

On the history of the Euclidean Steiner tree problem

Author(s): Marcus Brazil, Ronald L. Graham, Doreen A. Thomas and Martin Zachariasen

Source: *Archive for History of Exact Sciences*, Vol. 68, No. 3 (May 2014), pp. 327-354

Published by: Springer

Stable URL: <https://www.jstor.org/stable/24569605>

Accessed: 19-05-2020 09:59 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Archive for History of Exact Sciences*

On the history of the Euclidean Steiner tree problem

Marcus Brazil · Ronald L. Graham ·
Doreen A. Thomas · Martin Zachariasen

Received: 1 July 2013 / Published online: 27 August 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The history of the Euclidean Steiner tree problem, which is the problem of constructing a shortest possible network interconnecting a set of given points in the Euclidean plane, goes back to Gergonne in the early nineteenth century. We present a detailed account of the mathematical contributions of some of the earliest papers on the Euclidean Steiner tree problem. Furthermore, we link these initial contributions with results from the recent literature on the problem.

1 Introduction

The Euclidean Steiner tree problem asks for a shortest possible network interconnecting n points in the plane. This is a classic example of a problem that is easy to state and understand, but difficult to solve. Since the 1960s, an increasingly

Communicated by: Jesper Lützen.

M. Brazil (✉)
Department of Electrical and Electronic Engineering, The University of Melbourne,
Melbourne, VIC 3010, Australia
e-mail: brazil@unimelb.edu.au

R. L. Graham
Department of Computer Science and Engineering, UC San Diego,
La Jolla, CA 92093-0404, USA

D. A. Thomas
Department of Mechanical Engineering, The University of Melbourne,
Melbourne, VIC 3010, Australia

M. Zachariasen
Department of Computer Science, University of Copenhagen,
2100 Copenhagen Ø, Denmark

sophisticated mathematical theory of minimal networks has developed around this problem building on a combination of techniques from combinatorics, geometry and analysis. The interest in the Steiner tree problem stems not only from the challenge it represents mathematically, but also from its range of potential applications in areas such as communications, infrastructure networks and physical chip design.

The history of the Euclidean Steiner tree problem, however, is generally not well understood—particularly the history from before 1941 when the problem was exposed to a large audience through the book *What is Mathematics?* (Courant and Robbins 1941). This is, perhaps, not surprising given the highly nonlinear way the study of Steiner trees has developed since the early 1800s. The mathematical history of this problem is full of twists and turns and dead ends; on at least three occasions, the problem has been completely forgotten, and then “rediscovered” many years later. Furthermore, the modern attribution of the problem to Jakob Steiner (1796–1863) is misleading, if not to say erroneous.

In a letter to Schumacher from 1836, Carl Friedrich Gauss (1777–1855) briefly discussed the problem that is now called the Euclidean Steiner tree problem. For some time, this letter was believed to be the earliest source of the problem (Schreiber 1986). Today we know that the Euclidean Steiner tree problem was posed and analysed even earlier, in 1811, by Joseph Diaz Gergonne (1771–1859); this fact is mentioned briefly in a mathematics history book by Scriba and Schreiber (2010).

What is still missing in the literature is a detailed account of the mathematical contributions of the early papers on the Euclidean Steiner tree problem and an account of the historical trajectory from these early papers to the modern literature in the area. In this paper, we give such an overview, including a study of the numerous rediscoveries of the problem and the origins of the current misleading nomenclature. Our paper includes a number of direct quotations and figures from the earliest literature in the area; all passages quoted in this paper have been translated in a manner that attempts to be faithful to the meaning and intentions of the original text while also being comprehensible to a modern reader.

The paper is organised as follows. In Sect. 2, we give some background on the Euclidean Steiner tree problem (including a formal definition and important theory and algorithms), and in Sect. 3, we introduce the related Fermat–Torricelli problem and its history. In Sect. 4, Gergonne’s contributions from 1811 are presented. In Sect. 5, Gauss’ letter to Schumacher from 1836 and related German contributions from the late nineteenth century are discussed. In Sect. 6, we survey the first modern treatment of the problem, from 1936, by Jarník and Kössler, and in Sect. 7, we discuss the recent history of the problem—in particular the influential book by Courant and Robbins (1941) and the seminal modern study of the problem by Gilbert and Pollak (1968).

2 Background on the Euclidean Steiner tree problem

The *Euclidean Steiner tree problem* asks for a network of minimum total length inter-connecting a given finite set N of n points in the Euclidean plane. We present two mathematical definitions; the first definition is often used in earlier papers, while the latter is a more contemporary definition:

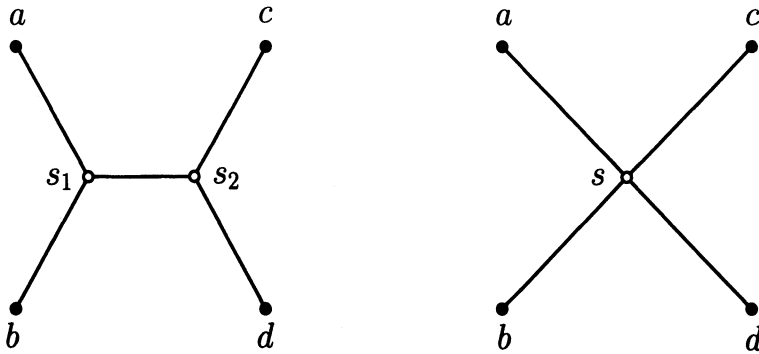


Fig. 1 A minimum Steiner tree (left) and a solution to the Fermat–Torricelli problem (right) for a set of 4 points $N = \{a, b, c, d\}$ on a unit square. The minimum Steiner tree has two Steiner points, s_1 and s_2 , and has length $1 + \sqrt{3} \approx 2.732$. The network resulting from the solution to the Fermat–Torricelli problem has length $2\sqrt{2} \approx 2.828$

1. Find a system of line segments such that the union of the line segments forms a connected set containing N , and such that the total Euclidean length of the line segments is minimised.
2. Find a geometric network $T = (V, E)$ (i.e. a connected graph embedded in the plane, with vertex set V and edge set E) such that $N \subseteq V$ and $S = V \setminus N$ is a (possibly empty) set of points known as *Steiner points*, and such that $\sum_{e \in E} |e|$ is minimised (where $|e|$ denotes the Euclidean length of edge $e \in E$).

A solution to the Steiner tree problem can be assumed to be a tree T (otherwise, it would not minimise length), and it can be assumed that any Steiner points in T have at least three incident edges. Such a tree is referred to as a *minimum Steiner tree*. The given points N are often denoted *terminals*.

The (generalised and unweighted) *Fermat–Torricelli problem* also has a finite set N of n points as input, but the problem is to compute a *single point* s , such that the sum of Euclidean distances from s to each point in N is minimised. For $n \leq 3$, a solution to the Fermat–Torricelli problem also gives a solution to the Steiner tree problem. However, for $n \geq 4$, a solution to the Fermat–Torricelli problem does *not* (generally) lead to a solution to the Steiner tree problem (Fig. 1).

A minimum Steiner tree has a number of properties that we briefly summarise here (see, e.g. Gilbert and Pollak 1968):

Degree and angle properties A minimum Steiner tree with $n = |N|$ terminals has at most $n - 2$ Steiner points. Edges meeting at common vertices form angles that are at least 120° ; this implies that Steiner points have exactly three incident edges meeting at 120° angles (Fig. 1, left).

Full components A minimum Steiner tree can be decomposed into components in which every terminal is a leaf, known as *full components* or *full minimum Steiner trees*. This decomposition is unique for a given minimum Steiner tree, but is not unique for a given terminal set.

Full topologies A description of the interconnection pattern of a tree is called the *topology* of the tree. A *full topology* is the description of the tree topology for a full component, where terminals are leaves and Steiner points of degree 3 are interior

nodes. Full topologies can be described using a parenthesis structure. For example, for the set of terminals $N = \{a, b, c, d\}$, the representation $(ab)(cd)$ means that terminals a and b are both connected to some Steiner point s_1 and similarly, that terminals c and d are both connected to some other Steiner point s_2 ; finally, s_1 and s_2 are connected in the corresponding topology. Terminal pairs ab and cd are called *cherries* in the topology (Fig. 1, left).

The Euclidean Steiner tree problem is known to be NP-hard (Garey et al. 1977), even when the terminals are restricted to lie on two parallel lines (Rubinstein et al. 1997). The Melzak construction (Melzak 1961) forms the building block of the most successful algorithm for the problem—the GeoSteiner algorithm (Warme 1998; Warme et al. 1999, 2001; Winter and Zachariasen 1997). Using the GeoSteiner algorithm, optimal solutions to the Euclidean Steiner tree problem for thousands of terminals can be computed in reasonable time. Approximate solutions can be computed efficiently in theory and practice; the so-called polynomial time approximation scheme of Arora (1998) provides solutions that are a factor of $1 + \epsilon$ away from optimum in polynomial time for any fixed $\epsilon > 0$.

3 The Fermat–Torricelli Problem [1638–1834]

The origins of the Euclidean Steiner problem can be traced back to the closely related Fermat–Torricelli problem, which can be thought of as the simplest non-trivial case of the Steiner problem on n terminals for the case where $n = 3$. The history of the Fermat–Torricelli problem has already been researched in great depth elsewhere, so here we simply give a brief overview of the early historical development of this problem and refer the reader to the comprehensive historical study of Kupitz and Martini (1997) for more details and further references.

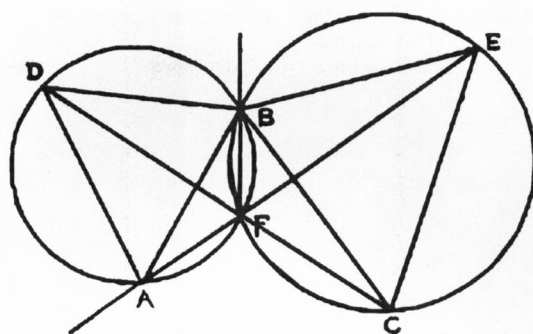
Pierre de Fermat (1661–1665) described the problem in about 1643 in his work: “Method for Determining Maxima and Minima and Tangents to Curved Lines” (de Fermat 1891, p. 135). In the original Latin, the statement of the problem reads: “datis tribus punctis, quantum reperire, a quo si ducantur tres rectae ad data puncta, summa trium harum rectarum sit minima quantitatis” or “given three given points, a fourth is to be found, from which if three straight lines are drawn to the given points, the sum of the three lengths is minimum”. The problem appears to have been inspired by a question posed by René Descartes in 1638 (see Descartes 1896, p. 324), who invited Fermat to investigate curves of the form:

$$\sum_{i=1}^4 \|p_i - x\| = c$$

for given points p_1, \dots, p_4 in the plane, and a constant c . These can be thought of as level curves for the function measuring the sum of distances of a point x from the given points p_i .

Fermat’s problem for three points p_1, p_2, p_3 can usefully be divided into two cases: the first where every angle of the triangle $\triangle p_1 p_2 p_3$ is strictly less than 120° ; the second

Fig. 2 Torricelli's construction of the Torricelli point F for three points A , B and C , from (Torricelli 1919). The construction shows that F is both the intersection of two circumcircles and the intersection of the two line segments AE and CD



[Fig. 38].

where the $\triangle p_1 p_2 p_3$ contains an angle of 120° or greater. The earliest known solution to Fermat's problem (at least for the first case) was a geometric construction due to the Italian physicist and mathematician Evangelista Torricelli (1608–1647) (Torricelli 1919), shown in Fig. 2. The method is to construct equilateral triangles on the sides of the triangle $\triangle p_1 p_2 p_3$ and outside this given triangle. Then, circumcircles are found for the three equilateral triangles and their intersection gives the solution to the Fermat–Torricelli problem, which we will refer to as the *Torricelli point*.

The construction by Torricelli in Fig. 2 also shows that any pair of the lines found by joining the furthest vertex of one of the equilateral triangles, in the construction, to the given point opposite the equilateral triangle, intersect in the Torricelli point. This alternative method of solving the Fermat–Torricelli problem was rediscovered more than a century later (in Simpson 1750) by Thomas Simpson (1710–1761), from whom we get the name “Simpson line” for each of these constructed line segments. However, it was not until 1834 that Heinen (1834) proved that in the first case of Fermat's problem, the three Simpson lines each have the same length as the sum of the distances to the Torricelli point.

In his book “*Exercitationes Geometricae Sex*” (Cavalieri 1647), Bonaventura Cavalieri, an Italian mathematician and Jesuite (1598–1647), characterised the three angles at the Torricelli point subtended by the sides of the given triangle $\triangle p_1 p_2 p_3$ as all being equal to 120° , and thus the isogonal point—a fact also known by Torricelli. Cavalieri also explicitly solved the second case of Fermat's problem, stating that in this case the minimising point is simply the vertex of $\triangle p_1 p_2 p_3$ with an obtuse angle.

In France, by the early 1800s, knowledge of the Fermat–Torricelli problem seems to have vanished, but was revived in 1810 when the problem and several generalisations were posed again by Joseph Gergonne. One of these generalisations was the Euclidean Steiner tree problem. Gergonne's contributions are the subject of the next section.

4 The first studies of the Steiner tree problem [1810–1819]

The earliest known statement and analysis of the Euclidean Steiner tree problem was by the French mathematician and logician Joseph Diaz Gergonne (1771–1859) in 1811 (Fig. 3). After a brief career in the French military, Gergonne had been appointed to

Fig. 3 Portrait of Joseph Diaz Gergonne (1771–1859), the originator of the Steiner tree problem



the chair of transcendental mathematics at the École Centrale in Nîmes in about 1795. He was particularly interested in geometry, but found it difficult to get his mathematics papers published. Gergonne attributed this to the lack of any independent specialist mathematics journals at the time. In 1810, Gergonne established his own mathematics journal entitled the *Annales de mathématiques pures et appliquées* but more generally known as the *Annales de Gergonne*. The journal, which was published monthly for 22 years, had a strong emphasis on projective, synthetic, and algebraic geometry. Much of the content was by Gergonne himself, but many other prominent mathematicians also published there, including Heinrich Christian Schumacher, Jakob Steiner, Gabriel Lamé and Évariste Galois.

In the “Prospectus” (preface) for the first issue of the journal, it was made clear that the journal would have a particular emphasis on problem solving. This editorial, almost certainly by Gergonne, states: “Each issue of the Annals will offer one or more theorems to prove, one or more problems to solve. The Editors, in the choice of these theorems and problems, give preference to statements that can be identified by their correspondents, and they will record proofs and solutions that are received in their compilation”.

Amongst the problems featured in the first volume of the *Annales de Gergonne* was the Fermat–Torricelli problem, and several generalisations of the Fermat–Torricelli problem, including a version of what is now known as the Steiner tree problem.

4.1 Problems from the Annales de Gergonne

We now examine the relevant problems posed in volume 1 of the *Annales de Gergonne*. The statement of the first problem appears on page 196:

PROPOSED QUESTIONS.

Problems of Geometry

I.

An engineer wishes to establish a communication between three cities, not located in a straight line, by means of a network composed of three branches, leading at one end to the three cities, and meeting at the other end at a single point between these three cities. The question is, how can one locate the point of intersection of the three branches of the network, so that their total length is as small as possible? (*)

This is the Fermat–Torricelli problem, discussed in Sect. 3, in an engineering guise. Its publication predates electrical telegraphy which was not in use until the 1830s, so the “communication” network may simply refer to roads, or possibly optical telegraphy, based on a semaphore system, which operated in France from 1792 to 1846.

The footnote to the problem is as follows:

(*) One can generalise this problem by asking how to determine, on a plane, a point whose sum of distances to a number of arbitrary points located in this plane is minimal. One can even extend to points located in any manner in space.

In other words, the footnote extends the Fermat–Torricelli problem to more than three given points (but only one intersection point) and higher dimensions. As mentioned in the previous section, this was solved in a later number of the journal by Tédénat (1810).

The next related problem appears on page 232:

PROPOSED QUESTIONS.

Problem of Geometry

Two straight-line canals intersect at a determined angle, and a city is situated, in a known manner, in one of the four regions formed by the intersection.

One wants to build two bridges over the canals, and build a communication network from these two bridges to the city for whose use they are intended.

The problem is to determine the locations in which both bridges must be built, and how one should position the branches of the network, so that the total length thereof is as small as possible? (*)

(*) One can generalise this problem, assuming the two canals to be of any form whatsoever.

This question asks for the minimum length network interconnecting a given point and two given straight lines. Although not made explicit, there is an assumption that the network can contain vertices other than those corresponding to the position of the city and the two bridges; otherwise, the problem would be equivalent to asking for the distance of a point from a straight line. The solution to the problem later in the volume includes a degree-three junction in the communication network, making it clear that the use of extra vertices is allowed.

A final pair of related problems appear on page 292:

PROPOSED QUESTIONS

Problems of Geometry

I.

Two cities are located, in a known manner, on the same side of a straight-line canal (*).

One wants to build a bridge over the canal and build a communication network from that bridge to the two cities for whose use it is intended.

The question is to determine at what location one should build this bridge and how one should position the branches of the network, so that the total length thereof is as small as possible?

II.

A number of cities are located at known locations on a plane; the problem is to link them together by a system of canals whose total length is as small as possible? (**)

(*) More generally, one can assume that the canal is curved.

(**) One should not confuse this question with the one on page 285. Indeed, in the former problem the number of branches of the network must equal the number of cities to which they lead on the one hand, and it is required, on the other hand, that these branches meet in the network at a single point; here, on the contrary, this condition is not required, and one should not enforce it if it does not result in a minimum.

Question I asks for the shortest network interconnecting two given points and a given straight line; question II is the *first known statement of the Steiner tree problem*, asking for the shortest length network interconnecting a set of known points in the plane. The footnote to the second question makes it clear that this is a different problem from the Fermat–Torricelli problem for multiple points—the network is not restricted to having only one extra vertex. The reference to page 285 is to a solution to the general Fermat–Torricelli problem in the plane, given by M. Tédenat, Rector of the Academy of Nîmes (Tédenat 1810).

4.2 Gergonne's solution of the Steiner tree problem

On page 375 of volume 1 of the *Annales de Gergonne* is a ten page article titled “Résolues” (Resolved) addressing all the problems listed in the previous subsection.

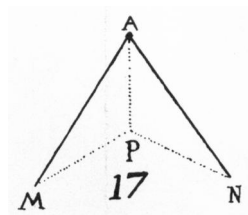
The article is subtitled:

PURELY GEOMETRIC SOLUTIONS to the problems of minimums proposed on pages 196, 232 and 292 of this volume, and various other analogous problems;
By a SUBSCRIBER.

The “subscriber” is definitely Gergonne. This is confirmed by the NUMDAM current director, Professor C. Gérini, a recognised expert on the *Annales de Gergonne*, who states that: “All papers signed ‘Par un abonné’ [i.e. “By a subscriber”] come from Gergonne, as well as the notes signed ‘JDG’”.¹ The figures for this article are signed “JDG”. Furthermore, in the digitised copy available online through the NUMDAM program, the name “GERGONNE” has been handwritten (apparently by Gergonne himself) next to “ABONNÉ” (subscriber). The identification of the “subscriber” as Gergonne is further confirmed in Guggenbuhl (1959) and Stigler (1976).

¹ Quoted in an unpublished paper by P. J. H. Piñeyro: “Gergonne; the isoperimetric problem and the Steiners symmetrization”, January 2012.

Fig. 4 Gergonne's Figure 17



The article addresses eleven problems connected with the Fermat–Torricelli problem. The first ten of these problems consider how to construct a minimum length network interconnecting three given geometric objects, each of which is either a point, a circle, or a straight line; the ten problems correspond to the ten possible combinations of these objects.

The eleventh problem is the Steiner tree problem, and we present a translation and discussion of Gergonne's solution here.

PROBLEM XI. Connect any number of given points by a system of lines whose total length is as small as possible.

The problem is stated here for the first time in abstract terms, rather than in terms of cities and canals. Earlier in the article Gergonne notes that: "...all points, lines and circles which will be discussed are always supposed to belong to the same plane"; hence, although planarity is not explicitly mentioned in the problem statement, it is nevertheless understood.

The solution is divided into an "Analysis" section and a "Construction" section. We will examine each in turn.

Analysis. I. One should not assume that each of the given points are reached by several of the sought lines; indeed assume that A is one of these points (Fig. 17) [Fig. 4] and two of the candidate lines, MA and NA , finish at that point—one could, in general (5), replace the system consisting of these two lines by a system of three lines PA , PM , PN , with a total length less, so that MA and NA does not fulfil the conditions of problem. In fact, it may well happen, in particular cases that PA should be zero, but it is during the construction of the problem that this circumstance should be indicated.

The analysis begins by observing that the minimum network is not necessarily a subnetwork of the complete network on the given points. The reference "(5)" refers to a set of inequalities derived in the solution to Problem I, which is that of finding the minimum length network interconnecting three points.

2. It may be noted, secondly, that if there are points of intersection of the lines sought other than the given points (and there cannot fail to be such, according to the above) the number of lines meeting there cannot be fewer than three. For if, in fact, a point M (Fig. 18) [Fig. 5, left] other than the given points, was reached by only two of the lines sought, MN and MP , then instead of linking points N and P by these two straight lines, one could link them by the single and shorter line NP ; hence the lines MN and MP would not meet the conditions of the problem.

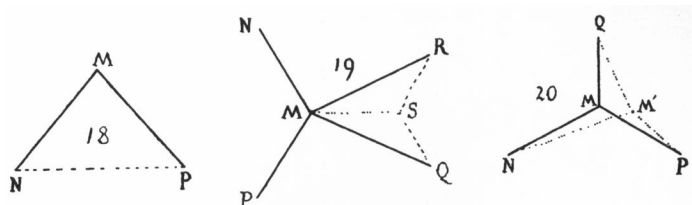


Fig. 5 Gergonne's Figures 18 to 20

Here, Gergonne observes that any additional vertices in a minimum network can be assumed to have degree at least three. The argument is correct apart from ignoring the trivial case where N, M and P are collinear.

3. One must not assume either that those sought lines that meet in a common point other than one of the given points, meet there in a number more than three; for, if one assumes just four of these lines MN, MP, MQ, MR (Fig. 19) [Fig. 5, middle] meet at the same point M , it would be possible (5), at least in general, to replace the system of two of these lines, MQ and MR for example, by the three lines SM, SQ, SR , with smaller total length, so that MQ and MR do not fulfil the conditions of the problem. In truth the relative positions of points M, Q, R , may well, as previously, in special cases, make SM zero, so that the point S coincides with the point M , but it is during the construction of the problem that this circumstance should be indicated.

Gergonne argues that the degree of an extra vertex in a minimum network can never be more than three. The argument, using proof by contradiction, is essentially correct, but is incomplete. The idea is to replace a pair of edges meeting at the extra point M by a shorter subnetwork including another extra vertex S (Fig. 5, middle). The problem, as acknowledged in the last sentence of the quoted passage, is that the edge SM may be degenerate, resulting in no change in the original network. Because no conditions are given as to when this degeneration occurs, there is no guarantee that it will not apply for every pair of edges incident with M . The missing condition for a suitable S to exist is that the pair of edges incident with M meet at an angle of less than 120° , which is clearly true for at least one pair of incident edges.

4. Finally, it is easy to see that the lines sought, meeting in threes at the one point, must form around this point equal angles between them, each $4/3$ of a right angle; for if M (Fig. 20) [Fig. 5, right] is the point of intersection of the three lines MN, MP, MQ , and if the angles between these lines, around this point, were not equal, then replacing the point M (5) by a point M' which satisfies this requirement means one replaces the three lines MN, MP, MQ by the three lines $M'N, M'P, M'Q$, with a smaller total length, so that the former do not fulfil the conditions of the problem.

This shows that A, B, C, D, \dots being the given points, the lines sought for can only form a kind of tree of the same nature as those shown in Figs. 21, 22, 23, 24, 25 [Fig. 6], so that the points M, N, P, Q, \dots where three lines meet, are, in general, two less in number than the number of given points, and that the

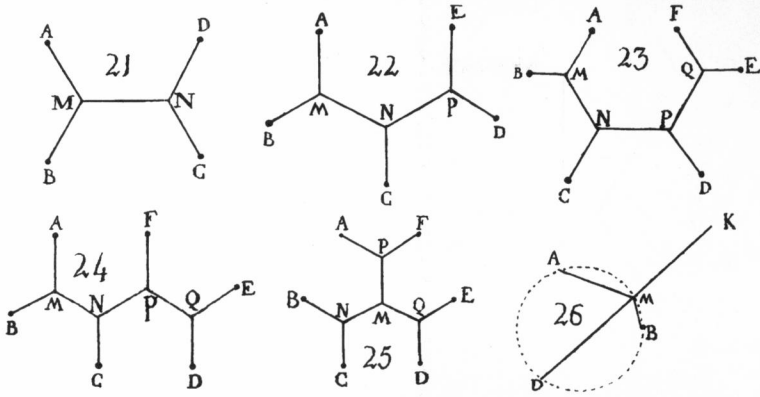


Fig. 6 Gergonne's Figures 21 to 26

angles enclosed by these lines at each point in turn are all equal to each other and are $4/3$ of a right angle. The only remaining question now is to learn how to construct [a solution to] the problem.

The first paragraph correctly observes that in a minimum network the angles between incident edges at any extra vertex are 120° . The argument here refers again to the proof of the Fermat–Torricelli problem on three points, which is Problem I in the article.

In the second paragraph, and in Figs. 21–25, Gergonne shows the different possible “forms” for full minimum Steiner trees with four, five or six terminals. Here, the term “forms” approximately corresponds to “unlabelled topologies” (or underlying graph structure), though not exactly, since Figs. 23 and 24 both have the same unlabelled topology; more precisely, we should think of the “forms” as being “unlabelled combinatorial embeddings” (in modern nomenclature). The problem of determining when a full topology does or does not exist is not addressed here; however, this is consistent with his statement at the beginning of the article that: “My purpose here is not to discuss the various circumstances which may affect the solution of problems which I intend to teach how to construct, and even make them impossible; I always assume in everything that follows, that the data is chosen so that these problems can be solved, and can supply all the solutions that their nature includes”.

The article next addresses the problem of constructing full minimum Steiner trees.

Construction. One already knows the solution for two given points, because then there is no line to construct other than that between the two points; one also knows how to solve the same for three given points (5), so if one succeeds in reducing its solution, in the case where the given points are n in number, to the correct one if these points are only $n - 1$ in number, one knows how to solve it in its generality; and this is what can be achieved very simply by proceeding as follows:

Arbitrarily take (Fig. 26) [Fig. 6, bottom-right] two points A, B , out of the given n points, such that joining them by an infinite straight line leaves the remaining $n - 2$ points on one side. The line AB can be thought of as a chord (lying on the other side to the remaining given points) of an arc AMB , inducing

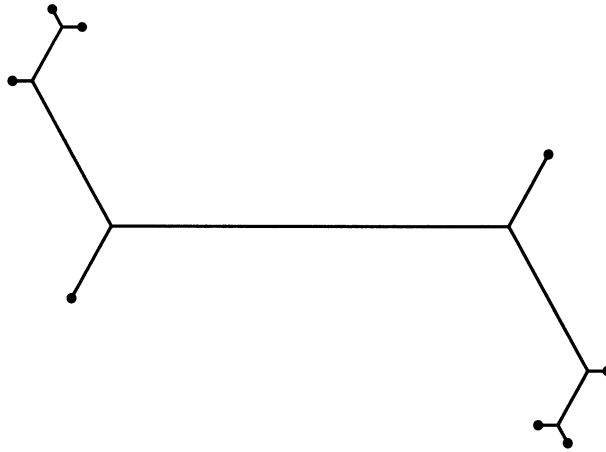


Fig. 7 An example of a full minimum Steiner tree where neither of the two cherries lies on the boundary of the convex hull. The minimality of this tree has been validated using GeoSteiner

an angle of 120° . Let D be the middle of the rest of the circumference; this D is substituted for the two points A, B ; we no longer have more than $n - 1$ points. Once we have resolved the problem on these $n - 1$ points, then the straight line KMD is sought for that ends at the point D ; by constructing MA, MB , and substituting these two chords for the MD part of KMD intercepted by the circle, the problem of finding the solution for the n given points is solved. It may be noted, moreover, that the lines of the system can only have three distinct directions; it follows that, three of them being determined, all others are determined by constructing lines parallel to these three.

Note that the construction here is an iterative construction, very much along the lines of the construction given by Melzak (1961) almost exactly 150 years later. The reduction step is to replace a pair of terminals belonging to a *cherry* (that is, a branch of the Steiner tree consisting of two terminals and a Steiner point adjacent to both) by an equilateral point, though the correctness of such a replacement is not proved here.

The condition on A and B at the beginning of the second paragraph is equivalent to saying that both points lie on an edge of the convex hull of the set of given points and that the equilateral point D lies on the opposite side of AB to the rest of the convex hull. The implicit assumption (made slightly clearer by the Remark I, below) appears to be that one can repeat the construction for all such pairs A, B at each stage ensuring one will eventually generate a full minimum Steiner tree on the given points (if it exists). Unfortunately, while this strategy seems reasonable for full trees on up to six given points, it is not in general correct. There does not necessarily exist a cherry both of whose terminals lie on the boundary of the convex hull of the terminals; this is illustrated by Fig. 7.

The error, however, can be easily remedied by attempting the reductive construction on every pair of available terminals A, B and for each pair trying both possibilities for the equilateral point D . This is guaranteed to find the full minimum tree when it exists, but at considerable expense to the computational efficiency of the procedure.

Gergonne concludes his study of this problem with two remarks.

Remark I. This construction allows for arbitrary choices of the points to be used successively, such that, if more than three points are given, the problem will have several solutions, and such that when these points are more than five in number, it can be solved by essentially different systems of lines. These systems are three in number for six given points (Figs. 23, 24, 25) and there are four systems for seven points, eight points provide thirteen, and so on. As for the number of solutions, it will be, in general, one for three points, two for four, five for five, fourteen for six, forty-two for seven, one hundred and thirty-two for eight, and so on.

Here, Gergonne makes it clear that there are many locally minimal solutions that need to be constructed in order to be certain of finding the global minimum. Gergonne distinguishes between the number of “systems” and the number of “solutions”. The former is the number of unlabelled *combinatorial embeddings* (also known as *planar embeddings*), which are equivalence classes of embedded graphs that take into account not only the graph topology, but also the cyclic order of edges around each vertex in the embedding.² The latter sequence is the number of non-self-intersecting full labelled topologies assuming that all terminals are on their convex hull; this gives the sequence of Catalan numbers, as was later shown by Cockayne (1967).

Remark II. If, for any number of given points A, B, C, D, \dots , one assumes that the lines that solve the problem are segments, each three of them meeting at nodes M, N, P, Q, \dots , and if, at points A, B, C, D, \dots , one applies equal forces directed along any segment that ends at these points, it is clear that these forces will form an equilibrium system.

In this remark, Gergonne gives a physical, mechanical interpretation of some of the geometric properties of a minimum Steiner network. This was a theme that strongly interested Gergonne, who elsewhere in his writing foresaw the future possibility of quasi-mechanical methods being used to uncover new mathematical results (Dahan-Dalmedico 1986).

4.3 The solution of Gallicus

The next known mention of the Steiner tree problem was by Gallicus in 1819, in *The Mathematical Repository*, a journal of mathematical problems and essays published in England, edited by Thomas Leybourn of the Royal Military College (now the Royal Military Academy, Sandhurst). In No. XV of that journal, Question 397 by Gallicus reads as follows: “Having given the position of any number of towns whatever, in the same plane, to connect them by a system of canals, of which, the total length shall be the least possible.” This requires only a brief mention as the solution “by the proposer” given in No XVII (pp. 134–136) is simply a slightly condensed version of Gergonne’s

² This appears to be the same sequence as the number of trivially fully gated graphs on n nodes (Colbourn and Huybrechts 2008); however, the equivalence between these two sequences is still to be established.

analysis, including figures that are identical to Gergonne's. Indeed, the pseudonym "Gallicus", from the Latin for "Gallic", suggests that the writer was French, or had a strong interest in French mathematics, and so would almost certainly have been familiar with the *Annales de Gergonne*. In a private communication, Douglas Rogers has suggested that Gallicus may have been the Scottish mathematician William Wallace, who was a colleague of Leybourne's at the Royal Military Academy. Wallace had a particular interest in geometry and actively followed French mathematics, making him a likely candidate.

As far as we know, this was the final discussion of this problem in the mathematical literature until it came to the attention of Gauss in 1836.³

5 The four-point problem and its generalisations [1836–1890]

In this section, we present the developments on the four-point Steiner tree problem in nineteenth-century Germany—beginning with Gauss' now famous letter to Schumacher from 1836, and concluding with two detailed treatments of the four-point problem, including generalisations to the n -point problem, by Karl Bopp and Eduard Hoffmann from 1879 and 1890, respectively.

5.1 Gauss' letter to Schumacher [1836]

The extensive epistolary correspondence between the great mathematician and scientist, Carl Friedrich Gauss (1777–1855), and the Danish–German astronomer, Heinrich Christian Schumacher (1780–1850), published in a series of volumes in the 1860s, has been a valuable source for historians and mathematicians. Gauss' connection to the Steiner tree problem was discussed by Schreiber (1986), and the story has been told in a number of other papers and books (Cheng et al. 2004; Cieslik 2004a,b; Du and Wu 2007; Gander et al. 2008; Scriba and Schreiber 2010).

In a letter from 19 March 1836, Schumacher presented an apparent paradox related to the Fermat–Torricelli problem on four points. Schumacher knew that the Fermat point x for the vertices a , b , c and d of a convex quadrilateral appears at the intersection of the diagonals of the quadrilateral. Now, if we move d towards c , then in the limit x also converges to c . However, in the example provided by Schumacher, the limit of x (which is c) is not the Fermat point for the three points a , b and c . In his answer from 21 March 1836, Gauss explains the apparent paradox—essentially that the solution to the four-point Fermat–Torricelli problem does not result in a minimum-length interconnection of the four points, as is the case for the solution to the Fermat–Torricelli problem for three points. Then, Gauss makes the following interesting remark:

³ It was thought by a number of sources, such as (Kupitz and Martini 1997, p. 101), that the Steiner tree problem had been rediscovered and discussed by Lamé and Clapeyron in 1827 or 1829 in the paper (Lamé and Clapeyron 1829). It is clear, however, from the 1989 paper of Frankzen and Grattan-Guinness (1989), which includes a complete translation of Lamé and Clapeyron (1829), that although Lamé and Clapeyron studied a number of generalisations of the Fermat–Torricelli problem, they did not work on any problem that was equivalent to the Steiner tree problem.

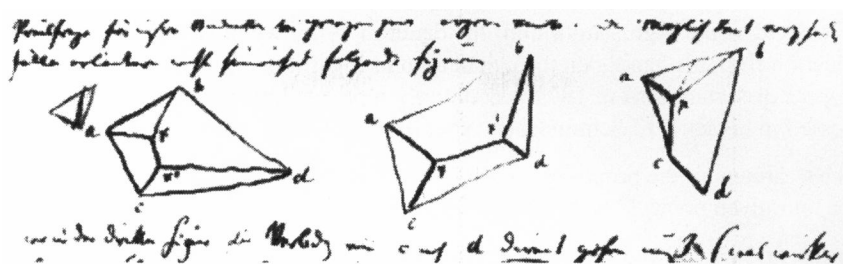


Fig. 8 Drawings in Gauss' letter to Schumacher from 21 March 1836. (Note that some of the reproductions of these drawings are inaccurate, see e.g. Gander et al. 2008)

If in a quadrilateral one asks for, in contrast to the question above, the shortest connection system in the plane, then several individual cases have to be distinguished, and you obtain quite an interesting mathematical problem, which is not foreign to me; as a matter of fact, I have on occasion considered the railroad connection between Harburg, Bremen, Hannover and Braunschweig, and I myself have thought that this problem would be an excellent prize problem for our students. The following drawings [Fig. 8] illustrate the different cases sufficiently; where in the third drawing the connection must go directly from c to d (which actually is the case in the above example). But time is pressing, hence no more of this today.

One may speculate as to why Gauss chose to illustrate the problem using a railroad connection example. In March 1836, only one railroad link existed in Germany (between Nürnberg and Fürth in Bavaria—opened in December 1835). In fact, the first railroad link to Gauss' home town Göttingen appeared almost 20 years later in 1854, and the opening of the link turned out to be Gauss' last public appearance (he died in February 1855). One explanation for Gauss' familiarity could be that his son, Joseph Gauss (1806–1873), had become involved in railroad construction in the 1830s as part of his service as military officer. In 1836, Joseph went to the USA to study the construction of railroads, and he later became director of the railway system in Hanover.

It is unlikely that Gauss was familiar with the first volumes of *Annales de Gergonne*, but Schumacher should already have been familiar with the Steiner tree problem as he had a paper published (Schumacher 1810) in the same volume of the *Annales de Gergonne* where Gergonne's statement of the problem and solution appear. Neither Gauss nor Schumacher returned to the (general) Steiner tree problem in their correspondence.

5.2 Karl Bopp's contributions on the four-point problem and its generalisation [1879]

In 1879, Karl Bopp (1856–1905) published a dissertation entitled "On the shortest connection system for four points" at the Georg-Augustus-Universität Göttingen (Bopp 1879). Karl Bopp was born in Frankfurt am Main in Germany, and after finishing his *dr. phil.* studies in Göttingen in 1879, he returned to Frankfurt am Main, where he became teacher at the local Goethegymnasium (high school). He appears to

have worked at the high school until his death in 1905; we only know of one other publication from his hand—on the calculation of logarithms from 1897.⁴

Bopp's dissertation from 1879 was directly motivated by the question that Gauss addressed in his letter to Schumacher. After quoting Gauss' question, Bopp states:

The treatment of the proposed problem "What is the shortest connection system for four given points?" is the focus of the following study.

In the first part of the dissertation, Bopp systematically enumerates the possible "complete connection systems" (or full topologies) for four given points a , b , c and d . This is done by considering the "curves" (or paths) between pairs of the given points. Bopp concludes that there exist three different full topologies for four points, each having two "help points" (or Steiner points) x_1 and x_2 that may overlap with the given points or with each other. Hence, he arrives at the following reformulation of the problem:

When asked for the shortest connection system—excluding superfluous parts of course—the question becomes: What is the shortest complete connection system? or: firstly, under what conditions is $ax_1 + x_1d + x_1x_2 + x_2b + x_2c$ a minimum, secondly, $ax_1 + x_1c + x_1x_2 + x_2b + x_2d$ thirdly, $ax_1 + x_1b + x_1x_2 + x_2c + x_2d$ a minimum, and finally, which of these three relative minima is the smallest and therefore the absolute minimum?

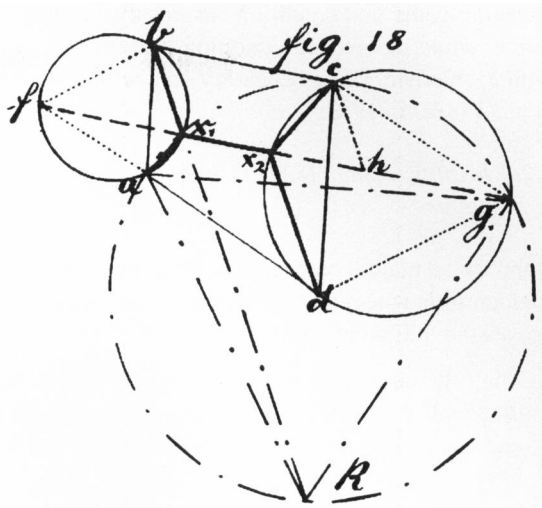
Bopp distinguishes between two cases of the problem: One where the four points a , b , c and d form the vertices of a convex quadrilateral, and one where they do not. Bopp considers the former, the convex quadrilateral case first; his contributions for this case are also the most interesting ones. The latter case involves detailed (careful) case analysis, but Bopp's analysis does not contain results of significant interest. In the following, we present Bopp's result for the convex quadrilateral case.

Assuming that the points appear in order a , b , c and d on the convex quadrilateral, Bopp argues carefully that the full topology $(ac)(bd)$ —where each pair a , c and b , d share a Steiner point as neighbour—need not be considered, since a relatively minimal tree for this topology will always be dominated by a relatively minimal tree having one of the two other full topologies. This leaves two full topologies instead of three—namely $(ab)(cd)$ and $(bc)(da)$.

5.2.1 Construction of a full Steiner tree for four points

Given a full topology, Bopp then addresses the problem of constructing a relatively minimal tree for the case where both Steiner points x_1 and x_2 exist—in modern terms, he considers the construction a *full Steiner tree* for each topology. He shows that if x_1 and x_2 can be located in the plane such that each of their three incident edges make

⁴ Personal record on Karl Bopp obtained from *Research Library for the History of Education at the German Institute for International Educational Research* (<http://bbf.dipf.de>).

Fig. 9 Bopp's Figure 18

angle 120° with each other, then the resulting tree is relatively minimal. The proof is mostly geometric and is an elegant application of Viviani's theorem (Viviani 1659), which states that the sum of distances from a point inside an equilateral triangle to each of the three sides of the triangle is a constant. (Interestingly, Bopp incorrectly attributes this theorem to Jakob Steiner!)

Bopp then shows that a full Steiner tree for topology $(ab)(cd)$ can be constructed via a construction (Fig. 9) similar to the approach developed by Gergonne (1810). Equilateral triangles are raised on edges ab and cd (and outside the convex hull), and a line is drawn between the third corners f and g of the two triangles. Steiner point x_1 (respectively x_2) is located at the intersection of the Simpson line fg with the circle through a, b and f (respectively c, d and g). Bopp proves that the length of the Simpson line is identical to the length of resulting tree; he even shows that alternative constructions exist, e.g. by constructing an equilateral triangle on the edge ag and connecting the third corner k of this triangle to b (Fig. 9).

The existence of a full Steiner tree for a given set of four points is discussed next. First Bopp notes that a necessary condition for the existence is that the four given points form a convex quadrilateral (a fact that was proved again almost a century later by Pollak 1978). After a lengthy and careful analysis, Bopp arrives at the following conclusion:

The necessary and sufficient conditions for the existence of a pair of points x_1 and x_2 , such that the edges ax_1 , bx_1 and x_1x_2 , and the edges x_1x_2 , cx_2 and dx_2 make angles $2\pi/3$ at x_1 and x_2 is, firstly, that the inequalities on page 20 are fulfilled—or equivalently that the line that connects the third corners of the outside equilateral triangles with sides ab and cd intersects ab and cd themselves, and not their extensions—and secondly, that the opposite angles made at ab respectively cd with the diagonals is not greater than $2\pi/3$.

Essentially, the first condition ensures that neither Steiner point coincides with a terminal, while the second condition prevents the two Steiner points from coinciding with each other. These necessary and sufficient conditions were rediscovered by Du et al. (1987a) in 1987.

5.2.2 Identification of the best topology based on the relative position of the four points

Bopp's final major contribution on the four-point problem is to give a simple condition for identifying the full topology that results in the shortest relatively minimal tree. Bopp proves the following theorem:

The following curious theorem holds: If $abcd$ is a convex quadrilateral whose diagonals ac and bd meet at angles $\angle azb$ and $\angle azd$, and we raise firstly at ab and cd , and secondly at bc and ad equilateral triangles outside $abcd$, then we have if

$$\angle dab < \pi, \angle bac < \pi, \angle bdc < \pi$$

that the distance between the third corners of the two former equilateral triangles is smaller than, equal to or greater than the distance between the third corners of the two latter equilateral triangles if and only if [respectively]

$$\angle azb < \angle dza, \angle azb = \angle dza, \angle azb > \angle dza$$

or equivalently [respectively]

$$\angle azb < \pi/2, \angle azb = \pi/2, \angle azb > \pi/2.$$

In short, we can use the meeting angle of the diagonals of the quadrilateral to decide which of the full Steiner trees is the shortest. The theorem does not guarantee the existence of the two full Steiner trees, but it gives the correct result when both trees exist—which has to be tested separately using, for example, the previous theorem. A similar result was proved by Pollak (1978); see also Du et al. (1987a).

5.2.3 Generalisation to the n point problem

In an appendix of the dissertation, Bopp considers the n point Steiner tree problem. In a simple and elegant way he proves that the number of complete connection systems (or full topologies) for n points is $a_n = 1 \cdot 3 \cdot 5 \dots (2n - 7) \cdot (2n - 5)$. His main observation is the following:

If we consider the connection system for the first $n - 1$ points, that is, if we remove the part of the connection system that connects the n th point to the rest of the connection system, then, on the one hand, a complete connection system for $n - 1$ points remains, and, on the other hand, a complete connection system for n points can be constructed by connecting the last point to some point on any line segment in the connection system for the first $n - 1$ points.

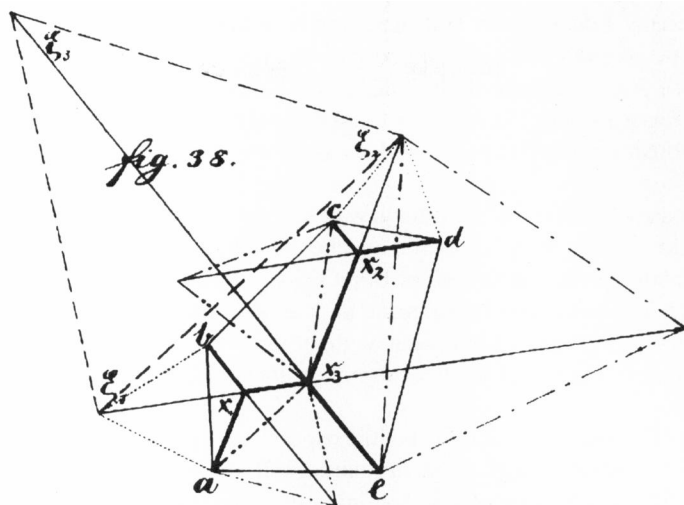


Fig. 10 Bopp's Figure 38

Bopp shows by induction that the number of edges in a complete connection system for n points is $s_n = 2n - 3$; then using the fact that the number of complete connection systems for n points is $a_n = a_{n-1}s_{n-1}$, the final result follows.

A finite-time algorithm is then sketched for the n point problem:

For each of these a_n complete connection systems (for which many only have a minima under degeneracies), a relative minima must be determined, and from this finite set of relative minima the shortest must be picked out. Determining and constructing the relative minima can be achieved without the slightest difficulty; the construction can be carried out with a ruler and a fixed angle of $\pi/3$.

Bopp presents a detailed example of constructing a full Steiner tree for five points (Fig. 10). Although the main idea is correct, Bopp does not really address the side problem—namely that equilateral triangles can be constructed on both sides of the defining base. Also, the treatment of degeneracies (construction of non-full Steiner trees) is incomplete. Bopp appears to be aware of this issue and notes that the number of different cases is so large “that a general investigation is hardly conceivable”.

As a final remark in his dissertation, Bopp presents some observations concerning the *three-dimensional Euclidean Steiner tree problem*. He argues that edges also meet at angles that are at least 120° and that a Steiner point and its three neighbours must reside in the same plane in 3-space. This appears to be the *first discussion of the three-dimensional Euclidean Steiner tree problem* in the mathematical literature.

5.3 Eduard Hoffmann's contributions on the four-point problem and its generalisation [1890]

In 1890, Eduard Hoffmann (1858–1923) wrote a paper with the title “On the shortest connection system for four points in the plane” (Hoffmann 1890). The paper was included in the program for the yearly celebration at the gymnasium (high school) in

Wetzlar, Germany. Eduard Julius Hoffmann was born in Homburg (near Saarbrücken), and finished his *dr. phil.* studies in Marburg (near Frankfurt am Main) in 1884 with a dissertation on conic sections. He became a high-school teacher, first in Wiesbaden, and later at Oberrealschule Hanau (near Frankfurt am Main). It is unclear why he was invited to publish his Wetzlar paper, and we know of no other publications from his hand.⁵

As in the case of Karl Bopp, Hoffmann was also directly motivated by the question that Gauss addressed in his letter to Schumacher. Hoffmann also quotes Gauss' question and notes that he became aware of Bopp's dissertation during the completion of his own paper. The selection of results and the methods employed do indeed indicate that Hoffmann wrote most of his paper without prior knowledge of Bopp's dissertation. Hoffmann's paper has neither the rigour nor the structural clarity of Bopp's dissertation.

Like Bopp, Hoffmann devotes most of the paper to the case where the four points a , b , c and d form a convex quadrilateral. In contrast to Bopp, he uses differential calculus to show that edges at Steiner points (assuming that they exist) meet at 120° angles. He then proves that if a full Steiner tree with two Steiner points exists, then this tree will be shorter than any solution having one or zero Steiner points. Hoffmann gives a detailed account on how to construct such a full Steiner tree using a construction similar to that developed by Gergonne (1810), and he also shows that the length of the Simpson line is identical to the length of the constructed tree. Finally, he gives necessary and sufficient conditions for the existence of a full Steiner tree.

Hoffmann presents a number of simple conditions on how to identify the full topology that results in the shortest relatively minimal tree. In addition to the condition discovered by Bopp, Hoffmann gives the following nice (but not very useful) condition which was rediscovered by Ollerenshaw (1978).

From this it follows that among the two [relative] shortest connection systems, the one with minimum length is the one for which the connection between the two additional points [Steiner points] has maximum length.

Hoffmann concludes his paper with a discussion on the n point problem. He gives some informal arguments to show that a shortest connection system is a tree with at most $n - 2$ Steiner points, and argues using differential calculus that edges at Steiner points have the same angle properties as in the four-point case. Then, he presents constructions (similar to those of Gergonne and Bopp) for two six-point problems: the first having a "chain-like" topology with two cherries, and the second having a "wide" topology with three cherries. Even though he provides no detailed analysis, Hoffmann is aware that the number of possible full topologies increases rapidly as the number of points increases. Interestingly, an underlying assumption—which was shared, at least to some extent, by both Gergonne and Bopp—appears to be that full Steiner trees can only exist for point sets that form vertices of their own convex hull; a counter example to this assumption is given in Fig. 7.

⁵ Personal record on Eduard Hoffmann obtained from *Research Library for the History of Education at the German Institute for International Educational Research* (<http://bbf.dipf.de>).

6 A modern treatment of the Steiner tree problem: Jarník and Kössler [1934]

After Hoffmann's paper, interest in the Steiner Tree problem seems to have vanished until the paper of Vojtěch Jarník and Miloš Kössler (Jarník and Kössler 1934). These Czech mathematicians produced an impressive and modern treatment of the problem, not only in the plane, but also for general higher-dimensional Euclidean spaces. Their motivations for studying the problem, however, are far from obvious, and they were clearly unaware of any previous work in this area. Unfortunately, their paper was ignored by the mathematical community for more than 50 years despite Jarník's strong mathematical reputation, probably because it was written in Czech and was outside Jarník's usual mathematical topics of number theory and analysis (see Korte and Nesetril (2001) for more background on this paper). The first significant citation was by Du et al. (1987b);⁶ here, the authors completed Jarník and Kössler's study of minimum Steiner trees for the vertices of a regular polygon. In the final section of their paper, Jarník and Kössler constructed minimum Steiner trees for the n -vertex regular polygons where $n = 2, 3, 4$ and 5 , and showed that a minimum Steiner tree for $n = 6$ or $n \geq 13$ coincides with a minimum spanning tree on the same set of vertices (i.e. contains no Steiner points). Du et al. (1987b) completed the remaining cases ($n = 7, 8, 9, 10, 11, 12$), showing that again in these cases, the minimum Steiner trees and minimum spanning trees coincide.

In 2001, Korte and Nesetril (2001) produced a translation into English and detailed discussion of the whole of Jarník and Kössler's paper, other than this final section. Because of this, we only give a brief overview of their results here. One of the most notable features of their paper is the rigour and precision underlying their treatment. Their formulation of the problem is as follows: "given n points C_1, C_2, \dots, C_n , find a connected set consisting of finitely many [line] segments, which contains the points C_1, C_2, \dots, C_n , so that 'the total length' of this set is the least possible". The setting in which the authors study this problem is general k -dimensional Euclidean space, and the main focus of the first half of the paper is the existence of such minimum length sets. The proof of this result is an elegant argument using elementary techniques—much of the modern machinery of graph theory and convex functions was not well developed at this time.⁷ The approach makes full use of the higher-dimensional setting; in fact, the proof for the planar case relies on embedding the problem into three-dimensional space and showing that the minimum solution there is also a minimum solution in two dimensions.

As a consequence of this existence proof, Jarník and Kössler derive some of the fundamental structural properties of minimum Steiner trees. Most importantly, they show that the angle at any vertex is at least $2\pi/3$ and that: "Every branching point [ie, Steiner point] of G has degree 3. The three sides of the graph incident to a branching point lie in a two-dimensional plane and any two have angle $2\pi/3$." This latter result generalises the similar result of Bopp for three-dimensional space. The authors use

⁶ There were at least four earlier papers citing (Jarník and Kössler 1934), mostly in the Czech mathematical literature; however, none of these papers deal directly with the Euclidean Steiner Tree problem.

⁷ In particular, Denes König's seminal textbook on graph theory, "Theorie der Endlichen und Unendlichen Graphen" was not published until 1936, two years after this paper.

these fundamental properties in the final section of the paper to begin a study of the solution of the Steiner problem for small numbers of terminals. After showing that the problem on three points has a unique solution, they make the following statement: “For $n > 3$ the situation is too complicated; we restrict ourselves, therefore, to a discussion of the case where the basic points [i.e. terminals] form the vertices of a regular n -gon.” There follows a detailed treatment of minimum Steiner trees for n -vertex regular polygons, as discussed earlier.

6.1 Gustave Choquet’s Road Network problem

It appears that the only other contribution to the Steiner problem in the twentieth century, before Courant and Robbins’ book (Courant and Robbins 1941), was a brief paper by the French mathematician Gustave Choquet (Choquet 1938) published in 1938 (when he was about 23 years old).⁸ The paper includes no proofs and is essentially an extended abstract. Here, Choquet discusses the problem of minimising the total length of a road network designed to interconnect a given set of n cities. Throughout most of the paper he assumes that there are no junctions between roads, except at the cities, in order to allow high volumes of traffic to travel unimpeded (this is the minimum spanning tree problem). However, in the final section, Choquet relaxes this condition (meaning that the network is now a minimum Steiner tree) and discusses some of the consequences, most of which are very similar to those derived by Jarník and Kössler. In particular, he notes that two roads meet at an angle of at least 120° , junctions not at cities have degree 3, and there are at most $n - 2$ such “bifurcation points”. He also observes, like Jarník and Kössler, that if the problem is generalised to higher-dimensional Euclidean spaces, then Steiner points still have degree 3 and the three incident edges are co-planar.

Choquet’s most interesting and original contribution is a suggestion that the length of a minimum Steiner tree can be used as a metric for measuring any closed point set, where the length of such a set is defined as: “the upper bound of the absolute length of the [minimum length] networks joining any finite number of points of the set.”⁹ This unusual way of thinking of Steiner trees reflects Choquet’s interest in measure theory, for which he would later become renowned.

7 Recent results and the development of current terminology [1941–1968]

The work that has had the most important influence on the modern literature on the Steiner tree problem, which has developed steadily since the 1960s, is the 1941 book by Richard Courant and Herbert Robbins *What is Mathematics?* (Courant and Robbins 1941). In this section, we discuss the treatment of the Steiner tree problem in this book,

⁸ In a private communication, when asked why he worked on the minimum spanning tree and Steiner tree problems, Choquet simply replied that when he was young he worked on all kinds of different problems (and that he had nothing more to contribute on these two problems).

⁹ This, in fact, corrects a mistake made by Menger (1931), who attempted to establish the same result using minimum spanning trees. It is unknown whether Choquet was aware of Menger’s work.

and then briefly trace the subsequent Steiner literature through to the seminal paper of Gilbert and Pollak (1968).

7.1 Courant and Robbins and the street network problem [1941]

What is Mathematics?: An Elementary Approach to Ideas and Methods by Courant and Robbins was a successful and highly influential account of a range of mathematical areas from topics in analysis and number theory to geometry and topology. A mark of its success is the fact that it is still available today, having been republished in a new edition in 1996. This is not a book written specifically for the professional mathematician, rather it is an expository, informal survey of mathematics, designed to be accessible to students and the interested layperson with a “certain degree of intellectual maturity”.

The authors’ discussion of the Steiner problem appears in Chapter VII: “Maxima and Minima”. Section 5 of that chapter is titled “Steiner’s Problem”, and begins with an account of the three-point Fermat–Torricelli problem. The problem is described in terms of minimising the total length of a system of roads connecting three villages. The origins of the problem are ascribed to Jakob Steiner in the early nineteenth century, although his contributions to the Fermat–Torricelli problem were marginal. Kupitz and Martini (1997) argue convincingly that Courant and Robbins had read about the Fermat–Torricelli problem in a short paper of Jakob Steiner’s appearing in pages 729–731 of Steiner (1882). This is, in fact, a posthumous publication of one of Steiner’s private manuscripts written by Steiner for himself only, and not surprisingly, it contains no references and some errors. Courant and Robbins’ familiarity with this paper is confirmed by the fact that they repeat one of these errors on page 358, a mistake also noted by Krarup and Vajda (1997). The authors appear to be unaware of the role of Fermat, Toricelli and other early geometers in the development of solutions to the problem.

In the final subsection of Sect. 5, the authors briefly discuss “Steiner’s problem” (i.e. the Fermat–Torricelli problem) on more than three points, which they state “does not lead to interesting results.”¹⁰ Instead they suggest that a more interesting generalisation is that of the “street network problem” which they define as follows: “Given n points $A_1, \dots, A_n, [\dots]$ find a connected system of straight-line segments of shortest total length such that any two of the given points can be joined by a polygon consisting of segments of the system”. The use of the word “polygon” is somewhat unusual here, but appears to mean a path composed of straight-line segments. Note that there is no suggestion by the authors that Jakob Steiner ever considered this generalisation, but ironically, it is this street network problem to which Steiner’s name has become attached.

In the last page of the section, the authors briefly mention (without proof or explanation) some of the basic properties of these minimum networks, including that there are at most $n - 2$ “multiple intersections” (i.e. Steiner points) at each of which three segments meet at angles of 120° . They also demonstrate that solutions to the

¹⁰ Kupitz and Martini (1997) argue that this dismissal of the generalised Fermat–Torricelli problem is based on a misreading of Steiner’s paper in pages 729–731 of Steiner (1882).

problem are not necessarily unique, giving the networks for the vertices of a square as an example. The Steiner problem is also mentioned twice later in the book. The first of these is in Section 9 of Chapter VII, where it is shown that some Steiner trees can be thought of as limiting cases of an isoperimetric problem. The other mention is in Section 11 of the same chapter, where it is shown that locally minimal solutions to the Steiner problem can be found using soapfilms, via a mechanism consisting of two parallel glass sheets joined together by perpendicular bars (representing the terminals). The use of soapfilms to study minimal surfaces was a particular interest of Richard Courant's (see, for example, Courant 1940) and may have helped spark his interest in the Steiner problem.

7.2 The development of the Steiner tree problem through to the 1960s

We conclude by briefly outlining the development of the theory and terminology for the Steiner tree problem in the literature since Courant and Robbins' book, through to the publication in January 1968 of the seminal paper by Gilbert and Pollak (1968).

The first paper on Steiner trees to be published after Courant and Robbins' book was by Miehle (1958), some 17 years later. Miehle proposes a mechanical method for solving the Steiner problem using strings which can be tightened and moveable pegs (for the Steiner points). He also sets up a numeric analytic model based on this mechanical system and compares the merits of the mechanical system, the mathematical model and Courant's soapfilm method, mentioned above. Miehle gives no particular name to the minimisation problem or the minimum networks, but, taking his lead from Courant and Robbins, he credits the early work on the three-point problem to Jakob Steiner.

The following year Beardwood, Halton and Hammersley (Beardwood et al. 1959) published their well-known paper showing that the length of the shortest closed path through n points in a bounded plane region of area v is "almost always" asymptotically proportional to \sqrt{nv} as $n \rightarrow \infty$. The authors show that this asymptotic bound also applies to the Steiner tree problem as the number of terminals in a bounded region increases. They call the Steiner tree problem "Steiner's street network problem", again referencing Courant and Robbins; this seems to be the first time Steiner's name is formally used as part of the name of the general network problem. In 1961, there appeared a short follow-up paper by Hammersley (1961) in which the problem was again called "Steiner's street network problem"; however, the title of that paper is simply "On Steiner's Network Problem". After this paper, the word "street" never again appears in the name of this problem.

In the same year, Zdzislaw Melzak published a paper titled "On the Problem of Steiner" (Melzak 1961). The paper outlines a finite algorithm for generating a minimum Steiner tree for any given set of points. The idea underlying the algorithm is similar to the method devised by Gergonne, but shows a better appreciation of the "side problem" (as discussed in the second-last paragraph of Sect. 5.2). Melzak consistently uses the terminology "Steiner problem" throughout the paper, but refers to the minimum trees as "S-trees" and Steiner points as "S-points". Melzak's exposure to this problem appears to have come from his links to the Bell Telephone Laboratories; he acknowledges in the paper the encouragement of Robert C. Prim (of Bell

Laboratories), and he apparently also knew Edgar N. Gilbert.¹¹ Six years later, Melzak's student Cockayne wrote a paper (Cockayne 1967) improving Melzak's algorithm and discussing the Steiner problem in other metrics. Cockayne's paper uses the same terminology as Melzak's.

A little earlier than this, in 1966, Maurice Hanan published his highly influential paper on the Steiner problem in the rectilinear metric (Hanan 1966). This paper uncovers most of the basic properties of minimum Steiner trees where distance is measured using the rectilinear norm (also known as the Manhattan or ℓ_1 norm), and correctly predicts the importance of this theory to the emerging area of "printed circuit technology" (i.e. physical design of microchips). Hanan cites Melzak, and like him uses the phrase "Steiner problem" throughout the paper.

Another significant and widely cited Steiner paper that appeared a year later was that by Cavalli-Sforza and Edwards (1967), giving an application of the Steiner tree problem to phylogenetic analysis (i.e. the construction of evolutionary trees). This paper not only uses the phrase "Steiner problem" but is also the first to refer to the resulting minimal trees as "Steiner Trees". Another important paper in the same year was the study by Edgar Gilbert (of Bell Laboratories) of weighted Steiner trees, with applications to communications networks (Gilbert 1967).

Soon afterwards Edgar Gilbert and Henry Pollak published their comprehensive examination of the properties of minimum Steiner trees, entitled "Steiner Minimal Trees" (Gilbert and Pollak 1968). This paper was remarkable for setting up, with great clarity, almost all of the geometric framework that was to play a key role in the following decades in gaining a detailed understanding of the properties of minimum Steiner trees. This framework led the way to the construction of fast algorithms for solving the Steiner problem, most notably the GeoSteiner algorithm (Warme et al. 2001). The paper also sets in place most of the modern terminology for this problem, including "Steiner minimal trees", "Steiner minimal tree problem" and "Steiner points". The authors were aware, however, that Steiner's contribution was only to the three-point problem, and that that problem itself was originally proposed by Fermat, but they presumably felt that Steiner's name was too closely associated with the problem in the recent literature to suggest any other alternatives. Their paper helped kindle a fascination with this problem in the mathematical community that has lasted to the present day.

8 Conclusion

One of the intriguing questions to emerge from the history of the Steiner problem is why the literature on this problem was so discontinuous in nature until the 1960s. There is no easy answer to this question, but it is worthy of some speculation. It may simply be that the popularising of the problem in *What is Mathematics?* eventually led to a sufficient critical mass of researchers actively working on the problem to keep it continuously alive, but this explanation is unsatisfactory, given that there were no published papers on the Steiner tree problem in the 16 years following the book's publication. Part of the answer appears to be that the period from the 1960s onwards

¹¹ This is according to one of Melzak's students, Raymond Booth, in a private communication.

provided, for the first time, a proper context for the Steiner Problem. The properties of minimum Steiner trees, rather than simply being geometric curiosities, could now be viewed as providing the ingredients for relatively efficient algorithms for generating such trees; this tied in with the formalisation of the concept of an algorithm in computer science in the 1950s and 1960s, and the ability to effectively implement algorithms as computer technology advanced. (Prior to this, the large number of topologies made the problem computationally intractable for more than four or five terminals). The Steiner tree problem could also be seen as belonging, to some extent, to the burgeoning field of operations research, being closely related to the area of facility location.

Real-world applications appear to have also had a role in maintaining an active interest in this problem in the research community since the 1960s. Although some of the early statements of the problem in the 1800s were couched in terms of applications such as the design of canals or railway tracks, there is no evidence that the problem was ever studied as anything other than a problem in pure geometry. In the late 1950s, however, a genuine application emerged for the Euclidean Steiner tree problem. As Henry Pollak relates in Pollak (1978), the Bell Telephone Company faced the issue that its tariff for private line service was written in terms of the length of the minimum network connecting the customers' stations (that is, the customer would not be charged for redundancy or inefficient interconnections in the network). In theory, a customer could insist that the Bell Service adds a new terminal at a location that would reduce the cost of the minimum network (even if it was of no direct use to the customer), or at least that the customer be charged as though such a terminal existed. Hence, there was a real motivation for understanding how to compute the minimum network cost where such extra terminals (Steiner points) were allowed, which engaged the attention of a number of researchers working for or associated with Bell Laboratories.

In the 1960s, there was also an emerging appreciation of the potential applications of the rectilinear version of the Steiner tree problem in the efficient physical design of microchips, following the paper of Hanan (1966); today, this is still an important application of the Steiner tree problem. By the end of the 1960s, the Steiner tree problem had established itself as a rewarding field of study, both from a pure and applied point of view. From that time onwards, there has been a continual interest in a range of different forms of the Steiner tree problem in the mathematical, computer science and engineering literature.

Acknowledgments Two of the authors, Marcus Brazil and Doreen Thomas, were partially supported in the writing of this paper by a grant from the Australian Research Council. We would also like to thank: Francois Lauze (University of Copenhagen) and Morgan Tort (The University of Melbourne) for their generous assistance with the French translations required for this paper; Henry Pollak for useful commentary on Gauss' letters; Pavol Hell (Simon Fraser University) for assistance with translating (Jarník and Kössler 1934); Jakob Krarup (University of Copenhagen) for help with providing some original sources; Donald Knuth for helpful comments and suggestions on an earlier draft of this paper; and Konrad Swanepoel for alerting us to the existence of the paper of Menger (1931).

References

- Arora, S. 1998. Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *Journal of the ACM* 45(5): 753–782.

- Beardwood, J., J.H. Halton, and J.M. Hammersley. 1959. The shortest path through many points. *Proceedings of Cambridge Philosophical Society* 55: 299–327.
- Bopp, K. 1879. Über das kürzeste Verbindungssystem zwischen vier Punkten. PhD thesis, Universität Göttingen.
- Cavalieri, B. 1647. *Exercitationes Geometricae Sex*.
- Cavalli-Sforza, L.L., and A.W.F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. *Evolution* 21: 550–570.
- Cheng, X., Y. Li, D.Z. Du, and H.Q. Ngo. 2004. Steiner trees in industry. In *Handbook of combinatorial optimization*, vol. 5, ed. D.Z. Du, and P.M. Pardalos, 193–216. Dordrecht: Kluwer.
- Choquet, G. 1938. Étude de certains réseaux de routes. *Comptes Rendus Acad Sci* 206: 310–313.
- Cieslik, D. 2004a. The essential of Steiner's problem in normed planes. Technical report 8, Ernst-Moritz-Arndt-Universität Greifswald, Preprint-Reihe Mathematik.
- Cieslik, D. 2004b. *Shortest connectivity—Introduction with applications in phylogeny, combinatorial optimization*, vol. 17. New York: Springer.
- Cockayne, E.J. 1967. On the Steiner problem. *Canadian Mathematical Bulletin* 10: 431–450.
- Colbourn, C.J., and C. Huybrechts. 2008. Fully gated graphs: Recognition and convex operations. *Discrete Mathematics* 308: 5184–5195.
- Courant, R. 1940. Soap film experiments with minimal surfaces. *The American Mathematical Monthly* 47: 167–174.
- Courant, R., and H. Robbins. 1941. *What is mathematics?*. London: Oxford University Press.
- Dahan-Dalmedico, A. 1986. Un texte de philosophie mathématique de Gergonne. *Revue d'histoire des sciences* 39: 97–126.
- de Fermat, P. 1891. *Oeuvres*, vol. 1.
- Descartes, R. 1896. *Oeuvres de Descartes*, vol. II. Paris: J. Vrin.
- Du, D.Z., and W. Wu. 2007. Approximations for Steiner minimum trees. In *Handbook of approximation algorithms and metaheuristics*. Chapman and Hall/CRC.
- Du, D.Z., F.K. Hwang, G.D. Song, and G.Y. Ting. 1987a. Steiner minimal trees on sets of four points. *Discrete and Computational Geometry* 2: 401–414.
- Du, D.Z., F.K. Hwang, and J.F. Weng. 1987b. Steiner minimal trees for regular polygons. *Discrete Computational Geometry* 2: 65–84.
- Franssen, O.I., and I. Grattan-Guinness. 1989. The earliest contribution to location theory? Spatio-economic equilibrium with Lamé and Clapeyron, 1829. *Mathematics and Computers in Simulation* 31: 195–220.
- Gander, M.J., K. Santugini, and A. Steiner. 2008. Shortest road network connecting cities. *Bollettino dei docenti di matematica* 56: 9–19.
- Garey, M.R., R.L. Graham, and D.S. Johnson. 1977. The complexity of computing Steiner minimal trees. *SIAM Journal on Applied Mathematics* 32(4): 835–859.
- Gergonne, J.D. 1810. Solutions purement géométriques des problèmes de minimis proposés aux pages 196, 232 et 292 de ce volume, et de divers autres problèmes analogues. *Annales de Mathématiques pures et appliquées* 1: 375–384.
- Gilbert, E.N. 1967. Minimum cost communication networks. *Bell System Technical Journal* 46: 2209–2227.
- Gilbert, E.N., and H.O. Pollak. 1968. Steiner minimal trees. *SIAM Journal on Applied Mathematics* 16(1): 1–29.
- Guggenbuhl, L. 1959. Gergonne, founder of the Annales de Mathématiques. *The Mathematics Teacher* 52(8): 621–629.
- Hammersley, J.M. 1961. On Steiner's network problem. *Mathematika* 8: 131–132.
- Hanan, M. 1966. On Steiner's problem with rectilinear distance. *SIAM Journal on Applied Mathematics* 14(2): 255–265.
- Heinen, F. 1834. *Über Systeme von Kräften*. G. D. Bädeker.
- Hoffmann, E. 1890. Über das kürzeste Verbindungssystem zwischen vier Punkten der Ebene. In *Program des Königlichen Gymnasiums zu Wetzlar für das Schuljahr von Ostern 1889 bis Ostern 1890*. Schnitzler.
- Jarník, V., and M. Kössler. 1934. O minimálních grafeth obehajících n daných bodů. *Cas Pest Mat a Fys* 63: 223–235.
- Korte, B., and J. Nešetřil. 2001. Vojtech Jarník's work in combinatorial optimization. *Discrete Mathematics* 235: 1–17.
- Krarup, J., and S. Vajda. 1997. On Torricelli's geometrical solution to a problem of Fermat. *IMA Journal of Management Mathematics* 8(3): 215–224. doi:10.1093/imaman/8.3.215.

- Kupitz, Y.S., and H. Martini. 1997. Geometric aspects of the generalized Fermat–Torricelli problem. In *Bolyai Society Mathematical Studies*, vol. 6, ed. I. Barany, and K. Boroczky, 55–127. Budapest: Intuitive Geometry, Janos Bolyai Mathematical Society.
- Lamé, G., and B. Clapeyron. 1829. Mémoire sur l'application de la statique à la solution des problèmes relatifs à la théorie des moindres distances. *Journal des Voies et Communications* 10: 26–49.
- Melzak, Z.A. 1961. On the problem of Steiner. *Canad Math Bull* 4(2): 143–148.
- Menger, K. 1931. Some applications of point-set methods. *Annals of Mathematics, Second Series* 32: 739–760.
- Miehle, W. 1958. Link-length minimization in networks. *Operations Research* 6(2): 232–243.
- Ollerenshaw, D.K. 1978. Minimum networks linking four points in a plane. *Bulletin of the Institute of Mathematics and its Applications* 15: 208–211.
- Pollak, H. 1978. Some remarks on the Steiner problem. *Journal of Combinatorial Theory, Series A* 24(3): 278–295.
- Rubinstein, J.H., D.A. Thomas, and N.C. Wormald. 1997. Steiner trees for terminals constrained to curves. *SIAM Journal on Discrete Mathematics* 10(1): 1–17.
- Schreiber, P. 1986. *Zur Geschichte des sogenannten Steiner-Weber-Problems*. Technical report, Wissenschaftliche Zeitschrift der Ernst-Moritz-Arndt-Universität Greifswald, Mathematisch-Naturwissenschaftliche Reihe. 58.
- Schumacher, H.C. 1810. Géométrie analytique. Solution analytique d'un problème de géométrie. *Annales de Mathématiques pures et appliquées* 1: 193–195.
- Scriba, C.J., and P. Schreiber. 2010. *5000 Jahre Geometrie*. Berlin: Springer.
- Simpson, T. 1750. *The Doctrine and applications of fluxions*. London: John Nourse Publisher.
- Steiner, J. 1882. *Gesammelte Werke*, vol. II. Berlin: Reimer.
- Stigler, S.M. 1976. The anonymous Professor Gergonne. *Historia Mathematica* 3(1): 71–74.
- Tédenat, M. 1810. Questions résolues. Solution du premier des deux problèmes proposés à la page 196 de ce volume. *Annales de Mathématiques pures et appliquées* 1: 285–291.
- Torricelli, E. 1919. De maximis et minimis. In *Opere di Evangelista Torricelli*, ed. G. Loria, and G. Vassura. Faenza, Italy.
- Viviani, V. 1659. *De maximis et minimis geometrica divination in quintum Conicorum Apollonii Pergaei*, Vol. II. Florence, Italy.
- Warne, D.M. 1998. *Spanning Trees in Hypergraphs with Applications to Steiner Trees*. PhD thesis, Computer Science Department, The University of Virginia.
- Warne, D.M., P. Winter, and M. Zachariasen. 1999. Exact solutions to large-scale plane Steiner tree problems. In *Proceedings of the tenth annual ACM-SIAM symposium on discrete algorithms*, 979–980.
- Warne, D.M., P. Winter, and M. Zachariasen. 2001. *GeoSteiner 3.1*. Department of Computer Science, University of Copenhagen (DIKU), <http://www.diku.dk/geosteiner/>.
- Winter, P., and M. Zachariasen. 1997. Euclidean Steiner minimum trees: an improved exact algorithm. *Networks* 30: 149–166.