Shortest Road Network Connecting Cities

Martin J. Gander, Kévin Santugini and Antonio Steiner

21st May 2008

Abstract

The problem of finding the shortest road network is very old, the first written record goes back to a letter of Gauss replying to a question posed by Schumacher. After a historical review, we study the minimal road network problem using elementary mathematics, and show solutions for several examples.

1 Introduction

The problem of finding the shortest network connecting cities can be traced back to a correspondence between Schumacher and Gauss in 1836. In a first letter [1], Schumacher asks Gauss about a simpler, but related problem (see also Figure 1):

> Mir ist neulich ein Paradox vorgekommen, das ich so frei bin Ihnen vorzulegen. Bekanntlich ist, wenn man bei einem Vierecke $ABCD$ einen Punct sucht, von dem die Summe der an die Winkelpuncte gezogenen Linien ein Minimum sey, der gesuchte Punct der Durchschnittspunct der Diagonalen E. Lässt man nun die Puncte A, B in den Linien DA, BC immer mehr hinaufrücken, bis sie am Ende in F zusammenfallen, so fällt auch E zugleich in F, das Viereck verwandelt sich in das Dreieck DFC, und man hätte den Punct F als denjenigen, von dem die Summe der an die Winkelpuncte F, C, D des Dreiecks gezogenen Linien ein Minimum sey. Das ist aber bekanntlich nur wahr, wenn der Winkel $F \geq 120^o$. [1]

Gauss replied only two days later [2], with the following explanation:

[1] I recently came across a paradox I would like to explain to you. It is well known that if one searches in a quadrangle $ABCD$ the point which, when connected with the corners, leads to the shortest sum of lines, one finds the intersection of the diagonals E. If we

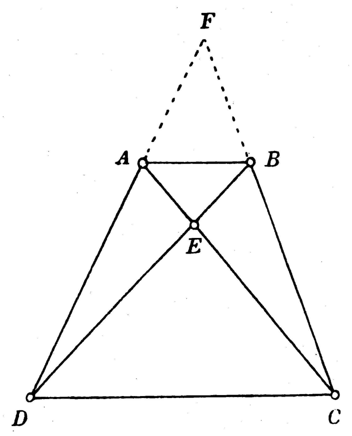

Figure 1: Example sent to Gauss by Schumacher in 1836

Was Ihr Viereck betrifft, so heisst doch die Aufgabe so: Vier Pkte a, b, c, d sind gegeben, man soll einen 5ten x finden, so dass $ax + bx + cx + dx$ ein Minimum wird, und das ist von den 3 Durchschnittspunkten ab mit cd, ac mit bd, ad mit bc der eine, wo man für die Auswahl die Bedingung entweder leicht auf Anschauung reduciren, oder analytisch einkleiden könnte. Lassen Sie nun a, b, c fest sein, und d dem c immer näher rücken, so bleibt diese Auflösung noch immer so lange richtig, als Sie nicht c mit d zusammenfallen lassen. Fällt aber c mit d zusammen, so erfordert Geist und Buchstabe der mathematischen Aufgabe, als solcher, dass Sie dann c zweimahl zählen, also in dem Dreieck abc $ax + bx + 2cx$ zu einem Minimum machen, wo sich die allgemeine Auflösung noch immer als richtig ausweiset. [2]

let now the points A and B follow the lines DA, BC up to F, then E also arrives at F, the quadrangle becomes the triangle DFC, and the point F becomes the point which minimizes the the sum of the connections of the corners F, C, D. It is however well known that this is only the case if the angle at F is less than 120^o.

[2]Concerning your quadrangle, the problem should be formulated like this: four points a, b, c, d are given, and one should find a 5th point x, such that $ax+bx+cx+dx$ is minimal, and this is of the intersection points ab with cd, ac with bd, ad with bc the one, which one can easily determine by either just looking at the problem, or by proceding analytically. If you keep now a, b, c fixed, and let d approach c, the solution remains correct, as long as c and d do not coincide. Once however c and d coincide, the mathematical nature of the problem requires that you count c twice, and hence you have to minimize in the triangle abc the quantity $ax + bx + 2cx$, in order for the solution to remain correct.

The problem Schumacher was interested in, namely the problem of finding one point that connects a given set of points with a shortest network, is actually older, it goes back to 1638, when Descartes asked Fermat to study curves, whose points have a constant sum of distances to four given points. Motivated by this question, Fermat asked in 1643 for the case of three given points which point would minimize the sum of distances [3], and Torricelli was the first to solve the three point case [4], which gave the problem the name "Fermat-Torricelli Problem". The solution for four points was given by Fagnano [5], and a generalization to n points was given independently by Tédenat [6] and Lhuilier [7], who was professor of mathematics at the imperial academy of Geneva. While one can construct the solution for up to four points with ruler and compass, it was shown using Galois theory that for more points in general position, such a construction is not possible.

In the letter where Gauss replied to Schumacher [2], he also proposed a different, but related problem, namely the one of really finding the shortest network:

> Ist bei einem 4 Eck nicht von der stricten mathematischen Aufgabe, wie sie oben ausgesprochen ist, sondern von dem kürzesten Verbindungssystem die Rede, so werden mehrere einzelne Fälle von einander unterschieden werden müssen, und es bildet sich so eine recht interessante mathematische Aufgabe, die mir nicht fremd ist, vielmehr habe ich bei Gelegenheit einer Eisenbahnverbindung zwischen Harburg, Bremen, Hannover, Braunschweig sie in Erwägung genommen und bin selbst auf den Gedanken gekommen, dass sie eine ganz schickliche Preisfrage für unsere Studenten bei Gelegenheit abgeben könnte. Die Möglichkeit verschiedener Fälle erläutern wohl hinreichend folgende Figuren, wo in der dritten Figur die Verbindung von c nach d direct gehen muss (was wirklich bei obigem Beispiel der Fall wird). [3]

We can see that Gauss had worked on this problem for the very practical reason of the construction of the shortest possible rail network linking four important cities in Germany. This type of problem appears in many applications, in particular in circuit layout and network design, and it is now known

[3]If in a quadrangle one asks in contrast to the question above what the shortest network is, then one needs to distinguish several cases, and we obtain quite an interesting mathematical problem, which I am familiar with, since I had the opportunity to study it in the context of train connections between the cities of Harburg, Bremen, Hannover and Braunschweig, and I came to the conclusion that this problem would be an excellent prize problem for our students. The different possibilities are illustrated in Figure 2, where in the third drawing the connection from c to d must be a direct one (like in the example above as well).

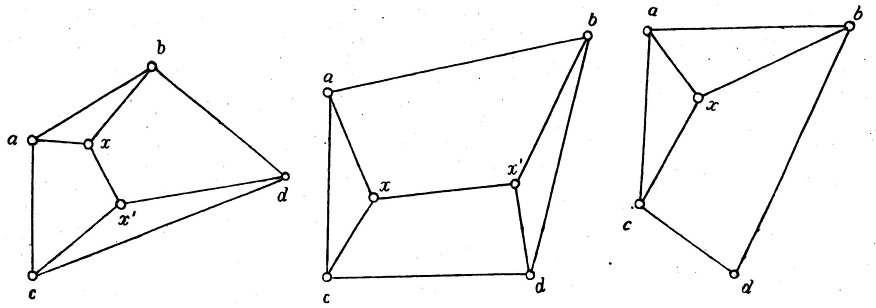

Figure 2: Examples for the shortest network problem by Gauss.

under the name "Steiner Tree Problem" [8, 9], although it is not clear what the eminent geometer Jakob Steiner (born in Utzenstorf, Switzerland, not to be confounded with A. Steiner, the third author of this article) contributed to this problem. The basic Steiner tree problem is to find the shortest network that connects a set of planar points. For some sets of points, adding an intersection point, known as Steiner point, can reduce the length of the network. Du and Hwang [10] showed that adding an intersection point can not reduce the length of the network by a factor of more than $1-\frac{\sqrt{3}}{2}$, about 13%, using an equilateral triangle with the center of the triangle added as an intersection. In 1977, Graham, Garey and Johnson showed that for the general case, the problem of picking the optimal intersection points is NP-complete [11], see also the article by Kolata [12].

In this paper, we study, using elementary mathematics, the problem of finding the shortest network linking n cities in a $2d$ plane. Throughout this paper, we call

A city: a point of the given initial set of cities.

A Steiner point: a point in a network that is not a city and where two or more roads connect (with a non flat angle between the roads).

A node: either a Steiner point or a city.

2 A Limit on the Number of Steiner Points

Between two points in a flat plane, the curve of minimal length is a straight line. The network of minimal length is therefore a union of straight lines or segments, and we obtain a first relation between the number of roads and the number of nodes:

Lemma 2.1. *We consider n cities and a network of minimal length connecting those cities. If p is the number of Steiner points and r the number of roads, then*

$$r = p + n - 1.$$

Proof. A network of minimal length made of a collection of segments cannot contain a cycle. If there was a cycle, we could remove any road from the cycle and achieve a shorter connecting network. A connected network without cycles is a tree and the number of roads r in a tree is equal to the number of nodes minus 1. □

The next lemma shows that Steiner points are real intersections, i.e. they connect more than two roads.

Lemma 2.2. *In a shortest network connecting n cities, any Steiner point is connected to at least three roads.*

Proof. If a Steiner point P were connected to only two roads, then we could remove the Steiner point and connect directly the two nodes that were connected to P thus obtaining a network of smaller or at most equal length. □

Lemma 2.3. *We consider n cities and a network of minimal length connecting those cities. If p is the number of Steiner points and r the number of roads, then the inequality*

$$r \geq \frac{3p + n}{2}$$

holds.

Proof. We count the roads connected to each node. Each city is connected to at least one road and each Steiner point to at least three. We sum over the nodes and since each road is counted twice (once for each extremity of the road), we obtain $r \geq \frac{3p+n}{2}$. □

We will later refine this result, once we have studied the network of minimal length for three cities. Using the previous lemmas however, we can already give an upper bound on the number of Steiner points in a shortest network.

Theorem 2.4. *We consider n cities and a network of minimal length connecting those cities. Let p be the number of Steiner points. Then we have the inequality*

$$p \leq n - 2.$$

Proof. This result is obtained by combining the results of Lemma 2.1 and Lemma 2.3. □

3 Minimal Network Connecting Three Cities

It is instructive to study the special case of three cities, since we obtain geometric information about the shape of the network near Steiner points.

Proposition 3.1. *Let us consider three cities A, B and C. Then, there are two possibilities:*

1. *if one angle in the the triangle is more than 120^o, then the shortest network contains no Steiner points and is made of two roads that connect at the obtuse angle.*

2. *Otherwise, the shortest network is the one that connects all three cities to the unique Steiner point O that is inside the triangle ABC and such that the angles $(OA, OB) = (OB, OC) = (OC, OA) = \pm 120^o$.*

Proof. Let O be the Steiner point connecting A, B and C. The network length associated with O is $l(O) = OA + OB + OC$. To find an optimal position for O, we need to know the variation of $l(O)$ with respect to O, i.e. we need to compute the first derivative of the network length when O moves in the $\boldsymbol{h}$ direction,

$$\frac{l(O + t\boldsymbol{h})}{\mathrm{d}t} = (\frac{\boldsymbol{OA}}{OA}, \boldsymbol{h}) + (\frac{\boldsymbol{OB}}{OB}, \boldsymbol{h}) + (\frac{\boldsymbol{OC}}{OC}, \boldsymbol{h}).$$

For this expression to vanish for all $\boldsymbol{h}$, we thus need to have

$$\frac{\boldsymbol{OA}}{OA} + \frac{\boldsymbol{OB}}{OB} + \frac{\boldsymbol{OC}}{OC} = 0. \tag{1}$$

This can however only happen if the vectors $\boldsymbol{OA}$, $\boldsymbol{OB}$ and $\boldsymbol{OC}$ are at angle 120^o from each others. If ABC has an angle superior to 120^o, then O cannot exists and the shortest network has no Steiner point. If no angle is superior to 120^o, there is a unique point O satisfying 1, which can be constructed by drawing three circular arcs, one over each side of the triangle, see Figure 3, which contain all points inside the triangle ABC, such that the angle with the two corner points under the arc equals 120^o. The radius of each circular arc is $\frac{2\sqrt{3}}{3}$ times the length of the side, which is constructible. The Steiner point O is then located at the intersection of the three circular arcs, see Figure 3.

The length of the network passing by O can also easily be computed: if

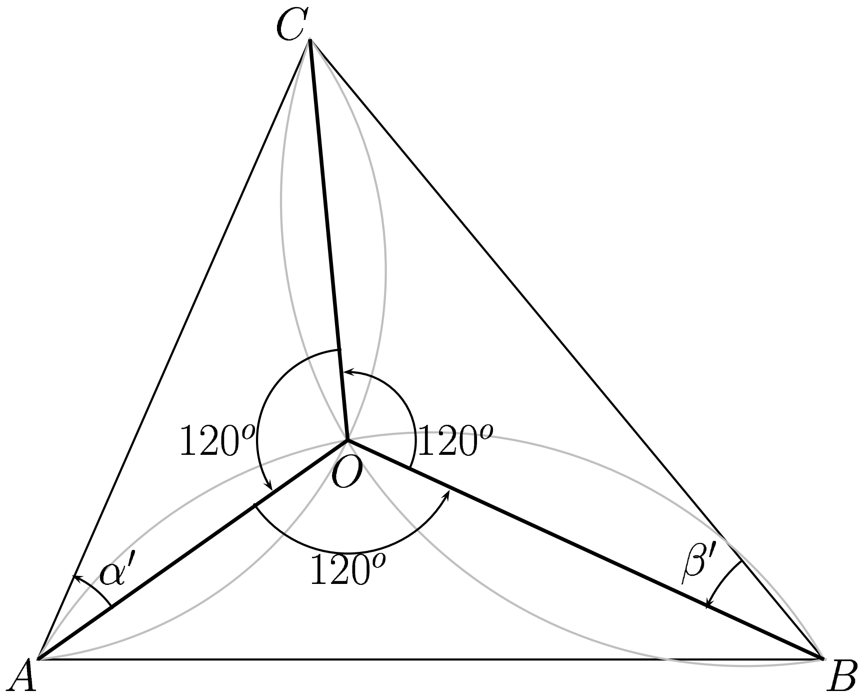

Figure 3: Construction of the optimal Steiner point for a triangle

we note $\alpha' = (\boldsymbol{AO}, \boldsymbol{AC})$ and $\beta' = (\boldsymbol{BC}, \boldsymbol{BO})$ then

$$
\begin{aligned}
AO &= \frac{\sin 60^o - \alpha'}{\sin 120^o} AC \\
BO &= \frac{\sin 60^o - \beta'}{\sin 120^o} BC \\
CO &= \frac{\sin \alpha'}{\sin 120^o} AC \\
CO &= \frac{\sin \beta'}{\sin 120^o} BC
\end{aligned}
$$

Therefore,

$$AO + BO + CO = \cos(\alpha')AC + \cos(\beta')BC.$$

The network with one Steiner point in O is therefore smaller than any of the three networks with no Steiner point. □

4 Some results for the minimal network in the general case

From the case of three cities, we infer some interesting properties that the shortest network for n cities must satisfy.

Proposition 4.1. *The shortest network connecting n cities has the following properties:*

1. *Two roads connecting a common node form an angle superior or equal to* 120^o.
2. *No node is connected to more than three roads.*
3. *Any node connected to three roads has roads forming* 120^o *angles.*
4. *All Steiner points are connected to exactly three roads.*

Proof. If two roads connecting three nodes, A, B, C form an angle smaller than 120 degrees at their common node B, then by Proposition 3.1, we could construct another Steiner point O such that $OA+OB+OC$ is smaller than $AB+BC$.

As the full circle is only 360 degrees, if a node N were to be connected to 4 or more roads, there would be at least two roads connecting N that would form an angle inferior to 120 degrees. For the same reason, if a node is connected by three roads, no two roads connecting it can form an angle larger than 120 degrees: if that were the case, the remaining road would form an angle smaller than 120 degrees with at least one of the two roads.

A Steiner point is connected to at least three roads by Lemma 2.2. We just proved a node could not be connected to more than three roads. Therefore, a Steiner point is always connected to exactly three roads. □

In turn, this allows us now to improve Theorem 2.4:

Theorem 4.2. *We consider n cities and a network of minimal length connecting those cities. Let p be the number of Steiner points, r be the number of roads, and let N be the number of connections to cities (a road connecting two cities will be counted twice). We have $N \geq n$, and*

$$r = \frac{3p+N}{2}, \quad \textit{and} \quad p = 2n - N - 2.$$

In particular, that means that if all cities are connected to only one road, then $p = n - 2$.

5 Examples

In this section, we give examples of shortest length networks.

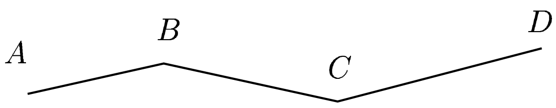

Figure 4: A linear network

5.1 Shortest network for the quadrangle

By Theorem 2.4, there can only be 0, 1 or 2 Steiner points. We enumerate all the possibilities for the shortest network:

No Steiner point, degenerate case One of the cities is exactly placed as the Steiner point of the other three cities. Connect this city to the three other ones to have the shortest network.

No Steiner point, normal case Two cities are connected to two roads and the last two cities are connected to only one road. This means the network is a linear succession of roads, see Figure 4.

One Steiner point The Steiner point is connected to three cities. The last city is directly connected to one of the other cities, see for example the third drawing by Gauss in Figure 2.

Two Steiner points Each Steiner point is connected to two cities and to the other Steiner point, see the first and second drawing by Gauss in Figure 2.

We consider four points of a square and we are going to compute the shortest network. Let a be the length of the edges. With no Steiner point the length of the shortest network is $3a$. With one Steiner point, the Steiner point is connected to three cities and the network length is $\frac{a}{2}(2+\sqrt{6}+\sqrt{2}) \approx 2.93a$. With two Steiner points, the Steiner points are connected to each other and to two cities and the network length is $(\sqrt{3}+1)a \approx 2.73a$, see Figure 5.

5.2 Shortest network for the regular pentagon

We now compute the shortest network for the regular pentagon with the maximum number of Steiner points: three. No two cities can be connected directly when the number of Steiner points is maximal. So each city is connected to a Steiner point. In turn, a Steiner point cannot be connected to three cities, as it would leave the network disconnected. So there are two Steiner points S_1 and S_2 connected to two cities and one Steiner point S_3

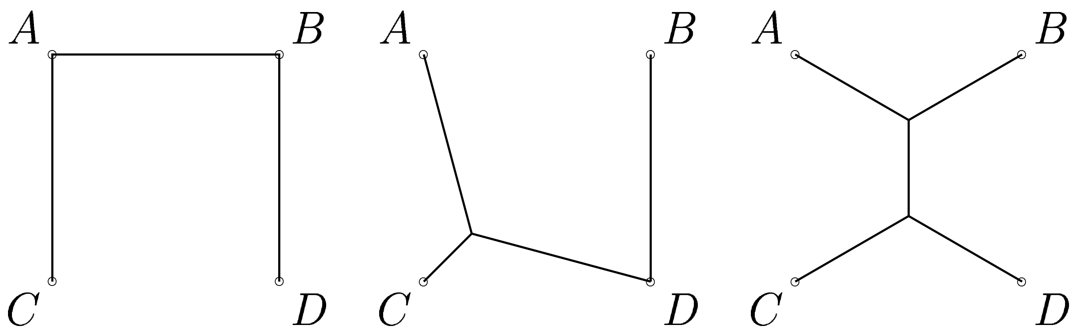

Figure 5: Shortest network for a square

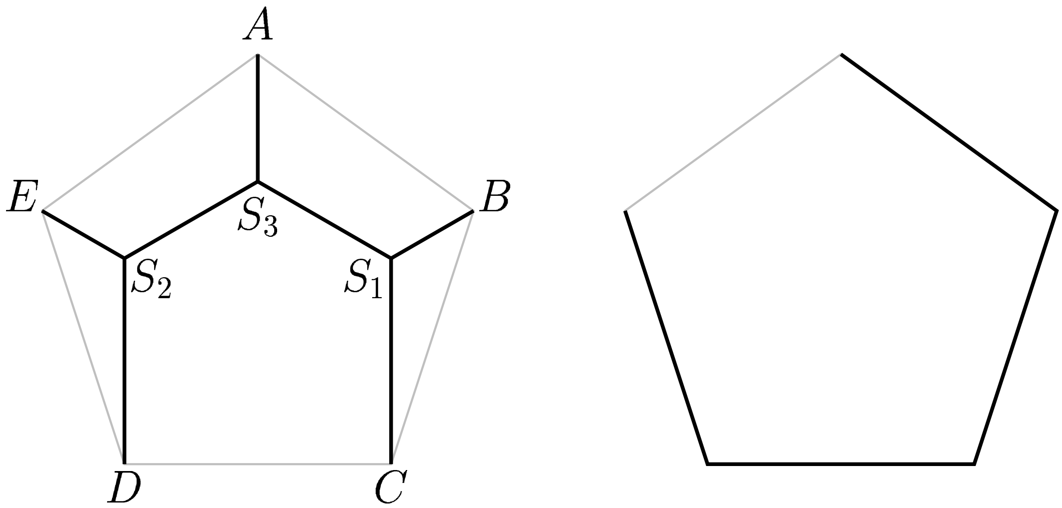

Figure 6: Networks with three and no Steiner points for the regular pentagon

connected to only one city and to the two other Steiner points, see Figure 5.2. We denote by R the length between the center of the pentagon and any of its vertices. The length of each road of this network can easily be computed:

$$CS_1 = DS_2 = 2R\frac{\sin(36^o)\sin(42^o)}{\sin(120^o)} \approx 0.9083R,$$
$$BS_1 = ES_2 = 2R\frac{\sin(36^o)\sin(18^o)}{\sin(120^o)} \approx 0.4195R,$$
$$S_2S_3 = S_1S_3 = 2R\sin(36^o) \approx 0.6787R,$$
$$AS_3 = (1 + cos(36^o))R - CS_1 - R\sin(36^o)/\sqrt{3} \approx 0.5614R.$$

The total length of the network is thus $2CS_1 + 2BS_1 + 2S_2S_3 + AS_3 = 4.5743R$. The network with no Steiner points cannot be shorter, since the

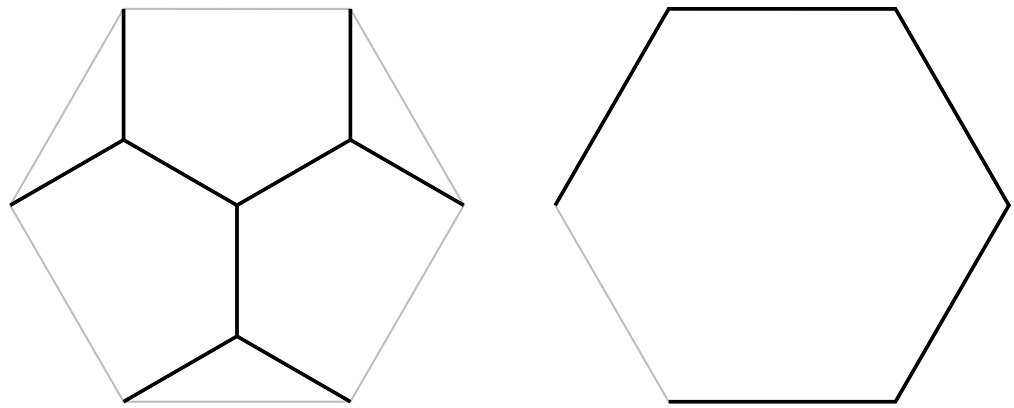

Figure 7: Networks with four and no Steiner points for the regular hexagon

angle between two adjacent edges is $112^o < 120^o$. Indeed, the length is $8\sin(36^o)R = 4.7022R$.

5.3 Shortest network for the regular hexagon

We compute the shortest network for the regular hexagon with the maximum number of Steiner points: four. The intuitive way is to first add three Steiner points each of them connected to two adjacent cities. Then, we connect those three Steiner points through the fourth Steiner point located at the center of the hexagon, see Figure 7. If each edge has length R, then the length of each road connected to a city is $\frac{\sqrt{3}}{3}R$. The length of a road connecting a Steiner point to the Steiner point located at the center of the hexagon is $(\sqrt{3}/2 - \sqrt{3}/6)R = \frac{\sqrt{3}}{3}R$. Therefore the total network length is $3\sqrt{3}R = 5.196R$. This is longer than the network of length $5R$ with no Steiner points.

5.4 Shortest network for the regular polygon with $3 \cdot 2^n$ vertices

We only consider the shortest network with the full number of Steiner points. Consider the outer layer of Steiner points. If $n > 1$, every two cities on the polygon are connected to a Steiner point. We place the Steiner point S connected to two adjacent cities A and B such that $AS = BS$ and $(\boldsymbol{SA}, \boldsymbol{SB}) = 120^o$. There are $3 \cdot 2^{n-1}$ Steiner points: they form a regular polygon and we can consider them to be the cities of another network

Number of vertices	Length of network
3	3
6	5.19615
12	7.82894
24	10.7892
48	13.9837
96	17.3390

Table 1: Length of network with full number of Steiner points for regular polygons

and connect them via other Steiner points. By induction, this allows us to compute the length of such a network.

Let R be the length between the center of the polygon and any city (located at the vertices of the regular polygon). First, we compute the distance of the outer layer of Steiner points to the center of the polygon. This distance is $R\cos(180^o/N) - R\sin(180^o/N)/\tan(60^o)$ with $N = 3 \cdot 2^n$. The distance of a road connecting a Steiner point to a city is $R\sin(180^o/N)/\sin(60^o)$.

Therefore, if we denote by l_n the length of such a network (for $R = 1$), then we have the relation $l_{n+1} = N\sin(180^o/N)/\sin(60^o) + (\cos(180^o/N) - \sin(180^o/N)/\tan(60^o))l_n$ and $l_0 = 3$. Can this configuration always be the shortest network ? It cannot! For a triangle, we get a length of 3. For an hexagon, we get a length of 5.196 and for a dodecagon, we get a length of 7.83 which is superior to 2π: the length of the full circle. This means that for the dodecagon and for the polygons with more vertices, the network with no Steiner point is shorter. See Table 5.4.

6 Conclusion

To conclude, we use our results to compute the solutions of the three problems Gauss used as illustration in his letter to Schumacher [2]. We show in Figure 8 in green the shortest network we obtained, and in red the optimized shape of the network topology which Gauss indicated in his drawing, which leads however to a network longer than the green one: in the first case, the red network is of length 12.6, and the green one 11.6 (we measured the coordinates to millimeter accuracy), in the second the red network is not really a stationary point, whereas the green one is of length 15.5, and in the last case the red network is of length 12.44, while the green one is 12.39.

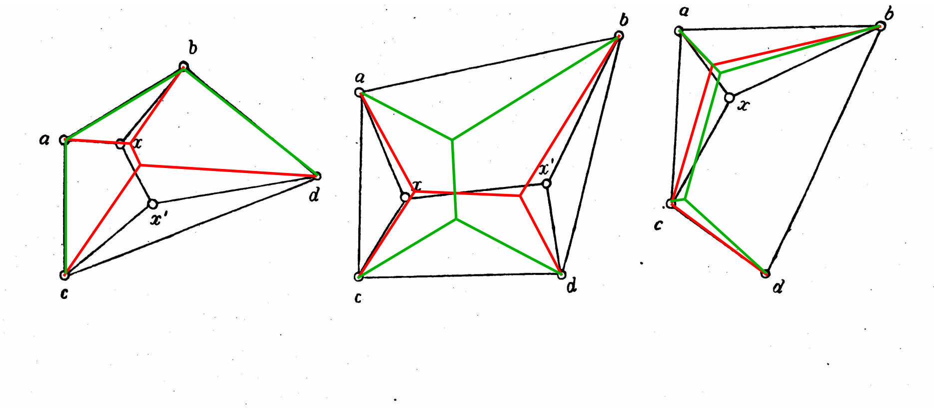

Figure 8: Correct solutions in green for the three examples presented by Gauss, and in red the optimized solution Gauss indicated by his drawing, which is however longer than the green one.

References

[1] H.C. Schumacher, Letter to Gauss, Altona, 19. März, 1836, in Briefwechsel zwischen Gauss und Schumacher III, Altona 1861, pp. 7–8.

[2] C.F. Gauss, Reply to Schumacher, Altona, 21. März, 1836, in Briefwechsel zwischen Gauss und Schumacher III, Altona 1861, pp. 13–14.

[3] P. de Fermat, Œvres, I, edited by H. Tannery, Paris 1891, Supplement: Paris 1922.

[4] E. Torricelli, Opere, I/2, edited by G. Loria and G. Vassura, Faënza, 1919, pp. 90–97.

[5] G.F. Fagnano, Problemata quaedam ad methodum maximorum et minimorum spectantia, Nova Acta Eruditorum, anni 1776, Lipsine 1779, p. 281.

[6] P. Tédenat, Questions résolues. Solution du premier des deux problèmes proposés à la page 196 de ce volume, Annales de mathématiques pures et appliquées, publiées par Gergonne 1, 1810-1811, pp. 285–291.

[7] S. Lhuilier, Questions résolues. Solution du premier des deux problèmes proposés à la page 196 de ce volume, Annales de mathématiques pures et appliquées, publiées par Gergonne, 1, 1810-1811, pp. 297–301.

[8] F.K. Hwang, D.S. Richards, and P. Winter, The Steiner Tree Problem, Annals of Discrete Mathematics, vol. 53, Elsevier, North-Holland, 1992.

[9] D. Cieslik, Steiner minimal trees, Kluwer Acad. Publ. 1999.

[10] D. Z. Du, and F. K. Hwang. A proof of the Gilbert-Pollak conjecture on the Steiner ratio. Algorithmica, 7:121-135, 1992.

[11] M. R. Garey, R. L. Graham, and D. S. Johnson. The complexity of computing Steiner minimal trees. SIAM Journal on Applied Mathematics, 32:835-859, 1977.

[12] G. Kolata, Solution to old puzzle: How short a shortcut? New York Times, 30 October 1990.