T_0E'_1} &= (1-v^2)^{-1/2}\,\mathbf{T_0E_1} + v(1-v^2)^{-1/2}\,\mathbf{T_0T_1}\\
\mathbf{T_0T'_1} &= v(1-v^2)^{-1/2}\,\mathbf{T_0E_1} + (1-v^2)^{-1/2}\,\mathbf{T_0T_1}
\end{aligned}$$

Now let $X : (x, t)$ be the coordinates of the event X in the frame $(T_0,\, E_1,\, T_1)$ and $X : (x', t')$ in the frame $(T_0,\, E'_1,\, T'_1)$. The previous relations give us

$$\begin{aligned}
x &= (1-v^2)^{-1/2}\,x' + v(1-v^2)^{-1/2}\,t'\\
t &= v(1-v^2)^{-1/2}\,x' + (1-v^2)^{-1/2}\,t'
\end{aligned}$$

which are the well known *Lorentz transformations.*

If we solve this system with respect to x' and t' we obtain:

$$\begin{aligned}
x' &= (1-v^2)^{-1/2}\,x - v(1-v^2)^{-1/2}\,t\\
t' &= -v(1-v^2)^{-1/2}\,x + (1-v^2)^{-1/2}\,t
\end{aligned}$$

These relations show that the velocity of h in $(T_0,\, E'_1,\, T'_1)$ is $-v$.

2.16 Doppler effect

Let h and h' be two observers not at rest with respect to each other and having a common event T_0 and endowed with standardized clocks (h, T_0, T_1) and (h', T_0, T_1').

Suppose that in T_1' a light signal is emitted to h, recieved in the event U_1. (See Figure 2.39.) We want to determine the length of the time interval T_0U_1 as measured in the clock (h, T_0, T_1). We shall see that in the case where h and h' are not at rest the event U_1 is posterior to T_1.

Let (T_0, E_1, T_1) be the inertial frame such that the light signal emitted from E_1 to T_1 has the same direction as the one emitted from T_1' to U_1. This is the direction of the vector $\mathbf{T_0S_{-1}}$.

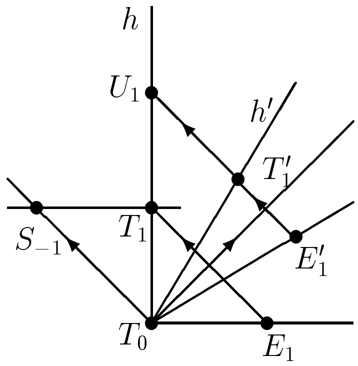

Figure 2.39: Doppler effect.

Let v be the velocity of h' in (T_0, E_1, T_1). The event T_1' has coordinates

$$T_1' : \left(\frac{v}{\sqrt{1-v^2}}, \frac{1}{\sqrt{1-v^2}}\right)$$

and the line through T_1' and parallel to the vector $\mathbf{T_0S_{-1}}$ has equation

$$x + t = x(T_1') + t(T_1')$$

The event U_1 common to this line and to h has coordinates $x(U_1) = 0$ and $t(U_1) = (1+v)/\sqrt{1-v^2}$, that is

$$t(U_1) = \sqrt{\frac{1+v}{1-v}}$$

The vector $\mathbf{T'_1U_1}$ is given by

$$\mathbf{T'_1U_1} = \frac{v}{\sqrt{1-v^2}}\mathbf{T_0S_{-1}}$$

and bacause the coeficient of $\mathbf{T_0S_{-1}}$ is positive we conclude that $v > 0$. Therefore $t(U_1) > 1$.

Since the time of U_1 can be directly measured in the clock of h we have described here a process to measure the velocity v.

2.17 Rulers

A ruler can be represented in the spacetime by two observers at rest with respect to each other. They represent its extremities. So we represent a ruler by a pair ab where a and b are the observers representing its extremities.

The *proper length* of a ruler ab is the distance of these observers (see section 2.8).

We can also measure the length of a ruler from other observers, in general obtaining a different value. In order to define the length of a ruler ab as measured from an observer h we start choosing an event A in a and cosider the observer h_A through A and at rest with respect to h. Then we consider the event B in the extreminty b of the ruler which is simultaneous to A with respect to h_A (see section 2.5 and Figure 2.40).

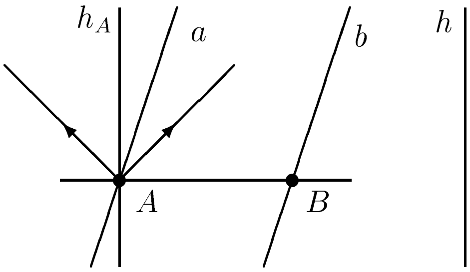

Figure 2.40: Starting the measure of the ruler ab.

The distance of a and b as observed by h is the distance of the events A and B (see section 2.8).

Using inertial frames we will compare these two measures of distance. So we consider a standardized clock (h_A, T_0, T_1) in h_A, such that T_0 coincides with A. Choose an inertial frame (T_0, E_1, T_1) such that the light signals from B and E_1 to h have the same direction. (See figure 2.41.)

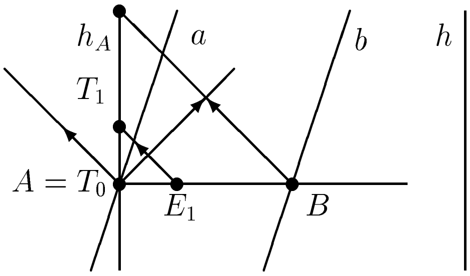

Figure 2.41: An inertial frame on h_A.

Let v be the velocity of a in (T_0, E_1, T_1). Then the event $V : (v, 1)$ is in a. The coordinates of B are $B : (\beta, 0)$ and the choice of E_1 gives $\beta > 0$. It is easy to see that the number β is the length of the ruler ab as observed in h_A (and therefore in h).

The line b goes through $B : (\beta, 0)$ and $B' : (\beta + v, 1)$. The light signal emitted in T_0 in the direction of the vector $\mathbf{T_0S_1}$ meets the line b in the event $U : (\beta/(1-v), \beta/(1-v))$. The light signal emitted in U in the direction of the vector $\mathbf{T_0S_{-1}}$ meets the line a in the event U' : $(2v\beta/(1-v^2), 2\beta/(1-v^2))$. (See Figure 2.42.)

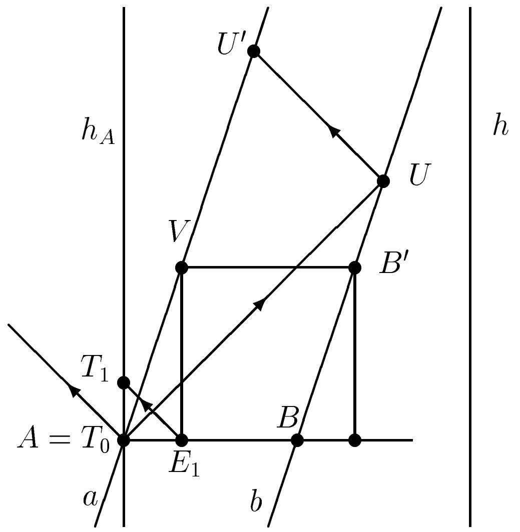

Figure 2.42: Computing the proper length of ab.

Consider now a standardized clock in a (a, A_0, A_1) such that A_0 coincides with T_0 and the time intervals T_0T_1 and A_0A_1 have the same duration.

The coordinates of A_1 in (T_0, E_1, T_1) are (see section 2.14)

$$A_1 : \left(\frac{v}{\sqrt{1-v^2}}, \frac{1}{\sqrt{1-v^2}} \right)$$

From this we obtain for the segment ratio T_0U'/T_0A_1 the value $2\beta(1-v^2)^{-1/2}$. Therfore the distance of a and b is then the proper length of the ruler ab and has the value $\beta(1-v^2)^{-1/2}$.

Thus the observed length of a ruler is the product of its proper length by the factor $(1-v^2)^{-1/2}$, where v is the velocity of this ruler in an inertial frame defined by a standardized clock of h.

Notice that if $v = 0$ then both lengths coincide. And in general the observed length is less than, or equal to, the proper length.

2.18 Mass

According to the Priciple of Inertia already mentioned in section 2.4 the trajectory of a material particle is represented in spacetime by a timelike straight line (see cetion 2.3).

We know that particles can interact in collision processes. We say that two particles collide in an event K if it is a common event to both trajectories. In this case we consider as representatives of these particles only the events anterior to K. From a collision there may result other particles whose trajectories contain the event K. We consider their representatives only the events posterior to K. (See Figure 2.43.)

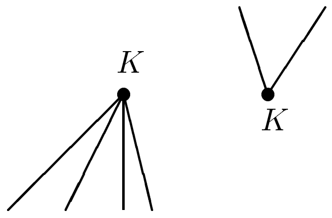

Figure 2.43: Collision of four particles resulting in two particles.

The number of particles resulting from the collision do not necessarily coincide with the number of colliding particles. In this text we restrict ourselves to collisions resulting in just one particle. These are known as *fusion processes.*

We know from experimental facts that only the trajectories are not enough to describe collision processes. In fact we can have the following situation.

Let p and q be the trajectories of two particles colliding in the event K and resulting a third particle with trajectory r. Consider two other particles with trajectories p' and q' at rest with respect to p and q respectively. It may happen that the resulting particle of this second collision has a trajectory r' not at rest with respect to r. (See Figure 2.44.)

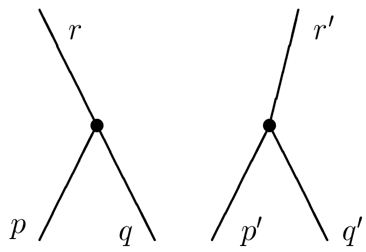

Figure 2.44: Trajectories are not enough to describe collisions.

In order to complete the description of a particle, at least for the study of collisions we consider its trajectory together with a vector parallel to it which is, by definition, the *impulse* of the particle. A particle will be denoted by the pair $(p, \mathbf{p})$ consisting of its trajectory p and its impulse vector $\mathbf{p}$. (See Figure 2.45.)

We use the *Principle of Conservation of Impulses* which says that the impulse of the resulting particles is the vector sum of the impulses of the colliding particles.

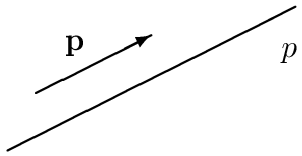

Figure 2.45: Collision of two particles resulting in one particle: fusion process.

So in our case we say that the impulse of the resulting particle is the vector sum of the impulses of the colliding particles. (See Figure 2.46.)

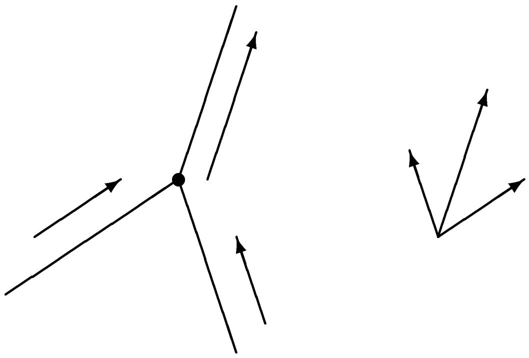

Figure 2.46: Impulse vectors in a collision.

Timelike vectors are non null and the sum of two timelike vectors is a timelike vector and so it is non null. Therefore, in a collision of two particles resulting in a third particle the impulse of this one is non null and determines the trajectory of it.

Now we describe a notion of equivalence of particles which will allow us to recognize particles of equal mass.

We say that two distinct particles (a, $\mathbf{a}$) and (b, $\mathbf{b}$) are *equivalent* if from their collision in an event K there results a particle (p, $\mathbf{p}$) whose trajectory p is the observer equidistant of a and b. (See Figure 2.47.)

The impulse vectors $\mathbf{a}$ and $\mathbf{b}$ are timelike vectors, therefore there are events A in a and B in b such that $\mathbf{a} = \mathbf{KA}$ and $\mathbf{b} = \mathbf{KB}$.

We will show now that if the two particles are equivalent then the events A and B are symmetric with respect to p.

Let B' be the event symmetric of A with respect to p. From the hypothesis that the particles (a, $\mathbf{a}$) and (b, $\mathbf{b}$) are equivalent there follows that B' is in b and that the middle point M' of A and B' is in p. On the other hand, from the equality $\mathbf{p} = \mathbf{a} + \mathbf{b}$, there follows that the middle point M of A and B is in p. Therefore M and M' must coincide and so B and B' also coincide. (See Figure 2.48.)

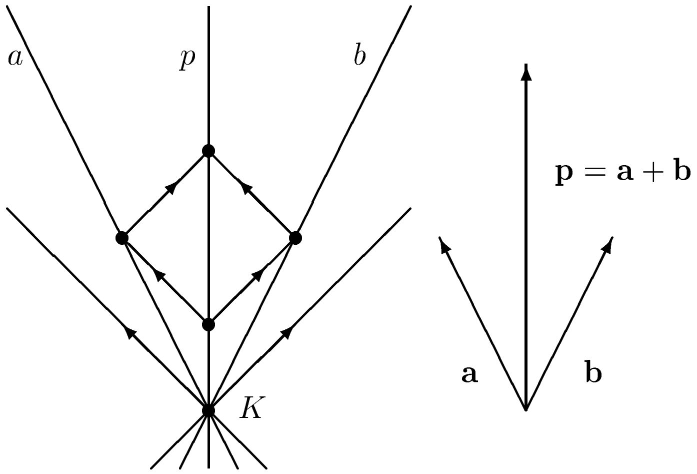

Figure 2.47: Equvalence of impulses.

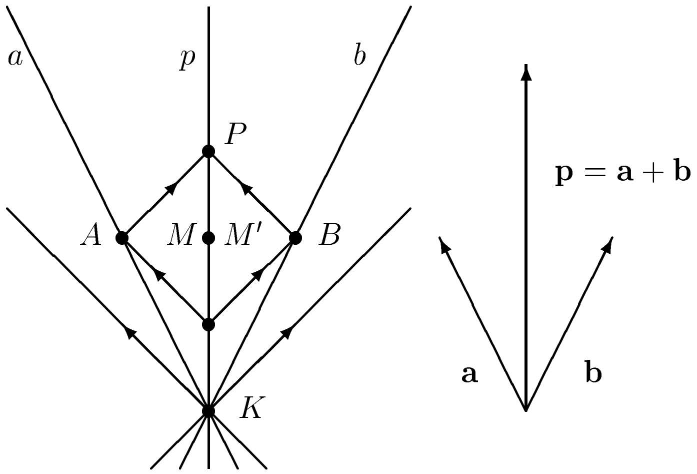

Figure 2.48: Equivalence of impulses and symmetry.

Thus in the geometrical model a timelike vector can be interpreted in two ways: one as the vector determined by the extremities of a time interval, or as the impulse of a particle. We also see that the process to decide if two particles are equivalent is geometrically the same as the process to decide if two time intervals have the same duration.

Let us define the *mass* of a particle. Let $(p,\ \mathbf{p})$ be a given particle and assume that $(p,\ P_0,\ P_1)$ is a standardized clock of p. Since the impulse vector $\mathbf{p}$ is parallel to, and has the same direction as, $\mathbf{P_0P_1}$ then there is a number $m > 0$ such that $\mathbf{p} = m\mathbf{P_0P_1}$. This number is called the *mass* of that particle.

We will show now that two equivalent particles have the same mass.

Let $(p,\ \mathbf{p})$ and $(q,\ \mathbf{q})$ be two equivalent particles colliding in P_0 and let $(p,\ P_0,\ P_1)$ and $(q,$

P_0, Q_1) be standardized clocks of p and q, respectively. Suppose that U and W are events such that $\mathbf{p} = \mathbf{P_0U}$ and $\mathbf{q} = \mathbf{P_0W}$. Since ($p$, $\mathbf{p}$) and (q, $\mathbf{q}$) are equivalent the events U and W are symmetrical with respect to the observer equidistant of p and q and the same is true of P_1 and Q_1. Let m_p and m_q be the masses of (p, $\mathbf{p}$) and (q, $\mathbf{q}$) respectively. Then we have $m_p = P_0U/P_0P_1$ and $m_q = P_0W/P_0Q_1$. By Thales Theorem we conclude that $m_p = m_q$. (See Figure 2.49.)

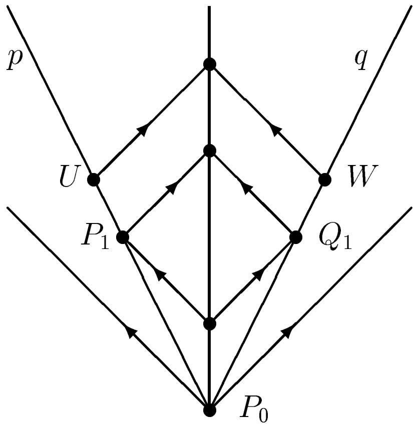

Figure 2.49: Equivalence of impulses implies equal mass.

In order to define mass of particles we have used the fact that timelike vectors have two interpretations in the geometrical model. Since this interpretation is inherent to the model and not to relativity we need to describe a process to measure mass using relativistic concepts.

Firstly recall that we have chosen (arbitrarily) a class of time intervals of same duration to standardize the clocks. Now choose (arbitrarily) a class of equivalent particles whose impulse vectors are given by timelike vectors given by standardized time intervals. Each of these will be called a *test particle* or *standardized mass*. From the definition of mass it follows that the mass of such test particle is 1.

We measure the mass of a particle by using an experiment of collision of that particle with a test particle and using the Principle of Conservation of Impulse. With this idea we show that in an inertial frame the impulse of a particle can be expressed as a function of its mass and its velocity.

Let (p, $\mathbf{p}$) be a particle and (T_0, E_1, T_1) and inertial frame. Let (p, P_0, P_1) be a standardized clock of p. We have seen in section 2.14 that the vector $\mathbf{T_0P_1}$ is given by:

$$\mathbf{T_0P_1} = \frac{v}{\sqrt{1-v^2}}\mathbf{T_0E_1} + \frac{1}{\sqrt{1-v^2}}\mathbf{T_0T_1},$$

where v is the velocity of p in the frame (T_0, E_1, T_1). Let m be the mass of that particle. Then $\mathbf{p} = m\mathbf{T_0P_1}$, so from the previous expression we obtain

$$\mathbf{p} = \frac{mv}{\sqrt{1-v^2}}\mathbf{T_0E_1} + \frac{m}{\sqrt{1-v^2}}\mathbf{T_0T_1}.$$

Let us write down the equations expressing the conservation of impulse in this frame. Consider two particles (a, $\mathbf{a}$) and (b, $\mathbf{B}$) colliding in T_0 and resulting in a particle (q, $\mathbf{q}$). By the Principle of Conservation of Impulses, $\mathbf{q} = \mathbf{a} + \mathbf{b}$. But

$$\mathbf{a} = \frac{m_a v_a}{\sqrt{1 - v_a^2}} \mathbf{T_0 E_1} + \frac{m_a}{\sqrt{1 - v_a^2}} \mathbf{T_0 T_1},$$

$$\mathbf{b} = \frac{m_b v_b}{\sqrt{1 - v_b^2}} \mathbf{T_0 E_1} + \frac{m_b}{\sqrt{1 - v_b^2}} \mathbf{T_0 T_1},$$

$$\mathbf{q} = \frac{m_q v_q}{\sqrt{1 - v_q^2}} \mathbf{T_0 E_1} + \frac{m_q}{\sqrt{1 - v_q^2}} \mathbf{T_0 T_1}.$$

Therefore we have:

$$m_q v_q (1 - v_q^2)^{-1/2} = m_a v_a (1 - v_a^2)^{-1/2} + m_b v_b (1 - v_b^2)^{-1/2},$$

and

$$m_q = m_a (1 - v_a^2)^{-1/2} + m_b (1 - v_b^2)^{-1/2}.$$

Notice that these equations are true in any inertial frame containing T_0.

In particular, consider an inertial frame (T_0, E_1, T_1) defined by a clock (h, T_0, T_1) such that (q, $\mathbf{q}$) is at rest with respect to h, that is, the observer q coincides with h. (See Figure 2.50.)

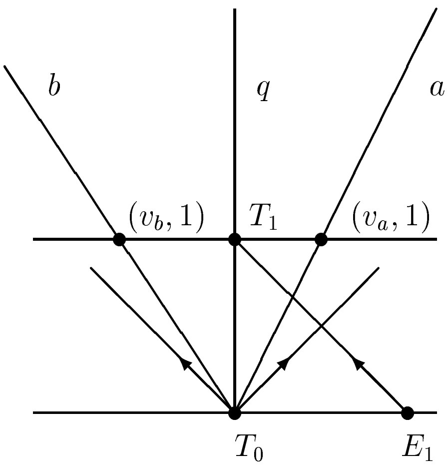

Figure 2.50: Measure of mass.

In these conditions we have $v_q = 0$ and the expressions above become

$$m_a v_a (1 - v_a^2)^{-1/2} + m_b v_b (1 - v_b^2)^{-1/2} = 0$$

from which we obtain

$$\frac{m_b}{m_a} = \frac{v_a (1 - v_a^2)^{-1/2}}{v_b (1 - v_b^2)^{-1/2}}$$

and in the case where $m_a = 1$,

$$m_b = \frac{v_a(1 - v_a^2)^{-1/2}}{v_b(1 - v_b^2)^{-1/2}}.$$

Notice that in the second member of these equations there appear only velocities, which can be measured in clocks.

Still in this case, the above relations imply

$$m_q = m_a(1 - v_a^2)^{-1/2} + m_b(1 - v_b^2)^{-1/2}.$$

From this we see that the mass m_q is bigger than the sum of the masses m_a and m_b, that is,

$$m_q > m_a + m_b.$$

If the velocities v_a and v_b are near 1 (the velocity of light) the mass m_q is considerably bigger than $m_a + m_b$.

Chapter 3

The 3-dimensional model

3.1 The 3-dimensional spacetime

The considerations about linear movement were used to motivate the use of an affine plane to describe the relativistic model where these movements were represented.

The next step in classical kinematics is to study plane movements which can be geometrically described (with respect to position and time) in a 3-dimensionl affine space. We recall that the word *affine* means that we consider no *a priori* metric notion.

So we fix a 3-dimensional affine space and call its points *events*.

3.2 Light rays, time intervals and observers

Light rays will be represented by some oriented straight lines. We now state (or "axiomatize") its properties.

Axiom 1 Through each event there are at least two distinct (i. e. non parallel) light rays.

If there is a light ray through two distinct events A and B we say that these events communicate. If the vector $\mathbf{AB}$ has the same direction as that light ray we say that A emits a light signal to B and that B recieves a light signal from A.

Axiom 2 If B recieves a light signal from A then A does not recieve a light signal from B.

Axiom 3 If A, B and C are three distinct events not in the same light ray and two pairs of them communicate then the third pair do not communicate.

An ordered pair of events PQ constitute a *time interval* if there is an event S such that P emits a light signal to S and S emits a distinct light signal to Q (compare with section 2.3).

From Axiom 3 we see that the events P and Q do not comunicate.

Axiom 4 Let PQ be a time interval and T an event which comunicate with both P and Q. Then T recieves a light signal from P and emits another one to Q.

A straight line r represents an *observer* if for any event X not in r there are unique events X' and X'' in r such that X' emits a light signal to X and X'' recieves a light signal from X.

Axiom 5 If PQ is a time interval then the straight line determined by these events is an observer.

Axiom 6 If a straight line s is parallel to an observer r then s is also an observer.

In this case we say that s is *at rest* with respect to r.

3.3 Two-dimensional sections

Let a and b be two observers at rest with respect to each other. The events meeting light rays which meet both observers constitute what is called a *2-dimensional section* of the spacetime (or also a *timelike plane*). It is easy to see that this is the plane determined by a and b. In this section we will show that we can naturally recover the 2-dimensional model in this plane.

Notice that no light ray in this plane can be parallel to a or b (because of Axiom 3).

Choose an event A in a and let s' and s'' be the light signals which A recieves from b and A sends to b, respectively, and let B be the event in b which recieves the signal s''. Now let q be another light ray and suppose that it is detected in the events Q_a of a and Q_b of b. Assume that Q_a emits a light signal to Q_b. (See Figure 3.1.)

We show now that q is parallel to, and has the same direction as, s''. Firstly notice that if q goes through A then q must coincide with s'' because of Axiom 4 applied to the time interval determined by the events B and Q_b.

Suppose that q and s'' are distinct and that they are not parallel. Let S be the common event to these lines. Then either S emits light signals to a or recieves light signals from a (see Figure 3.2). In both cases we have a contradiction to Axiom 4 applied to the time interval determined by A and Q_a.

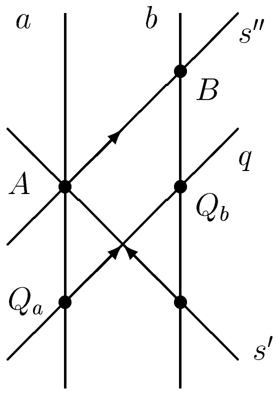

Figure 3.1: Light rays in the plane determined by a and b.

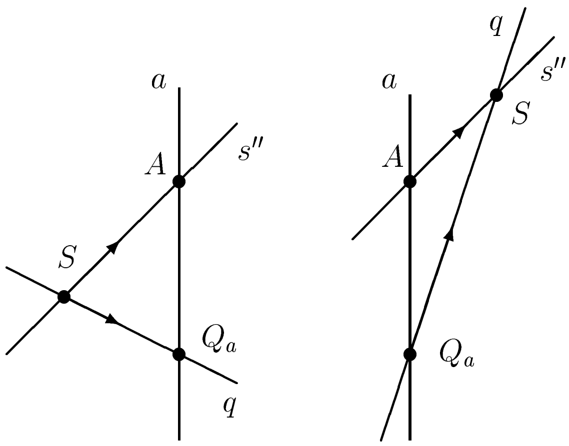

Figure 3.2: S is the common event to q and s'': two possibilities.

3.4 Pencils of parallel light rays

We will prove here that if r is a straight line parallel to a light ray s and oriented with the same direction as s then r also represents a light ray.

Let s be a light ray and P an event not in s. We will show that the straight line u through P and oriented with the same direction as s also represents a light ray.

In order to do this we start with an arbitrarily chosen event A which meets the light ray s and an observer a containing A. Let b be another observer distinct from a, at rest with respect to a and meeting the light ray s and such that there is an event B in b wich recieves a light signal from

a throught s. (Notice that Axiom 5 guarantees the existence of such a b.) Let w be the other light signla through A in the 2-dimensional section determined by a and b. (See Figure 3.3.)

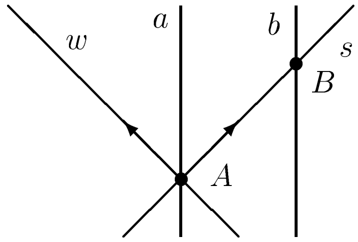

Figure 3.3: Two-dimensional section determined by a and b.

Let B_n be a sequence of events in b posterior to B and converging to B. This notion of convergence can be interpreted in terms of a light clock in this 2-dimensional section. The intervals AB_n are all timelike and we denote b_n the observer determined by AB_n. (See Figure 3.4.)

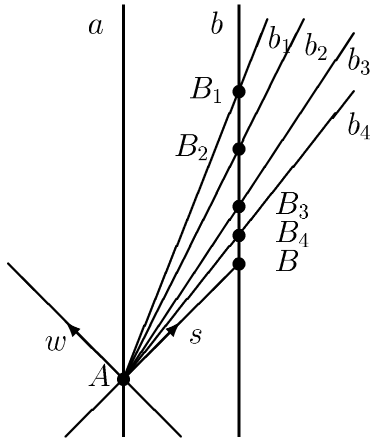

Figure 3.4: Sequence B_n converging to B.

Now choose another event P (not necessarily in the plane determined by a and b) and let a' be an observer through P and at rest with respect ot a, and let b'_n be observers through P at rest with respect to b_n and b' the observer through P and at rest with respect to b. From parallelism properties we can see that a', b' and the b'_n lie in the 2-dimensional section determined by a' and b'.(See Figure 3.5.)

We have that the sequence of events B'_n (in both b'_n and b') determines a sequence of time intervals PB'_n haveing the same duration as the intervals AB_n, respectively (as measured in the sense of the 2-dimensional section determined by b_n and b'_n, respectively). Again from parallelism properties, we conclude that the events converge to the event b' in b' and the line u through P and

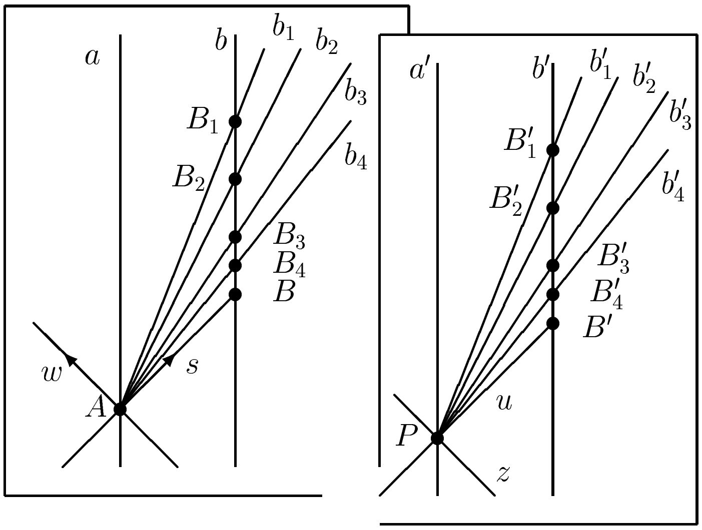

Figure 3.5: The section determined by a' and b'.

B' is parallel to s. Since PB'_n are time intervals, PB' is either a timelike or a lightlike interval. Because u is parallel to the light ray s, Axioms 4 and 6 imply that u cannot be an observer, therefore representing a light ray. Let us determine its direction.

Let z be a line through P parallel to the light ray w. Then z also represents a light ray. Suppose that the direction of u is opposite to that of s. Then the direction of z must be opposite to that of w (because a' is an observer). This implies that the pair B'_1P is a time interval, and we know that PB'_1 is a time interval. This contradicts Axiom 4.

3.5 Proximity of events

In this section we show how the geometrical concept of proximity can be formulated in terms of relativity.

Firstly, let us discuss the notion of *convergence* of events in a line. Given two distinct points V and W, the *open segment* $]VW[$ is the set of all points R in the line VW such that the vectors $\mathbf{VR}$ and $\mathbf{VW}$ satisfy $\mathbf{VR} = \lambda\mathbf{VW}$ for some λ with $0 < \lambda < 1$. We say that a sequence of points X_n in a line r converge to the point X if for all points V and W in r such that X lies in the open segment $]VW[$ there is some n_0 such that for all $n > n_0$, the points X_n all lie in the segment $]VW[$.

Let p be an observer and P an event in p and P_n a sequence of events. If P_n is not in p we define the events P'_n and P''_n in p as in section 2.4 (see also Figure 3.6), and if P_n is in p we define the events P'_n and P''_n as P_n.

The convergence of the sequences P'_n and P''_n in p can be detected (at least theoretically) by the

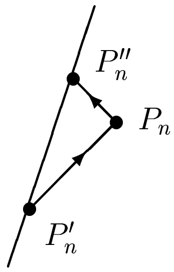

Figure 3.6: Definition of the events P'_n and P''_n.

use of a clock of p.

We say that the sequence P_n *converges* to P if the sequences P'_n and P''_n (in any given observer p containing P) converge to P (as detected with a clock in p).

We will impose two conditions on this notion.

Axiom 7 Let S and P two distinct events and P_n a sequence of events converging to P. If each P_n can recieve a light signal from S then P can recieve a light signal from S.

Axiom 8 If PP_n is a sequence o time intervals and if the events P_n converge to an event Q distinct from P, then either PQ is a time interval or Q can recieve a light signal from P.

Finally, let us discuss the notion of *convergence* of events. Choose three distinct noncollinear points O, A and B in the plane. These points define an affine coordinate system with origin O such that, for any point P in the plane we can write the vector $\mathbf{OP}$ as the sum $\mathbf{OP} = x(P)\mathbf{OA}+y(P)\mathbf{OB}$, with $x(P)$ and $Y(P)$ real numbers. We say that a sequence of points P_n converge to a point P if the coordinates $x(P_n)$ and $y(P_n)$ converge to $x(P)$ and $y(P)$, respectively.

We assume that for all pair of real numbers x and y, there is a point P in the plane such that $x = x(P)$ and $y = y(P)$ in this coordinate system.

We recall the so called *completeness* property of real numbers: given a sequence of real numbers r_n such that there are two real numbers $a < b$ such that for all $n \geq 0$, $a < r_n < b$ (we say that the sequence is bounded) , then there is a subsequence r_{i_n} wich converges to some r, $a \leq r \leq b$.

As a consequence, we have the following property of points. We say that a sequence of points P_n is *bounded* if there are real numbers $a < b$ and $c < d$, such that for all n, $a < x(P_n) < b$ and $c < y(P_n) < d$. In this case, there is a subsequence P_{i_n} which converges to some point P with $a \leq x(P) \leq b$ and $c \leq y(P) \leq d$.

3.6 Light cones

The *light cone* of an event P is defined as the union of two sets, one called the *future light cone* consisting of the events which can recieve a light signal from P and the one called the *past light cone* consisting of the events which can send a light signal to P.

If r is an observer through P then any plane containing r represents a 2-dimensional section. We know that there are exactly two light rays in this section going through P. (See Figure 3.7.) Consequently, any plane containing r intersects the light cone in two distinct straight lines.

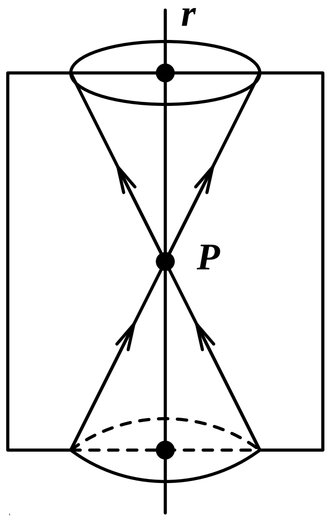

Figure 3.7: Light cone and 2-dimensional section.

The *interior of the light cone of* P is the set of events which can recieve from, or send to, P a light signal.

The set of events X such that PX constitute a time interval is called the *interior of the future cone of* P and the set of events Y such that YP constitute a time interval is called the *interior of the past cone of* P.

The following two properties are easy to prove. (See Figure 3.8.)

1. If U and V are two distinct events in the interior of the future cone of P, then all the points in the segment UV lie in the interior of this cone.

2. If U can recieve a light signal from P and V is in the interior of the future cone of P, then there exists a unique event W which can recieve a light signal from P such that V belongs to the segment UW.

There are analogous properties for the past cones.

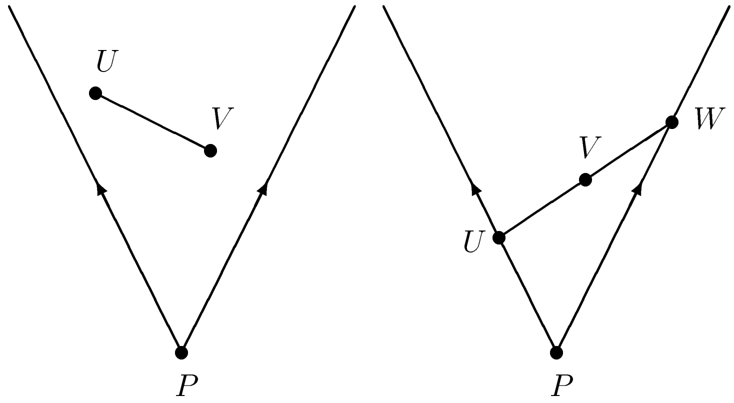

Figure 3.8: Two properties of future cones.

Notice that because of parallelism properties distinct light cones are related by a parallel translation.

3.7 Simultaneity revisited

In the 2-dimensional model of relativity clocks at rest (with respect to each other) were set using the idea of simultaneity of event with respect to observers. Thus clocks at rest are cosidered synchronized if their respective beats are simultaneous. Rephrasing this we say that clocks at rest are synchronized if their n^{th} beats belong to the same space at rest.

Using these ideas we introduce a process of synchronization of clocks in the 3-dimensional spacetime.

Firstly we define the *space at rest* of an event U in a given observer h as the set $E_{U,h}$ (or simply E_U) of points X which are simultaneous to U with respect to h in the sense of the 2-dimensional section determined by h and X. (See section 2.9.)

We impose the following coherence condition on this notion.

Axiom 9 If a, b and c are three distinct observers at rest with respect to one another and A, B and C are events in a, b and c, respectively, such that both A and B are simultaneous to C then A is simultaneous to B. (See Figure 3.9.)

From the definition of simultaneity for the 2-dimensional model we conclude that the line AC is common to E_A and E_C and the line BC is common to E_B and E_C. The coherence condition implies that all events in the line AB are in E_C. From this we easily conclude that E_C is an affine plane. This fact gives the following crucial property of light cones.

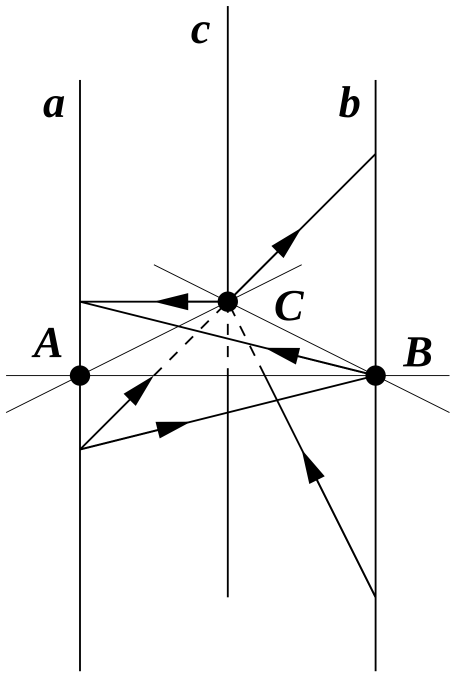

Figure 3.9: Coherence condition.

Let P be an event and U an event posterior to P and let σ_U be the set of events in E_U (the space at rest of U with respect to the observer which contains the time interval PU) which can recieve a light signal from P. (In other words, σ_U is the intersection of E_U with the light cone at P. See Figure 3.10.)

For each X in σ_U its symmetrical event X' (in the 2-dimensional section containing P, X and U) with respect to the observer determined by PU also belongs to σ_U. Thus U is the center of symmetry of the set σ_U.

We can restate this fact thus:

For any event U in the interior of the future cone of P there is a unique plane E_U which intersects the cone in the set σ_U which has U as its center of symmetry.

In the next section we will see that this property allow us to recognize the light cones as they are usually defined in relativity.

3.8 Properties of the light cones

Let P be an event and U an event posterior to P and let σ_U be the set of events in E_U (the space at rest of U with respect to the observer which contains the time interval PU) which can recieve a light signal from P. (In other words, σ_U is the intersection of E_U with the light cone at P. See

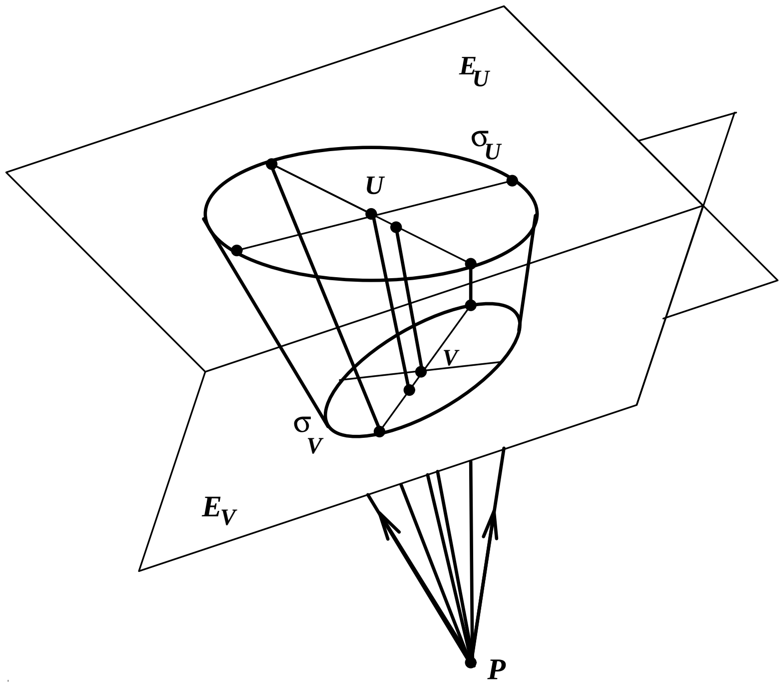

Figure 3.10: Symmetry property of light cones.

Figure 3.10.) Let η_U be the set of events in E_U posterior to P and τ_U the union of the sets σ_U and η_U. This last set has important properties, namely:

- Let s be a straight line in E_U containing a point in η_U. Then the intersection of s and η_U is a segment whose extremities are in σ_U and its interior is contained in η_U.
- If s contains U then U is the middle point of that segment. In this case we say that U is the *center* of τ_U.
- If a line contains two distinct points of σ_U then the line determined by these points meet η_U in a segment.

Our main goal now is to prove that the set of events σ_U is an ellipsis in the plane E_U. This will be done in the following sections.

3.9 Projections

Let α and α' be two (not necessarily distinct) planes in a three dimensional affine space and P a point not contained in these planes. Let X a point in α such that the line PX meets the plane α' in the point X'. We call X' the *projection* of X on α'. If the planes are parallel then any point of the first plane can be projected in the second. Otherwise, the points which cannot be projected are the ones (and only those) which are in the line that is the intersection of α with the plane α'' parallel to α' and contains P. (See Figure 3.11.) We call this line the *horizon* of α with respect to α'. Analogously, we can project the points of α' on α and we can have the horizon of α' with respect to α.

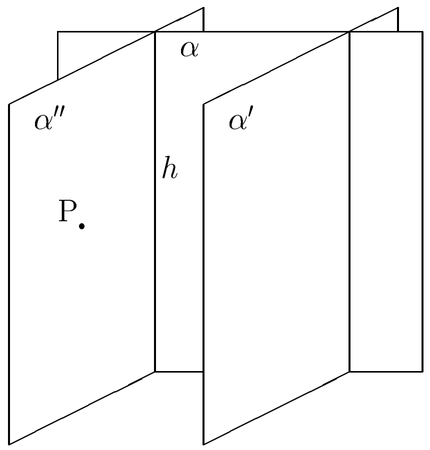

Figure 3.11: The horizon of α with respect to α'.

Henceforward, except otherwise explicitly mentioned, we assume that the points to be projected are not in the horizon of its plane.

Let h and h' be the horizons of α and α', respectively. The projections establish two bijections between the sets obtained from the planes by removing their respective horizons.

The following two claims are easy to verify and are left to the reader:

1. If X, Y and Z are non collinear points of α not in the horizon then their respective projections X', Y' and Z' are non collinear.

2. If $[A, B]$ is a segment of α not meeting the horizon then its projection is the segment $[A', B']$.

Let r be a straight line in α. Then there exists one, and only one, line r' in α' such that each point of r is projected in a point of r'. In this case we say that r' is the projection of r. Two lines r and s have parallel projections if, and only if, both meet at the same point in the horizon or are parallel to the horizon.

3.10 Ellipses

Ellipses are known since Antiquity and was exhaustively studied by the geometers of old Greece. These curves appear also in Astronomy, as orbits of planets, being an essential part of Kepler's laws. In Special Relativity they also appear, related to light cones. We will deal here with some properties of ellipses needed to characterize Lorentz cones.

Among the several definitions of an ellipsis we choose one which is not the most familiar but is the most adequate to our purposes.

We call a subset ε of an affine plane an *ellipsis* if there are three non collinear points A, B and C such that the points X belonging to ε are exactly those points whose coordinates $x(X)$ and $y(X)$ in the coordinate system $(C,\, A,\, B)$ satisfy

$$x^2(X) + y^2(X) = 1.$$

Recall (see section 2.10) that these coordinates are given by the vectorial relation

$$\mathbf{CX} = x(X)\,\mathbf{CA} + y(X)\,\mathbf{CB}.$$

We easily see that A and B belong to ε whereas C does not belong to it. (See Figure 3.12.) It is also easy to see that if X belongs to ε then the point X' whose coordinates are $-x(X)$ and $-y(X)$ belongs to ε. We say that X' is the *symmetrical* to X with respect to C and that C is the *center* of the ellipsis.

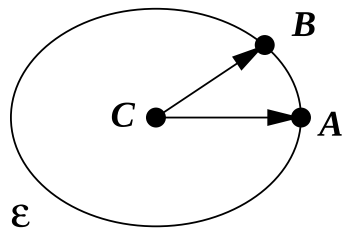

Figure 3.12: Ellipsis determined by the points A, B and C.

Let us consider now some facts about lines tangent to a ellipsis.

Given an ellipsis ε we say that the line r touch the ellipsis in a point V if $r \cap \varepsilon = \{V\}$ (i. e. the only common point is V). If r is unique with such property then we say that V is a *tangent point* and that r is the tangent line to ε at V.

We show now that any point of ε is a tangent point.

The parametric equations of a line through a point V are

$$\begin{aligned} x &= x(V) + v_A\lambda \\ y &= y(V) + v_B\lambda \end{aligned}$$

where the vector $\mathbf{v} = v_A\,\mathbf{CA} + v_B\,\mathbf{CB}$ represents the direction of the line.

Let us study the values of λ for which $x^2 + y^2 = 1$ holds (assuming that V is in ε, that is, $x(V)^2 + y(V)^2 = 1$). We can easily see that the solution $\lambda = 0$ is unique if, and only if,

$$x(V)\, v_A + y(V)\, v_B = 0.$$

In other words, the line tangent to ε at the point V is the line through V whose direction is given by the vector

$$\mathbf{v} = y(V)\, \mathbf{CA} - x(V)\, \mathbf{CB}.$$

Using the tangent lines we can derive some basic properties of ellipses.

We first establish the following fact:

Suppose that an ellipsis ε is given in a coordinate system $(A,\, B,\, C)$ as above. Let V be a point of ε whose coordinates are $x(V)$ and $y(V)$. The point V' whose coordinates are $x(V') = y(V)$ and $y(V') = -x(V)$ also lies in ε. Moreover, the equation of ε in the coordinate system $(C,\, V,\, V')$ is $x^2 + y^2 = 1$ also. (This is a simple calculation and is left to the reader.) Notice that the line CV' is parallel to the line tangent to ε at V.

We have adopted the definition of an ellipsis determined by three noncollinear points A, B and C where the first two points lie in the ellipsis and the third one does not. We will give another description of an ellipsis which is more adequate to our purposes.

Firstly notice that given two parallel lines r and s, the center of any ellipsis tangent to r and s at R and S respectively has the middle point of the segment RS as its center.

Notice also that given a third point W not in r, s or RS, and between r and s, there is a unique ellipsis tangent to r and s at R and S, respectively, and containing W. (See figure 3.13.)

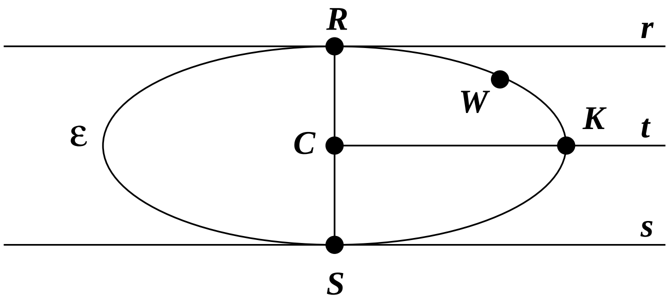

Figure 3.13: Uniqueness of ellipsis.

In order to see this, let t be a line through the center C of the ellipses and parallel to r and s. Consider two of these ellipses, ε and ε' say. Let K and K' be points in t meeting ε and ε' respectively, such that are in the same side of the segment RS. Then we can write $\mathbf{CK'} = k\, \mathbf{CK}$, for some $k > 0$. The equation $x^2 + y^2 = 1$ describe ε and ε' in the coordinate systems $(C,\, S,\, K)$

and $(C,\ S,\ K')$ respectively, and since W is in both of them, we have

$$\begin{aligned} \mathbf{CW} &= w_S\,\mathbf{CS} \ + \ w_K\,\mathbf{CK} \\ \mathbf{CW} &= u_S\,\mathbf{CS} \ + \ u_{K'}\,\mathbf{CK'} \end{aligned}$$

so $u_S = W_s$ and $w_K = k\,u_{K'}$. Therefore $k = 1$ and $\varepsilon = \varepsilon'$.

3.11 A construction of ellipses

In this section we give a geometrical construction of an ellipsis which will be useful later on. We start with the case where the center of the ellipsis is given and then we treat a more general situation.

Given a point F contained in a line f, another point C not in f, let G a point in the line FC such that C is the middle point of the segment FG, and let g be a line parallel to f and through G. Given another point Q between the lines f and g and not in the line FG, then we show that there is a unique ellipsis passing through Q and having tangent lines f at F and g at G.

Let c be the line through C and parallel to f, and D any point in c distinct from C.

We start with the affine coordinate system (C, D, F), with origin C and abscissas axis CD and ordinate axis CF. So the points C, D and F have coordinates $(0, 0)$, $(1, 0)$ and $(0, 1)$, respectively. Let (a, b) be the coordinates of Q. Notice that $a \neq 0$ and $-1 < b < 1$, by the hypotheses on Q.

We describe now the constructions pictured in Figure 3.14 and write the relevant equations.

An ellipsis has equation $x^2/\alpha^2 + y^2/\beta^2 = 1$ for some $\alpha, \beta > 0$ in a coordinate system. In our case, the ellipsis has the equation

$$\frac{x^2}{a^2/(1-b^2)} + y^2 = 1$$

and f, the tangent line at F, has equation $y = 1$.

Let X be a variable point in the segment $]F, G[$ and let $(0, \xi)$ be its coordinates.

The line FQ has equation

$$y - 1 = \frac{b-1}{a}\,x.$$

The line FQ and the line $y = \xi$ (parallel to the line f) meet at the point R with coordinates $x(R) = a(\xi - 1)/(b - 1)$ and $y(R) = \xi$.

The line GQ has equation

$$y + 1 = \frac{b+1}{a}\,x.$$

The line GQ and the line $y = \xi$ meet at the point S with coordinates $x(S) = a(\xi + 1)/(b + 1)$ and $y(S) = \xi$.

The line FS has equation

$$y - 1 = \frac{(\xi - 1)(b+1)}{a(\xi + 1)} x.$$

The line GR has equation

$$y + 1 = \frac{(\xi + 1)(b-1)}{a(\xi - 1)} x.$$

The lines FS and GR meet at the point T with coordinates

$$x(T) = \frac{a(1-\xi^2)}{\xi^2 - 2b\xi + 1}$$

and

$$y(T) = \frac{-b\xi^2 + 2\xi - b}{\xi^2 - 2b\xi + 1},$$

which satisfy the equation

$$\frac{x(T)^2}{a^2/(1-b^2)} + y(T)^2 = 1.$$

The line QX has equation

$$y\xi = \frac{b - \xi}{a} x.$$

The line QX meets the line $y = y(T)$ (parallel to the line f) at the point Q' with coordinates

$$x(Q') = \frac{a(\xi^2 - 1)}{\xi^2 - 2b\xi + 1}$$

and

$$y(Q') = \frac{-b\xi^2 + 2\xi - b}{\xi^2 - 2b\xi + 1},$$

which also satisfy the equation

$$\frac{x(T)^2}{a^2/(1-b^2)} + y(T)^2 = 1.$$

3.12 Convex plane sets

Given two distinct points P and Q, the open (respectively, closed) *ray* determined by these points is the set $]PQ)$ (respectively, $[PQ)$) of points X such that the vector $\mathbf{PX} = \lambda\mathbf{PQ}$, with $\lambda > 0$ (respectively, $\lambda \geq 0$). We say that P is the origin of the ray.

Given two distinct points P and Q we can define the segments

$$]PQ[=]PQ) \cap]QP), \quad [PQ] = [PQ) \cap [QP),$$

$$[PQ[= [PQ) \cap]QP), \quad]PQ] = [PQ) \cap [QP).$$

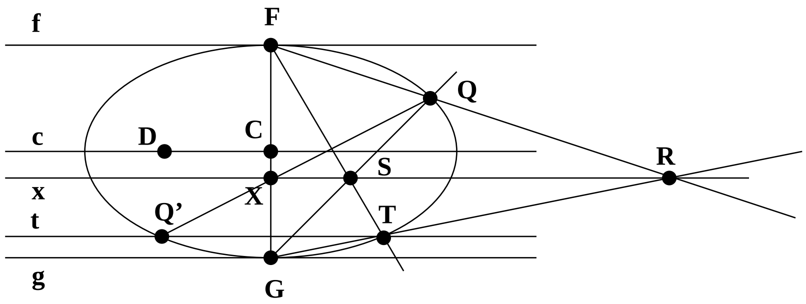

Figure 3.14: The construction of the ellipsis.

If, instead of two points P and Q we are given a point P and a non null vector $\mathbf{v}$ we define the the open (respectively, closed) *ray* determined by P and v as the set $]P, \mathbf{v})$ (respectively $[P, \mathbf{v})$) of the points X such that $\mathbf{PX} = \lambda \mathbf{v}$ for $\lambda > 0$ (repectively, $\lambda \geq 0$).

The straight line PQ is called the *support line* of any of the rays and segments introduced above.

Given a line r and apoint Q not in r (or a vector $\mathbf{v}$ not parallel to r) we define the open (closed) semiplanes determined by the pair (r, Q) (or, respectively, by the pair $(r, \mathbf{v})$) as the set obtained by the union of the open (closed) rays determined by the points P in r and Q (or, respectively, P and $\mathbf{v}$).

Given three noncollinear points O, P and Q, the *angular region* defined by the open rays $]OP)$ and $]OQ)$ is the set of all point X such that there is a point V in $]OP)$ and W in $]OQ)$ such that X is in the open segment $]VW[$. Given two distinct lines r and s, both containing the point O, there are four angular regions determined by them, that is, let P and P' in r and Q and Q' in s be such that O is inside the segments $]PP'[$ and $]QQ'$. Then the four pairs of rays $]OP)$ and $]OQ)$, $]OP')$ and $]OQ)$, $]OP)$ and $]OQ')$ and $]OP')$ and $]OQ')$ determine the four angular regions.

We define a (plane, bounded, open) *convex set* as a set U in a plane α such that for all lines r in α, either r does not intersect U or it intersects U in an open segment $]PQ[$. The points p and Q of this segment are called *boundary points* of U.

Let us describe two important examples.

Firstly, let U be interior of the paralellogram with vertices A, B, C and D, where a point X is called an interior point if the line through X and parallel to the side AB meets the sides AD in V and BC in W and X is in the open segment $]VW[$. (See Figure 3.15.)

The second example is the interior of an ellipsis , wich is the union of all the open segments with extremities in the ellipsis. Observe that the center of the ellipsis lies in its interior.

Observe also that the boundary of the interior of the paralellogram (ellipsis) is the paralellogram (ellipsis).

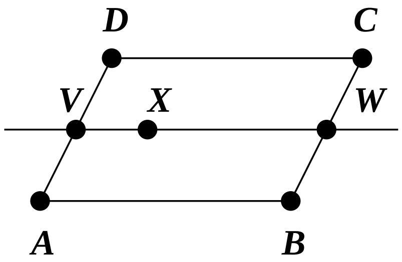

Figure 3.15: Interior of a paralellogram.

Convex sets, as defined here, share an interesting property, which will be relevant later on. Let X be a point in the convex set U and let r and s be two distinct lines containing X. Then these lines intersect U in the open segments $]PQ[$ and $]RS[$, respectively, and X belongs to both segments. Choose points A and C in $]PQ[$ and B and D in $]RS[$ such that X is the middle point of both intervals $]AC[$ and $]BD[$. (See Figure 3.16.) Then the interior of the paralellogram $ABCD$ is contained in U as the readers can verify by themselves.

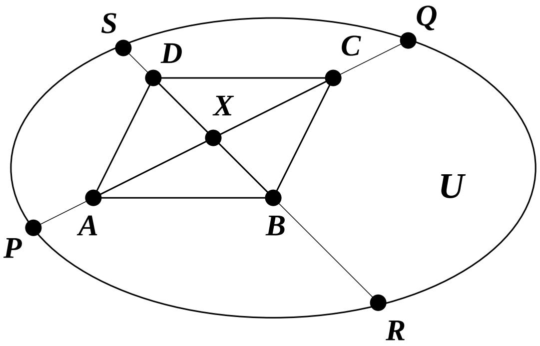

Figure 3.16: Paralellogram around X.

As a first consequence to this result is that convex sets cannot contain any of its boundary points.

Observe that in the case of the paralellogram $ABCD$, whose interior we will call U, the points A and B, for instance, are two boundary points such that the whole segment$[AB]$ consists of boundary points of U. The same does not happen with the ellipsis. If U is the interior of an ellipsis ε and A and B are two distinct boundary points of U, then the open segment $]AB[$ is contained in U.

Now, let U be a convex region and consider a sequence of points P_n in U converging to a point F not in U. Then F must be on the boundary of U. In order to see this, choose a point O in U and a line r through O and distinct from the line OF. Let W and W' be points of U lying in r and such that O is in the open segment $]WW'[$. We define a new sequence of points Q_n in U as follows. If P_n lies in the line OF then let Q_n be P_n. Otherwise, exactly one of the open segments $]WP_n[$ or $]W'P_n[$ intersect the line OF in a point which we call Q_n (see Figure 3.17). The sequence Q_n is in U and also converges to F. Thus F must be a boundary point.

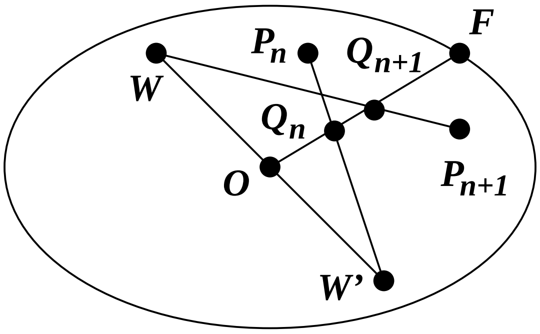

Figure 3.17: Convergence to the boundary.

Now let us prove that if V is a bounded open convex set then there is a parallelogram $ABCD$ such that V is contained in the interior of $ABCD$. In order to see this, choose a point U inside V and two distinct lines r and s through U. The following constructions are depicted in Figure 3.18. Choose points R and R' in r and S and S' in s, such that the point U is the middle point of the open segments $]RR'[$ and $]SS'[$. Let $R_1 = R$, $R'_1 = R'$, $S_1 = S$ and $S'_1 = S'$. Let R_n, R'_n, S_n and S'_n be the points in the lines r and s, respectively, such that $\mathbf{UR}_n = n\mathbf{UR}_1$, $\mathbf{UR'}_n = n\mathbf{UR'}_1$, $\mathbf{US}_n = n\mathbf{US}_1$ and $\mathbf{US'}_n = n\mathbf{US'}_1$. Let K_n be the parallelogram $R_nS_nR'_nS'_n$. Suppose that the stated result is not true. So, for each n there is a point U_n in V not in the interior of K_n. Let T_n be the meeting point of the line UU_n and the parallelogram $RSR'S'$.

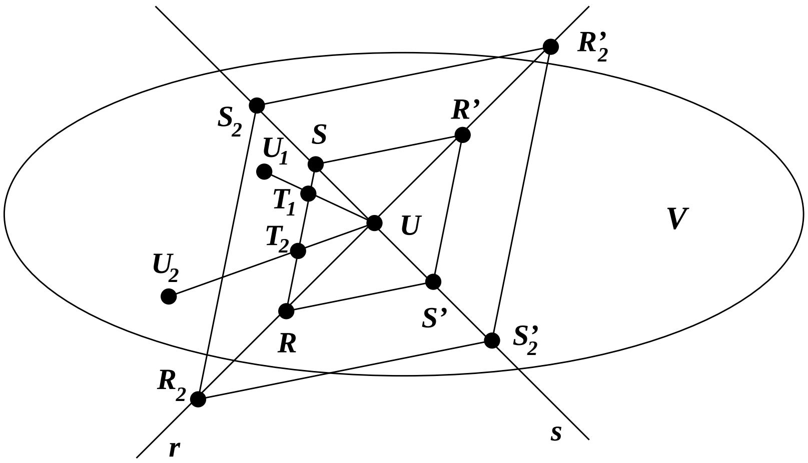

Figure 3.18: Proof of convex inside a parallelogram.

Since there are infinitely many points U_n and only four sides in $RSR'S'$, there is a side, RS say, that contains T_n for infinitely many $n > 0$. The sequence T_n is bounded, so there is subsequence T_{i_n} in the segment $[RS]$ which converges to a point T in this segment. The line UT contains the point U which is inside V and so it meets V in the open segment $]FF'[$, for some boundary points F and F'. We may assume that the point F lies in the open ray $]UT)$. Let G be a point in the ray $]UT)$ not in the closed segment $[UF]$. Thus $UG/UF > 1$. Let g be the line through G and parallel to RS. Let G_n be the meeting point of g and UT_{i_n}. Then G_n converge to G. By the definition of the points T_{i_n} and G_n, there is some n_0 such that, for all $n > n_0$, the point G_n is in the open segment $]UU_{i_n}[$ and therefore, the G_n, $n > n_0$, all belong to V (by convexity). By the comments above, either G is a boundary point of V or lies in V. In both cases, since F is in the open segment $]UG[$, F should be inside V, a contradiction. This proves that there must be some n such that V is in the interior of K_n.

Support lines

Now we will treat the main topic of this section.

Let F be a boundary point of the convex set V. A line s through F which intersects V is called a *secant line*. If s does not intersect V then we call s a support line. If the line s is the unique support line through F, then s is called *tangent line* and, in this case, F is called a *tangency point*.

Clearly through any boundary point F of the convex V there is at least one secant line FP, for P inside V. We are going to prove that through any boundary point F of a convex set V either there is a tangent line s through F or two distinct support lines s and s' such that there are point S in s and S' in s' distinct from F such that all rays $]FX)$ for X in the segment $]SS'[$ intersect V and no other ray intersect V. In the latter case, we say that F is a *vertex* of V.

Let F be a boundary point of V and suppose that it is not a tangency point.

Let U be a point in V, u a line through U distinct from the line UF. let W and W' be points of V lying in u and such that U is in $]WW'[$. Let w and w' be the lines through W and W', respectively, and parallel to UF. Let H and H' be the meeting points of the lines w and $W'F$, and w' and WF, respectively. If X is a point in the semiplane (UF, W), let X_w be the common point of the lines w and FX. Let P_w be the set of all points X_w, for X in the ray $]UW)$.

The readers can easily verify that H does not belong to P_w; if J is any point of U in the semiplane, then the ray $(J_w, \mathbf{FU})$ is contained in P_w; P_w is contained in the ray $(H, \mathbf{FU})$; for each point Y_w in P_w, there is a point Z_w in P_w which is in the open segment $]HY_w[$.

Now we claim that there is a point S in w such that P_w is contained in the ray $(S, \mathbf{FU})$ and for all other point $\bar{S}$ in w such that the ray $(\bar{S}, \mathbf{FU})$ contains P_w, then it also contains the ray $(S, \mathbf{FU})$. In order to see this, let $S_0 = H$ and $T_0 = W$, and for all $n \geq 0$ let R_n be the middle point of $]S_nT_n[$. If R_n is in P_w then let $T_{n+1} = R_n$ and $S_{n+1} = S_n$. Otherwise, let $S_{n+1} = R_n$ and $T_{n+1} = T_n$. Then

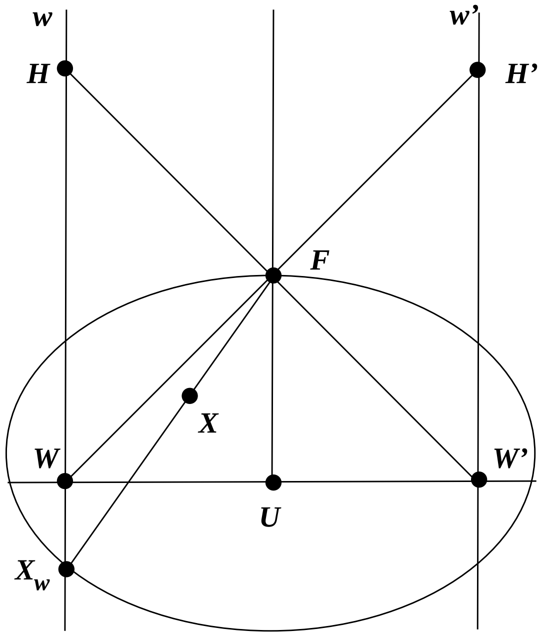

Figure 3.19: Proof of existence of support lines.

both sequences are bounded and it is easy to see that both converge to the same point S in $[HW]$. Firstly notice that S does not belong to P_w, by the properties listed above. If X_w is a point of P_w, then for all n, X_w belongs to the rays $(S_n, \mathbf{FU})$, so it must belong to the ray $(S, \mathbf{FU})$. Now let Z be a point of w such that P_w is contained in the ray $(Z, \mathbf{FU})$. Then this ray contains all the rays $(T_n, \mathbf{FU})$, and so it must contain the ray $(S, \mathbf{FU})$. Therefore P_w must be the ray $(S, \mathbf{FU})$.

Analogously, we can define the set P'_w in the line w' and obtain the point S' such that P'_w is the ray $(S', \mathbf{FU})$.

Let s and s' be the lines FS and FS', respectively. We now show that these are the support lines we are looking for.

Let K be the common point of FU and SS'. We claim that K is in the open ray $(F, \mathbf{FU})$. We will prove this by showing that it cannot belong to the open ray $(F, \mathbf{UF})$, and that $K \neq F$.

Let f be the line through F and parallel to SS'. Let f be the line through F and parallel to SS'. Let L and L' be the common points of f and w, and f and w', respectively. (See Figure 3.20.)

If K were in the open ray $(F, \mathbf{UF})$ then the points S and S' would be in the semiplane $(f, \mathbf{UF})$ and so L would be in the ray $(S, \mathbf{FU})$ and L' in the ray $(S', \mathbf{FU})$. that is, L and L' would be in P_w and P'_w, respectively, and so in V. Therefore, F (which lies in $]LL'[$) would be in V, a contradiction.

If $K = F$, then $s = s' = SS' = LL'$. So this line is a tangent to V, again a contradiction.

So K is in the open ray $(f, \mathbf{FU})$. This implies that neither of the lines s and s' meets V. Now let V be a point in the ray $]FS)$ and W a point in the ray $]FS')$ and X a point in the open segment

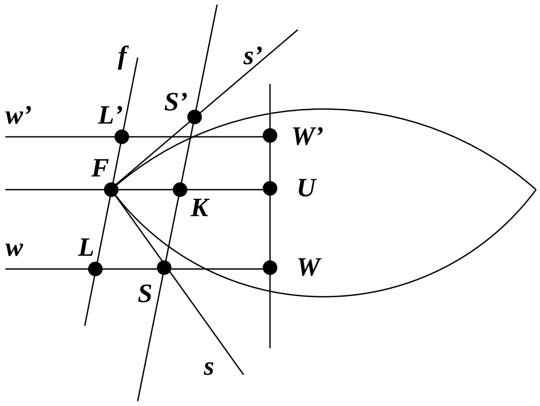

Figure 3.20: Proof that s and s' are the support lines.

$]VW[$. Then the ray $]FX)$ meets V, by the definition of the sets P_w and P'_w. Also we can conclude that the convex V is contained in the angular region determined by the rays $]FS)$ and $]FS')$.

3.13 Convex sets with symmetry

Given a point C in an affine plane, for each point X of that plane we can define the *symmetric point* of X with respect to C as the point X' such that the vectors $\mathbf{CX}$ and $\mathbf{X'C}$ are equal. So X' is the point such that C is the middle point of X and X'. We define the *symmetric set* of a set U with respect to C as the set of symmetrical points of the points X in U with respect to C.

A point C is called a *center* of a set U if U is its own symmetric set with respect to C.

Now let U be a (bounded) convex set with a center C. If D is also a center of U, then $D = C$ because, otherwise, letting $D_1 = D$ and supposing that we have defined D_n $(n \geq 1)$, we define D_{n+1} as the symmetrical point of K_n with respect to C, K_n being the symmetrical point of D_n with respect to D (see Figure 3.21). These points belong to the same line r and cannot be all contained in a bounded segment. So they cannot be all inside U, a contradiction.

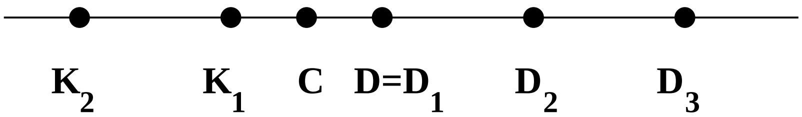

Figure 3.21: The construction of sequence of points D_n.

From the definition of support and tangent lines to U, we see that the symmetrical of support lines are support lines, and the symmetrical of a tangent line is a tangent line (which is parallel to the first one).

Now consider a convex set U with center C and let F be a boundary point of U such that there are two distinct support lines a and b through F. Let G be the symmetric to F with respect to C. Then the lines a' and b' symmetric to a and b respectively with respect to C are the support lines to U through G. (See Figure 3.22.) The lines a and a' (respectively b and b') are parallel. Let M be the meeting point of a and b' and N the meeting point of a' and b. Then the points F, G, M and N form the vertices of a parallelogram with center C, so M and N are symmetric with respect to C.

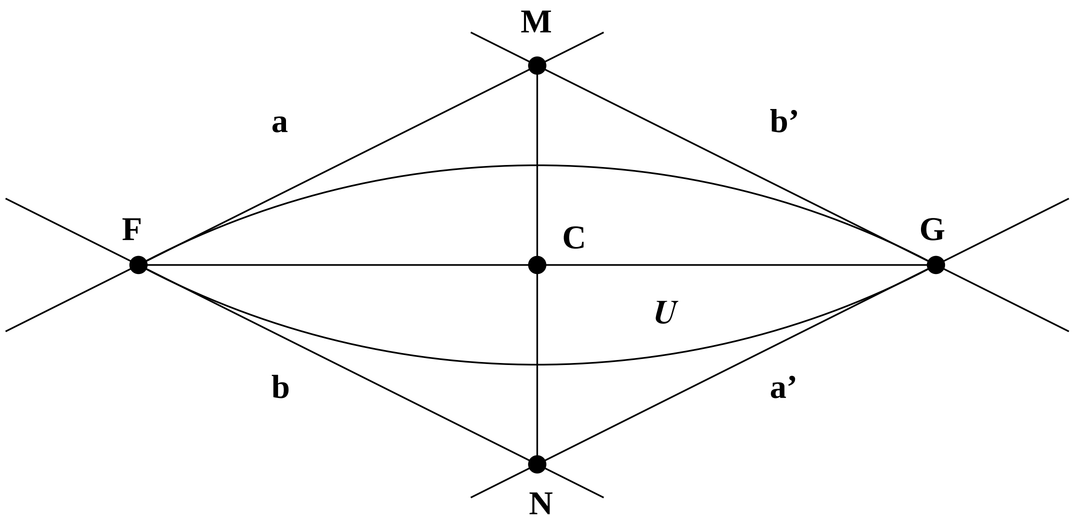

Figure 3.22: Symmetry of intersections of symmetric lines.

3.14 Totally symmetric convex sets

Let us fix a point of a three dimensional affine space P and a plane α in this space not containing P. This point P is intended to be the vertex of several projections.

Let U be a convex set in α. We say that U is *totally symmetric* if for each point X in U there is a plane α_X not containing P such that

1. all points in U and its boundary are projected in α_X;

2. the point X is projected in the center X' of the image U' of U. (See Figure 3.23.)

If h_X denotes the horizon line of α for such projection, the first condition says that it does not touch neither U nor its boundary points.

The **main result** of this section is to prove that U must be an ellipsis.

As a first step, let us prove that every boundary point of U is a tangency point.

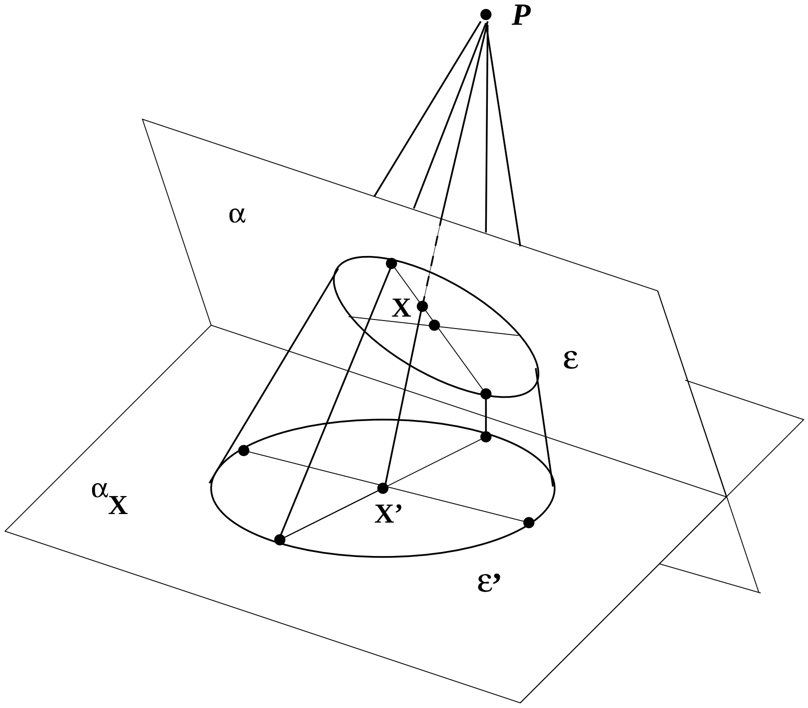

Figure 3.23: Projection of totally symmetric convex set.

Let C be the center and F be a boundary point of U. Let G be the symmetric of F with respect to C. Suppose that F is not a tangency point, so there are two distinct support lines a and b through F and let a' and b' be their respective simmetric lines. Let M be the meeting point of a and b' and N the meeting point of a' and b (refer back to Figure 3.22).

Choose a point D in the open segment $]F, C[$. Since U is totally symmetric, there is a plane α_D not containing P such that the projection D' of D is the center of the projection U' of U in α_D. Observe that α and α_D cannot be parallel because of the uniqueness of the center. So there is a line h_D which is the horizon line in α of this projection. This line does not intersect U, so this line cannot contain both M and N. If it contains one of them, M say, then the projections $\bar{a}$ and $\bar{b}'$ of a and b' respectively, in α_D, must be parallel. Let $\bar{N}$, $\bar{a}'$ and $\bar{b}$ be the projections of N, a' and b respectively. Then $\bar{N}$ is the meeting point of $\bar{a}'$ and $\bar{b}$. On the other hand, D' is the center of U', and by the choice of D, the projections F' and G' of F and G, respectively, are symmetric with respect to D'. Since support lines are projected onto support lines, by symmetry, $\bar{b}$ should be parallel to $\bar{b}'$ and $\bar{a}$ to $\bar{a}'$. But this implies that $\bar{a}'$ and $\bar{b}$ should also be parallel, an absurd. So F is a tangency point and, by symmetry, G is a tangency point and the tangent lines through F and G are parallel.

Now we are ready to prove the main result.

For the following constructions refer to Figure 3.24.

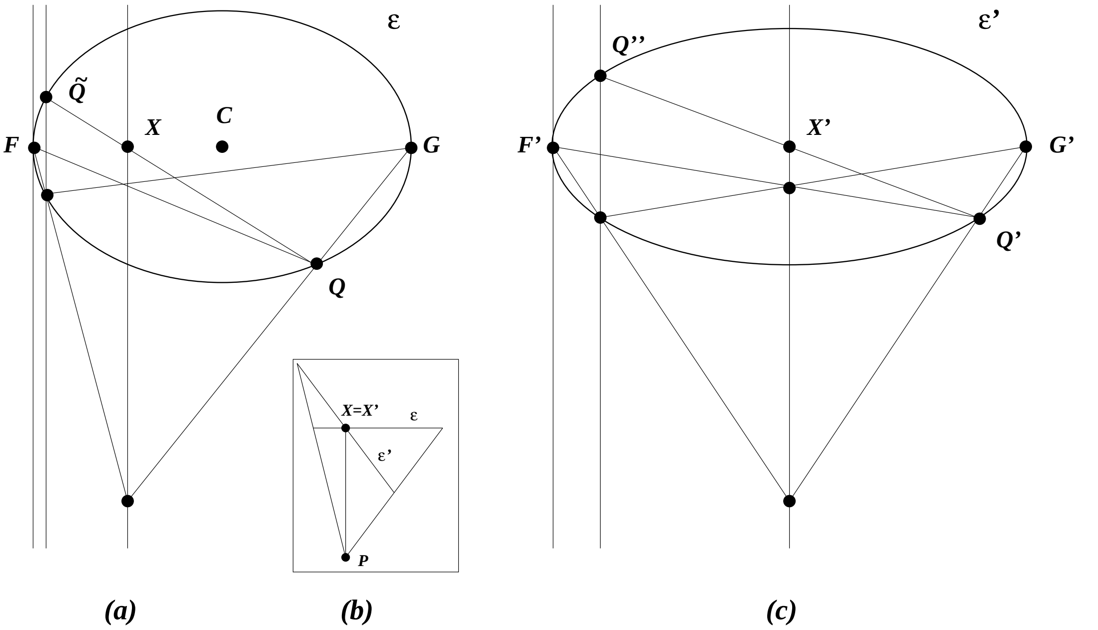

Figure 3.24: Projection of the construction of the ellipsis: (a) plane α, top view; (b) profile (side view); (c) plane α_X, top view.

Let C be the center and F be a boundary point of U and f the tangent to U at F. Let G and g be the symmetric of F and f, respectively, with respect to C. Let Q be another boundary point of U distinct from F and G. Let ε be the ellipsis passing through F and Q, having f as tangent at F and C as center, as constructed in Section 3.11 (see also Figure 3.14). Let X be any point in the segment $]F,G[$ and distinct from C. Let α_X be a plane not containing P (as above) and such that the projection U' of U has center X', the projection of X. Let Q' be the projection of Q and Q'' its symmetric with respect to X'. Observe that the horizon in the plane α of this projection is parallel to f and so the construction of the ellipsis ε projects in the construction of a unique ellipsis ε' in the plane α_X with center X', the tangent at F' being the projection f' of f, and passing through Q', so it also passes through Q''. But ε' is the projection of ε and Q'' lies in both ε' and the boundary of U' (because X' is the center of U'). So its preimage $\tilde{Q}$ lies in both ε and the boundary of U. Now, varying the point X, and keeping fixed Q the varying points $\tilde{Q}$ obtained are in both ε and the boundary of U. With this we prove the coincidence of these sets.

3.15 Euclidean geometry given by an ellipsis

The ellipses which are produced by the intersections of the light cones with the planes at rest of an observer can be used to develop the Euclidean geometry of that plane. We can regard these ellipses as the circles of the plane and will call them *circles*. These represent the compass we use to make the usual constructions of Euclidean plane geometry.

As an example, let us describe the construction of the perpendicular line s to a given line r at a given point A in r. (See the Figure 3.25.) Choose a *circle* centered in A. It will meet r in two points. Let B be one of them and take the translated *circle* with center B. Then this *circle* will meet r in two points, one of them being A and will meet the first *circle* in two points, one of them we call C. Again, take the translated *circle* with center C. It will meet the previous *circle* in two points, one of them being B and the other we call D. Finally, take the translated *circle* with center D. It will meet the previous *circle* in two points, one of them being A and the other we call E. Then the line AE is the line we wanted.

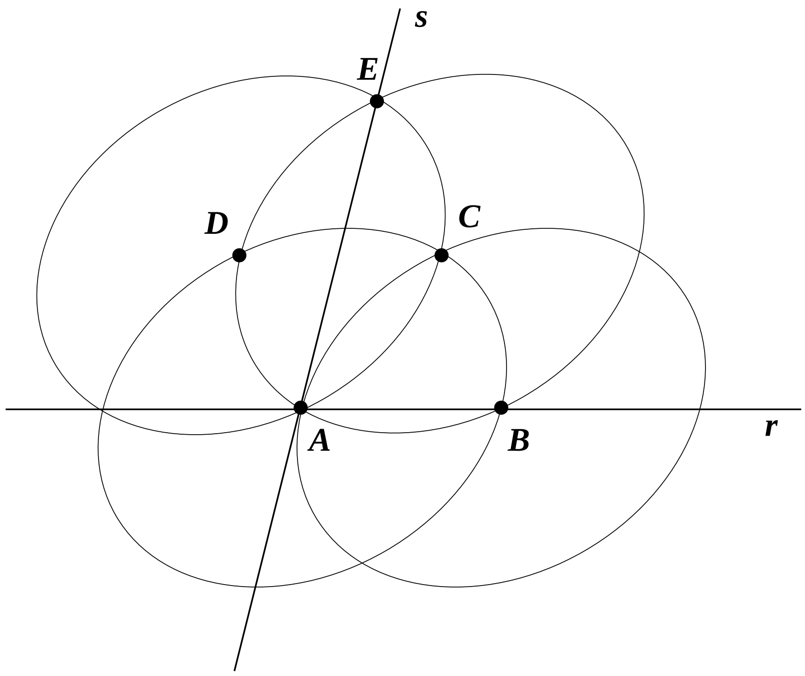

Figure 3.25: Construction of the perpendicular.

As another example, let us describe the construction of the bissectrix of an angle $A\hat{O}B$ (as in Figure 3.26). Again we start choosing one of the *circles* centered in O. It will meet the line OA in the point C and OB in the point D. Draw a *circle* centered in C and other in D, both meeting in two points such that the line determined by these points contain the point O (in the Figure 3.26

these *circles* meet in O). Let E be one of these meeting points, distinct from O. Then the line OE is the sought bissectrix.

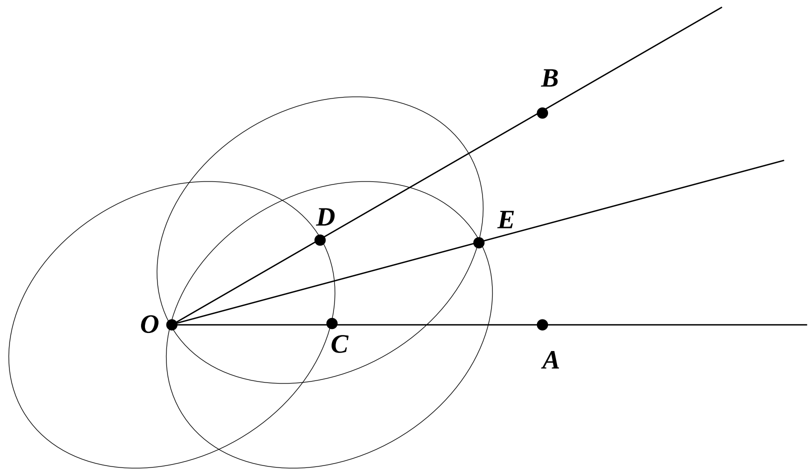

Figure 3.26: Construction of the bissectrix.

Index