

# Extremal problems in geometry

In this lecture we will study problems in geometry, where the goal is to find a configuration that minimizes or maximizes a certain quantity. For instance we will answer the following question: find the point  $P$  in the plane for which the sum of the distances to the vertices of a given triangle is minimal.

## 0.1 Physical and mathematical arguments

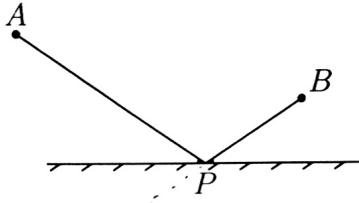
First we will study the following question from different perspectives: find a point  $P$  on a given line  $l$  so that the sum of the distances from two given points  $A$  and  $B$  in the plane to  $P$  is minimal.

Such questions arise naturally in the study of geometric optics — light rays move along paths that extremize the time to get from the initial point to the final point. This experimental fact is called “Fermat’s principle”.

Let’s look at the light ray that gets reflected from a mirror  $l$ :

If  $A$  and  $B$  are two points on the path of the ray and the light first goes through  $A$ , then gets reflected in  $l$  at point  $P$  and then arrives at  $B$ , we see that the path of the light ray  $APB$  should, according to Fermat’s principle, solve the extremal problem we formulated. Indeed, the time it takes the light to get from  $A$  to  $P$  to  $B$  is proportional to the distance  $APB$ , so the point of reflection  $P$  should extremize the sum of the distances  $|AP| + |PB|$ . We also know another experimental fact — when light ray gets reflected in a mirror, the angle of incidence is equal to the angle of reflection.

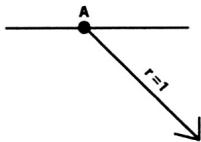
Since mathematics deals with models of reality, rather than with reality itself, arguments that rely on experimental observation cannot be considered mathematically rigorous. Moreover, mathematics allows one to notice that different experimental observations might be related to each other: for instance the fact that reflection angle equals the incidence angle follows from Fermat’s principle.



Before giving a mathematically rigorous proof of this, we will present another physical argument.

Consider a string, which exerts unit force on any object attached to it:

**FIGURE 2**

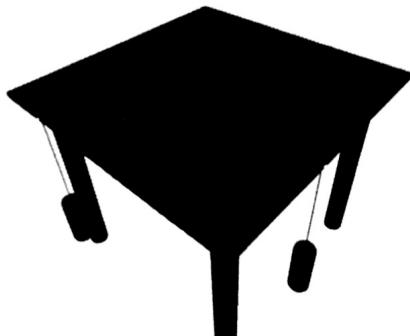


One can check that for such a string the potential energy stored in it, when it is stretched between points  $A$  and  $P$  is proportional to the distance between the points  $A$  and  $P$ .

Now consider such a string attached to point  $A$ , going through a small ring sliding freely on line  $l$  and ending at point  $B$ . Such system should have some stable point. At this stable point the energy, which is proportional to the length  $AP + PB$ , should be minimized.

On the other hand the force exerted on the ring at the stable point should be perpendicular to the line  $l$  (if it had some component in the direction of the line  $l$ , it would slide in this direction). Since this force is the sum of two unit vectors pointing along  $PA$  and  $PB$ , it points in the direction of the bisector of the angle  $APB$ .

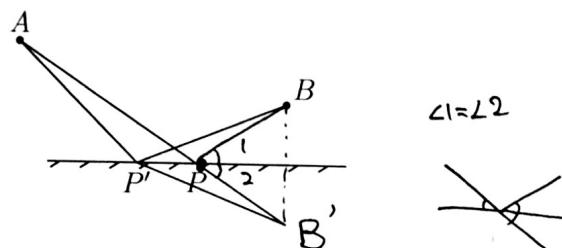
The conclusion is then that at the point  $P$  on line  $l$  that minimizes the sum  $AP + PB$ , the angle bisector of angle  $APB$  is perpendicular to the line  $l$ . We could make this argument mathematically precise by defining the notions of "forces", "potential energy", "stable point" and proving relevant theorems,



but this takes some time and effort. Instead we will treat the problem using the mathematical tools we already have.

Consider again two points  $A$  and  $B$  lying in the same half-plane defined by a line  $l$ . Let  $B'$  be the reflection of point  $B$  in line  $l$ .

For any point  $P$  on the line  $l$ , the distance  $PB$  is the same as the distance  $PB'$  (by symmetry).



Hence the distance  $AP + PB$  is the same as the distance  $AP + PB'$ . Now the latter distance is minimized at the point of intersection of segment  $AB'$  with the line  $l$  - the shortest way to go from  $A$  to  $B'$  is in a straight line. Thus we not only proved that at the point  $P$  on  $l$  which minimizes  $AP + PB$ , the angle between  $AP$  and  $l$  and the angle formed by  $PB$  and  $l$  are equal, but we also gave a very simple way to construct this point: reflect point  $B$  in line  $l$  to get point  $B'$  and connect  $B'$  to  $A$  by a straight line segment. The intersection of  $l$  and this line segment  $AB'$  is the desired point  $P$ .

Even though we have solved our problem completely, we won't stop here. Instead we will give one more solution, which shows an idea with a wide range of applications. The idea should be familiar from calculus - to extremize a

function one should look at points where its derivative vanishes.

Before we proceed further, we will remind the reader of the multidimensional notion of derivative - the gradient.

$$\nabla f(\vec{x}) = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

Consider a function  $F$  of  $n$  variables  $x_1, x_2, \dots, x_n$ . Consider also a curve  $\vec{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ . The function  $F$  restricted to this curve becomes the function  $F(x_1(t), \dots, x_n(t))$  of a single variable  $t$ . Its derivative  $\frac{dF(x_1(t), \dots, x_n(t))}{dt}$  is equal to  $\frac{\partial F}{\partial x_1} \frac{dx_1(t)}{dt} + \dots + \frac{\partial F}{\partial x_n} \frac{dx_n(t)}{dt}$ . This expression is equal to  $\langle \nabla F, \dot{\vec{x}}(t) \rangle$ , the scalar product of the gradient of  $F$  (the vector  $\nabla F = (\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n})$ ) evaluated at the point  $x(t)$  and the velocity vector  $\dot{\vec{x}}(t)$  of the curve  $\vec{x}(t)$  at time  $t$ .

In particular if the curve lies in the level set  $F(x) = c$ , then its velocity vector is everywhere perpendicular to the vector  $\nabla F$ . In short, the gradient of the function is always perpendicular to its level set.

Also if the restriction of the function  $F$  to the curve  $x(t)$  has an extremum, then the gradient of  $F$  is perpendicular at this point to the curve.

Now consider a function of point  $P$  defined by  $F(P) = |AP| + |PB|$ . Its gradient is the sum of the unit vectors in directions from  $A$  to  $P$  and from  $B$  to  $P$ :  $\nabla F = \frac{\vec{AP}}{|AP|} + \frac{\vec{BP}}{|BP|}$ . At the point  $P$  where the minimum of the restriction of  $F$  to line  $l$  is obtained, this gradient should be orthogonal to the line  $l$ .

This is the mathematical reformulation of the physical argument about the strings.

Note that the solution we found using gradients is not complete - all it gives is a condition for the sum of distances  $AP + PB$  to have a local extremum at a point. The geometric solution using reflection gives us the full answer - we find that there is (only one) extremal point which is the global minimum. This relative weakness of the method from calculus is explained by the fact that it is very general - it applies to a vast amount of extremization problems. The geometrical method of reflections is specific to a narrower class of problems, where it gives more precise results.

Before we switch our attention to conical sections, we will give several applications of the geometrical method we've seen.

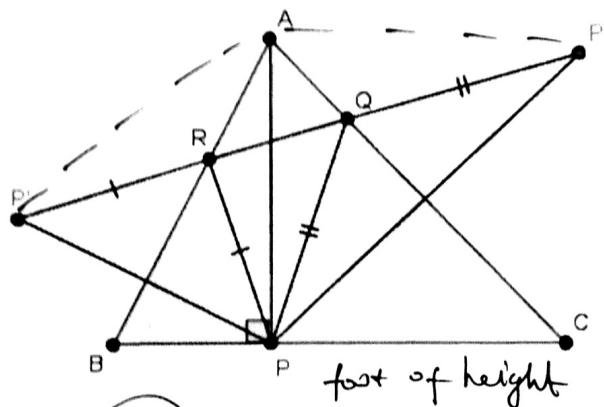
Our first example will be the following problem proposed and solved by Laurent Schwartz, a famous 19th century analyst: given an acute triangle  $ABC$  find the triangle inscribed into it with minimal perimeter, i.e. find points  $P, Q, R$  on the sides  $BC, CA, AB$  respectively which minimize the sum  $PQ + QR + RP$ .

First of all, let's see what the calculus method gives us: if  $P, Q, R$  is the

configuration that extremizes the perimeter of  $PQR$ , then in particular the point  $P$  is the point on line  $BC$  that extremizes the sum  $QP + PR$ , the points  $Q$  and  $R$  being kept fixed. We know that for such a point  $P$  the angles  $\angle RPB$  and  $\angle QPC$  must be equal. Similarly we find that the angles  $\angle PQC$  and angle  $RQA$  must be equal as well as angles  $\angle QRA$  and  $\angle PRB$ .

While this is already something, this information isn't enough to locate the points  $PQR$ .

So we will proceed differently. First we will solve a simpler problem - suppose that we are given the location of point  $P$  on  $BC$ . How can we find the location of the points  $R$  and  $Q$ ? We know what to do: we should reflect the point  $P$  in lines  $AB$  and  $AC$  to get points  $P'$  and  $P''$ . For any point  $R$  on the line  $AB$  the distances  $RP$  and  $RP'$  are equal to each other and likewise for any point  $Q$  on  $AC$ ,  $QP = QP''$ . Hence the sum  $PQ + QR + RP$  is equal to the length of the broken line  $P'RPQP''$ . This length is obviously minimal when the points  $R$  and  $Q$  are on the line  $P'P''$ .



$$\begin{aligned} PQ + QR + RP &= PR + RQ + QP' < P'A + AP'' \\ &= 2PA \end{aligned}$$

so  $PA \downarrow$ , sum  $\downarrow$   
 $PA \downarrow \Rightarrow$  height,

so  $P$  is the foot of  
height.

So for a given point  $P$  we found an answer: to construct  $R$  and  $Q$  that minimize the perimeter  $RPQ$  we should reflect point  $P$  in lines  $AB$  and  $AC$  to get points  $P'$  and  $P''$  and intersect the segment  $P'P''$  with segments  $AB$  and  $AC$ . The intersection points are the points  $R$  and  $Q$  we were looking for. Moreover, the perimeter in this case is equal to the length of the segment  $P'P''$ .

Now we should understand how this answer varies as we move the point  $P$  and find the point  $P$  for which it is minimal.

Notice that for any choice of point  $P$  the lengths  $AP$ ,  $AP'$  and  $AP''$  are equal. This follows because  $P'$  and  $P''$  are both reflections of  $P$  in lines that

Key!

pass through the point  $A$ . Also note that  $\angle BAP = \angle BAP'$  and  $\angle CAP = \angle CAP''$  (again this follows because  $P'$  and  $P''$  are reflections of  $P$  in  $AB$  and  $AC$ ). Thus for any choice of point  $P$  the angle  $\angle P'AP''$  is the same and is equal to  $2\angle BAC$ . So all the triangles  $P'AP''$  for all the possible choices of point  $P$  are similar to each other. The smallest of them corresponds to the location of  $P$  that minimizes the distance  $AP$ , i.e. to the foot of the height dropped from the vertex  $A$  of triangle  $ABC$ .

The answer we found is clearly unique - we introduced an explicit construction of it. But we could perform the same kind of construction with point  $Q$  or  $R$  instead of  $P$ . Since the answer must be the same, we can conclude that the triangle  $PQR$  that solves our problem is the triangle whose vertices are the feet of the altitudes of triangle  $ABC$ .

Note that the information about the angles that calculus gave us proves an amusing fact about triangle  $PQR$  without any angle chasing - the heights  $AP$ ,  $BQ$ ,  $CR$  of triangle  $ABC$  are angle bisectors in the triangle  $PQR$ . We can prove another nice and simple result using nothing else but the fact that a straight line is the shortest distance between two points: if a convex polygon lies inside another one, then the perimeter of the first one is smaller than the perimeter of the second one.

Note that without the convexity assumption the result doesn't hold:

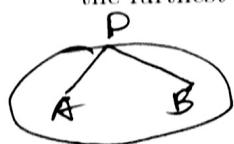
In higher dimensions one can prove that the surface area of a convex polyhedron is bigger than that of any convex polyhedron inside of it.

So how can we prove such a result? The idea is that if we cut a convex polyhedron by a line into two parts, then the perimeter of each of them is smaller than that of the original polyhedron - this can be proved easily using the triangle inequality. What remains is to realize that one can get any convex polyhedron contained inside the original polyhedron by a series of such cuts (and forgetting about the unnecessary parts).



## 0.2 Conic sections

We define an ellipse with foci  $A$  and  $B$  and major axis  $L$  as the locus of all points  $P$  in the plane with the property that  $\overline{AP} + \overline{PB} = L$ . Its interior is the set of points where this sum of distances is smaller:  $\overline{AP} + \overline{PB} < L$ . This description gives us a way to construct an ellipse: one should attach a thread of length  $L$  to two nails located at the foci and trace with a pencil the furthest points one can reach by stretching the thread.

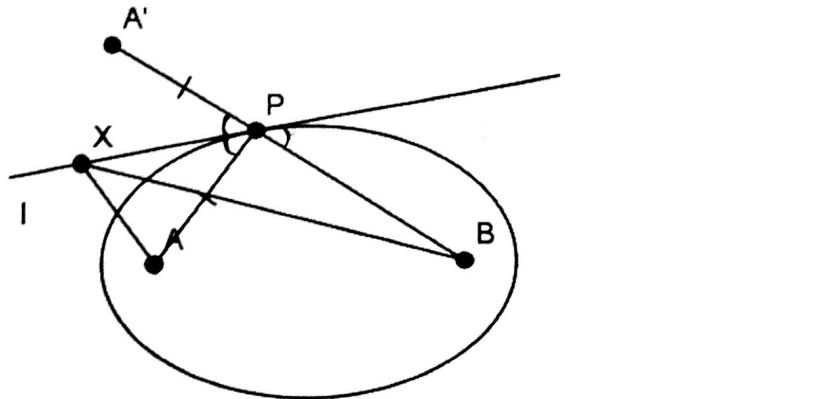


We can easily prove that the interior of an ellipse is a convex figure. Indeed, the restriction of the function  $AR + RB$  of the point  $R$  to some segment  $PQ$  doesn't have any local maxima (we've seen this already). Hence its maximum must be obtained at one of its endpoints. So if we know that  $AP + PB \leq L$  and  $AQ + QB \leq L$ , then we can conclude that for any point  $R$  on the segment  $PQ$  the inequality  $AR + RB \leq L$  holds. This is exactly what we have to verify to prove that the interior of the ellipse is convex.

We will use this convexity to show that the ellipse has a remarkable optical property: if light ray comes out of a focus  $A$  and gets reflected in an elliptical mirror with foci  $A$  and  $B$ , then its reflection passes through the other focus  $B$  of the ellipse. This optical property was used sometimes in designs of medieval castles - by building two separate rooms at foci of an ellipse and shaping the walls around them in the form of an ellipse, medieval people could create a way to listen to conversations in one of the rooms from the other one, without being noticed.

*optical property*

What we should prove mathematically is that for any point  $P$  on the ellipse the angle formed by the tangent line  $l$  at point  $P$  and the line  $AP$  is equal to the angle formed by  $l$  and the line  $BP$ . Now since the ellipse is convex, it lies to one side of the tangent line  $l$ . Hence the point  $P$  on the line  $l$  minimizes the sum of distances  $AX + XB$  for all points  $X$  on the line  $l$ . But this implies the desired equality of the angles, as we already noted.



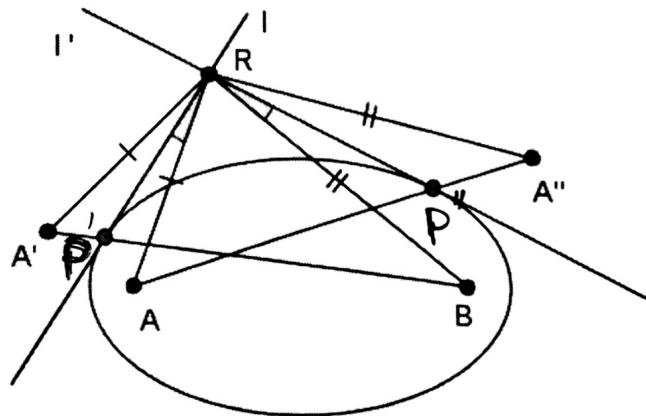
Alternatively we could prove this optical property in a slightly different way - since the ellipse is a level set of the function  $AP + PB$  of point  $P$ , the gradient of this function must be perpendicular at all points to the ellipse. But the gradient points in the direction of the angle bisector of angle  $APB$ ,

so we get the desired result once again.

Now we will present a way to construct the two tangent lines to an ellipse from any point  $R$  outside the ellipse with foci  $A$  and  $B$  and length of major axis  $L$ . This construction will be used later to prove a certain generalization of the optical property of the ellipse we just proved.

Suppose that the line  $l$  passing through the point  $R$  and tangent to the ellipse is given. Let  $P$  denote the point of tangency of  $l$  with the ellipse. Denote by  $A'$  the reflection of the focus  $A$  in the line  $l$ . Then clearly  $L = BP + PA = BP + PA'$ . Now the point of tangency  $P$  must lie on the line  $BA'$ , because it is the point where the minimum of  $AX + XB = A'X + XB$  is attained for points  $X$  on the line  $l$ . Hence the distance  $BA'$  is equal to the given distance  $L$ . The distances  $RA' = RA$  and  $RB$  are also known.

From what we said above we can construct the point  $A'$  without knowledge of the tangent line  $l$ . All we have to do is to construct the two triangles  $RBA'$  and  $RBA''$  with side  $RB$  and other two sides of length  $RA$  and  $L$ . The two possible triangles will lead us to the construction of the two tangent lines from the point  $R$ . Once we find the position of the points  $A'$  and  $A''$ , it is easy to find the tangent lines  $l'$  and  $l''$  - they bisect the angles  $A'RA$  and  $A''RA$  respectively.



If we denote by  $P'$  and  $P''$  the points of tangency of the ellipse with the lines  $l'$  and  $l''$ , then we can find the angles  $\angle ARP'$  and  $\angle BRP''$ : the angle  $\angle ARP'$  is equal to  $\frac{\angle ARA'}{2} = \frac{\angle BRA' - \angle BRA}{2}$ . Similarly,  $\angle BRP'' = \angle ARP'' - \angle ARB = \frac{\angle ARA''}{2} - \angle ARB = \frac{\angle BRA'' + \angle ARB}{2} - \angle ARB = \frac{\angle BRA'' - \angle ARB}{2}$ . Since  $\angle BRA'$  is equal to angle  $\angle BRA''$  by construction, the angles  $\angle ARP'$  and  $\angle BRP''$  are equal as well. This is the generalization of the optical property

of the ellipse we referred to earlier - for any point  $R$  outside the ellipse the angles between the segments  $AR$  and  $BR$  connecting  $R$  to the foci and the tangent lines from  $R$  to the ellipse are equal. To see that this is indeed a generalization of the optical property of the ellipse we let  $R$  tend to a point on an ellipse. Then the two tangent lines tend to one line, which is tangent at the point  $R$  on the ellipse and our theorem reduces to the optical property of the ellipse.

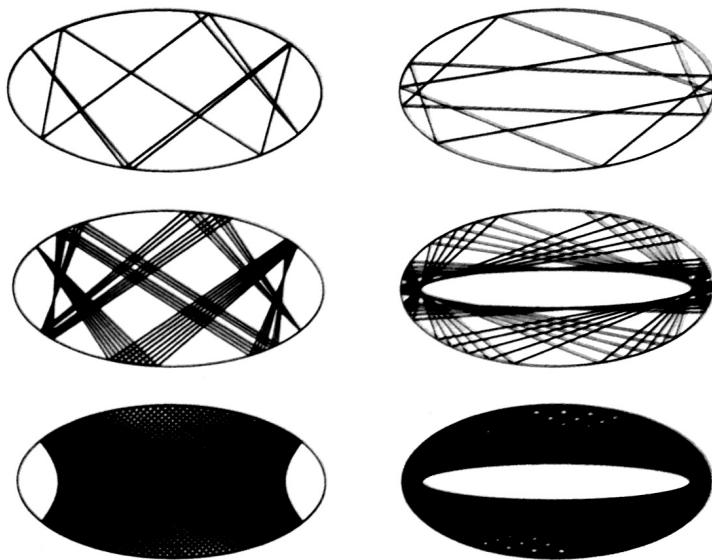
The fact we just proved leads to a striking, almost unbelievable result. Let's take an elliptic mirror and shine a light ray inside of it. This ray will undergo infinitely many reflections. The optical property tells us that if the ray passes through one of the foci, all the reflections will also pass through one of the foci. Now we claim that even if the initial ray didn't pass through a focus, we can still say a lot about all its reflections. Namely we shall now prove that if the ray doesn't intersect the segment  $AB$ , then there exists a smaller ellipse inside our ellipse, to which all the reflections of the ray will be tangent. In particular the light will never pass any point inside of this ellipse. In exercise ... the reader will prove a similar statement for a ray that does intersect the segment  $AB$ .

So let's prove this theorem: the first ray  $l_1$  is tangent to some ellipse with foci  $A$  and  $B$ . Indeed, the point of tangency  $P$  can be found by reflecting the focus  $A$  in the line  $l_1$  and intersecting the line  $l_1$  with the segment connecting the reflection to the other focus  $B$ . We know that the only ellipse with foci  $A$  and  $B$  that passes through the point  $P$  thus constructed will be tangent to the line  $l_1$ . Now all we have to do is to show that the ray reflected from the ellipse at point  $R$  will still be tangent to the ellipse we just constructed. Indeed, if the ray stays tangent to this ellipse after one reflection, we can repeat the argument to find that it will stay tangent after the second one and so on.

Now let  $l_2$  be the other line passing through  $R$  and tangent to the smaller ellipse. We will show that  $l_2$  is exactly the reflection of  $l_1$  in the larger ellipse. From the optical property of the larger ellipse we know that the angle between  $AR$  and the ellipse is equal to the angle between  $BR$  and the ellipse. Also we have just proved that the angle between  $AR$  and  $l_1$  is equal to the angle between  $BR$  and  $l_2$ . Subtracting these equalities from each other, we get that the angles between the ellipse and lines  $l_1$  and  $l_2$  are equal, which means exactly that  $l_2$  is the reflection of the light ray  $l_1$  in the elliptic mirror.

Now we will talk a little bit about a close cousin of the ellipse, the hyperbola. The hyperbola is defined as the locus of points  $P$  in the plane with

The distance difference is fixed.



the property that the absolute value of the difference of the distances to two foci is constant:  $|AP - PB| = L$ , where  $A$  and  $B$  are the foci and  $L$  is a real number, which is smaller than the distance from  $A$  to  $B$ .

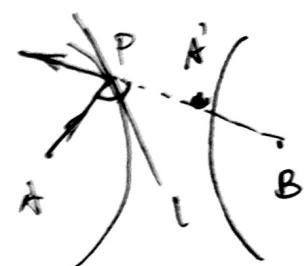
Unlike the ellipse, the hyperbola is not connected - it has two branches, corresponding to  $AP - PB = L$  and  $AP - PB = -L$ . The two connected pieces are called branches.

The hyperbola has an optical property of its own: if we shine light from one of the foci of the hyperbola and it gets reflected in the branch of the hyperbola closest to the focus, then the reflected light looks like it had been sent from the other focus.

We can formulate this property in a different way - for every point  $P$  on the hyperbola the angle  $\angle APB$  is bisected by the tangent line at  $P$ .

In the case of the ellipse the proof of the optical property came from studying the behavior of the function  $AP + PB$  restricted to a line. Similarly for hyperbola the relevant extremal problem is: given a line  $L$  and two points  $A$  and  $B$  lying on different sides of it, find the point  $P$  on the line  $L$  which maximizes the function  $|AP - PB|$ .

To solve the problem, let's reflect the point  $A$  in line  $L$ . Denote the reflected point  $A'$ . Then the triangle inequality for the triangle  $A'PB$  gives us that  $BP - PA' \leq A'B$  and  $PA' - BP \leq A'B$ . This, together with the fact that  $PA' = PA$  implies that  $|AP - PB| \leq A'B$  with equality if and



only if  $A'$ ,  $B$  and  $P$  lie on one line. Thus if the line  $l$  is not parallel to  $A'B$ , the maximum is attained and exactly at one point. Note also that the line line  $l$  bisects the angle  $APB$  (for the point  $P$  where maximum is attained).

Now let's apply this knowledge to the hyperbola. Let  $P$  be a point on the hyperbola with foci  $A$  and  $B$ . Let  $l$  denote the tangent line to hyperbola at  $P$ . Since at the point  $P$  the absolute value of the difference  $|AP - PB|$  attains its maximum on the line  $l$  (\*I'm using something that I haven't proved - that this is the maximum, rather than a minimum or local maximum - add explanation about the convexity of a branch of the hyperbola\*), the line  $l$  bisects the angle  $APB$ , as we proved before.

Similarly to the case of the ellipse we can construct the two tangent lines to the hyperbola from a point  $R$  lying outside of it, given its foci  $A$  and  $B$  and the length  $L$  appearing in its definition. Here's how we do it: If the line  $l$  is tangent to the hyperbola at point  $P$ , then the reflection of the point  $A$  in  $l$ ,  $P$  and the point  $B$  lie on one line (again we are using that the maximum of  $|AP - PB|$  on the line  $l$  is attained at the point of tangency). Thus the distance  $A'B$  is equal to  $|AP - PB| = L$ . So in triangle  $RBA'$  all the lengths are known:  $RB$  is given,  $RA'$  is equal to  $RA$  and  $A'B$  is equal to  $L$ . Thus we can construct the two triangles  $RBA'$  and  $RBA''$  with side  $RB$  and two other sides of lengths  $RA$  and  $L$  (these triangles exist - the triangle inequality that assures this follows from the fact that the point  $R$  lies outside the hyperbola). Once these two triangles are constructed and we know the location of the points  $A'$  and  $A''$ , it's easy to reconstruct the tangent lines - they bisect the angles  $\angle ARA'$  and  $\angle ARA''$ . We leave it as an exercise for the reader to derive from this construction that if  $P'$  and  $P''$  are the two points of tangency of the two tangent lines to the hyperbola from  $R$ , and  $l'$  and  $l''$  are the two tangent lines, then the angles formed by the lines  $l'$  and  $l''$  and the rays to the foci, lying in the branches on which  $P'$  and  $P''$  sit are equal to each other (for some points  $R$  the two points  $P'$  and  $P''$  lie on different branches, while for others they lie on the same branch).

Once we know this last property, we can understand the elliptic mirror better. Namely we know what happens to the ray of light in the elliptic billiard if it doesn't intersect the segment between the foci - we proved that it stays always tangent to the same ellipse. Now we can understand what happens in the other case as well - if the ray  $l$  intersects the segment  $AB$ , then there exists a hyperbola with foci  $A$  and  $B$  that is tangent to it (to find this hyperbola we should reflect the focus  $A$  in the ray  $l$  and connect the resulting point to  $B$  - the intersection of  $l$  and the line we constructed

is the point of tangency of  $l$  with the hyperbola we are looking for). Now if  $R$  is the point where the ray  $l$  gets reflected in the elliptical mirror and  $l'$  is the other line passing through the point  $R$  and tangent to the hyperbola, then the angles between the lines  $l$  and  $AR$  and the angle between the lines  $l'$  and  $BR$  are equal - this is the exercise we have left to the reader. The angles between the ellipse and the lines  $AR$  and  $BR$  are equal as well - this is the optical property of the ellipse. Summing these equalities we get that the angles between the ellipse and the lines  $l$  and  $l'$  are equal, meaning that  $l'$  coincides with the reflection of the ray  $l$ . Thus the ray will stay tangent to the same hyperbola after one reflection. Since we can repeat the argument, it will stay tangent to the same hyperbola forever.

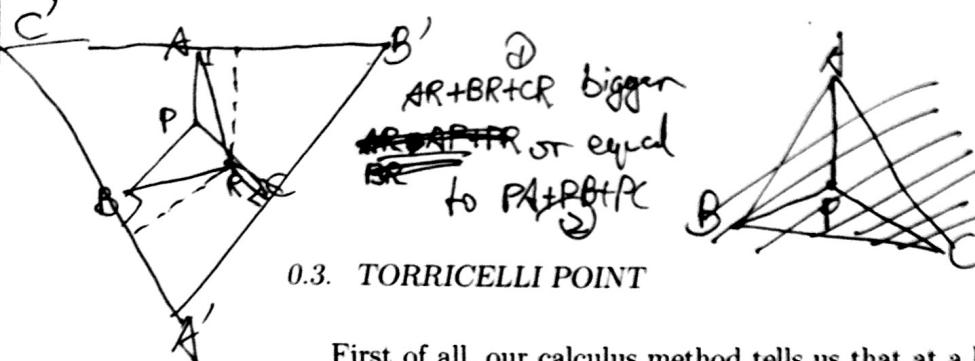
At this point we should also make a comment about another remarkable curve, the parabola. The parabola is defined by one focus and one line called the directix. It is defined as the locus of the points in the plane whose distance from the focus is the same as the distance from the directix.

One can think of the parabola as a limit case of both an ellipse and a hyperbola as one of the two foci goes to infinity and the other one stays fixed. The parabola inherits the optical properties of the ellipse and the hyperbola in the following form: all the rays that go out of the focus get reflected to parallel rays, all of them perpendicular to the directrix. This property is extremely useful in applications - if we want to gather a bunch of parallel rays coming from far away to a point, we can use a mirror in the shape of paraboloid - the figure of revolution of the parabola around its axis of symmetry. This is of course used in telescopes and satellite dishes. The generalizations of the optical properties of the ellipse and hyperbola we've seen have their analogues for the parabola. Namely if  $R$  is a point outside a parabola, and lines  $l$  and  $l'$  are two tangent lines to the parabola passing through point  $P$ ,  $A$  is the focus of the parabola and  $r$  is the ray from the point  $R$  and pointing in the direction perpendicular to the directrix, then the angles formed by lines  $AR$  and  $l$  and by lines  $r$  and  $l'$  are equal.

### 0.3 Torricelli point

This section is devoted to the question "given a triangle  $ABC$  with angles smaller than  $120^\circ$  find the point inside the triangle with the minimal sum of distances to the vertices".

We will present again several solutions.



### 0.3. TORRICELLI POINT

13

First of all, our calculus method tells us that at a local minimum of the function  $AP + BP + CP$  of the point  $P$  the gradient necessarily must be equal to zero. The gradient is the sum of three unit vectors pointing in directions  $AP, BP, CP$ . For the sum of three unit vectors to be zero, the angles between them must be equal to  $120^\circ$  (if we put them head to tail, they should form an equilateral triangle). Thus at local extrema of the function  $AP + BP + CP$  we should have  $\angle APB = \angle BPC = \angle CPA = 120^\circ$ .

This information is not enough of course. To get more information, consider a point  $P$  inside the triangle with the property that  $\angle APB = \angle BPC = \angle CPA = 120^\circ$  (one can construct this point by drawing equilateral triangles on the sides of the triangle  $ABC$ , inscribing them in circles and taking the point of intersection of these three circles; these circles intersect. Indeed if  $P$  is the point of intersection of the first two, then  $\angle APB = 120^\circ$  and  $\angle BPC = 120^\circ$ . But then  $\angle CPA = 120^\circ$  as well, so the point  $P$  must lie on the third circle as well). Construct three lines passing through the vertices  $A, B, C$  and perpendicular to the segments  $PA, PB, PC$ . Denote the triangle these three lines form  $A'B'C'$  ( $A'$  being the intersection of the lines through  $B$  and  $C$  etc). The triangle  $A'B'C'$  is equilateral - all its angles are equal to  $60^\circ$ .

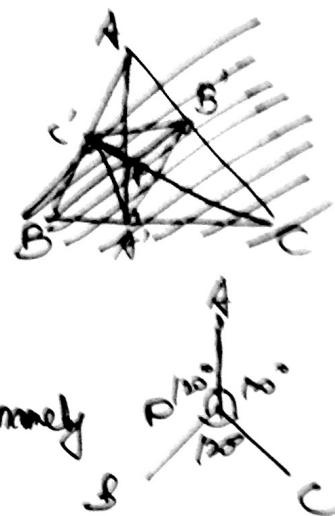
We will now prove a lemma: for all points  $P$  inside an equilateral triangle, the sum of distances to the sides is the same.

Proof: If we multiply the sum of distances from the point  $P$  to the sides by the lengths of the side, we will get twice the sum of the areas of triangles  $APB, BPC$  and  $CPA$ . But this latter sum is independent of the choice of point  $P$  - it is just the area of the triangle  $ABC$ .

This lemma gives us the answer almost immediately: for any point  $R$  inside triangle  $A'B'C'$  the sum of distances to the points  $A, B, C$  is bigger or equal to the sum of distances to the sides  $B'C', C'A', A'B'$  (with equality only if  $R = P$ ). But the latter sum is equal to the sum  $AP + BP + CP$  according to the lemma we just proved. Thus the sum of distances to points  $A, B, C$  is minimal at the point  $P$  where  $\angle APB = \angle BPC = \angle CPA = 120^\circ$ .

Let's give another proof of the same fact, using the idea that the shortest path between two points is the straight line segment between them. To use this idea we will have to "straighten" somehow the sum  $AP + BP + CP$  in a way analogous to what we did for the inscribed triangle of minimal perimeter.

Namely, let's rotate the triangle  $BPA$  by  $60^\circ$  in the direction from  $BC$  to  $BA$ . Denote by  $A'$  and  $P'$  the images of the points  $A$  and  $P$  under this rotation. Since the angle  $ABA'$  is equal to  $60^\circ$  and the sides  $AB$  and  $A'B$



~~but also in this~~  
~~triangle BPC~~  
~~AP~~



This is easy because  
we can always use  
explicit algebraic  
method to calculate

~~as  $PA, PB, PC$  are  
fixed by the  
size / side length of  
 $\triangle ABC$~~

are equal, the triangle  $ABA'$  must be equilateral. Similarly the triangle  $PBP'$  is equilateral. This shows that  $BP = PP'$ . Finally  $P'A' = PA$ , because one is the rotated copy of the other. Thus the sum  $AP + BP + CP$  is equal to the sum  $CP + PP' + P'A'$ . Such a sum is clearly larger than the length  $CA'$  and is equal to it if and only if both  $P$  and  $P'$  lie on the segment  $CA'$ . This is the case if and only if  $\angle CPB + \angle BPP' = 180^\circ$  and  $\angle BP'P + \angle BP'A' = 180^\circ$ . Since we know that  $\angle BPP' = \angle BP'P = 60^\circ$  this is so if and only if  $\angle CPB = 120^\circ$  and  $\angle BP'A' = 120^\circ$ . Finally the angle  $\angle BP'A'$  is equal to the angle  $\angle BPA$  (one is a rotated copy of the other) so we get once again that the sum of distances from point  $P$  to the vertices  $A, B, C$  is minimized if and only if  $\angle APB = \angle BPC = \angle CPA = 120^\circ$ .

## 0.4 Isoperimetric inequality \* *not covered in class)*

In this section we will outline a solution to the following very natural question: among all figures of a given perimeter find the figure with the largest area. We can reformulate the question in the following "practical" way: given a fence of some length, what is the largest area we can enclose using it?

It turns out that the answer is a circle.

We will give now a flawed proof of this fact and ask the reader to spot the mistake in the proof.

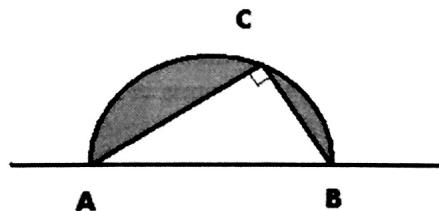
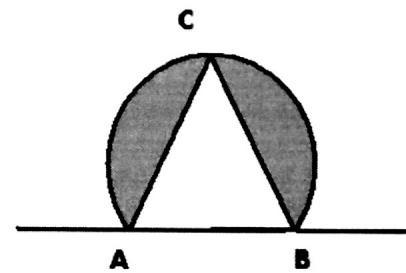
Let  $R$  be the figure of given perimeter with the largest area. The figure  $R$  must be convex. Indeed, if it isn't, there is a segment with endpoints in the figure, but not contained entirely in it. By looking at a connected component of the part of the segment outside the figure we will find a segment whose endpoints are on the boundary of the figure  $R$  and the interior is outside. Now reflect the part of the boundary of the figure  $R$  lying between the endpoints of the segment we described. Clearly by doing so we don't change the total length of the boundary, but we do increase the area - some part of the plane that used to be outside the figure is now inside and we haven't removed anything.

[remark by Uyen Hoang: in fact the argument doesn't always work - holes in the non-convex set can overlap after reflection. We should prove instead that the perimeter of convex hull is smaller than the perimeter of the original set]

Next take a line that intersects the boundary of our convex body in two points and divides the perimeter into two equal parts (such a line exists by

continuity arguments - if we start from a line far above the figure and drag it to a line far below, the perimeter of the part of  $R$  above the line should increase from 0 to full perimeter, so at some point the line should have divided the perimeter in two equal parts). Then this line divides the area of  $R$  into two equal parts. Indeed, if it weren't so, then we could replace the part with smaller area by the reflection of the part with larger area without changing the perimeter, but increasing the area.

By replacing one of the parts by a reflection of the other, we can assume that our figure is in fact mirror symmetric with respect to the line above.



Now denote by  $A$  and  $B$  the points where the line intersects the boundary of  $R$  and let  $C$  be any point on the boundary of  $R$ . Then the angle  $ACB$  must be equal to  $90^\circ$ . Indeed, we can divide the half of the figure to three parts: the "mountain" on top of  $AC$ , the "mountain" on top of  $BC$  and the triangle  $ACB$ . Then we can slide the segments  $AC$  and  $BC$  together with

their mountains so that the point  $C$  is fixed without changing the perimeter of the half-figure. But if angle  $ACB$  is not  $90^\circ$ , we can increase the area of triangle  $ACB$  by making the angle  $ACB$  equal to  $90^\circ$  (area of  $ACB$  is equal to  $1/2ACCB \sin \angle ACB$ , so it is maximized when  $\angle ACB = 90^\circ$ ).

Now we proved that for every point  $C$  on the boundary the angle  $ACB = 90^\circ$ . But this means that the boundary is a half-circle. Since the other half of the figure is the reflection of this one, we conclude that our figure is a circle.

Have you noticed the flaw in the arguments?

If not, try the following "proof":

"Claim": One is the largest natural number.

"Proof": Let  $N$  be the largest natural number. Since for every natural number  $N$  we have  $N^2 \geq N$  and  $N$  is the largest one, equality must hold:  $N^2 = N$ . Thus  $N = 1$ , QED.

So the flaw in our argument was right in the beginning - we didn't justify that the figure of maximal area among all figures of given perimeter exists.

One way to justify this is to use some sort of compactness of the set of convex figures inside a large box - all figures of given perimeter that are competing to be of maximal length must be convex and can be put in a large enough box (whose dimensions can be determined a priori from the perimeter). Then the functions "perimeter" and "area" on this set of convex figures can be shown to be continuous and so the maximum of the area should be obtained somewhere on the level set of the perimeter. This is a very rough outline, however, and to make it precise requires quite a lot of work.

Thanks to Sergio Da Silva for numerous corrections to the text of this note.