OPTIMIZATION METHODS IN PLANAR GEOMETRY

VIPUL NAIK

Abstract. Optimization problems frequently arise in the context of geometric configurations. In this article, we discuss the formulation and broad strategies used to handle optimization problems, and apply these in the context of geometric optimization problems. In the process, we apply geometric results of great diversity.

1. Prebeginnings

1.1. Optimization – a cross cutting concern.

The tendency or desire to optimize – whether it is on our money, our time, or our other resources, is built into us. In fact, the very basis of economics is that "wants are unlimited, but resources are limited". Much of technology is devoted to finding better and more efficient, ways of doing things.

Optimization is practised not only by living creatures but also by nature. A physical system tries to rearrange itself to a locally stable configuration. For instance, a soap bubble endeavours to occupy the minimum possible surface area for its given boundary. Left to itself, a drop becomes spherical so as to minimize its surface tension. The intestines use fractal geometry to maximize their surface and thus increase the effectiveness of absorption.

In most cases, optimization simply occurs at the endpoints. For instance, the lowest gravitational energy is possessed when the object is placed at the least height. To maximize one's wealth, it makes sense to take as much as possible.

But some situations arise where there are conflicting considerations, each of which is important to the optimization process. A suitable trade off is then required to find the optimal point. To maximize the product of two numbers with a fixed sum, we need to balance them out so that they are equal.

In this article, we look at optimization problems that are essentially geometrical in nature, and are related to the more regular figures (such as polygons, circles, and ellipses) as opposed to general closed curves. We shall also confine ourselves, for the most part, to the plane (that is, to two dimensions). Further the quantities we try to maximize or minimize have direct geometrical significance. As we shall see, geometrical as well as other tools play a crucial role in solving these problems, though there is no "hard-and-fast rule".

1.2. Optimization and calculus.

Bulk of optimization requires a formal understanding of calculus and differentiation, and an intuitive feel for the so called ϵ-δ arguments. These arguments use a very basic principle – a globally optimal solution is also locally optimal, and it is thus highest in some immediate neighbourhood. Derivatives are a first test that any locally optimal solution must pass.

The geometric approaches to optimization do not explicitly use the formulas or the arguments of calculus, but they are based on this principle of optimality. Some of the (optional) alternative proofs on vector calculus also link geometric and calculus methods for optimization.

The main aspect to remember is that calculus usually only helps us narrow down the problem to a discrete collection of points, and hence, usually, a finite collection of possibilities. Checking for the final solution within the discrete collection usually requires a mix of many techniques.

1.3. **The Steiner problem.** In problem 2, we shall ask for a point in a triangle such that the sum of its distances from the vertices is minimum. The problem, first posed by Fermat, was studied by **Steiner** in greater detail. Steiner studied the more general problem:

Problem 1 (Steiner network)**.** Given a collection of points in the plane, construct a network of minimum length that connects all these points.

Here, a **network** means a collection of nodes (points in the planes) including the points already given and a collection of line segments that may join two nodes. The **length** of the network is the sum total of the lengths of the line segments. The network *connects* all the points if there is a path, comprising edges, from every vertex in the original collection to every other.

The optimal Steiner network for a triangle is always constructed by adding, as a node, the vertex whose total distance from all the points is minimum, and then making line segments joining this to each vertex. However, it is nontrivial to establish that this network is indeed optimal. For instance, how can we be sure that there is no better network with two new nodes?

This problem was, in fact, given to a combinatorialist Ronald Graham by the telephone company he was working in, which wanted to find out how to place its stations so as to minimize the length of the telephone network, connecting three major stations. Graham was unable to solve the problem fully, and, in the true traditions of mathematics, offered a prize for the solution. The problem was finally settled affirmatively around 1990.

2. The core problem

2.1. **Motivation.** In this article, we will analyze **optimization problems**. Some commonly encountered optimization problems are:

- Given a geometric configuration, *find a point* in the plane (or in some given region of the plane) so that some geometric expression is maximized (or minimized). For instance, the problem of finding a point in the plane whose sum of distances from the vertices of a given triangle is minimum (Problem 2).
- Given certain constraints, *find a geometric configuration* satisfying those constraints and optimizing some expression. We shall see an example of this in section 7, namely Problem 5.
- Given a geometric configuration, *find a further geometric configuration* relative to it that satisfies certain constraints and optimizes an expression. For instance, given a triangle, the problem of finding an ellipse inside it with maximum area (Problem 4)

Many tools are used to attack these problems. To be able to *best* make use of these tools, we will find it useful to fix some general terms for optimization problems. This terminology is used in **linear programming** and **non-linear programming**.

2.2. **The general format.** Every optimization problem has the following :

- A collection of **parameters** each of which can vary over some set. Each tuple of parameter values corresponds to one configuration. The number of parameters needed denotes the number of **degrees of freedom** possessed by the system.

  For instance, in the problem of finding a point in the plane optimizing some expression, the parameters are the coordinates of the point chosen in some suitable coordinate system. The usual Cartesian coordinate system has two parameters, each varying over $\mathbb{R}$, and every 2 tuple of parameter values corresponds to one point.

  The manner of choosing the parameters plays a crucial role in how the problem is solved. A **change of parameter** involves expressing the system of parameters in terms of some other system of parameters.

  When we are looking for a geometric configuration upto similarity transformations, then we are interested in parameterizing it intrinsically rather than extrinsically. Such choices of parameter spaces are termed **moduli spaces**. More on this is to be found in
- **Constraints** placed on the parameters, that disallow certain configurations, and hence, certain tuples of parameter values. The **feasible solutions** or feasible configurations (tuples of parameter values) are those that satisfy all the constraints. Optimization problems without constraints are called **unconstrained optimization problems**. Sometimes, a suitable change of parameters can change a constrained optimization problem to an unconstrained optimization problem.

  Constraints may or may not decrease the effective number of degrees of freedom. Typically, **equality constraints** decrease the number of degrees of freedom, while **inequality constraints** do not.
- An **objective function** or **criterion function** expressed in terms of the parameters. This function needs to be maximized (or minimized).The configurations for which this is attained are termed **optimal solutions**. If more than one solution exists, they are collectively termed **alternative optimal solutions**.

Typically, the objective function is quite evident from the statement of an optimization problem. However, the nature of the parameters as well as the constraints operating on them are more subtle. By clearly stating all of these at the outset, we make the task of cracking the problem much easier.

2.3. **The strategies available.** What does an optimum look like? How can we locate it?

Key Point 1 (Perturbations and Local Optimality)**.** Slight perturbations to an optimal point make the corresponding solution less optimal.

This principle suggests that to determine optimal points, we should look for points that are optimal in their immediate neighbourhoods. This is known as *local analysis*. On the other hand, methods where the entire collection of feasible solutions is considered come under the heading of *global analysis*.

Thus the two broad themes are:

- **Local analysis** : This studies the *local* properties, viz, the properties in the immediate neighbourhood, of an optimal solution. Local analysis introduces ideas like local minima/maxima, and strong local minima/maxima. Derivatives play an

important role in determining local maxima/minima when the objective function is differentiable.
- **Global analysis** : This compares properties across all feasible solutions (configurations) simultaneously to establish the optimal solution. It involves the use of *inequalities*. As problems in this article are geometric in nature, both geometric and algebraic inequalities come up in global analysis.

Usually, a mix of both these themes is at work.

3. Minimizing length sums – Fermat

We begin with the famous Fermat problem :

Problem 2 (Fermat, to Torricelli)**.** Given a triangle $\triangle ABC$ find the point P on its plane such that $PA + PB + PC$ is minimum.

In variable free language, the problem can be restated as: "given a triangle, locate a point that has minimum total distance from the vertices".

This point is nowadays variously called the **Fermat point** and the **Torricelli point** of the triangle. One of its characterizing properties is as follows :

- When all the angles of the triangle are less than $2\pi/3$, then it is the point subtending equal angles at all three sides, or equivalently, the point P such that $\angle APB = \angle BPC = \angle CPA = 2\pi/3$.
- When one of the angles is $2\pi/3$ or more, then that vertex itself is the required Fermat point.

In the coming subsections, we shall establish this characterization and also some other important properties of the Fermat point.

3.1. A proper problem formulation.

Let us look again at the variable free formulation of the problem: "given a triangle, locate a point that has minimum total distance from its vertices".

A better way of stating this is: "given a triangle, locate a point that minimizes the following function: the sum of distances from the vertices".

For every point, we are computing its total distance from the vertices, and then we are looking for the point where this is minimum. Thus, it is clear that the objective function is the sum of distances of the point from the vertices, and that this needs to be minimized.

We now need to choose a suitable parameterization, or coordinate system, for the point. Because the point is free to vary over the plane of the triangle, there are two free parameters. In case we want to do local analysis, we need to choose some way of parameterization where we are able to apply Key Point 1 (on Perturbations and Local Optimality) effectively. In case of global analysis, we must make a choice by which we can effectively make use of inequalities.

Thus far, we have obtained that :

- The parameters are the coordinates of the point according to some convenient parameterization of points in the plane.
- There are no constraints.
- The objective function is the sum of distances from three given points (that form the vertices of a triangle).

Before exploring the issue of coordinates, we look at a beautiful "physical solution" of the problem.

3.2. **A physical proof.** The physical proof that we now present is essentially the same as the differential geometry proof that shall follow it, but it appeals directly to physical intuition. The argument runs – make holes at the vertices of the triangle, tie a knot and hang three weights from it via the three vertices. If the weights are equal, they will attain equilibrium precisely when the forces at the knot are in equilibrium. This will occur when they make angles of $2\pi/3$. But by physical reasoning again, the equilibrium will be attained only when the total rope length above the plane of the triangle is minimum (so as to make the hanging length maximum).

When one of the angles is greater than $2\pi/3$, the knot will actually shift to that vertex and may even fall through it if allowed!

The beauty of this physical proof is that it immediately convinces us as to *why* the angle must be $2\pi/3$. But let us examine what is going on here, more precisely. All we are saying is that the point where the three ropes at the knot make an angle of $2\pi/3$ is a point where there is *local stability*. That is, it is a point where there is no tendency to move nearby, and if there is a slight perturbation in any direction, there will be a tendency to pull it back the position.

This is *equivalent to the assertion* that the point is a local optimum. It does not prove rigourously that the point is a global optimum.

Observation 1. The statement in physics that the force acting at a configuration of minimum energy is zero, is essentially a physical equivalent of the statement that the derivative of a function is zero at the minimum. This is what physicists call a **stable equilibrium** point.

Do you see it? The **force** that acts as the knot represents its tendency to move towards a stabler position. So when it is already at its stablest position, this force is zero. The force plays the role of the derivative of potential energy.

Having seen this physical proof whose idea is essentially mathematical, we would like to get hold of a completely mathematical formulation of it. The advantage of a mathematical formulation is that we can apply it even in situations where we cannot go around hanging masses and tying knots! After a short detour into coordinate systems, we shall come back to the problem with renewed vigour, and in the process, introduce the concept of the **vector gradient** of a **scalar function**.

3.3. **Coordinate systems on the plane.** The objective function is a map from the plane to $\mathbb{R}$. But dealing directly with the plane is difficult, as there is no natural way of performing algebraic manipulations on points in the plane. We need a way to refer to these points using real numbers. This is the *raison d'etre* of coordinates.

There are a number of coordinate systems in use for the plane, each suited for a different kind of modelling. Here, we consider some of the more well known coordinate systems:

(1) A **Cartesian coordinate system** that depends on two intersecting directed lines. The point of intersection is the origin and the two lines are the **axes**. Typically, we choose a **rectangular coordinate system**, that is, a coordinate system where the axes are perpendicular. The coordinates of a point are given

(2) A **polar coordinate system** that depends on the choice of a point (the **pole**) and a ray starting at it (called the **initial line**). The coordinates of a point are given as (r, θ) where r is the distance from the pole and θ is the angle from the initial line to the ray from the pole passing through the initial point, measured counterclockwise.[1]

These choices rely on a 2 parameter description where both parameters vary over $\mathbb{R}$ or over intervals in $\mathbb{R}$, which is natural as the plane is 2 dimensional.

Sometimes, it is profitable to view the point in terms of a single parameter rather than in terms of many coordinates. Two approaches for this are:

(1) **Vector treatment** where the point is treated as a single parameter, the parameter being its **position vector**. This parameter is not a real valued parameter, unlike in the previous case. The next section demonstrates that for the Fermat point problem vector treatment is fitting.
(2) **Complex numbers** where the point is treated as a single parameter, that parameter being a complex number. Relative to some choice of origin and choice of real line, the plane gets mapped bijectively with the complex numbers (that is, it gets treated as an **Argand plane**). The advantage of complex numbers over vectors in general is the ease with which **rotation** can be effected by means of multiplication by complex numbers.

Let us now summarize the situation so far: the problem at hand (The Fermat problem) is an unconstrained optimization problem in the plane with the objective function being the sum of distances from three given points (that are vertices of a triangle). It is not clear how to parameterize the plane, but four possible ways of doing so are – Cartesian coordinates, polar coordinates, vectors, and complex numbers.

3.4. **Local analysis – differential geometry.** We apply local analysis to the problem. The objective function is a map from the Euclidean plane to $\mathbb{R}$ defined by $P \mapsto PA + PB + PC$.

In Key Point 1 we had observed that slightly perturbing around an optimal point makes the corresponding solution less optimal. In fact, this holds for any solution that is optimal in some sufficiently small neighbourhood. Such points (solutions) are termed **local optimal solutions** and the value of the objective function at such a point is a **local optimum**.

In more precise terminology, if f is a function from a region S (in some Euclidean space) to $\mathbb{R}$ we define:

(1) **Optimal solution** : A point in S for which the function attains its maximum/minimum. The value attained is the **optimum**.
(2) **Local optimal solution** : A point in S that has an open neighbourhood in which it is a optimal solution. The value attained is a **local optimum**.
(3) **Strong local optimal solution** or **Isolated local optimal solution** : A point in S that has an open neighbourhood in which it is a unique optimal solution. The value attained is a **strong local optimum**.

When S is $\mathbb{R}$ or a suitably nice subset of $\mathbb{R}$ (such as an open interval or a closed interval) :

[1]There are other coordinate systems such as bipolar coordinates (distances from two points), pedal coordinates (used for curves), trilinear coordinates (with a reference triangle) and areal or barycentric coordinates (again with a reference triangle). We shall not need these coordinate systems here

(1) If f has a well defined LHD at a, and a is a local maximum point, then the LHD must be nonnegative at a. If a is a local minimum point then the LHD is nonpositive.
(2) If f has a well define RHD at a, and a is a local maximum point, then the RHD must be nonpositive at a. If a is a local minimum point then the RHD must be nonnegative.
(3) Thus, if f is differentiable at a, and a is a local maximum point, then the derivative at a is 0. The same holds if a is a minimum point.

The corresponding converse results require further assumptions, such as the continuity of the derivative.

Key Point 2. If a function is defined and differentiable in some open neighbourhood of a point, its derivative at the point is 0.

Key Point 2 has been mentioned in the context of functions on $\mathbb{R}$ or intervals in $\mathbb{R}$. Geometrical situations, such as the Fermat problem which is at hand, have functions defined on the plane or on regions in the plane. Our goal is thus to find some notion of differentiation for functions from sets S in the plane to $\mathbb{R}$, for which Key Point 2 continues to hold.

The idea is that we need to describe the rate at which the function changes at some point in the plane. But at this point, there are many directions in which it could be perturbed in the plane. So how do we make sense of the rate of change?

Interestingly, when the function is sufficiently nice, then the following holds: there are two orthogonal directions such that an infinitesimal[2] change in one of these directions leaves the function invariant, and thus all the change occurs in the perpendicular direction to that. Thus, moving in *any* direction produces a change roughly proportional to the change in the component along that perpendicular direction. The direction along which all the change occurs is termed the **gradient direction** of the function. And the value of the derivative in that direction is termed the **gradient** of the function.[3]

Thus, the **gradient** ∇ of a **scalar function** on a set in $\mathbb{R}^n$ is a vector valued function on the set, that, at each point, gives the gradient of the function at the point. Note that while the original function was scalar, the gradient function is not – it is a vector valued function. This means that we cannot apply gradient twice in succession. The gradient of f is denoted as ∇f.

A scalar function is said to be **differentiable** at a point if its gradient is defined at that point. A scalar function is **differentiable** if it is differentiable at all points in its domain. And it is **continuously differentiable** if its gradient function is a continuous vector valued function. [4]

3.5. **The differential geometry solution.** With adequate preparation in the basics, we are ready to embark on what we originally intended – obtain a preuly mathematical formulation of the beautiful "physical solution" obtained earlier.

[2]infinitesimal means very small, approaching zero

[3]The concept of gradient, or, more generally, exterior derivative, makes sense in higher dimensional spaces as well.

[4]Even for functions $\mathbb{R} \to \mathbb{R}$, differentiable functions are not always continuously differentiable. For instance $x \mapsto x^2 \sin(1/x)$ when $x \neq 0$ and 0 when $x = 0$ is differentiable but not continuously differentiable at 0

We are basically interesting in computing the gradient of the function $P \mapsto |PA| + |PB| + |PC|$. Before using mathematics, let us see what the physical reasoning had told us.

What was the force coming to P due to the mass hanging at A. The tension in the rope towards A. This suggests that the gradient of the functio We observe that the vector gradient of $P \mapsto f|OP|$ where f is a map $\mathbb{R} \to \mathbb{R}$ and O is a fixed point, is given by $f'|OP|$ times a unit vector in the outward direction OP. This implies that the vector gradient of the function mapping P to the value $|PA|$ is a unit vector pointing outward along AP. Thus the vector gradient of $PA + PB + PC$ is simply the vector sum of unit vectors AP, BP and CP.

This clearly means that at the points where the derivative vanishes, the lines PA, PB and PC makes angles of $2\pi/3$ with each other.

A maximum or minimum occurs at points where :

- The derivative vanishes.
- We reach some sort of an extreme point.

When the derivative vanishes how do we determine whether the function is at a maximum or a minimum? The idea is that when a minimum is attained, the derivative all around the point is outward. Thus the divergence of the gradient at the point is expected to be positive. A little thought reveals that when the point we start with is in the interior of the triangle, the divergence of the gradient is always positive.

This suggests that if we can find a point P in the interior of the triangle satisfying the condition of equal angles, that is a minimum point. In fact, as any minimum point must be inside the triangle, P is the unique point satisfying the required conditions.

However, if one of the angles is $2\pi/3$ or more, we will not be able to find a point making the requisite angles inside the triangle. In that case, a little thought shows that the minimum is attained at the vertex.

We note the following key point :

Key Point 3. Analyzing the vector gradient of the scalar function allowed us to apply principles of calculus to simplify a problem of geometric optimization.

We shall continue to see applications of this idea.

3.6. **A proof by Ptolemy's Theorem.** This is an interesting solution to solving the Fermat point problem using Ptolemy's Theorem and some inequalities. The importance of this approach is not just its pure geometric nature but its generalization to the problem of finding **Steiner networks** (problem 1) for collections of four or more vertices.

Before proceeding with this solution, it may benefit us to have a look at the problem again:

- The problem is an unconstrained problem aobut optimization in the plane.
- The objective function to be minimized is the sum of distances of the given point (that we call P) from three given non-collinear points.

The approaches we tried so far have been local, that is, they have used principles of local perturbation. We now seek a global solution, making use of geometric inequalities. So, before understanding the particular proof, it might be beneficial to understand how global proofs work.

Consider the problem of proving that $x^2 + x + 1$ is positive. The expression can be written as $(x + 1/2)^2 + 3/4$. Now we know that $(x + 1/2)^2 \geq 0$ for all real x, giving the requied result.

The idea in the above was to use the fact that a *square is nonnegative* to obtain a *globally true inequality*. Further, the globally true inequality also gave us some insight into *when equality holds*.

In the geometric context, the only real and powerful global inequality we have is the **triangle inequality**. This states that, for any point P, we have:

$$|PA| + |PB| \geq |AB|$$

with equality holding if and only if P lies on the line segment joining A and B.

Thus, if we only wanted to minimize the sum of distances from *two* points, the triangle inequality would have sufficed. However, we have *three* points at hand. The geometric approach that we follow works in two steps:

- First, it finds an auxilliary point A' such that $|PB| + |PC| \geq |PA'|$ for all points P, also explicitly giving the set of points P at which equality holds. This set should be nonempty.
- Then, we have, by the triangle inequality, that $|PA| + |PA'| \geq |AA'|$ for all points P. The locus of P for which equality holds is the line segment AA'. Thus, the minimum point(s) P must be at the intersection of the line segment AA' and the curve given above.

The problem now is to determine the auxiliary point A'. It is here that Ptolemy's Theorem comes in.

The statement of Ptolemy's Theorem is as follows:

> The sum of products of pairs of opposite sides of a quadrilateral is greater than or equal to the product of the diagonals. Equality occurs if and only if the quadrilateral is cyclic, with the vertices in either clockwise or anticlockwise order. Equivalently, four points U, V, W, and X in a plane satisfy
>
> $$|UV|.|WX| + |VW|.|UX| \geq |UW|.|VX|$$
>
> Equality holds if and only if U, V¡ W and X are concyclic in cyclic order or collinear in linear order.

What we want is to find a point A' such that there is a clear locus for P satisfying the condition that $PB + PC \geq PA$ with equality occurring if and only if P lies on the locus. It is clear that the way to do this is to make P, A, B and A' concyclic with $\triangle ABA'$ equilateral.

Here is the formal proof:

Proof. Construct an equilateral triangle $\triangle A'BC$ on the side of BC opposite A. Now construct the circumcircle of this triangle. If P be any point in the plane, we have, by Ptolemy's inequality, $PB + PC \geq PA'$, with equality occurring only if P is on the arc BC not containing A'. Further we have $PA + PA' \geq AA''$ with equality occurring only if P is on the line segment AA'. Putting these two facts together we deduce that the point P minimizing $PA + PB + PC$ is the second intersection of AA' with the circumcircle of $\triangle A'BC$ provided that this is in the arc of BC opposite A'. This holds only when $\angle A < 2\pi/3$. When $\angle A \geq 2\pi/3$, we can show that the optimum must occur at A, through a direct argument involving triangle inequalities. $\square$

Some important geometric facts that are revealed by this analysis are :

- The circumcircles of the three equilateral triangles formed with base as the sides and on the side opposite the third vertex, are concurrent.
- The lines joining each vertex to the vertex of the associated equilateral triangle also concur at the same point. All of them have equal length, say l.
- The point is the Fermat point and has the minimum distance property when all the angles of the triangle are less than $2\pi/3$. The total distance in this case is equal to l.
- The angles subtended by the Fermat point at the sides are all $2\pi/3$.

3.7. **An interesting interpretation in terms of circles and ellipses.** The following is yet another interesting way of establishing that, in general, the angles subtended by the Fermat point at the sides are equal. This approahc requires a rudimentary familiarity with conic sections and can be skipped without loss of continuity.

The idea here is to view the objective function $PA + PB + PC$ as composed of two parts, PA and $PB + PC$, whose sum needs to be minimized.

Let P be a point where the minimum occurs. Consider the loci $PA = k_1$ and $PB + PC = k_2$. The former is a circle centered at A while the latter is an ellipse with B and C as foci. The circle and ellipse intersect at P. Suppse they intersect at another point as well. The, there is some area common to both. But if Q is a point in the area of overlap, we shall get $QA < PA$ and also, by an elementary use fo triangle ineuqalities, $QB + QC < PB + PC$. This gives us that Q is a point where the objective function attains a lower value than at P, contradicting the minimality of P.

Thus, the circle and ellipse can intersect at only one point – in other words, they are tangent.

We now use the **reflection property** of an ellipse which states that the normal bisects the angle between the lines joining the given point to the foci. More on the reflection property can be found in the appendix.

With the aid if the reflection property we can see that $\angle BPA = \angle APC$. Similar reasoning gives us that all the three angles are equal.

3.8. **A look at all the proofs.** We have provided various proofs to establish the properties of the Fermat point:

- A proof relying on physical intuition.
- A rigourization of that, which involved the language of differential geometry.
- A proof by Ptolemy's Theorem.
- A proof in terms of circles and ellipses.

The first two proofs employed *local* conditions while the third proof was global in nature. The fourth proof, while relying primarily on local intuition, could also be expressed in a global fashion.

Applying the local arguments required little knowledge of geometry, but a good intuitive feel for local behaviour and perturbations, as well as some formal understanding of vectors. While this may be thought of as more advanced, the advantage with the proof is that arrivign at it is a *straightforward* process and does not involve too much groping. The proof relying on geometric inequalities, on the other hand, required a thorough understanding of the structure of inequalities to, first of all, come up with the appropriate question (of locating an A' and a P such that some conditions were held) and then drawing upon the storehouse of geometric knowledge to realize that Ptolemy's Theorem was the right thing to use.

For most of the remaining problems that we discuss, we concentrate on developing a single solution. As expected, the solutions will have a mix of both local and global reasoning. In some cases, geometric results will help in transforming a hopeless looking problem into a doable one, which can be subject to either local or global analysis. However, we must keep in mind that the heuristics developed so far need to be carried forward and applied when attempting any optimization problem.

4. A challenging national olympiad problem

4.1. **Problem statement and preliminary observations.** The problem that we am going to discuss is a fairly challenging geometry problem, and even giving it a decent try requires a good deal of familiarity with the geometry of the triangle. It has the flavour of the **scientific method** – data is collected, observations are made, then some clever hypothesis is made, and a strategy to prove the hypothesis is chalked out. The steps involving data collection in themselves require a decent degree of comfort with triangle geometry, though these can be dispensed with in the final proof. (Those who are not *into* triangle geometry too much may find it better to skip this section and read it later, at more leisure.)

Problem 3 (Not yet filled in the source)**.** Let $\triangle ABC$ be a triangle with circumcircle Γ. Find all positions of the point M inside $\triangle ABC$ such that $\frac{BM.CM}{MD}$ takes the least possible value, where D is the point where AM, when extended, meets Γ. Also find this minimum value.

This is a *constrained* optimization problem with the constraint being that the point must lie in the interior of the triangle. The objective function is the expression $\frac{BM.CM}{MD}$.

A little exploration reveals to show that the objective function is symmetric in the three vertices, and can be expressed as [5] :

$$\text{Objective function (to minimize)} = M \mapsto \frac{AM.BM.CM}{\text{Power of } M \text{ wrt } \Gamma} \tag{1}$$

However, neither of the formulations of the objective function gives us many hints about how to proceed with solving the problem.

Both Cartesian and polar coordinatizations seem to complicate the problem rather than simplify it. Further, local analysis via a vector treatment (as was done for Problem 2, the Fermat Problem) gives a rather complicated collection of vectors whose sum needs to be 0 at the minimum point. The primary hope in this is to hit upon some geometric insight that transforms this problem to a far more straightforward one.

We begin by collecting some geometric data :

- The objective function takes identical values on **isogonal conjugates**. The isogonal conjugate of a point P with respect to a triangle $\triangle ABC$ is the intersection point of the reflections of the lines AP, BP and CP in the internal angle bisectors at A, B and C respectively.[6] It could also be described as the point whose

[5]The power of a point P with respect to a circle with center O and radius r is given by $OP^2 - r^2$. It is equal to the signed products of the distance of P from the intersection points of any secant through P with the circle

[6]concurrence follows directly from the trigonometric form of Ceva's theorem

trilinear coordinates are inverse of those of P. If Q is the isogonal conjugate of P, then P is the isogonal conjugate of Q.

The incenter and three excenters are self-isogonal, that is, each of them coincides with its isogonal conjugate.

The orthocenter and circumcenter are isogonal conjugates of each other. The Fermat point and isodynamic point are isogonal conjugates of each other (when all angles are less than $2\pi/3$). The isogonal conjugate of the centroid is termed the **symmedian point** or the **Lemoine point**.

- Of all the important triangle centers, the minimum value appears to be taken at the incenter, where it comes as $2r$. At the circumcenter and orthocenter, it comes as R.

Careful analysis of these values is suggestive of something. What?

4.2. **Making the crucial connection.** We need to choose some geometric attribute that is :

(1) Same for a pair of isogonal conjugates
(2) Minimum for the incenter
(3) Gives values $2r$ for the incenter and R for the circumcenter and orthocenter

Once we have chosen such an attribute, we will try to show that this is indeed equal to the objective function.

The most visible property with respect to which the incenter is minimum is the radius of the circle passing through the projections (feet of perpendiculars) on the sides. This follows because the incircle is smaller than any other circle intersecting all the sides of the triangle.[7]

We next show that the pedal circle of a point, and of its isogonal conjugate, are identical. This follows by showing that for two isogonal conjugates, the projections on two sides are concyclic and then using the **six point trick** discussed in the article on co-incidence problems and methods.[8] Thus, any expression dependent on the pedal circle is the same for a pair of isogonal conjugates.

We have thus guessed that the expression to be minimized is related to the pedal radius of the given point. The pedal radius for the incenter is r and that for the circumcenter and orthocenter is $R/2$. So the objective function seems to be the *diameter of the pedal circle*. This, then, becomes the claim we need to prove :

Claim. The objective function is the diameter of the pedal circle of the given point.

We now need to see what geometric and trigonometric methods we require to be able to prove the above claim. Note that once the claim goes through, we would have shown that the minimum is indeed $2r$ and is attained *only* at the incenter (because the pedal radius is minimum only at the incenter).

[7]To prove this, we must show that for any other point, at least one of its perpendicular distances from the sides is more than the inradius. This is easily done by expressing the area of the triangle ABC as the sum of areas of PBC,PCA and PAB

[8]The pedal circle of a pair of isogonal conjugates is the auxiliary circle of the ellipse touching the sides, with the two isogonal conjugates as its foci. The product of the lengths of perpendiculars of the foci on each side is the same.

4.3. **The geometry that remains.** In the last subsection we used the scientific method of exploration to formulate a reasonable hypothesis. In a sense, obtaining the hypothesis was the hard part. Verifying it, as we shall see, is not a tough job.

Let us look at the structure of the objective function:

$$\frac{BM.CM}{MD}$$

How do we relate this to the pedal circle? Let us, for the moment, consider the pedal triangle. We begin with an observation:

> The pedal triangle of M is similar to the triangle formed by the points where AM, BM and CM again meet the circle.

This result could prove important becasuse it establishes a link between the pedal circle (which is the circumcirlce of the pedal triangle) and the circumcircle of the original triangle. In particular, it provides a way of using side lengths and areas of the two triangles to compute the radius (or diameter) of the pedal circle.

We are now in a position to obtain the formal proof.

Proof. Let P, Q and R be the projections of M on the sides BC, CA and AB of $\triangle ABC$. Let BM and CM meet the circumcircle again in E and F. We first observe that the triangles $\triangle PQR$ and $\triangle DEF$ are similar. This goes through by angle chasing.

What we are now intent on determining is the ratio of similitude. For this purpose, it suffices to determine, say, $\frac{QR}{EF}$.

By similarity of triangles, we have $\frac{EF}{BC} = \frac{EM}{MB}$. We also have $BC = 2R\sin A$ and $QR = AM\sin A$. Here R is the radius of Γ. Combining all these we get $\frac{QR}{EF} = \frac{AM.MB}{2R.EM}$ which, by the cyclic symmetry becomes $\frac{1}{2R}\frac{BM.CM}{MD}$.

In particular, the ratio of circumradii of the triangles $\triangle PQR$ and $\triangle DEF$ is also the same, giving us that $\frac{BM.CM}{MD}$ is twice the circumradius of the pedal triangle of M, as required. □

Interestingly, though perhaps nt unexpectedly, these configurations play an important role in more advanced triangle geometry.

5. Maximal and minimal inscribed and circumscribed ellipses

5.1. **The problem statement.** At the outset, we had talked of three classes of problems: those where we have to *find a point*, those where we have to *find a configuration*, and those where we have to *find a configuration relative to another configuration.* So far, the problems we have seen are of the *find a point* flavour. The problem we shall now discuss is of the *find a configuration* flavour. In these kinds of problems, as we shall see, deciding the parameterization becomes a lot tougher.

The problem is as follows :

Problem 4 (Inellipse and circumellipse)**.** Given a triangle, determine the circumscribed ellipse of minimum area and the inscribed ellipse of maximum area.

Before we begin developing the tools that will help us flay the problem, it may be worthwhile for some time to appreciate the problems that will beset any naive approach.

In problems of the *find a point* flavour, all we had to do to convert the problem to an algebraic one was to choose a suitable coordinate system for the point, and then formulate all conditions in terms of conditions between the values of the coordinates. Thus, the problem could easily be translated into a **nonlinear program**. But in this case, the sheer choice of a coordinate system is extremely daunting.

To parameterize an ellipse in a plane, we need the **center**, the **orientation**, and the lengths of the **semimajor axis** and **semiminor axis**. This totals up to 5 parameters (The center being a point requires two parameters to psecify it). So, we have an optimization problem in terms of these five variables.

But we also have three constraints – namely, those of touching the three sides of the triangle. Each side of the triangle is a line, so that these constraints all trnalsate to equations asserting that a given line is tangent to the ellipse. ANd these are not all. We also need to ensure that the ellipse is in the *interior* of the triangle. Actually, with a little care, we may be able to rid ourselves of that assumption by showing that a minimum will always occur when the ellipse is in the interior of the triangle.

Thus, our nonlinear program is being fed with a 5 variable optimization with 3 constraints that are fairly complicated. This is a challenging system to solve in general terms – it is challenging even when values are plugged in. Thus, even though the actual number of degrees of freedom is 2, we have to handle 5 variables.

Earlier, we observed that with a **change of parameter**, a constrained optimization problem may become unconstrained, or, at any rate, the number and complexity of the constraints may be significantly reduced. We are interested in effecting something of that kind in this problem. As usual, we need to expand our toolkit to find a way of achieving the goal.

5.2. **First steps.** To solve this problem, we make use of the following basic fact : under an **affine transformation** (that is, an invertible transformation preserving collinearity) :

- Collinearity is preserved
- Ratios of areas within a plane are preserved
- Ratios of lengths along a line are preserved
- Any two non-degenerate triangles are equivalent. Equivalence here means that there is an affine transformation taking one triangle to the other. Moreover, fixing the image of a triangle fixes the entire affine transformation on its plane.
- Any two ellipses are equivalent.

The idea is to transform the given *awkward* geometry to something more amenable to manipulation. The two typical *nice* cases are :

- When the triangle is equilateral.
- When the ellipse is a circle.

In this case, we prefer to make the ellipse a circle, as our comfort with non-equilateral triangles is more than our comfort with arbitrary inscribed ellipses.

When the ellipse becomes a circle, it in fact becomes the incircle of the given triangle. The triangle itself changes. However, the ratio of the areas of the inellipse to the triangle remains the same. Thus, the original information about the way the ellipse was inscribed is now captured in the angles of the resultant triangle.

The question now becomes :

> *For what triangle shape is the ratio of the area of the incircle to the area of the triangle maximum?*

By some elementary trigonometric considerations, this can be shown to be the equilateral triangle. The expression to be maximized comes as $\tan A/2 \tan B/2 \tan C/2$, or simply $\prod \tan A/2$. Here the $\prod$ notation indicates that the product is taken cyclically.

Thus the inscribed ellipse with maximum area for any given triangle must be such that an affine transformation takes the triangle to an equilateral triangle and the ellipse to a circle.

What does this mean back in the original triangle? As the incircle touches the midpoints of the sides in an equilateral triangle, and ratios of lengths along a line remain unchanged upon performing affine transformations, the inscribed ellipse with maximum area also touches the sides at their midpoints. This is the well known **Steiner inellipse**.

A very similar approach establish that the minimal circumscribing ellipse is the **Steiner circumellipse** that becomes a circle in the case of the equilateral triangle.

The key idea here was :

Heuristic 1. The variable parameters should be changed from tougher-to-handle parameters (in this case, those specifying ellipse shape) to easier-to-handle parameters (in this case, those specifying triangle shape) by performing some suitable (usually affine) transformation. This has the effect of reducing the number and complexity of the constraints in the formulation of the problem.

6. A minimax problem

This problem was G5 in the IMO 2002 Short List :

> Let $M(S)$ and $m(S)$ denote the maximum and minimum areas of triangles of a pentagon S. What is the minimum possible value of $M(S)/m(S)$?

6.1. **Regular pentagon – anybody's guess?**

The regular pentagon may be the most *natural* guess, and more so when it is seen that the answer in case of the regular pentagon is the Golden Ratio $\Phi = 1.618\ldots$. The challenge is to try and *prove* it.

First of all the pentagon must be convex, as any concave pentagon has $M(S)/m(S)$ more than Φ.

As the quantity $M(S)/m(S)$ is a *ratio* of areas, it is invariant under all affine transformations. If our guess is that the minimum case is attained for a regular pentagon, then the following seems like a reasonable conjecture :

> Any pentagon for which the minimum is attained can be converted, via affine transformations, into a regular pentagon.

To go about this, we use a trick similar to that used previously – make some attribute of the pentagon easy to handle. Earlier on, we had considered two options – convert the ellipse to a circle, or make the triangle equilateral. The former was chosen based on our degree of comfort. Here again there are two choices :

- Project it so that the circumellipse (defined uniquely for a convex pentagon) becomes a circle.
- Project it so that three of the vertices (of the smallest, or the largest?) triangle become the vertices of the regular pentagon we hope to convert the pentagon into.

The first choice definitely appears more symmetric and natural, but is computationally more inconvenient, as can be seen, once we try to reformulate the problem for a cyclic pentagon.

This suggests that we try out the second approach.

6.2. **Chalking out the details.** Let A, B, and C be three of the five vertices forming a triangle with largest area. Then the other two vertices lie in the triangle whose medial triangle is $\triangle ABC$ (that is, the triangle formed by drawing parallels to the sides through the opposite vertices). Call this triangle $A'B'C'$ with A' opposite to A and so on.

The other two vertices, say D and E, thus lie in either 1 or 2 of the three corner triangles of this bigger triangle. Let us say, again without any loss of generality, that neither D nor E lies in the corner triangle $\triangle A'BC$.

The way to project is now clear : we project the figure in such a manner that $\triangle ABC$ becomes an isosceles triangle with vertex angle $\pi/5$. This means that there are points P and Q such that $APBCQ$ is a regular pentagon. What remains to establish is that the minimum is attained precisely when D and E are P and Q in some order.

First, the geometry of the regular pentagon establishes that P lies on BC' and Q lies on CB'. Clearly, if either D or E lies inside either of the triangles $\triangle AQC$ or $\triangle APB$ the ratio will be more than that for the regular pentagon, so the only case we are left to consider is when both D and E are in $\triangle AB'Q \cup AC'P$. But in this case we have $\mathsf{Ar}\,\triangle ADE = \frac{1}{2}|AD||AE|\sin\angle DAE$ where either $\angle DAE \leq \pi/5$ or $\angle DAE \geq 3\pi/5$. Observing that $|AD|, |AE| \leq |AP|, |AQ|$ and $\sin\angle DAE \leq \sin\angle PAQ$ with equality holding iff P and Q coincide with D and E, we have $\mathsf{Ar}\,\triangle ADE \leq \mathsf{Ar}\,\triangle APQ$ and we are done.

Aside 1.
A comment accompanying this problem in the IMO Short List booklet states that the above method does not generalize to higher n, and the answer is unknown to the authors even for $n = 6$. The general problem is also related to the famous **Heilbronn problem** on the smallest triangle formed from n points in the unit square.

7. A problem from Romania

The following problem appeared in a Romanian selection test :

Problem 5. Let P be a point inside $\triangle ABC$ so that $PA = 1$, $PB = 2$, $PC = 3$. Suppose $\angle BAC = \pi/3$. What is the maximum possible area of $\triangle ABC$?

7.1. **Formulating the problems and deciding parameters.** To understand what precisely this problem asks for, we first observe that out of the 6 real parameters needed to specify completely a triangle in $\mathbb{R}^2$, 5 have been used up :

- A fixed location for A, using up 2 parameters.
- A circular locus for B, using up 1 parameter.
- A circular locus for C, using up 1 parameter.
- An additional fact that $\angle BAC$ is $\pi/3$.

A good choice of the varying parameter is the pair of angles $\alpha = \angle BAP$ and $\beta = \angle PAC$ because this pair easily encodes the fact that $\angle BAC = \pi/3$ (as $\angle BAC$ is $\alpha + \beta$).

The next step is to choose some convenient expression for $\mathsf{Ar}\,\triangle ABC$, in a manner that makes good use of the given data, in particular the value of $\angle BAC$.

The best choice for that is :

$$
\begin{aligned}
\mathsf{Ar}\,\triangle ABC &= \frac{1}{2}|AB||AC|\sin\angle BAC \\
\implies \mathsf{Ar}\,\triangle ABC &= \frac{\sqrt{3}}{4}|AB||AC| \\
\implies \mathsf{Ar}\,\triangle ABC &\propto |AB||AC|
\end{aligned}
$$

Thus the value of $|AB||AC|$ needs to be maximized.

The next crucial step is to observe that AB is exclusively a function of α, while AC is exclusively a function of β. This reminds us of a basic result in extremizing functions of several variables.

7.2. A brief recall of optimizing multi variable functions.

Lemma 1 (Separable function optimization)**.** If $\sum_{i=1}^{n} x_i$ is a constant, then the local extrema of $\sum_{i=1}^{n} f_i(x_i)$ are attained where $f_i'(x_i)$ are equal $\forall i$. When all $f_i''(x_i)$ are negative, the tuple corresponds to a local maximum and when all $f_i''(x_i)$ are positive, the tuple corresponds to a local minimum. The global maximum is either a local maximum, or is attained when one or more of the variables is at its extreme value.

Here, instead of a summation, there is a product, but that is easily remedied by taking a logarithm, and observing that the corresponding observation for products is :

And thus in this case, the condition becomes :

$$
\frac{d\,\log|AB|}{d\,\alpha} = \frac{d\,\log|AC|}{d\,\beta} \tag{2}
$$

Or equivalently :

$$
\frac{d\,|AB|}{|AB|d\,\alpha} = \frac{d\,|AC|}{|AC|d\,\beta} \tag{3}
$$

Heuristic 2. Whenever any kind of optimization needs to be done it typically makes sense to observe the special structure of the expression we seek to optimize. This will help us to apply ideas and techniques already well developed for those structures. Directly going in for differentiation may yield the correct results but may be computationally tedious and may not give the required insights for the next step.

7.3. **Polar coordinates.**

The locus of B is a circle centered at P. We need to observe how the distance of B from A varies with the angle that AB makes with the line AP.

This suggests that we are interested in determining relations involving the (r, θ) polar coordinates where the pole (origin) is A and the initial line is AP. A basic result in the polar coordinates of loci is :

$$
\frac{d\,r}{r\,d\,\theta} = \cot\rho
$$

Where (r, θ) are the polar coordinates at a given point on the locus, and ρ is the angle made by the tangent to the curve at the given point with the radial from the origin.

Conceptually the numerator corresponds to the rate of movement along the outward radial, and the denominator represents the rate of movement perpendicular to the radial. Thus, their ratio should give the direction of actual motion.

We now come back to the two loci we were originally considering : the circles about P.

The value of ρ corresponding to the locus of B is $\pi/2 - \angle ABP$, and the ρ corresponding to the locus of C is $\pi/2 - \angle ACP$.

Which finally gives :

$$\tan \angle ABP = \tan \angle ACP \tag{4}$$
$$\implies \angle ABP = \angle ACP \tag{5}$$

To round off this exercise, the following needs to be established :

- This is indeed a local maximum, not a local minimum.
- The endpoint case gives a smaller answer.

Key Point 4. The choice of the right coordinate system needs to be taken keeping in mind the kind of quantities we are interested in measuring. This means that we should be comfortable with the important coordinate systems so that we are able to spot the situations where they may come in handy. This is just like in the last section, a good feeling of varied optimization problems helped us spot the right way to go about optimizing.

7.4. **The final calculations.** Let us summarize where we have reached so far. In our search for the maximum area, we chose a suitable expression, and realized that this was the product of functions of two variables whose sum was a constant. Applying some basic principles of calculus and the ingenious use of polar coordinates, we avoided much computation to reach the geometric fact that the optimum configuration is attained when $\angle ABP = \angle ACP$. It now remains to compute the area of the triangle from this. The final calculations now involve simplification using the sine rule. The following sequence of computations are a route to the answer :

$$\begin{aligned} \frac{\sin\alpha}{2} &= \frac{\sin\beta}{3} \\ \implies \frac{\sin\alpha}{2} &= \frac{\sin(\pi/3-\alpha)}{3} \\ \implies & \end{aligned}$$

The use of polar coordinates is not a must, and an alternative approach would be to compute AB in terms of α by some trigonometry and algebra and then take derivatives. The approach, however, is geometrically less illuminating and computationally more tedious.

8. The technique toolkit

In the first two examples – the inellipse example and the pentagon example, our guiding theme was the idea of performing a suitable affine transformation, and using the resulting new geometry effectively. In the third example, we looked at how a suitable choice of parameters, and a careful application of differentiation principles as well as polar coordinates, did the trick.

Affine transformations must be used with care :

Key Point 5. Affine transformations do not preserve similarity, they do not preserve angles. The important things that they do preserve have already been mentioned at the outset. Thus, in cases where the original geometry is very crucial to the optimization, an affine transformation is likely to be of little use.

Affine transformations generalize similarity preserving transformations by relaxing the assumption that angles are preserved, while preserving the assumptions that collinearity is preserved. The reverse question may be : are there transformations where angles are preserved, though collinearity may be destroyed? There are, but they cannot be defined over the whole plane, rather, they can be defined over the plane along with a point at infinity (known as the **Riemann sphere**). An example of such a transformation is an inversion.

Thus, these transformations are not of much direct use in area optimizations, but they come of use in various other kinds of geometrical problems.

Appendix A. Terms and definitions

A.1. General terminology on optimization.

(1) The **objective function** in an optimization problem is the function whose value needs to be optimized (maximized or minimized.
(2) The **parameters** are the quantities in terms of which the objective function is defined. These are also the **degrees of freedom**.
(3) A **configuration** is a particular tuple of values of the parameters. The set of all possible configurations is termed the **configuration space**. A configuration is thus a **point** in the configuration space. It is also sometimes termed a **solution**.
(4) A **change of parameter** is a reparamterization of the configuration space. The set of points remains the same but the way they are described using coordinates changes.
(5) The **constraints** are relationships that the configuration must satisfy. A configuration satisfying all the constraints is termed a **feasible configuration**, or a **feasible solution**.
(6) An **unconstrained problem** is one where there are no constraints. Thus, all configurations are feasible. A constrained problem may be changed to an unconstrained problem via a suitable change of parameter.
(7) An **optimum** is an optimum value of the objective function, subject to the constraints. An **optimal point** is a point at which the function takes the optimum value.
(8) A **unique optimum** is an optimum value that is attined only at one point.
(9) A **local optimum** is a value attained at a point, such that there is an open neighbourhood of the point in which the value is optimum. The point is a **locally optimal point**.
(10) A **strong local optimum** or **isolated local optimum** is a value attained at a point, such that, for some open neighbourhood of that point, it is the unique optimum in that neighbourhood. The point is called a **strongly locally optimally point**.

(11) The **derivative** of a function from a subset of $\mathbb{R}$ to $\mathbb{R}$ is a function that, to each point, associates the rate of change of the function at that point. The derivative of a function f is denoted as f'. The process of going from a function to its derivative is termed **differentiation**. The **left hand derivative** is the derivative where approach to the point is from the left, and the **right hand derivative** is the derivative where approach to the point is from the left.

A function is termed **differentiable** at a point if its derivative at that point exists.

(12) The **perturbation principle** states that if a point is optimal, or even locally optimal, then sufficiently small perturbations to it render it less optimal. When the point is astrongly locally optimal, the perturbations make it strictly less optimal.

(13) The **first derivative test** states that if a function from some interval in $\mathbb{R}$ to $\mathbb{R}$ is differentiable at a point, and is also a local optimum at that point, the derivative at the point is 0.

(14) The **vector gradient** ∇ of a function $\mathbb{R}^n \to \mathbb{R}$ is a vector function that, at each point, represents the rate of change of the scalar function near the point. The direction of the vector gradient indicates the direction along which the change occurs and the magnitude indicates the rate at which it changes.

A scalar function is **differentiable** at a point if the vector gradient at the point is well defined.

(15) The **derivative (gradient) test** for scalar functions on sets in $\mathbb{R}^n$ states that if a function attains a local optimum at a point and is differentiable at that point, the vector gradient at the point is zero.

A.2. **Coordinate systems.**

(1) A **Cartesian coordinate system** comprises a choice of a point called **origin**, and directed lines through these origins called **axes**, such that the directions are linearly independent and span the whole space. The coordinates of a point in a Cartesian coordinate system are lengths along each of the coordinate axes, so that, if we add up the vectors corresponding to these lengths, we get the position vector of the point started with.

Cartesian coordinates are typically chosen as **rectillinear Cartesian coordinates**, that is, the coordinate axes are chosen to be pairwise orthogonal.

(2) A **polar coordinate system** for the plane comprises a choice of a point, called the **pole** or **origin** and a ray starting at it, called the **initial line**. The polar coordinates of any point include the **radial coordinate** that specifies its distance from the pole, and the **argument**, that specifies the angle made from the initial line to the line segment directed from the pole to the given point.

Polar coordinate systems do generalize to higher dimensions. The **radial coordinate** remains, and the remaining coordinates are angular quantities. These measure angles made with each member of a flag of subspaces going to the full space. For instance, in three dimensions, the polar coordinates are given by – distance from the origin, angle made with the xy plane, and the angle made in the xy plane by its projection on the xy plane, with the x axis.

A.3. **Geometric terminology.**

(1) The **Fermat point** or **Torricelli point** of a triangle is the point that minimizes the following objective function: "sum of distances from vertices of the triangle".

(2) A **Steiner network** for a collection of points is a collection of vertices (nodes) and edges (line segments joining pairs of nodes), such that all the points in the original

collection are connected via the network. The length of the Steiner network is the total length of all the edges. Steiner posed the problem of finding the Steiner network of minimum length for a given set of points in the plane.

(3) Two points are **isogonal conjugates** with reference to a triangle if the following holds: The lines joining them to the vertices are reflections of each other in the angle bisectors at the vertices. That is, points P and Q are isogonal conjugates with respect to $\triangle ABC$ provided that $\angle ABP = \angle CBQ$, $\angle BCP = \angle ACQ$ and $\angle CAP = \angle BAQ$. Two of the three conditions automaticlally imply the third, by the trigonometric form of Ceva's theorem. The orthocenter and circumcenter are isogonal conjugates, the inceter is its own isogonal conjugate. The Fermat point, whne inside the triangle, is the isogonal conjugate of one of the isodynamic points.

(4) The **reflection property** of an ellipse says that the normal to the ellipse at any point on it is the internal angle bisector of the line segments joining it to the foci.

This reflection property can be deduced in a variety of ways, of which two are suggested below :

- Determine the gradient of the scalar function given by the sum of distances from the foci.
- Consider a tangent line through P to the ellipse with foci B and C. The point P is now the point on the tangent line minimizing $PB + PC$. To determine this point reflect C in the tangent, make the straight line joining this to B, and determine the intersection point with the tangent. This is P.

Appendix B. Problems

The problems given here are somewhat indirect. Unlike the Concept Testers, these problems are not to be directly solved based on the material covered in the text. However, they are based on ideas already encountered within the article text.

(1) Using the differential geometry methods outlined in the beginning of the section, determine the points P in the plane of triangle ABC that minimize the following expressions :
- $PA.BC + PB.CA + PC.AB$
- $PA^2 + PB^2 + PC^2$

(2) Let D, E and F be the projections from P onto the sides BC, CA and AB respectively. Find points P in the plane that minimize the following expressions :
- $PD + PE + PF$
- $PD^2 + PE^2 + PF^2$

(3) This is a modification of the area optimization problem discussed earlier. Consider $\triangle ABC$ with $\angle BAC = \pi/3$ and let P be a point in the interior so that $PA = 1$, $PB = 2$ and $PC = 3$. Determine the maximum and minimum possible lengths of the line segment BC. Try to follow a similar approach to that used here.

(4) These are problems regarding optimizations for a point between two lines. Let l_1 and l_2 be two rays with a common endpoint O. Let P be a point in the smaller angle between l_1 and l_2. Find points $A \in l_1$ and $B \in l_2$ such that :
- $AP + PB$ is minimum : show that for such A and B, if Q is the projection of O on AB, then P and Q are isotomic conjugates on AB.
- $PA.PB$ is minimum : give a geometric construction for determining such A and B.
- $1/PA + 1/PB$ is maximum : find a condition similar to that in the first part for the projection of O on AB.

Index

alternative optimal solution, 3

Cartesian coordinates, 5
constraint, 3
criterion function, 3

equality constraint, 3

Fermat point, 4

inequality constraint, 3
initial line, 6
isogonal conjugate, 11

linear programming, 2

non-linear programming, 2

objective function, 3
optimal solution, 3
optimization problem, 2

parameter, 3
polar coordinates, 6
pole, 6

six point trick, 12
Steiner circumellipse, 15
Steiner inellipse, 15

Torricelli point, 4

unconstrained optimization problems, 3