# 2

# GAUSS' QUESTION

At first glance it seems that our two problems in the previous chapter have not many facets in common. Fermat's problem is a typical one in the class of geometric, and the problem of a minimum spanning tree in the class of combinatorial optimization problems. And moreover, we used very different methods to find solutions. Now, consider:

## 2.1 GAUSS' QUESTION AND THEIR CONVERSION TO STEINER'S PROBLEM

On March 19., 1836 the astronomer Schuhmacher wrote a letter to his friend the mathematician Gauß, in which he expressed surprise about a specific case of the Fermat problem: He considered four points $v_1, v_2, v_3, v_4$ in the plane which forms a quadrilateral such that the segments $\underline{v_1v_2}$ and $\underline{v_3v_4}$ are parallel and the lines of the segments $\underline{v_1v_3}$ and $\underline{v_2v_4}$ meet in one point $v$ outside. Now the Torricelli point $q$ of these four points is the intersection point of the diagonals $\underline{v_1v_4}$ and $\underline{v_2v_3}$. Schuhmacher did not understand the fact that, if the segment $\underline{v_3v_4}$ runs to the point $v$ then the point $q$ runs in the same way to $v$, but this cannot be, since the Torricelli point of three points is not necessarily one of the given points.[1]
Gauß [181] answered on March 21.[2] that Schuhmacher did not consider the

[1] See our considerations in 1.1.2.

[2] It is remarkable that in age a letter inside Germany was delivered in one day.

Fermat problem; instead, he looked for a solution of

$$||v_1 - w|| + ||v_2 - w|| + 2 \cdot ||v_3 - w|| = \min! \tag{2.1}$$

More important was the next remark of Gauß. He said that it is natural to consider the following, more general problem :

> Ist bei einem 4Eck ... von dem kürzesten Verbindungssystem die Rede ..., bildet sich so eine recht interessante mathematische Aufgabe, die mir nicht fremd ist, vielmehr habe ich bei Gelegenheit eine Eisenbahnverbindung zwischen Harburg, Bremen, Hannover, Braunschweig...in Erwägung genommen ....

In English: "How can a railway network of minimal length which connects the four German cities Bremen, Harburg (today part of the city of Hamburg), Hannover, and Braunschweig be created?"[3]

In such a generalized Fermat problem we are given a certain number of points in the plane which are to be connected by a system of curves of smallest total length. The concrete problem by Gauß was completely discussed by Bopp [50] in 1879.[4] Today it is easy to see that a solution is given by a network in which Bremen, Harburg and Hannover are interconnected by their Torricelli point and Hannover and Braunschweig are connected by a straight line [5].

Perhaps starting with the book *What is Mathematics* [116] by Courant and Robbins in 1941, Gauß' Problem became popularized under the name of

**Steiner's Problem**
**Given:** A finite set of points in a plane (or in another metric space).
**Find:** A network which connects all points of the set with minimal length.

Steiner's Problem for three points, however, is also a special case of Fermat's Problem. If four or more points are given, then Steiner's Problem is independent from Fermat's, and asserts its own character.

[3]A picture of this letter can be found on the cover of the book *Approximation Algorithms* [435].

[4]Martini in [48] names an older source for this generalized problem, namely Lame and Clapeyron in 1827, but he doesn't give an exact reference. Moreover, Scriba and Schreiber [391] give a discussion of the origins of the problem.

[5]In our figure this is marked with bold line style; the thin line style depicts an MST.

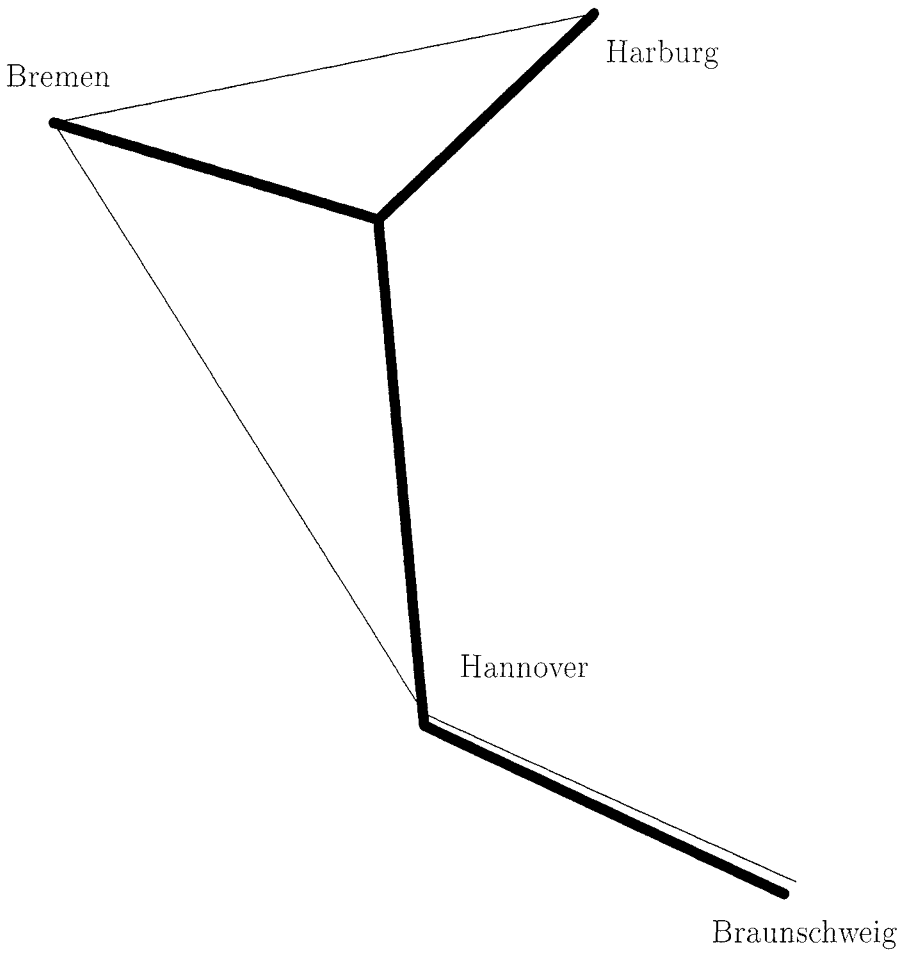


**Figure 2.1** Gauss' problem

Now, we mainly refer to the misleading title "Steiner's Problem", which was attributed by Courant and Robbins. They referred in their book to neither Fermat for the $n = 3$ problem nor Gauß for the general problem. An extensive discussion of reasons for this incorrect naming is contained in [271], [386] and [391].[6]

[6]It seems that Courant and Robbins knew of a report by Steiner on "Fermat's Problem"(!) to the Prussian Academy of Sciences in 1837.

Notice that the authors generated both the mistake in priority, and the great interest of many scientists in this problem. The popularity of their book has been the main reason that the misnomer "The Steiner Problem" or "Steiner's Problem" has stuck, and that the interest in this problem has spread.

Geometric optimization problems are inherently not pure combinatorial problems since the optimal solution often belongs to an infinite feasible set, the entire real Euclidean space. Steiner's Problem is typical for a large number of so-called "Geometric Network Design Problems" which act in a geometric structure. But often it is necessary to combine geometric and combinatorial methods to find a solution, and this is the approach we taken here.

Any network solving Steiner's Problem must be a tree, which is called a Steiner Minimal Tree (SMT). It may contain vertices different from the points which are to be connected. Such points are called Steiner points.
Given a set of points, it is a priori unclear how many Steiner points one has to add in order to construct an SMT. Furthermore, we are interested in the location of the Steiner points. A Steiner point is the Torricelli point of its neighbors. That means: Fermat's Problem is the local version of Steiner's Problem. On the other hand, if the number and location of the Steiner points are known then an SMT is an MST for the union of the given and the Steiner points.

Until 1961 it was not even known that Steiner's Problem is finitely solvable. There are infinitely many points in the plane, and even though most of them are probably irrelevant, it is not obvious that any algorithm exist. Then Melzak [305] established many basic properties of an SMT: Without loss of generality, the following is true for any SMT $T$ for a finite set $N$ of points in the Euclidean plane:

(i) The degree of each vertex is at most three.

(ii) The degree of each Steiner point equals three.

(iii) Any Steiner point is the Torricelli point of its neighbors; and two edges incident to a Steiner point meet at an angle of $120^o$. Consequently, a Steiner point is uniquely located in relation to its neighbors.

(iv) There are at most $|N| - 2$ Steiner points; equality holds if and only if the vertices from $N$ are the leaves of $T$ and the Steiner points are of degree three.

(v) An SMT has at most $2|N| - 3$ edges; equality holds if and only if the vertices from $N$ are the leaves of $T$ and the Steiner points are of degree three.

(vi) When there is a Steiner point in the tree $T$, then least one of these points has two given points as neighbours.

(vii) The SMT is an MST for the set $N \cup Q$, where $Q$ is the set of Steiner points of $T$.

As a consequence of all these statements it is sufficient to develop solution methods only for specific kinds of trees: Let $T = (V, E)$ be a tree for $N = \{v_1, \ldots, v_n\}$, $n > 2$, with

- $V = \{v_1, \ldots, v_{2n-2}\}$,
- $g_T(v_i) = 3$ for $i = n+1, \ldots, 2n-2$ and $g_T(v_i) = 1$ for $i = 1, \ldots, n$, where $g_T(v)$ denotes the degree of the vertex $v$ in $T$.

Such a tree will be called a full tree.

Second, Melzak gave a finite solution method to Steiner's Problem, using a set of Euclidean (that is ruler and compass) constructions. The central idea is given in the Torricelli construction given in the chapter before: In the three-point problem, a replacement point can be substituted for two of the given points without changing the length of the tree. In the general version of the problem the algorithm must guess which pair is to be replaced, which could potentially involve may trying all possible guesses. After one pair of points in the subset has been replaced by a single point, each subsequent step of the algorithm replaces either two given points, a given point and a replacement point or two replacement points with another replacement point until the subset is reduced to three points.[7]
Once the Steiner point for those three points has been found, the algorithm works backwards, attempting to determine the Steiner point corresponding to each replacement point. An attempt can fail because of contradictory constraints on the placement of Steiner points.
Now we give a complete list of the instructions of this method:

[7]Surprisingly, the Melzak algorithm cannot be extended to higher-dimensional Euclidean spaces, not even to spaces of dimension three. The reason is that for two given points there are an infinite number of replacement points.

**Algorithm 2.1.1** *(Melzak [305]) Let $T = (V, E)$ be a full tree for the finite set $N$ of points. Then do*

1. $T' = (V', E') := T$; $N' := N$;
2. *(Reduction stage)*
   $Q := V' \setminus N'$;
   *if $Q$ is empty then goto 4.;*
3. *Let $q$ be in $Q$ such that $q$ is adjacent to $v_1$ and $v_2$ in $N'$;*
   *Delete $v_1, v_2$ and $q$;*
   *Add a substitution point $v_{12}$ that forms an equilateral triangle with $v_1$ and $v_2$: $V' := V' \cup \{v_{12}\} \setminus \{v_1, v_2, q\}$;*
   $N' := N' \cup \{v_{12}\} \setminus \{v_1, v_2\}$;
   *goto 2.;*
4. *(Recovery stage)*
   *Connect the last two points of $N'$ by an edge;*
5. *Reserve the order of the reduction steps and bring back each pair of $v_1$ and $v_2$ at each recovery step;*
6. *Let $C$ be the circle circumscribing $v_1, v_2$ and $v_{12}$;*
   *If the arc $v_1 v_2$ of $C$ intersects the edge incident to $v_{12}$ at the point $v'$, then $v'$ is the Steiner point joining $v_1$ and $v_2$; in this case connect these points and discard $v_{12}$;*
   *goto 5.*

The proof of correctness is to apply the construction of 1.1.2(b): Let $q$ be a Steiner point adjacent to the given points $v_1$ and $v_2$. $v_1$, $v_2$ and $v_{12}$ form an equilateral triangle. Since the Steiner point $q$ is the Torricelli point for $v_1$, $v_2$ and $v_3$ it makes angles of $120^o$ with the edges to each of them. If a quadrilateral is inscribed in a circle, the sum of opposite angles equals $180^o$. Thus the Steiner point $q$ is necessarily located on the circle circumscribing $v_1, v_2$ and $v_{12}$.
The theorem of Ptolemaius says

$$||v_1 - v_{12}|| \cdot ||v_2 - q|| + ||v_2 - v_{12}|| \cdot ||v_1 - q|| = ||v_{12} - q|| \cdot ||v_1 - v_2||;$$

and consequently

$$||v_2 - q|| + ||v_1 - q|| = ||v_{12} - q||.$$

This means that

$$\sum_{i=1}^{3} ||v_i - q|| = ||v_{12} - q|| + ||q - v_3||$$

achieves a minimal value if and only if $q \in \underline{v_{12}v_3}$.

□

It is obvious that using Melzak's algorithm to find an SMT, although effective, is extremely redundant and inefficient, more exactly it takes exponential time. There are two causes of the exponential running time:

- The main reason is the large number of trees which are to be considered.
- Moreover, each step chooses one of two possible substitution points because there are two equilateral triangles for a given side. Since the correctness of the choice cannot be seen until the tree has been constructed or demonstrated to be impossible, backtracking is necessary. Hence, we require $O(2^k)$ time, where $k$ is the number of Steiner points in the given tree.

Hwang [229] has described a implementation of Melzak's construction which eliminates the second cause of exponential behavior.

In general, to determine an SMT for a given finite set of points we have to consider many different trees, and compare their lengths in order to single out the shortest ones. Unfortunately, this needs an astronomical number of computational steps. Although exponential-time algorithms have been found for Steiner's Problem, no polynomial-time algorithms have yet been found and the prospects for such an algorithm are not good.

## 2.2 EXAMPLES AND EXERCISES

For an introduction to Steiner's Problem it is helpful to investigate several specific cases to explore the difficulties and surprising twists of the problem.

**I.** Show that the degree of each vertex is at most three; and hence, that the degree of each Steiner point equals three. It is helpful to observe that any Steiner point is the Torricelli point of its neighbors. Moreover, we then have that two edges incident to a Steiner point meet at an angle of $120^o$.

**II.** Not every locally minimal tree, however, is a solution of minimal length overall - that is, an SMT. Large-scale rearrangements of the Steiner points may be necessary to transform a network into a shortest possible tree, which is a globally minimal tree. To see this we investigate the following example: Consider the four corners of a rectangle in the Euclidean plane measuring three units by four units. An MST for these points has length 10. There are two locally minimal trees with two Steiner points. Each arrangement forms a tree that has three edges connected to each Steiner point at $120^o$. If the Steiner points are arranged parallel to the width, the locally minimal tree that results is $9.928\ldots$ units long. If the Steiner points are arranged parallel to the length, a locally minimal tree results with a length of $9.196\ldots$. Consequently, only in the last case do we have an SMT.

Ollerenshaw (compare [147]) proved that if two full trees exist for the four points, the one with the longer edge between the two Steiner points is the shorter tree, i.e. the SMT. Moreover, this consideration shows that a solution of Steiner's Problem is not always uniquely determined: For four points forming a square, we have two equivalent (equal length) solution.

**III.** Let $N$ be the set of nodes of a regular $n$-gon in the plane, $n = |N| \geq 3$. Find an SMT for $N$. For $n = 3$ we seek a Torricelli point. For $n = 4$ the example above will be helpful, where, roughly spoken, the "Double Y" is shorter than the "X". It is not simple to see (compare [141]) that for $n \geq 6$ there is no Steiner point in the tree, meaning the SMT is an MST with length equal to $(n-1) \cdot l$, where $l$ is the length of a side. Jarnik and Kösler proved this result in 1934 for $n \geq 13$. It was another fifty years until the proof by Du et al., compare [314].

**IV.** A set $N = \{(i,0),(i,1) : i = 0,\ldots,n-1\}$ is called a ladder. Chung and Graham [84] examined ladders and determined the length of SMT's for these sets. Particularly, they demonstrated that there are arbitrarily large sets of points for which the SMT cannot be separated, that means cannot be divided in full trees.
Burkard et.al [60] describe a method that always finds a solution for Steiner's Problem for ladders of the kind $N = \{(i \cdot b, 0), (i \cdot b, 1) : i = 0,\ldots,n-1\}$, where $b \leq 1$.
The subject becomes more difficult if we consider grids of arbitrary dimension. A nice representation of this question has been given in [175] and [176].

**V.** Suppose we wish to find a network that will connect a set of given points. One way to do this is to use a MST, which uses only edges joining pairs of the given points. We saw that such a network is easy to find. Another is to use an SMT. Obviously, the length of the SMT is less than or equal the length of the MST. How much shorter can it get? Consider three points which form the

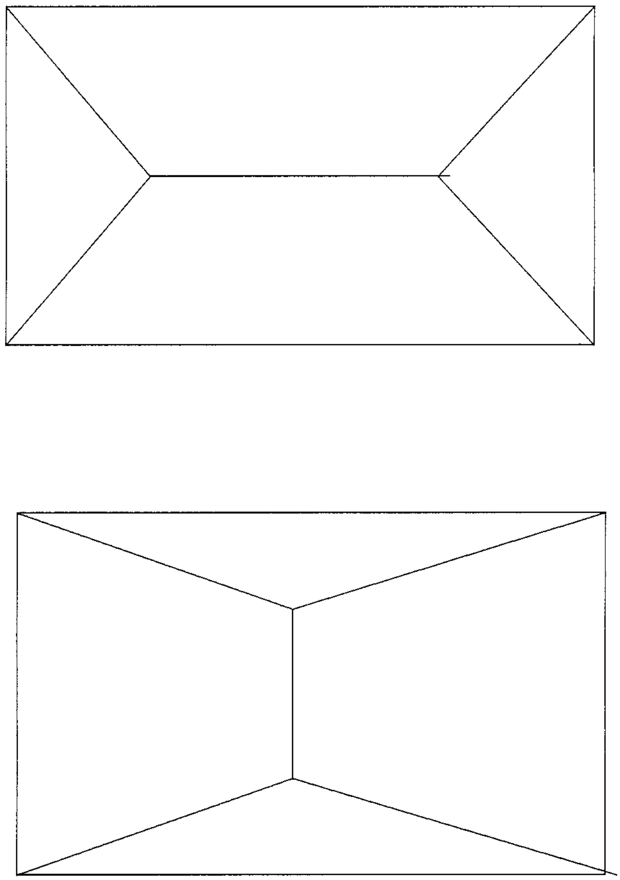

**Figure 2.2** Two locally minimal trees

corners of an equilateral triangle of unit side length. An MST for these points has length 2. An SMT uses one Steiner point, which is uniquely determined by the condition that the three angles at this point are equal, and consequently equal $120^o$. Consequently, with help of a simple calculation, using the cosine law, we find the length of the SMT in $3 \cdot \sqrt{1/3} = \sqrt{3}$. So we have the ratio of between the length of the both networks is $\sqrt{3}/2 = 0.866025....$ Is there a finite set of points for which the ratio is smaller?

**VI.** Related to Steiner's Problem, we will require that the minimal network has at most $k$ Steiner points, where $k \geq 0$ is a predetermined integer independent of the number of given points. Such a network must be a tree also, and is called a $k$-SMT. This problem was introduced independently by C. [87] in 1982 and Georgakopoulos and Papadimitriou [183] in 1987.
The combinatorial structures of $k$-SMT's and SMT's are quite different. Particularly, in contrast to I., we find Steiner points of degree 4 in $k$-SMT's.[8]

It is a difficult task to discuss all these examples in spaces other than the Euclidean plane.

## 2.3 REFERENCES

Steiner's Problem is one of the most famous combinatorial-geometrical problems. It is the core of the so-called Geometric Network Design, but has itself two origins: Fermat's Problem and the Minimum Spanning Tree problem. Consequently, in the last three decades the investigations into and, naturally, the publications about Steiner's Problem have increased rapidly. The articles that have been written on Steiner's Problem and its relatives are nearly countless.
The first survey of Steiner's Problem in the Euclidean plane was presented by Gilbert and Pollak in 1968 [186]; they christened the terms "Steiner Minimal Tree" for the shortest interconnecting network and "Steiner points" for the additional vertices.

It is well-known that solutions of network design problems depend essentially on the way in which the distances in space are determined. Clearly, this is true for Steiner's Problem. Consequently, there are many metric spaces[9] to be considered.

Surveys in form of monographs are given by

1. S.Voß: "Steiner-Probleme in Graphen", 1990, [439].
2. F.K.Hwang, D.S.Richards, P.Winter: "The Steiner Tree Problem", 1992, [231].
3. A.O.Ivanov, A.A.Tuzhilin: "Minimal Networks - The Steiner Problem and Its Generalizations", 1994, [238].

[8]But not Steiner points of higher degree, see [89] and [369].
[9]See the next section.

4. D.Cieslik: "Steiner Minimal Trees", 1998, [92].
5. A.O.Ivanov, A.A.Tuzhilin: "Branching Solutions of One-Dimensional Variational Problems ", 2000, [235].
6. D.Cieslik: "The Steiner Ratio", 2001, [99].
7. H.J.Prömel, A.Steger: "The Steiner Tree Problem", 2002, [355].

Surveys in journals are given by Harris [212], Hwang and Richards [230], and Winter [464]. There are several collections about Steiner's Problem and its relatives: [79], [143], [239], [335], [441] and [435]. A nice representation of the complete subject has been given in [44], [45], [108], [175], [176], [219], [234], [389] and [422].

In this sense it is strange that people "discover" Steiner's Problem again and again, and prove "facts" which have already been proven a dozen times.[10]

## 2.4 A FIRST ANALYSIS OF STEINER'S PROBLEM

We start with a general analysis of Steiner's Problem in arbitrary metric spaces. We describe several basic facts about the combinatorial and geometrical structure of SMT's. Later we will discuss more detailed facts that arise if we restrict ourselves to specific cases.

### 2.4.1 Metric spaces

Distance is the mathematical description of the idea of proximity, and consequently, we may assume (and it is not hard to see) that a solution of Steiner's Problem depends essentially on the way in which a distance in the space is determined.

The following term was introduced by Fréchet in 1906: A pair $(X, \rho)$ is called a metric space if $X$ is a nonempty set of elements called the points, and $\rho : X \times X \to \mathbb{R}$ is a real-valued function satisfying:

[10] One of these discoveries is the fact that the degree of a Steiner point in an SMT in Euclidean spaces of arbitrary dimension equals 3.

(i) $\rho(x, y) \geq 0$ for all $x, y$ in $X$;

(ii) $\rho(x, y) = 0$ if and only if $x = y$;

(iii) $\rho(x, y) = \rho(y, x)$ for all $x, y$ in $X$; and

(iv) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for all $x, y, z$ in $X$ (triangle inequality).

Usually, such a function $\rho$ is called a metric.[11] We will say that the quantity $\rho(x, y)$ is the distance between the points $x$ and $y$. If $\rho$ satisfies (ii) only in the weaker form

(ii') $\rho(x, x) = 0$ for all $x$ in $X$;

we say that $\rho$ is a pseudometric.
If the function $\rho$ satisfies the conditons (i),(ii') and (iii) it is called a dissimilarity.[12]
A metric, pseudometric or dissimilarity $\rho$ on a finite set $X$ of $n$ points can be specified by an $n \times n$ matrix of (nonnegative) real numbers. (Actually $\binom{n}{2}$ numbers suffice because of (ii') and (iii).)

Let $(X, \rho)$ be a metric space. If $X' \subseteq X$, then the restriction $\rho'$ of the metric $\rho$ on $X' \times X'$ is a metric on $X'$. In what follows we regard $(X', \rho')$ as a metric space and call it a subspace of $(X, \rho)$.

A graph $G = (V, E)$ is embedded in $(X, \rho)$ such that

(i) $V$ is a (finite) subset of $X$.

(ii) $E$ is the set of all unordered pairs $\underline{vv'}$ of points $v$ and $v'$ in $V$.

(iii) The metric $\rho$ induces a length function for the graph, so that for each edge $\underline{vv'}$ a length is given by $\rho(v, v')$.

---

[11] The axioms are not independent: (i) is a consequence of (iv). On the other hand,

**Observation 2.4.1** *A metric $\rho$ can be defined equivalently by*

*(ii) $\rho(x, y) = 0$ if and only if $x = y$; and*

*(iv') $\rho(x, y) \leq \rho(x, z) + \rho(y, z)$ for all $x, y, z$ in $X$.*

[12] We will give the reason for this name later. There are various measures of dissimilarity, and not all of them yield a metric, but many do.

(iv) We define the length of the graph $G$ in $(X, \rho)$ as the total length of $G$:

$$L(G) = L(X, \rho)(G) := \sum_{\underline{vv'} \in E} \rho(v, v'). \tag{2.2}$$

In general we will consider graphs and their embedding in a metric space at the same time. In each case it will be easy to see whether we use combinatorial or metric/geometric facts.

Steiner's Problem is the "Problem of Shortest Connectivity". Since the demand of shortness forces the network to be cycle free it is only necessary to consider trees:

**Observation 2.4.2** *Steiner's Problem is only interested in trees.*

Let $N$ be a finite set of points in the metric space $(X, \rho)$. For a given natural number $k$ and for $k$ points $v_1, ..., v_k \in X \setminus N$, let $T(k, v_1, ..., v_k)$ be a spanning tree of minimal length in the complete graph with the set $N \cup \{v_1, ..., v_k\}$ of vertices, where the length of the graph is induced by the metric $\rho$ as defined in (2.2).[13]
If there is both a number $k'$ and points $w_1, ..., w_{k'}$ such that the value

$$L(X, \rho)(T(k', w_1, ..., w_{k'})) \tag{2.3}$$

is minimal among all candidates $T(k, v_1, ..., v_k)$, then we call $T(k', w_1, ..., w_{k'})$ a Steiner Minimal Tree (SMT) for $N$, and the points $w_1, ..., w_{k'}$ are called Steiner points. That means, an SMT for $N$ is a minimum spanning tree on $N \cup Q$, where $Q$ is a set of additional vertices inserted into the metric space in order to achieve a minimal solution.

It is not true that there is an SMT for any given finite set in each metric space, but for all spaces considered in this book any given finite set has an SMT; this implies that the set $Q$ of additional vertices is a finite set as well.[14]

In the remainder of this section, we will discuss which properties an SMT possesses, under the assumption that an SMT exists.

[13] Recall that a minimum spanning tree can be found easily.

[14] Examples for spaces in which there does not always exist an SMT are given in the next chapter.

**Observation 2.4.3** *Let $(X, \rho)$ be a metric space and let $N$ be a finite set of points in $X$. Without loss of generality, the following is true for any SMT $T = (V, E)$ for $N$*

*(a)* $g_T(v) \geq 1$ *for each vertex $v$ in $V$;*

*(b)* $g_T(v) \geq 3$ *for each Steiner point $v$ in $V$.*

*Proof.* (a) is an obvious fact, since $T$ is a tree which connects all vertices.
It is impossible for a Steiner point $v$ to have degree one, since the edge $\underline{vv'}$ which joins $v$ with the remaining tree has a positive length, contradicts the minimality requirement.
The triangle inequality of the metric $\rho$ implies (b) in the following way: Let $v$ be a Steiner point of degree two. Then we may replace the two edges $\underline{vw}$ and $\underline{vw'}$ by the edge $\underline{ww'}$. Because

$$\rho(w, w') \leq \rho(w, v) + \rho(v, w'), \tag{2.4}$$

the new tree is not longer than the old.

□

Moreover, a Steiner point $v$ in an SMT $T$ can be of degree two. Then

$$\rho(w, v) + \rho(v, w') = \rho(w, w') \tag{2.5}$$

holds for $\{w, w'\} = N_T(v)$.[15]
Now, we will prove that the number of Steiner points cannot increase arbitrarily:

**Observation 2.4.4** *It is sufficient to consider only finite trees as candidates for an SMT.*

*Proof.* Let $T = (V, E)$ be a tree interconnecting a finite set $N = \{v_1, ..., v_n\}$ of points.
The number of vertices in $T$ is bounded, more precisely: If $|V| \geq 2n^2$ then there exists a tree interconnecting $N$ which is a proper subtree of $T$ and consequently has a shorter length. To show this we distinguish between two cases:

[15] This observation will be helpful in several investigations. In some proofs we will use Steiner points of degree two.

Case 1: For any two points $v$ and $v'$ in $N$ the path $T(v, \ldots, v')$ contains at most $2n$ vertices.
Then we define the graph $G$ by

$$G = \bigcup_{i=1}^{n-1} T(v_i, ..., v_{i+1}).$$

The graph $G$ interconnects all points of $N$ by edges of $T$ and contains at most $2n(n-1) = 2n^2 - 2n < 2n^2 = |V|$ vertices. Hence, a spanning tree of $G$ is a proper subtree of $T$ and must be shorter.

Case 2: There are two points $v$ and $v'$ in $N$ such that the path $T(v, \ldots, v')$ has more than $2n$ vertices.
Then $T(v, ..., v')$ contains at least $n+1$ Steiner points, each of which is of degree at least three, see 2.4.3. If $T(v, ..., v')$ is removed from $T$ we get the graph $G$. We observe that $G$ is a forest with at least $n + 1$ connected components. Hence, at least one component does not contain a point of $N$. If we remove this component in the tree $T$ we get a shorter tree.

□

We can determine a sharp upper bound for the number of Steiner points explicitly:

**Observation 2.4.5** *Let $(X, \rho)$ be a metric space and let $N$ be a finite set of points in $X$. Without loss of generality,*

$$|V \setminus N| \leq |N| - 2; \tag{2.6}$$

*hence*

$$|V| \leq 2 \cdot |N| - 2 \tag{2.7}$$

*and*

$$|E| \leq 2 \cdot |N| - 3 \tag{2.8}$$

*is true for any SMT $T = (V, E)$ for $N$.*
*Equality holds if and only if the vertices from $N$ are the leaves of $T$ and the Steiner points are of degree three.*

*Proof.* In 2.4.4 we found that it is sufficient to consider finite trees. Hence, the first assertion is a consequence of

$$2 \cdot |N| + 2 \cdot |V \setminus N| - 2 \quad = \quad 2 \cdot (|V| - 1)$$

$$
\begin{aligned}
&= 2 \cdot |E| \\
&= \sum_{v \in V} g_T(v) \\
&= \sum_{v \in V \setminus N} g_T(v) + \sum_{v \in N} g_T(v) \\
&\geq 3 \cdot |V \setminus N| + |N|.
\end{aligned}
$$

The number of edges in a tree is one less than the number of vertices. Consequently, the third inequality must hold.
The discussion of equality follows immediately from 1.2.5.

□

Another observation for trees with Steiner points will frequently be helpful:

**Observation 2.4.6** *Let $T = (V, E)$ be an SMT for $N$. If $V \setminus N$ is nonempty then it contains at least one Steiner point adjacent to two given points.*

*Proof.* Assume that each Steiner point is adjacent to at most one vertex in $N$. The set $V' = V \setminus N$ induces in $T$ a subgraph $G' = (V', E')$. It follows from 1.2.1 that

$$
\begin{aligned}
|E'| &= \frac{1}{2} \sum_{v \in V'} g_{G'}(v) \\
&\geq \frac{1}{2} \sum_{v \in V'} (g_T(v) - 1) \\
&\geq \frac{1}{2} \sum_{v \in V'} 2 \\
&= |V'|.
\end{aligned}
$$

This contradicts the fact that the forest $G'$ has at most $|V'| - 1$ edges, compare 1.2.6.

□

An SMT is a finite tree. The number of such trees for a finite set of given points (vertices) must also be finite. In other words,

**Observation 2.4.7** *It is sufficient to consider only a finite number of trees as candidates for an SMT.*

It will be helpful to associate a matrix to a graph: Let $G = (V, E)$ be a graph and assume that the vertices are labelled, i.e. $V = \{v_1, ..., v_n\}$. Then we define the adjacency matrix $A(G) = (a_{ij})_{i,j=1,...,n}$ with

$$a_{ij} = \begin{cases} 1 & : \text{ if the vertices } v_i \text{ and } v_j \text{ are adjacent} \\ 0 & : \text{ otherwise} \end{cases}$$

These matrices contain the complete information about the structure of graphs. Consequently, many matrix calculations have a meaning in the sense of graph theory.[16] The adjacency matrix of the graph $G$ does depend on the labelling of the vertices of $G$; that is, a different labelling of the vertices may result in a different matrix, but they are closely related in that one can be obtained from the other simply by interchanging rows and columns.
A matrix which contains entries only 0 or 1 is called a binary or Boolean matrix. Using adjacency matrices we can describe the length of a graph $G$ by

$$L(X, \rho)(G) = \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} a_{ij} \cdot \rho(v_i, v_j). \tag{2.9}$$

In other words,

**Observation 2.4.8** *For a given topology of a tree its length in a metric space is a linear function of the metric.*

Steiner point locations in the space are not prespecified from a candidate list of point locations, but we may assume that the set of Steiner points is contained in a suitably bounded subset of the space. Here, a set $W$ of points in a metric space $(X, \rho)$ is called bounded if

$$\sup\{\rho(x, y) : x, y \in W\} < \infty. \tag{2.10}$$

Equivalently, we consider balls in the space defined by

$$B_r(z) = \{x \in X : \rho(x, z) \leq r\}, \tag{2.11}$$

[16] We will discuss this further in the next section.

where $r \geq 0$ is a real and $z$ is a point of the space. Then it is easy to see that the set $W$ is bounded if and only if there exists a nonnegative real $r$ and a point $z$ such that

$$W \subseteq B_r(z). \tag{2.12}$$

**Observation 2.4.9** *Let $N$ be a finite set of points in a metric space $(X, \rho)$. Then we may assume that the set $V \setminus N$ of Steiner points of an SMT $T = (V, E)$ for $N$ is contained in a bounded subset of $X$:*

$$V \setminus N \subseteq B_r(v), \tag{2.13}$$

*where $v$ is an arbitrary point in $N$ and $r = L(X, \rho)(MST \text{ for } N)$.*

Note that it is not simple to describe a small set containing all Steiner points. Such a set is usually called a Steiner hull of $N$. A known Steiner hull allows confinement of the construction of the tree within a given set. Hence, the smaller a Steiner hull is, the better it is.

On the other hand, if the Steiner points in $Q$ have been localized, an SMT for $N$ is simple to find, since

**Observation 2.4.10** *Let $N$ be a finite set of points in a metric space. Then an SMT $T = (V, E)$ for $N$ is an MST for $V$.*

Comparing all these facts, the search for an SMT for a finite set of points in a metric space forces investigations of two specific questions:

- How many Steiner points are used in an SMT?
- Where are these Steiner points located in the space?

Unfortunately, these questions cannot solved independently from the construction of the shortest tree itself. For a complete discussion of these difficulties see [92], [230], [231] and [464], or the next chapter.

What are the spaces for which an SMT always exists? Such a tree necessarily exists if the bounded subset which contains the Steiner points is compact.[17] In

[17] Actually in several interesting cases it will be finite.

this case we must consider, for each tree of a finite number of trees, the value of the function (2.9). More precisely:

**I.** Considering "continuous" spaces, it is sufficient for an SMT to exist if the metric space has the following four properties:

(i) $(X, \rho)$ is complete;

(ii) $(X, \rho)$ is finitely compact, i.e. each bounded and closed subset is compact;

(iii) Each pair of points in $(X, \rho)$ can be connected with a geodesic curve, i.e. a curve of shortest length;[18]

(iv) For all points $x, x'$ in $(X, \rho)$, the distance $\rho(x, x')$ is equal to the length of a geodesic curve joining $x$ and $x'$.

The following classes of metric spaces satisfy the four properties and thus in each case we establish the existence of an SMT for a finite set $N$ of points with the help of a compactness argument:

(a) Euclidean spaces are classical examples for Steiner's Problem.

(b) Finite-dimensional Banach spaces. Since these spaces play an important role in both theoretical questions and in applications we will describe them more extensively.
In his book *Geometrie der Zahlen* [310], published in 1896, Minkowski proved a number of results by geometrical arguments, using the idea of normed spaces which is based on the assumption that to each vector can be assigned its "length" or norm satisfying some "natural" conditions.
A convex and compact body $B$ of the $d$-dimensional affine space $A_d$ centered in the origin $o$ is called a unit ball, and induces a norm $||.|| = ||.||_B$ in the corresponding $d$-dimensional linear space $A_d$ according to the so-called Minkowski functional:

$||v||_B = \inf\{t > 0 : v \in tB\}$ for any $v$ in $A_d \setminus \{o\}$,

and

$||o||_B = 0.$

On the other hand, let $||.||$ be a norm in $A_d$, which means that $||.|| : A_d \to I\!R$ is a real-valued function satisfying

(i) positivity: $||v|| \geq 0$ for any $v$ in $A_d$;

[18] This is the specific form of Steiner's Problem for two given points.

(ii) identity: $||v|| = 0$ if and only if $v = o$;

(iii) homogeneity: $||tv|| = |t| \cdot ||v||$ for any $v$ in $A_d$ and any real $t$; and

(iv) triangle inequality: $||v + v'|| \leq ||v|| + ||v'||$ for any $v, v'$ in $A_d$.

Then $B = \{v \in A_d : ||v|| \leq 1\}$ is a unit ball in the above sense. It is not hard to see that the correspondence between unit balls $B$ and norms $||.||$ is unique, that is, a norm is completely determined by its unit ball and vice versa. Consequently, such a space is uniquely defined by an affine space $A_d$ and a unit ball $B$. It is called a Banach-Minkowski space, and is abbreviated as $M_d(B)$.
A Banach-Minkowski space $M_d(B)$ is a complete metric linear space if we define the metric by

$$\rho(v, v') = ||v - v'||_B. \tag{2.14}$$

Usually, a (finite- or infinite-dimensional) linear space which is complete with regard to its given norm is called a Banach space. Essentially, every Banach-Minkowski space is a finite-dimensional Banach space and vice versa.[19]

**Observation 2.4.11** *Segments in a Banach-Minkowski space are shortest curves (in the sense of inner geometry). They are the unique shortest curves if and only if the unit ball is strictly convex.*[20]

Roughly speaking, the observation that a straight line is the shortest distance between two points is Steiner's Problem for a set of two points.

In particular, we consider finite-dimensional spaces with $p$-norm, defined in the following way: For $v = (x_1, ..., x_d)$ we define the norm by

$$||v||_{(p)} = \left( \sum_{i=1}^{d} |x_i|^p \right)^{1/p} \tag{2.15}$$

[19]An infinite dimensional Banach space is often called a Banach-Wiener space, compare [460]. The structure of such spaces is intrinsically more complicated than that of the finite dimensional ones.

[20]The fourth problem of Hilbert, is to characterize all geometries in which segments (convex hulls of two different points) are shortest curves (in the sense of inner geometry). In particular, Hilbert asks for the construction of all these metrics and the study of the individual geometries.
Hilbert's problem is a program of research about the foundations of geometry. The major contributions were the books *The Geometry of Geodesics* [61] by Busemann in 1955 and *Hilbert's Fourth Problem* [347] by Pogorelov in 1979. For a historical discussion compare [11] and [468].

where $1 \leq p < \infty$ is a real number. If $p$ runs to infinity then we get the so-called Maximum norm

$$||v||_{(\infty)} = \max\{|x_i| : 0 \leq i \leq d\} \tag{2.16}$$

In each case we obtain a Banach-Minkowski space written by $\mathcal{L}_p^d$.

(c) Compact manifolds.

About more facts of metric/geometric properties of several continuous spaces compare [262], [281], [297], [364], [381], [411] and [426].

**II.** Concerning "discrete" spaces we make the following definition: A metric space $(X, \rho)$ is called a discrete metric space if any bounded set is finite. In other words, if for a subset $W$ it holds that

$$\sup\{\rho(x, x') : x, x' \in W\} < \infty \tag{2.17}$$

then also

$$|W| < \infty. \tag{2.18}$$

Consequently, an SMT exists for any finite set of points in such spaces. Examples are:

(a) Finite metric spaces.

(b) Graphs (a specific case of finite metric spaces[21]).

(c) Let $\mathbf{Z}$ be the set of all integers, then $\mathbf{Z}^d$ equipped with a rectilinear, Euclidean or any other "desired chosen" distance is a discrete metric space.

(d) Spaces of words with phylogenetic (= space measured evolutionary) distances.

Note, that an infinite set with the so-called discrete metric, which defines the distance between two different points to be 1, is not a discrete metric space.[22]

For more facts about metric/geometric properties of several discrete spaces compare [246] and [476].

---

[21] An introduction to the theory of graphs we gave in the previous chapter; the representation as metric spaces we will describe at the end of the present chapter.

[22] But, of course, for any given set of points in such a space there exists an SMT, namely the MST.

### 2.4.2 More facts in the Euclidean plane

Of course, if we investigate a more specific metric space, we find further facts about Steiner Minimal Trees.

The Euclidean plane is defined in the affine plane with the Euclidean metric $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ between the points $(x_1, y_1)$ and $(x_2, y_2)$ derived from a norm $||.||$:

$$||(x, y)|| = \sqrt{x^2 + y^2}. \tag{2.19}$$

Steiner's Problem looks for a shortest network and in particular for a curve $\mathcal{C}$ of shortest length joining two points. For our purposes, we regard a geodesic curve as any curve of shortest length.
If we parametrize the curve $\mathcal{C}$ by a differentiable map $\gamma : [0, 1] \to I\!\!R^d$ we define

$$\text{length of } \mathcal{C} = \int_0^1 ||\dot{\gamma}||dt. \tag{2.20}$$

It is not hard to see that among all differentiable curves $\mathcal{C}$ from the point $v$ to the point $v'$ the segment

$$\underline{vv'} = \{tv + (1 - t)v' : 0 \leq t \leq 1\} \tag{2.21}$$

minimizes the length of $\mathcal{C}$. And, moreover, as a consequence of 2.4.11,

**Observation 2.4.12** *All segments and no other sets of points are geodesic curves in the Euclidean plane.*

Consequently, in the Euclidean plane SMTs always exist, and we may represent a graph $G = (V, E)$ embedded in the plane so that

(i) $V$ is a finite set of points;

(ii) Each edge $\underline{vv'}$ is a geodesic curve, which means a shortest curve in the sense of inner geometry. We may assume that $\underline{vv'}$ is a segment.[23];

(iii) Each edge $\underline{vv'}$ has length $||v - v'||$;

(iv) The length of $G$ is defined by

$$L(G) = \sum_{\underline{vv'} \in E} ||v - v'||. \tag{2.22}$$

[23] This justifies the double meaning of $\underline{vv'}$ as an edge of a graph and as a segment.

Using our first example in section 2.2, we have

**Observation 2.4.13** *Let $N$ be a finite set of points in the Euclidean plane. Without loss of generality, in an SMT $T = (V, E)$ for $N$ a given point can have degree 1, 2 or 3; a Steiner point always has degree 3.*

Moreover, paying attention 2.4.5, we find

**Observation 2.4.14** *An SMT for $n$ given points has exactly $n - 2$ Steiner points if and only if each given point is of degree one.*

A tree with the property described in the last observation is called a full tree. It is a binary tree, i.e. it contains only leaves and internal vertices of degree three.
The following property of full trees can be empirically observed: "Typical" sets of given points in the Euclidean plane usually do not have SMTs which are full trees. That is, its SMTs tend to be unions of small full trees.[24]
We decompose a given tree for $N$ into full trees by the following procedure:

**Procedure 2.4.15** *Let $T = (V, E)$ be a tree for $N$, that means $N \subseteq V$, and let $v$ be a point in $N$ with $g(v) > 1$.*

1. *Define $G = (V \setminus \{v\}, E \setminus \{\underline{vv'} : v' \text{ is a neighbor of } v\})$.*
   *($G$ is a forest with $g(v)$ components $G_i = (V_i, E_i)$, $i = 1, \ldots, g(v)$.)*
2. *Define for $i = 1, \ldots, g(v)$ the graph*
   *$G_{(i)} = (V_i \cup \{v_i\}, E_i \cup \{\underline{v_i v'} : v' \text{ is a neighbor of } v \text{ in } G \text{ and } v' \text{ is in } V_i\})$, where $v_i$ is not in $V$.*

If we repeat this procedure we obtain a family of trees in which for each tree, the degree of any vertex which is a given point equals one.

**Observation 2.4.16** *Let $T = (V, E)$ be a tree for $N$. Then the number of full trees of $T$ is*

$$1 + \sum_{v \in N} (g_T(v) - 1). \tag{2.23}$$

[24]The fastest exact algorithms (in practice) for Steiner's Problem use two phases: first a small but sufficient collection of full SMTs is generated and then an SMT is constructed from this collection. See [444].

To estimate the total number of full trees more exactly, denote by $f(n)$ the number of such trees with $n$ given and $n-2$ Steiner points. Then $f(2) = 1$. If one removes a given point and also its adjacent Steiner point, one obtains a full tree. This shows that every full tree with $n+1$ given points can be obtained from a full tree with $n$ given points by adding a Steiner point in one of the $2n-3$ edges and adding a new edge. Hence,

$$f(n+1) = (2n-3) \cdot f(n). \tag{2.24}$$

A solution of this recursive equation is given by

**Observation 2.4.17** *There are*

$$\frac{(2n-4)!}{2^{n-2}} \tag{2.25}$$

*pairwise distinct full trees with $n$ leaves.*
*Consequently, if we ignore the numbering of the internal vertices, we have to check*

$$\frac{(2n-4)!}{2^{n-2}(n-2)!} = 1 \cdot 3 \cdot 5 \cdots (2n-5) \tag{2.26}$$

*distinct full trees.*

Remember that it is not simple to describe a Steiner hull of $N$.

**Observation 2.4.18** *In the Euclidean plane the convex hull of the set of given points is a Steiner hull.*

In other words, there is a Steiner hull which is a polygon. Cockayne [110] was the first to find that an improved polygonal hull can be obtained by repeatedly deleting triangles from the boundary of the convex hull of the given set: In the following description, let $N$ be a finite set of points in the Euclidean plane.

1. Start with the convex hull conv$N$;

2. Let $v$ and $v'$ be two points of $N$ such that $\underline{vv'} \subseteq$ boundary of conv$N$. If there is a third point $w$ in $N$ such that the triangle conv$\{v, v', w\}$ contains no other point of $N$ and the angle at $w$ is not less than $120^o$ then no edge of the SMT is within conv$\{v, v', w\}$;

> The new boundary of the Steiner hull is obtained by replacing the segment $\underline{vv'}$ by the segments $\underline{vw}$ and $\underline{wv'}$. If the hull then becomes self-intersecting in some of the given points, the original problem can be decomposed into two or more smaller problems.

Weng [452] has generalized this concept and gives a method to construct Steiner polygons by repeatedly deleting $m$-gons, where $m$ is at most the number of given points. He has also shown the uniqueness of the Steiner polygons obtained by this method.

It is an interesting question to decide which of all these facts are true in higher-dimensional Euclidean spaces, or more generally, in metric spaces.[25]

[25] A helpful discovery in the investigations of Steiner's Problem is the the observation that the degrees of vertices of SMTs in finite-dimensional Banach spaces are bounded by a quantity which only depends on the space:

**Observation 2.4.19** *Consider d-dimensional Banach spaces with a smooth norm. Then it holds that*

(a) *(C. [92]) The degree of each vertex in an SMT is at most* $2d$.

(b) *(Lawlor, Morgan [275], Swanepoel [415])* $d+1$ *edges can meet at a Steiner point of an SMT, but never* $d+2$.

In particular, the degree of Steiner points in Euclidean spaces is independent of the dimension:

**Theorem 2.4.20** *In Euclidean spaces of any dimension the degree of a Steiner point in an SMT equals 3.*

*Proof.* The equation (1.11) also holds true in $d$ dimensions. Hence we have

$$-\frac{n}{2} \leq \left(-\frac{1}{2}\right) \cdot \binom{n}{2},$$

that is, an inequality which is satisfied only for $n \leq 3$.

For the planar case we know more about the vertex degrees.

**Observation 2.4.21** *Consider SMTs in a Banach planes equipped with a unit ball* $B$. *Then*

(a) *(C. [91], Swanepoel [416]) For the degrees of the vertices the following holds true: If* $B$ *is an affinely regular hexagon, then the degree is at most 6, otherwise at most 4.*

(b) *(Morgan et.al. [315]) At most four edges come together in a Steiner point.*

## 2.5 STEINER'S PROBLEM IN GRAPHS

Connectivity is also a very important concept in combinatorial optimization. We will discuss this concept in the sense of Shortest Connectivity in metric spaces.

### 2.5.1 The metric closure of a network

Here we consider networks. These are (connected) graphs $G = (V, E)$ equipped with a length function $f : E \to \mathbb{R}$. This function on the edges of $G$ is constrained to take only strictly positive values.[26]
The simplest question, which will be of great importance in further considerations, is to look for the "geodesic curves", which are the interconnecting chains of shortest length between vertices in the network:

**The Shortest Path Problem**
**Given:** A network $G = (V, E, f)$ and two vertices $v$ and $v'$ of $G$.
**Find:** A path connecting $v$ and $v'$ with minimal length.

A solution is called a shortest path (between the vertices $v$ and $v'$ in $G$).
With this in mind each network is a metric space, more precisely

**Observation 2.5.1** *Let $G = (V, E)$ be a connected graph equipped with a length function $f : E \to \mathbb{R}$. Define the distance function $\rho$ on $V$ so that*

$$\rho(v, v') = \text{the length of a shortest path between the vertices } v \text{ and } v' \text{ in } G, \tag{2.27}$$

*for two different vertices $v$ and $v'$, and $\rho(v, v) = 0$. Then $(V, \rho)$ is a metric space.*

The space $(V, \rho)$ is called the metric closure $G^f$ of a graph $G = (V, E)$ with length function $f : E \to \mathbb{R}$.
We can also define $G^f$ as the complete graph on $V$ such that the length of an edge $\underline{vv'}$ in $G^f$ is the length of a shortest path between $v$ and $v'$ in $G$. Then we call $G^f$ the distance graph of the network $G = (V, E, f)$. Note that $G$ is a subgraph of $G^f$, but the restriction of $\rho$ on $G$ must not be $f$.

[26]Nevertheless saying it explicitely, sometimes we will use a length function which has the value 0 for several edges.

The problem of finding shortest paths in a graph with a length function is easy to solve by the so-called dynamic programming technique, which is a rather general method for solving combinatorial problems having the property that their optimal solution can be computed recursively from solutions to subproblems. More precisely, we use the following observation, called Bellman's principle of optimality, which is indeed the core of dynamic programming:

**Observation 2.5.2** *(Bellman [37]) Let $G = (V, E, f)$ be a network, and let $v$ and $v'$ be two vertices of $G$. If $e = \underline{wv'}$ is the final edge of some shortest path $v, \ldots, w, v'$ from $v$ to $v'$, then $v, \ldots, w$ (that is the path without the edge $e$) is a shortest path from $v$ to $w$.*

Roughly speaking: An optimal strategy contains only optimal substrategies. The observation gives immediately

**Algorithm 2.5.3** *(Dijkstra [125]) Let $G = (V, E, f)$ be a network. A shortest path between the vertices $v$ and $v'$ can be found by the following procedure:*

1. *Start with the vertex $v$;*
   *Label $v$ with 0: $L(v) := 0$; all other vertices are unlabelled;*
2. *Determine $\min\{L(v_1) + f(\underline{v_1v_2})\}$ where $v_1$ and $v_2$ are adjacent vertices, $v_1$ labelled and $v_2$ not;*
   *Choose $\tilde{v}_1$ and $\tilde{v}_2$ which attain the minimum;*
   *Label $\tilde{v}_2$ by $L(\tilde{v}_2) = L(\tilde{v}_1) + f(\underline{\tilde{v}_1\tilde{v}_2})$;*
3. *Repeat the second step until $v'$ is labelled.*

*For all labelled vertices $w$ the quantity $L(w)$ is the length of a shortest path connecting $v$ and $w$:*

$$\rho(v, w) = L(w).$$

Now it is easy to construct the metric closure $G^f$: it is sufficient to apply 2.5.3 $|V|$ times.[27] When we are only interested in the metric $\rho$ we can find the metric closure in a simpler way:

[27] When the algorithm in 2.5.3 runs if all vertices are labelled then the algorithm creates a tree $T = (V, F)$ in which the unique path from $v$ to all other vertices $v'$ is a shortest path interconnecting these points in $G$. $T$ is called the distance tree related to $v$.

**Algorithm 2.5.4** *(Floyd [166]) Let $G = (V = \{v_1, \ldots, v_n\}, E, f)$ be a network. The metric closure $G^f = (V, \rho)$ can be found by the following procedure:*

1. *for $\underline{vv'} \notin E$ define $f(\underline{vv'}) = \infty$;*
2. *for $i := 1$ to $n$ do*
   *for $j := 1$ to $n$ do*
   *$\rho(v_i, v_j) := f(\underline{v_i v_j})$;*
3. *for $i := 1$ to $n$ do*
   *for $j := 1$ to $n$ do*
   *for $k := 1$ to $n$ do*
   *if $\rho(v_j, v_i) + \rho(v_i, v_k) < \rho(v_j, v_k)$ then $\rho(v_j, v_k) := \rho(v_j, v_i) + \rho(v_i, v_k)$.*

In particular, the function $f \equiv 1$ is a length function. It measures the distance by counting the number of edges in the path.

A first example: Let $A = A(G) = (a_{ij})_{i,j=1,\ldots,n}$ be the adjacency matrix for the graph $G = (V = \{v_1, ..., v_n\}, E)$. Then, obviously, the equation $a_{ij} = 1$ means that there is a chain of length 1 from $v_i$ to $v_j$. Now consider

$$A^k = (a_{ij}^{(k)})_{i,j=1,\ldots,n}, \tag{2.28}$$

the $k$-th power of $A$. Using induction it is not hard to see that the equation $a_{ij}^{(k)} = m$ means that there are $m$ different chains of length exactly $k$ from $v_i$ to $v_j$. Hence, the graph $G$ is connected if and only if for any pair of distinct vertices $v_i$ and $v_j$ there is a number $k = k(i,j)$ between 1 and $n-1$ such that $a_{ij}^{(k)} > 0$:

**Remark 2.5.5** *Let $G = (V = \{v_1, \ldots, v_n\}, E)$ be a connected graph, let $A = A(G)$ its adjacency matrix and let $A^k = (a_{ij}^{(k)})_{i,j=1,\ldots,n}$, $k = 1, 2, \ldots$. Then*

$$\rho(v_i, v_j) = \min\{k : a_{ij}^{(k)} > 0\}. \tag{2.29}$$

*holds true for any two distinct vertices $v_i$ and $v_j$.*

The quantity

$$\operatorname{diam} G = \max\{\rho(v, v') : v, v' \in V\} \tag{2.30}$$

is called the diameter of the graph $G = (V, E)$.[28] Of course, for any connected graph $G$ it holds that diam $G \le |V| - 1$. This implies that, using the adjacency matrix, we have to check only the powers up to $k = |V| - 1$ to decide if a graph is connected or not.

A complete overview about the theory of shortest paths in networks is given by Huckenbeck [227].

## 2.5.2 The Question

The central question of "Shortest Connectivity" in networks is

**Steiner's Problem in Graphs**
**Given:** A connected graph $G = (V, E)$ with a length-function $f : E \to I\!R$, and a nonempty subset $N$ of $V$.
**Find:** A connected subgraph $G' = (V', E')$ of $G$ such that

$$L(G') = \sum_{e \in E'} f(e) \tag{2.31}$$

is minimal.

This formulation is equivalent to our definition in the section before. This can be seen as follows: First, a solution $G' = (V', E')$ of Steiner's Problem must be a tree, because it is connected and acyclic. Consider the vertices in $V' \setminus N$. Such vertices $v$ with $g_{G'}(v) \ge 3$ are Steiner points, and the vertices $v$ with $g_{G'}(v) = 2$ lie on a shortest path between Steiner points and given points of $N$. In other terms, we consider Steiner's Problem in the metric closure $G^f$. The length of the SMT in both graphs must be the same.
In this sense, Steiner's Problem in graphs is a special case of the problem in metric spaces.[29] Since each finite metric space is equivalent to some network, compare [204], we have

**Observation 2.5.6** *Steiner's Problem in networks and in finite metric spaces are essentially the same.*

[28] For disconnected graphs this quantity is undefined, or $\infty$.
[29] In particular, there is no loss of generality in requiring that the length function satisfy the triangle inequality; if it does not, construct the metric closure.

Steiner's Problem in graphs was originally formulated by Hakimi [203] in 1971. Since then, the problem has received considerable attention in the literature. A collection of equivalent formulations for Steiner's Problem in graphs is given in [257].
Two specific cases are well-known:

$|N| = 2$**:** We search a shortest path interconnecting the two points in $N$. Here there does not exist a Steiner point, so any internal vertex on the path has degree 2. To find such paths we use the dynamic programming strategy of algorithm 2.5.3.

$N = V$**:** Here Steiner points are not necessary; we look for a minimum spanning tree. This is easy to do using the greedy strategy of algorithm 1.2.9.

Two algorithms, which generalize our specific cases, create an SMT in graphs. These algorithms are given by Dreyfus, Wagner and Hakimi.

The Dreyfus and Wagner solution method breaks the problem down into subproblems, and each of these subproblems themselves into subproblems etc., until the subproblems can be solved with help of a shortest path technique.

**Algorithm 2.5.7** *(Dreyfus and Wagner [133]) Let $G = (V, E, f)$ be a network. Let $N \subseteq V$ be a set of given points. Then an SMT for $N$ in $G$ can found by the following procedure:*

1. *(Initialization)*
   *For all vertices $v, v'$ compute $\rho(v, v')$ in $G$;*
2. *(Recursion)*
   *Perform the following calculations for all $k$ from 2 to $|N| - 1$:*
   *- For all $K \subseteq N$ such that $|K| = k$ and for all $v \in V \setminus K$, compute*
   $$L_v(K \cup \{v\}) = \min\{L(K' \cup \{v\}) + L(K \setminus K' \cup \{v\}) : \emptyset \neq K' \subset K\};$$
   *- For all $K \subseteq N$ such that $|K| = k$ and for all $v \in V \setminus K$, compute*
   $$L(K \cup \{v\}) = \min\{\min_{w \in K}\{\rho(v, w) + L(K)\}, \min_{w \notin K}\{\rho(v, w)) + L_w(K \cup \{w\})\}\},$$

*where $L(K)$ denotes the length of an SMT for $K$ and $L_v(K \cup \{v\})$ denotes the length of a shortest tree for $K \cup \{v\}$ satisfying the additional constraint that $v$ has degree at least two.*

The algorithm, as stated above, only computes the length of an SMT $T$. For the explicit construction of $T$, the algorithm has to be supplemented by a backtracking procedure.
The time complexity is $O(3^n k + 2^n k^2 + k^3)$, where $n = |N|$ and $k = |V|$. Hence, the algorithm is exponential in the number of given points and polynomial in the number of other vertices.

On the other hand, Hakimi proposed that a minimum spanning tree be calculated for each of the possible subsets of vertices, from just the set of given points through to the complete set of vertices:

**Algorithm 2.5.8** *(Hakimi [203], Lawler [274]) Let $G = (V, E, f)$ be a network. Let $N \subseteq V$ be a set of given points. Then an SMT for $N$ in $G$ can found by the following procedure:*

1. *Compute shortest paths between all pairs of vertices;*
   *Replace the edge lengths with the shortest path lengths, adding edges to the graph where necessary;*[30]
2. *For each possible subset $V' \subseteq V \setminus N$ such that $0 \leq |V'| \leq |N| - 2$, find a minimum spanning tree $T(N \cup V')$ in the induced subgraph $G^f[N \cup V']$;*
3. *Select the shortest spanning tree from the ones computed in step 2;*
   *Transform it into a tree of the original graph, i.e., replace each edge of the spanning tree with the edges of the shortest path between the vertices.*

The time complexity of the algorithm is $O(n^2 \cdot 2^{k-n} + k^3)$, where $n = |N|$ and $k = |V|$. Hence, the algorithm is polynomial in the number of given points and exponential in the number of the other vertices.

A polyhedral approach for Steiner's Problem in graphs is given by Aneja [16], Grötschel and Monma [195], and Lucena and Beasley [36]: For each edge $e \in E$, a variable $x_e$ is introduced. We consider the vector space $\mathbb{R}^E$. Each subset $F \subseteq E$ induces an incidence vector $\chi^F = (\chi^F_e)_{e \in E}$ in $\mathbb{R}^E$ by defining $\chi^F_e = 1$ if $e \in F$, $\chi^F_e = 0$ otherwise. Conversely, each 0/1 -vector $x$ in $\mathbb{R}^E$ induces a subset $F = \{e \in E : x_e = 1\}$ of the edge set $E$ of $G$. Then Steiner's Problem can be formulated as the following integer linear program:

$$\min \sum_{\underline{vv'} \in E} c_{\underline{vv'}} x_{\underline{vv'}}$$

[30] In other words, determine the metric closure of $G$.

subject to

| | |
|---|---|
| $\sum_{v\in W}\sum_{v'\in V\setminus W} x_{\underline{vv'}} \geq r_{\underline{uu'}}$ | for all pairs $u, u' \in V, u \neq u'$and for all $W \subseteq V$ with $u \in W, u' \neq W$ (where $r_{\underline{uu'}} = 1$ for all $u, u' \in N$, and $r_{\underline{uu'}} = 0$ otherwise); |
| $x_{\underline{vv'}} \in \{0, 1\}$ | for all $\underline{vv'} \in E$. |

Branch and bound is a technique for the complete enumeration of all possible solutions without having to consider them one by one. To apply this method to a combinatorial minimization problem, we need two steps:

**Branch:** A given subset of the possible solutions can be partitioned into at least two (nonempty) subsets;

**Bound:** For a subset obtained by branching iteratively, a lower bound on the length of any solution within this subset can be computed.

Such an algorithm for Steiner's Problem in graphs was first developed by Shore, Foulds and Gibbons [396]. Another branch and bound approach that uses heuristics to provide good lower bounds and is based on an integer programming formulation is given by Khoury and Pardalos [256].

Other approaches to solve Steiner's Problem in networks are given by C. et al. [102] and [103].

All known exact algorithms for Steiner's Problem in graphs are in some way enumerative algorithms. However, they differ in how the enumeration is done and how clever their strategies for avoiding total enumeration are.[31] Consequently, all of these algorithms need exponential time. But this is not a surprise, since

**Remark 2.5.9** *(Karp [251]) Steiner's Problem in graphs is $\mathcal{NP}$-hard.*[32]

Steiner's Problem remains $\mathcal{NP}$-hard if any of the following conditions hold:

[31] For the problem of enumerating all solutions see [104].

[32] For information about the complexity of problems see the next chapter.

- All edge lengths are equal, i.e. the length of a subgraph is its number of edges [251].
- The graph is bipartite [177].[33]
- The graph is a hypercube [169].
- The graph $G$ is planar [177], [355].[34]

[33]A graph $G = (V, E)$ is called bipartite if it is possible to partition $V$ into subsets $V_1$ and $V_2$ such that every edge joins a vertex of $V_1$ to a vertex of $V_2$. A well-known characterization is

**Theorem 2.5.10** *A connected graph is bipartite if and and only if it contains no cycle of odd length.*

*Sketch of the proof.* If a graph $G = (V, E)$ contains an odd cycle then it cannot possibly be bipartite.
Now suppose that $G$ contains no odd cycle, then choose any vertex $v$ of $G$ and create a partition by

$$V_1 = \{w \in V : \rho(v, w) \text{ is an even number}\} \tag{2.32}$$
$$V_2 = \{w \in V : \rho(v, w) \text{ is an odd number}\} \tag{2.33}$$

**Corollary 2.5.11** *All trees are bipartite.*

[34]A graph $G = (V, E)$ is called planar if it can be embedded into the plane such that no two curves which are the embeddings of the edges intersect each other outside of the vertices. More precisely, planarity asserts that it is possible to represent the graph in the plane in such a way that the vertices correspond to distinct points and the edges to simple Jordan curves connecting the points of its endvertices such that every two curves are either disjoint or meet only at a common endpoint.
An embedding of a planar graph determines a partition of the plane into regions. Exactly one of these regions is unbounded. The number of regions can be computed by the classical formula of Euler:

**Theorem 2.5.12** *Let $G = (V, E)$ be a connected and planar graph, and let $f$ denote the number of regions (including the single unbounded region) of an embedding of $G$ in the plane. Then*

$$|V| - |E| + f = 2. \tag{2.34}$$

Consequently, the number of regions is uniquely determined by the number of vertices and edges, i.e. by the combinatorial structure of the graph.

**Corollary 2.5.13** *Under the assumption of 2.5.12 it holds that*

$$|E| \leq 3|V| - 6 \text{ and} \tag{2.35}$$
$$f \leq 2|V| - 4. \tag{2.36}$$

- The graph is a grid [178].

Restricting $\mathcal{NP}$-hard algorithmic problems regarding arbitrary graphs to a smaller class of graphs will sometimes, yet not always, result in polynomially solvable problems.
For instance Steiner's Problem in graphs is polynomially solvable if any of the following conditions hold:

- The graph $G$ is planar and in addition all given points lie on the boundary of at most $m$ faces of the embedding of $G$ where $m$ is a number independent of the numbers of points [43], [231] or [356].
- The graph is strongly chordal; meaning that every cycle with four or more edges has a chord and every cycle with an even number of six or more edges has a chord dividing the cycle into two parts, each containing an odd number of edges [457].[35]
- The graph is a permutation graph [111].

Steiner's Problem in graphs can be solved in linear time if any of the following conditions hold:

- The graph is a series-parallel network [463].
- The graph is a Halin network. A Halin network is a graph formed by embedding a tree without degree-2 vertices into the plane and connecting its leaves by a cycle that crosses none of its edges [464].
- The graph is a partial 2-tree. Partial 2-trees are precisely those graphs which contain no subgraph homeomorphic to the complete graph with four vertices [440].
- The graph is a double tree [101].

Surveys on Steiner's Problem in graphs can be found in [231], [355], [439] and [464].

[35] This result cannot be extended to chordal graphs since then Steiner's Problem is $\mathcal{NP}$-complete [457].

http://www.springer.com/978-0-387-23538-7

Shortest Connectivity
An Introduction with Applications in Phylogeny
Cieslik, D.
2005, IX, 268 p., Hardcover
ISBN: 978-0-387-23538-7