Gradient-Constrained Minimum Networks, III. Fixed Topology

M. Brazil J.H. Rubinstein D.A. Thomas J.F. Weng
N. Wormald

Communicated by Reiner Burkard

Abstract

The gradient-constrained Steiner tree problem asks for a shortest total length network interconnecting a given set of points in 3-space, where the length of each edge of the network is determined by embedding it as a curve with absolute gradient no more than a given positive value m , and the network may contain additional nodes known as Steiner points. We study the problem for a fixed topology, and show that, apart from a few easily classified exceptions, if the positions of the Steiner points are such that the tree is not minimum for the given topology, then there exists a length reducing perturbation that moves exactly 1 or 2 Steiner points. In the conclusion, we discuss the application of this work to a heuristic algorithm for solving the global problem (across all topologies).

Key Words: Gradient Constraint; Steiner trees; minimum networks; optimization

Mathematics Subject Classification 2000: 05C05, 65K10

Research supported by the Australian Research Council, the Canadian Research Chairs Program and NSERC

M. Brazil (corresponding author)

Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia. Email: brazil@unimelb.edu.au

J.H. Rubinstein

Department of Mathematics and Statistics, The University of Melbourne, Victoria 3010, Australia

D.A. Thomas J.F. Weng

Department of Mechanical Engineering, The University of Melbourne, Victoria 3010, Australia

N. Wormald

Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, Canada N2L 3G1

1 Introduction

The Euclidean Steiner tree problem asks for a shortest total length network interconnecting a given point set N in Euclidean space. Such a network is necessarily a tree. In order to minimise length, it may include additional nodes not in N , called *Steiner points*. The points in N are called *terminals*. Each edge of the network can be thought of as a pair of nodes, with the length of an edge defined to be the length of a geodesic between its endpoints (which in the unconstrained case is simply a straight line segment). In 3-space, an embedded network is called *gradient-constrained* iff we restrict the embedding of each edge to a piecewise differentiable curve between the given endpoints such that the absolute gradient of any differentiable point is no more than a given positive value m . The length of an edge is then computed in the standard way. A *minimum gradient-constrained Steiner tree* T is a minimum gradient-constrained network interconnecting a given set of terminals [1]. For this paper, the set of given points is assumed to be finite. Higher dimensions could be considered, but here we treat only the 3-dimensional case.

An important application of the minimum gradient-constrained Steiner tree problem is in underground mine design. An underground mine can be viewed as a network of inclined tunnels which link the ore bodies, or their access points, to the surface portal or a vertical shaft. A constraint on this underground mine network is that it must be negotiable by large haulage trucks and drilling equipment. Hence there is a maximum gradient on the tunnels, generally about 1:7 (see [2]). Finding a minimum length network under such a gradient constraint corresponds (at least at a strategic level) to minimising the substantial mine infrastructure costs associated with the access tunnels.

In this paper, we build on the preliminary work on minimum gradient-constrained Steiner trees in [1] and [3], in order to study the problem of constructing such a Steiner tree for a fixed topology. Our main result shows that, apart from a few easily classified exceptions, if the positions of the Steiner points are such that the tree is not minimum for the given topology, then there exists a length reducing perturbation that moves exactly 1 or 2 Steiner points. This, along with the convexity of the problem for a fixed topology, means that there should exist a straight-forward descent method for solving the problem for any given topology. In the conclusion, we also discuss the application of this work to a heuristic algorithm for solving the global problem (across all topologies).

2 Preliminaries

A useful way of viewing gradient-constrained networks in 3-space is to treat them as general networks equipped with the so-called *gradient metric*, first introduced in [1]. Assuming that a maximum slope m , where $m < 1$, be given, the gradient metric for the distance between

two points is defined as the maximum of two metrics: the Euclidean metric and the *vertical metric*, which is $\sqrt{1 + m^{-2}}$ times the vertical distance between the two points. Note that the gradient metric is convex, since it is the maximum of two convex metrics.

The *topology* of a network is defined to be its underlying graph structure. We simplify the problem of finding a minimum gradient-constrained Steiner tree by restricting it to a given topology; that is, the locations of the terminals are given, the Steiner points are given in name but not in location, and all adjacencies between points are given. The goal is then to find a set of locations for the Steiner points that minimises the sum of distances (in the gradient metric) between all pairs of adjacent nodes. It is convenient to restrict our attention to topologies in which every Steiner point has degree 3, since that still allows for the possibility of higher degree Steiner points through degenerate edges of length zero.

It is well known that the Euclidean Steiner Problem (across all possible topologies) is NP-hard. This is proved for the planar case in [4], and this immediately implies the same result for higher dimensions and for the gradient-constrained case. For a fixed topology, the 2-dimensional Euclidean and gradient-constrained Steiner problems both have efficient polynomial-time solutions. In 3-dimensional space, the problem of solving the Euclidean Steiner Problem for a fixed topology is much more difficult. It has been shown in [5] that even on four terminals, a minimum Euclidean Steiner tree in 3-space cannot generally be solved by radicals. The same arguments also apply to minimum gradient-constrained Steiner trees in 3-space. This suggests that algorithmic methods are required to find good approximations of the solution for the general problem.

To solve any minimisation problem, one can attempt a descent method. This works well for the problem at hand if the gradient constraint is relaxed. In particular, it is well known that for the analogous problem in d -dimensional Euclidean space, the network is minimal iff there is no arbitrarily small perturbation of a single Steiner point that reduces the length of the network (see, for example, [6]). We express this by saying that, if the network is minimal with respect to one-point moves, then it is minimal. More briefly, we say that one-point moves suffice.

For the gradient-constrained minimum Steiner tree problem, given terminals and a topology, one can start with an initial non-optimal tree, and then iteratively move towards a global minimum by moving Steiner points. By the convexity of the gradient metric it follows that such a strategy will work, in theory, if we can determine the appropriate descent directions in which to move the Steiner points. This involves finding length-reducing perturbations of the positions of the Steiner points. One-point moves do not suffice for this problem, essentially due to the non-differentiability of the metric. This poses a difficulty algorithmically, and it helps to know which simultaneous multi-point moves might be required. One of our key results, in Section 5, shows that, apart from a few easily classified exceptions, one-point moves and two-point moves suffice; i.e., if the tree is not minimum for the given topology, then

there exists a length reducing perturbation that fixes all terminals, and perturbs only one or two Steiner points (Theorem 5.1). This makes the task of finding suitable length-reducing perturbations much simpler.

There are two situations where one-point moves or two-point moves may not suffice, i.e. where a consideration of all possible perturbations of one or two Steiner points are not sufficient for find a descent direction from a non-minimal configuration. The first is where the tree contains an *extreme point*, which we define to be a degree 3 Steiner point, whose incident edges all have the maximum permitted gradient and all lie in a single vertical plane. In this case, a special length-reducing perturbation involving arbitrarily many points may be required. However, we can precisely describe a limited number of such perturbations which need to be considered (as proved in Theorem 6.2). The other situation is where the tree degenerates to include zero-length edges. This case is very difficult to deal with in theory, but does not seem to cause a problem in practical applications. For this reason, we omit this case for most of the paper, and discuss it separately in Section 7.

In this paper, we consider only the minimisation of a gradient-constrained network with a given topology. It is, however, of great practical interest to find the minimum network joining a given set of nodes, without specifying the network topology, i.e. the minimum gradient-constrained Steiner tree. We have developed a software tool, UNO (Underground Network Optimiser), specifically for this purpose, which makes use of the theoretical results in the present paper. A description of an early version of UNO is given in [2], and a computational study using UNO is presented in [7].

3 Perturbations and Gradient-Constrained Steiner Trees

In this paper, we study Steiner trees in 3-space under the gradient metric. We assume the trees we deal with have a *full topology* (ie, a topology where all terminals are leaves and all Steiner points have degree 3 or more), since otherwise, we can decompose the problem into a number of smaller problems. Assuming that a maximum slope m , where $m < 1$, be given, the gradient metric defines a distance that correctly captures the lengths of edges in the network under a gradient constraint. It makes use of two metrics: the f-metric, which is the same as the Euclidean metric, and the b-metric, or vertical metric, which is $\sqrt{1 + m^{-2}}$ times the vertical distance between the two points. Edges in a gradient-constrained tree have different active metrics, depending on their Euclidean slope. If the edge has absolute slope greater than m , then the b-metric alone is active, and we label the edge “b”. If the edge has absolute slope less than m , then the f-metric alone is active, and we label the edge “f”. If both metrics are active, then the edge has slope m and can be labelled “m”. The gradient metric is convex

though it is not strictly convex.

A gradient-constrained tree T is said to be *minimal* iff its length cannot be reduced by any perturbation of its Steiner points. By the comments in the introduction, this is the minimum-length gradient-constrained Steiner tree with the topology of T . We need to be very precise about how we define perturbations, and the change of length under perturbations. A perturbation $\mathbf{v}$ on T is a collection of vectors in 3-dimensional Euclidean space, one for each Steiner point of T . We will consider smoothly moving all Steiner points with velocities specified by a given perturbation, and examine the resulting rate of change of the length of T .

Perturbations are used to examine arbitrarily small alterations in the positions of the Steiner points of T . We are interested in the change in the total length of T under such a perturbation. Note that the rate of change in length of T depends only on the directions and relative lengths of the vectors at the Steiner points — the length of each individual vector can be assumed to be arbitrarily small.

If T consists of a single edge uw , then we denote the length of T with respect to the gradient metric by $L(T) = L(uw)$, and we denote the derivative of this length with respect to a perturbation $\mathbf{v}$ by $\dot{L}(uw)_{\mathbf{v}}$. It is also useful to define the *active metric* for a perturbation. Suppose $L(uw) \neq 0$. If uw is a b-edge or an f-edge, then the active metric for $\mathbf{v}$ is the same as that of the edge. If uw is an m-edge, then the active metric for $\mathbf{v}$ is defined to be b iff the derivative of the gradient of uw is ≥ 0 , f iff the derivative of the gradient is ≤ 0 (and hence both b and f iff the derivative of the gradient equals 0). This is equivalent to saying that the active metric for the perturbation is the same as the metric active on the projection of the image of uw (after the perturbation has been applied) onto the vertical plane containing uw .

It is important to note that although a perturbation can make an active metric on an m-edge inactive, it cannot activate an inactive metric, since the perturbations are considered to be arbitrarily small.

For a larger tree T with edge set E , we define the derivative of $L(T)$ with respect to $\mathbf{v}$ to be

$$\dot{L}(T)_{\mathbf{v}} = \sum_{uw \in E} \dot{L}(uw)_{\mathbf{v}|_{\{u,w\}}},$$

where, as usual, $\mathbf{v}|_{\{u,w\}}$ is the perturbation that acts in the same way as $\mathbf{v}$ on u and w , and acts trivially on all other nodes of T . Since the terminals of a Steiner tree have fixed pre-determined positions, it is assumed that all perturbations act trivially on terminals.

Lemma 3.1. *Let T be a gradient-constrained Steiner tree with topology t . Then T is minimum with respect to t iff $\dot{L}(T)_{\mathbf{v}} \geq 0$ for every perturbation $\mathbf{v}$.*

Proof. The necessity of the condition that $\dot{L}(T)_{\mathbf{v}} \geq 0$ for every perturbation $\mathbf{v}$ in order for t to be minimum is immediate. This is also a sufficient condition since the gradient metric is convex [1]. $\square$

We now fix some of the basic definitions and notation required for our analysis. Most of this comes from or builds upon the notation in [1].

- For any point p , let $\mathcal{C}_p$ denote the cone generated by rotating a line through p with gradient m about the vertical line through p . So, $\mathcal{C}_p$ is a right vertical cone whose vertex is p .
- For any point p , let $\mathcal{H}_p$ denote the horizontal plane through p .
- For any m-edge pq , let $\mathcal{F}_{pq}$ denote the plane containing pq tangent to $\mathcal{C}_p$. We refer to $\mathcal{F}_{pq}$ as the *fundamental plane* of pq . The importance of the fundamental plane lies in the observation that a perturbation on pq that fixes p has active metric m iff q moves on $\mathcal{F}_{pq}$.
- All Steiner points have degree 3 in the topologies considered: zero length edges may be allowed, which may simulate points of degree 4 (if the zero length edge is between two Steiner points) or 2 (if the zero length edge goes to a leaf).
- If T is a tree and U is a subset of the nodes of T , then we define $T|_U$ to be the sub-forest of T induced by U . In other words, it is a forest whose set of nodes is U and whose edges are those edges of T which have both endpoints in U .
- Suppose s be a degree 3 Steiner point in T with two incident edges sa, sb lying on one side of $\mathcal{H}_s$, and the third sc lying on the other side of $\mathcal{H}_s$. (This includes the possibility that all three edges lie on $\mathcal{H}_s$.) Let g_a, g_b, g_c denote the respective labels of these edges. Then we say the *labelling* of this degree 3 Steiner point is $(g_a g_b / g_c)$.

We have shown in earlier papers that only a few different labellings can occur at a degree 3 Steiner point in a minimal tree.

Lemma 3.2. *Let T be a minimal gradient-constrained Steiner tree and s a degree 3 Steiner point of T . Then the labelling at s is (ff/f) , (ff/m) , (fm/m) , (mm/m) or (mm/b) . Furthermore, if the terminals of T lie in a vertical plane $\mathcal{P}$, then T lies in $\mathcal{P}$ and the labelling at s is (fm/m) , (mm/m) or (mm/b) .*

Proof. The general result is proved in [1], and the restriction to the vertical plane in [8]. $\square$

Note that, in the case of the vertical plane, the above labellings at a Steiner point allow us to directly compute its location in terms of the three neighbouring nodes. This follows from the fact that, in each case, there are at least two incident m-edges, which means that there are at most two possible locations for the Steiner point, the correct one of which is the location closest to the third neighbouring node.

For the general R^3 case, the following properties of minimal Steiner points will also prove useful.

Figure 1: A tree minimal with respect to 1-point perturbations but not with respect to 2-point perturbations.

Lemma 3.3. [1] *Let T be a minimal gradient-constrained Steiner tree and s a degree 3 Steiner point of T . Let v_1, v_2 and v_3 be the 3 nodes of T adjacent to s , and assume v_1s and v_2s be m -edges lying on opposite sides of $\mathcal{H}_s$.*

- (1) *If v_3s is an f -edge, then $v_3s \perp \mathcal{F}_{v_1s} \cap \mathcal{F}_{v_2s}$.*
- (2) *If v_1s and v_2s are collinear, then v_3s is an m -edge.*

4 Length-Reducing 1 and 2 Point

Perturbations.

Throughout this section and the two following, we assume T be a full gradient-constrained Steiner tree with no zero-length edges. In this section, we discuss conditions under which T is minimal with respect to 1-point and 2-point perturbations, ie, $\dot{L}(T)_{\mathbf{v}} \geq 0$ for all perturbations $\mathbf{v}$ that are zero on all nodes of the tree other than 1 or 2 Steiner points. The 1-point local minima for a single Steiner point were described and classified in our previous papers [1,3]. However, unlike the Euclidean case, the existence of a perturbation $\mathbf{v}$ on the Steiner points of T which decreases $L(T)$ does not imply that there exists a 1-point perturbation that decreases the length of T . This is clear from the example that follows.

Suppose T be the tree with terminals a, b, c, d , as shown in Figure 1. In the figure, we assume a, b, c, d all lie in a vertical plane, and that the edges labels be as shown. It is straightforward to confirm from Lemma 3.2 that T is minimal with respect to 1-point perturbations. In other words, no perturbation of s_1 or perturbation of s_2 can reduce the length of T . However, consider the 2-point perturbation $\mathbf{v}$, shown in the figure, where s_1 is perturbed along the line through s_1b away from b , s_2 is perturbed along the line through s_2d

towards d , and s_1s_2 remains an m-edge. Clearly, $\dot{L}(s_1b \cup s_1s_2 \cup s_2d)_{\mathbf{v}} = 0$, but it is easy to see that $\dot{L}(s_1a \cup s_2c)_{\mathbf{v}} < 0$, and hence that $\mathbf{v}$ is a length-reducing perturbation.

It is helpful to examine why $\mathbf{v}$ cannot be decomposed into 1-point perturbations on the original tree. Let $\mathbf{v}_1$ be the action of this perturbation on s_1 and $\mathbf{v}_2$ the action on s_2 . The question is why $\dot{L}(T)_{\mathbf{v}} \neq \dot{L}(T)_{\mathbf{v}_1} + \dot{L}(T)_{\mathbf{v}_2}$. The reason lies in the fact that the only active metric of s_1s_2 is the f-metric for $\mathbf{v}_1$ and the b-metric for $\mathbf{v}_2$.

If, in the above example, s_1s_2 had a common active metric for both $\mathbf{v}_1$ and $\mathbf{v}_2$, then there would exist a length-reducing 1-point perturbation. This would have occurred, for example, if the label of the edge s_1s_2 had been f or b.

This principle is generalised in the following lemma.

Lemma 4.1. *Let s_1s_2 be an interior edge of T , and let T_1 and T_2 be the two connected components of $T - s_1s_2$ (the tree T with the edge s_1s_2 removed) containing s_1 and s_2 , respectively. Let $\mathbf{v}$ be a perturbation on the Steiner points of T , and let $\mathbf{v}_1$ and $\mathbf{v}_2$ denote the induced perturbations on T_1 and T_2 respectively (ie, $\mathbf{v}_i = \mathbf{v}|_{V(T_i)}$). Suppose s_1s_2 has a common active metric for both $\mathbf{v}_1$ and $\mathbf{v}_2$. If $\mathbf{v}$ is a length-reducing perturbation for T , then so is either $\mathbf{v}_1$ or $\mathbf{v}_2$.*

Proof. We will show that

$$\dot{L}(T)_{\mathbf{v}} = \dot{L}(T)_{\mathbf{v}_1} + \dot{L}(T)_{\mathbf{v}_2}. \quad (1)$$

Then, if $\dot{L}(T)_{\mathbf{v}} < 0$, it follows that either $\dot{L}(T)_{\mathbf{v}_1} < 0$ or $\dot{L}(T)_{\mathbf{v}_2} < 0$, as required.

Suppose first that the common active metric for $\mathbf{v}_1$ and $\mathbf{v}_2$ on s_1s_2 be the f metric. From this it easily follows that the f metric is also active for $\mathbf{v}$ on s_1s_2 . To show that Equation (1) holds it suffices to consider the actions of the perturbations on s_1s_2 , since s_1s_2 is the only edge on which $\mathbf{v}_1$ and $\mathbf{v}_2$ both act in a non-trivial way. Let l be the line extending s_1s_2 . Then $\dot{L}(s_1s_2)_{\mathbf{v}}$, $\dot{L}(s_1s_2)_{\mathbf{v}_1}$ and $\dot{L}(s_1s_2)_{\mathbf{v}_2}$ are each equal to the derivative of length of the projection of s_1s_2 onto l under the corresponding perturbation (since the Euclidean metric always applies and the component perpendicular to l has derivative 0). Equation 1 then follows since $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ on l .

If, on the other hand, the common active metric for $\mathbf{v}_1$ and $\mathbf{v}_2$ on s_1s_2 is the b metric, then $L(s_1s_2)$ is proportional to the vertical distance between s_1 and s_2 (ie, the length of the projection onto the vertical axis), and so a similar argument applies. $\square$

In particular, the above decomposition can always be performed if s_1s_2 is a b-edge or f-edge.

Definition 4.1. *For a subtree T' of T , we say that T' is minimal with respect to its leaves iff T' is a minimal Steiner network interconnecting its leaves.*

Figure 2: The ellipses indicate the two subtrees T_1 and T_2 used for determining whether or not the edge s_1s_2 is decomposing with respect to s_2 .

Corollary 4.1. *Let s_1s_2 be an interior edge of T labelled f or b , and let T_1 and T_2 be the two connected components of $T - s_1s_2$ containing s_1 and s_2 respectively. If T is not minimal, then either $T|_{V(T_1) \cup \{s_2\}}$ or $T|_{V(T_2) \cup \{s_1\}}$ is not minimal with respect to its leaves.*

Proof. If s_1s_2 is labelled f or b , then, for any perturbation on T , the active metric on s_1s_2 is the f metric or b metric, respectively. Hence the result follows immediately from Lemma 4.1. $\square$

5 The Cases where 1 and 2 Point

Perturbations Suffice

The discussion in the previous section indicates that it is m -edges which cause difficulties when trying to decompose a length-reducing perturbation into a sum of 1-point perturbations, because of the possibility of active metrics being changed in the decomposition. However, if we include 2-point perturbations in the decomposition, the problem of m -edges can often be overcome. In this section, we examine the conditions under which such a decomposition is possible.

Throughout this section, let T be a full gradient-constrained Steiner tree with no zero-length edges. Suppose T has an interior node s_2 , whose three adjacent nodes are s_1 , v_1 and v_2 (each of which may or may not be a Steiner point). Let U_1 and U_2 be the two connected components of $T - s_1s_2$ containing s_1 and s_2 respectively. Let $T_1 = T|_{V(U_1) \cup \{s_2, v_1, v_2\}}$ and let $T_2 = T|_{V(U_2) \cup \{s_1\}}$. This is illustrated in Figure 2. Note that s_2 is an interior node of both

T_1 and T_2 . It is possible that a perturbation $\mathbf{v}$ may decrease $L(T)$ but increase both $L(T_1)$ and $L(T_2)$. For this reason, we study a special property of edges. We say that the edge s_1s_2 is *decomposing* with respect to s_2 iff, for any perturbation $\mathbf{v}$ acting on the interior nodes of T , there exist perturbations $\mathbf{v}_1$ and $\mathbf{v}_2$ acting on the interior nodes of T_1 and T_2 , respectively, such that

$$\dot{L}(T)_{\mathbf{v}} = \dot{L}(T)_{\mathbf{v}_1} + \dot{L}(T)_{\mathbf{v}_2}. \quad (2)$$

Furthermore, we say that s_1s_2 is *properly decomposing* with respect to s_2 iff s_1s_2 is decomposing with respect to s_2 , and $T_1 \neq T$ and $T_2 \neq T$ (ie, s_1 is not a terminal and at least one of v_1 and v_2 is not a terminal). Later in this paper, we will see how to use proper decomposability in showing that local minimality (with respect to 1-point and 2-point perturbations, for example) implies global minimality under certain conditions.

Lemma 5.1. *Let s_1s_2 be an f-edge or b-edge of T , where s_1 and s_2 are Steiner points. Then s_1s_2 is decomposing with respect to either endpoint. Furthermore, if T contains at least three Steiner points, then s_1s_2 is properly decomposing with respect to at least one of its endpoints.*

Proof. Let $\mathbf{v}$ be a perturbation on T . To see that s_1s_2 is decomposing with respect to s_2 , define T_1 and T_2 as above and let $\mathbf{v}_2$ be the induced perturbation on the interior nodes of T_2 and $\mathbf{v}_1$ the induced perturbation on the interior nodes of T_1 other than s_2 (which remains fixed). The first part of the result immediately follows from Corollary 4.1. The second statement is an easy consequence of the definition of ‘properly decomposing’. $\square$

The following lemma shows that, if all interior edges in T are properly decomposing, then only 1-point and 2-point perturbations need to be considered, when attempting to determine whether or not T is minimal.

Lemma 5.2. *Suppose every interior edge of T be properly decomposing with respect to at least one of its endpoints. If T is minimal with respect to all 1-point and 2-point perturbations, then T is minimal.*

Proof. Suppose, on the contrary, that T be not minimal. Then, by Lemma 3.1, there exists a perturbation $\mathbf{v}$ on T , such that $\dot{L}(T)_{\mathbf{v}} < 0$. We argue by induction on the number of Steiner points acted on (non-trivially) by $\mathbf{v}$. Suppose there be at least 3 such Steiner points. Then the condition that every interior edge of T is properly decomposing with respect to one of its endpoints guarantees that we can decompose $\mathbf{v}$ into perturbations $\mathbf{v}_1$ and $\mathbf{v}_2$, each acting on a smaller number of Steiner points, such that $\dot{L}(T)_{\mathbf{v}} = \dot{L}(T)_{\mathbf{v}_1} + \dot{L}(T)_{\mathbf{v}_2}$. One of these smaller perturbations must also decrease length. We conclude that there exists a length reducing 1-point or 2-point perturbation, giving the required contradiction. $\square$

We next examine conditions under which an edge of T is decomposing. If s is a Steiner point moved by a 1-point perturbation $\mathbf{v}$, we call $\mathbf{v}$ a *perturbation of s* .

We say that an ordered pair of adjacent Steiner points (s_1, s) in T is a *singular directed edge* iff there exists a perturbation $\mathbf{p}_1$ of s_1 such that there is no perturbation $\mathbf{p}$ of s for which the sum $\mathbf{p}_1 + \mathbf{p}$ preserves the active metrics of all edges incident with s .

Lemma 5.3. *Let T be minimal with respect to all 1-point perturbations. Let s_1s be an edge of T that is not decomposing with respect to s . Then (s_1, s) is a singular directed edge.*

Proof. Suppose, on the contrary, that (s_1, s) be not a singular directed edge. Let $\mathbf{v}$ be a perturbation of T acting non-trivially on s_1 and s (i.e, such that $\mathbf{v}|_{s_1} \neq \mathbf{0}$ and $\mathbf{v}|_s \neq \mathbf{0}$). Since s_1s is not decomposing, it follows from Lemma 5.1 that s_1s is an m-edge. Let $\mathbf{v}_1$ be the restriction of the perturbation $\mathbf{v}$ to the node set V_1 of the s_1 -component of $T - s_1s$. Since (s_1, s) is not a singular directed edge, $\mathbf{v}_1$ can be augmented to a perturbation $\mathbf{v}'_1$ which acts the same as $\mathbf{v}_1$ on V_1 , and, in addition, may perturb s (but no other nodes), such that all active metrics of edges incident with s are preserved. In particular, the active metric of $\mathbf{v}'_1$ on s_1s is m. Let $\mathbf{v}_2 = \mathbf{v} - \mathbf{v}'_1$. This perturbation fixes all nodes in V_1 ; it also retains the same metrics on the edge s_1s and on all edges of the s -component of $T - s_1s$ as are retained by $\mathbf{v}$. As in the proof of Lemma 4.1, $\dot{L}(T)_{\mathbf{v}} = \dot{L}(T)_{\mathbf{v}'_1} + \dot{L}(T)_{\mathbf{v}_2}$, and so s_1s is decomposing with respect to s , giving the required contradiction. $\square$

Define a degree 3 Steiner point s in T to be an *extreme point* iff s is labelled (mm/m), and the three edges incident with s all lie in a vertical plane.

Lemma 5.4. *Let T be minimal with respect to all 1-point perturbations. Let s be a Steiner point in T such that s is not extreme. Then each edge of T incident with s is decomposing with respect to s .*

Proof. Again, we argue by contradiction. Let s_1 be an adjacent node to s in T such that s_1s is not decomposing with respect to s . By Lemma 5.3, (s_1, s) is a singular directed edge, that is, there exists a perturbation of s_1 that forces one or more of the metrics active on the edges incident with s to change, no matter how s is perturbed. To prove the theorem, it now suffices to show that s is extreme.

Let $\mathbf{p}_1$ be a (non-zero) perturbation of s_1 , and $\mathbf{p}$ a perturbation of s . First note that the edge s_1s is an m-edge, by Lemma 5.1. We now consider three cases.

CASE 1: Suppose all three edges incident with s be m-edges, and v_1, s, v_2 be not collinear. Define $\mathbf{p}_1^s$ to be the perturbation $\mathbf{p}_1$ acting on s instead of s_1 . Observe that $\mathbf{p} + \mathbf{p}_1$ preserves the active metrics of all edges incident with s , contrary to (s_1, s) being a singular directed edge, if the following two conditions hold:

$$\mathbf{p} - \mathbf{p}_1^s \in \mathcal{F}_{s_1s}$$

and

$$\mathbf{p} \in \mathcal{F}_{v_1 s} \cap \mathcal{F}_{v_2 s} := L_1,$$

where L_1 is a line (the intersection of the two fundamental planes). Let L_2 be the line obtained by applying the perturbation $-\mathbf{p}_1$ to every point of L_1 . Note that L_2 is parallel to L_1 . Then the second condition can be written as

$$\mathbf{p} - \mathbf{p}_1^s \in L_2.$$

Hence, for (s_1, s) to be a singular directed edge, we require that

$$\begin{aligned} \mathcal{F}_{s_1 s} \cap L_2 &= \emptyset, \\ \Rightarrow L_1 &\parallel \mathcal{F}_{s_1 s}, \\ \Rightarrow L_1 &\subseteq \mathcal{F}_{s_1 s} \quad (\text{since } s \in L_1). \end{aligned}$$

If L_1 is not horizontal at s , then consider a perturbation $\mathbf{u} = \mathbf{p}$ or $-\mathbf{p}$ of s along L_1 . Since L_1 lies on each of the fundamental planes of the 3 edges incident with s , it follows that $|\dot{L}(ss_1)_{\mathbf{u}}| = |\dot{L}(sv_1)_{\mathbf{u}}| = |\dot{L}(sv_2)_{\mathbf{u}}| > 0$. Hence, for one of the two possible choices of $\mathbf{u}$, we have $\dot{L}(T)_{\mathbf{u}} = \dot{L}(ss_1)_{\mathbf{u}} + \dot{L}(sv_1)_{\mathbf{u}} + \dot{L}(sv_2)_{\mathbf{u}} < 0$, contradicting the minimality of T with respect to 1-point perturbations. Thus L_1 is horizontal, and it immediately follows that s is extreme.

CASE 2: Suppose all three edges incident with s be m-edges, and v_1, s, v_2 be collinear. In this case, we choose either $v_1 s$ or $v_2 s$ to be the non-decomposing edge, and then apply the argument in Case 1, since the other pair of incident edges will not be collinear.

CASE 3: Suppose that at least one of the edges $v_1 s$ and $v_2 s$ be not an m-edge. If neither $v_1 s$ nor $v_2 s$ is an m-edge then, applying the argument in Case 1, $\mathbf{p} - \mathbf{p}_1^s \in L_2$, where L_2 is all of $\mathbf{R}^3$. Hence, the requirement that $\mathcal{F}_{s_1 s} \cap L_2 = \emptyset$ necessarily fails, meaning that (s_1, s) cannot be a singular directed edge. On the other hand, if exactly one of these edges (say $v_2 s$) is an m-edge, then $\mathbf{p} - \mathbf{p}_1^s \in L_2$, where L_2 is a plane parallel to $\mathcal{F}_{v_2 s}$. But $v_2 s$ and $s_1 s$ are not collinear, by Lemma 3.3, hence L_2 is not parallel to $\mathcal{F}_{s_1 s}$. Again, the requirement that $\mathcal{F}_{s_1 s} \cap L_2 = \emptyset$ necessarily fails. $\square$

Theorem 5.1. *Let T be a full gradient-constrained Steiner tree, minimal with respect to 1-point and 2-point perturbations, and containing no extreme points or zero-length edges. Then T is minimum with respect to its topology.*

Proof. By Lemma 5.4, each interior edge of T is decomposing with respect to both endpoints. If T contains at least three Steiner points, then each interior edge e is incident with a Steiner point s_e which is adjacent to a third Steiner point (not an endpoint of e); thus e is properly decomposing with respect to s_e . Hence, by Lemma 5.2, T is minimal. On the

other hand, if T contains less than three Steiner points, then T is minimal by assumption. $\square$

Finally, in this section, we show that an extreme point still has at least one incident decomposing edge. Given an extreme point s of T , there exist two neighbours of s , say y and z , such that y , s and z are collinear. In this case, we call y and z the *collinear neighbours* of s .

Lemma 5.5. *Let s be an extreme point of T with collinear neighbours y and z . Let s_1 be the third neighbour of s . Then the edge s_1s is decomposing with respect to s .*

Proof. We show that (s_1, s) is not a singular directed edge. The result then follows from Lemma 5.3.

Let $\mathbf{p}_1$ be any perturbation of s_1 . Let $\mathbf{p}$ be the perturbation of s corresponding to the projection of $\mathbf{p}_1$ onto the line through y , s and z . It follows that $\mathbf{p} + \mathbf{p}_1$ preserves the active metrics of all edges incident with s , and hence (s_1, s) is not a singular directed edge. $\square$

Note: In the above lemma, if we add the conditions that s_1 is not a terminal and y and z are not both terminals, then it follows that the edge s_1s is properly decomposing with respect to s . This observation will be useful in the next section.

6 Extreme Point Arrangements and their Length-Reducing Perturbations

As before, let T be a full gradient-constrained Steiner tree containing no zero-length edges. Furthermore, for the remainder of this section, we assume T be minimal with respect to all 1-point and 2-point perturbations and contains no properly decomposing edges. (If there is a properly decomposing edge, then we can effectively reduce the size of the problem, by replacing T by a suitable subtree.) In this section, we define a class of easily identified perturbations we call backbone moves, and show that T is minimal iff none of these moves is length-reducing.

Let T_{ext} be the subnetwork of T induced by extreme points. In other words, T_{ext} is a forest whose nodes are the extreme points of T , and whose edges are the edges of T whose endpoints are both extreme. Note that each connected component of T_{ext} lies in a vertical plane.

We now study the structure of these components of T_{ext} .

Lemma 6.1. *If two extreme points of T are adjacent, then they cannot both have two neighbours below their horizontal planes.*

Proof. Up to reflection there are two cases, as shown in Figure 3. In each case, let s_1

Figure 3: Two non-minimal configurations for a pair of adjacent extreme points. All edges shown are m -edges lying on a vertical plane.

and s_2 be the two adjacent extreme points where, s_2 lies below s_1 . Let v_2 be the neighbour of s_2 that is not collinear with s_1 and s_2 , and let v_1 be the neighbour below s_1 other than s_2 . Neither of the two possibilities are minimal since in each case the 2-point perturbation, whose action on the two extreme points is a pair of equal length (and parallel) perturbations pointing towards v_1 and v_2 , respectively (as shown in the figure), is length-reducing. $\square$

Let s be an extreme point with collinear neighbours y and z . Let x be the third neighbour of s . It follows from Lemma 5.5 and the assumption that no edges of T are properly decomposing, that if x is extreme, then y and z are terminals. This implies that each extreme point of degree greater than 1 in T_{ext} has degree exactly 2 in T_{ext} , and that its neighbours in T_{ext} are the collinear ones. It follows immediately that all components of T_{ext} are paths whose nodes are collinear. Furthermore, the non-collinear neighbour of each extreme point must be a terminal (by the note following Lemma 5.5).

With these observations, we have the following simple corollary to Lemma 6.1.

Lemma 6.2. *Each connected component of T_{ext} is a path of collinear nodes in a vertical plane. Furthermore, the nodes in T adjacent to such a path but non-collinear with the path are all terminals and are located on alternate sides of the path.*

Definition 6.1. *Let P be a connected component of T_{ext} with node set $V(P)$, containing $k - 2$ nodes ($k \geq 3$). By Lemma 6.2, P is a path in a vertical plane $\mathcal{P}$ whose nodes are collinear. Let s_1 and s_k be the two nodes of T not in P that are neighbours of the two endpoints of P and are collinear with P (or are the collinear neighbours of the single node of P if $k - 2 = 1$). We call $T|_{V(P) \cup \{s_1, s_k\}}$ the backbone of P (ie, the backbone is the union of P and the two extra collinear edges of T incident with the endpoints of P). Furthermore, a k -point backbone move is defined to be a perturbation of $V(P) \cup \{s_1, s_k\}$ restricted to $\mathcal{P}$ such*

that the active metric on all edges in, or incident with, P remains m ; each node of $V(P)$ is perturbed an appropriate relative distance directly towards or away from its neighbouring terminals.

Note that a backbone remains collinear, if a k -point backbone move is applied. Furthermore, a k -point backbone move is determined entirely by its action on s_1 and s_k . Finding a length-reducing k -point backbone move is no more complex than finding a length-reducing 2-point perturbation.

In the following theorem, we show that there are situations in which k -point backbone moves are required to reduce the length of T .

Theorem 6.1. *For arbitrarily large $k \geq 3$, there exists a full gradient-constrained Steiner tree T_k such that*

1. T_k contains no zero-length edges;
2. T_k is minimal with respect to all $(k-1)$ -point perturbations;
3. no edges of T_k are properly decomposing; and
4. $\dot{L}(T_k)_{\mathbf{v}} < 0$ for some k -point backbone move $\mathbf{v}$.

Proof. Here we construct a specific full gradient-constrained Steiner tree T_3 and outline a proof showing that this tree satisfies the conditions of the theorem for $k = 3$. Similar methods can be used to construct other trees satisfying the theorem for higher values of k —this is briefly discussed at the end of the proof.

Consider a gradient-constrained Steiner tree T_3 with 5 terminals and 3 Steiner points, whose projection onto a vertical plane $\mathcal{P}$ is illustrated in Figure 4 (a). All edges of T_3 are m -edges. The Steiner points of T_3 are s_1, s_2, s_3 . Node s_2 is an extreme point with collinear neighbours s_1 and s_3 ; its other neighbour (to the left) is the terminal t , and $s_1, s_2, s_3, t \in \mathcal{P}$.

Node s_1 has two neighbouring terminals u_1, u_2 . These terminals do not lie on $\mathcal{P}$, but are symmetrically positioned with respect to $\mathcal{P}$, ie, they are equidistant from $\mathcal{P}$ and both project onto the same point $u \in \mathcal{P}$. Furthermore, $s_1 u$ is vertical and hence $\mathcal{F}_{u_1 s_1} \cap \mathcal{F}_{u_2 s_1} := L_1$ is a horizontal line in $\mathcal{P}$.

Similarly, node s_3 has two neighbouring terminals v_1, v_2 . Again, these terminals are symmetrically positioned with respect to $\mathcal{P}$ and both project onto a point $v \in \mathcal{P}$. In this case, however, v lies to the right of the vertical line through s_3 . The positions of terminals v_1 and v_2 are chosen so that the line in $\mathcal{P}$, $\mathcal{F}_{v_1 s_3} \cap \mathcal{F}_{v_2 s_3} := L_3$, has slope

$$\text{grad}(L_3) = \frac{1 + \varepsilon}{3 + \varepsilon} m \quad (3)$$

for some positive $\varepsilon \ll 1$.

By the symmetric positioning of the terminals around $\mathcal{P}$, it follows that if there is a length-reducing perturbation on T_3 , then its projection onto $\mathcal{P}$ is also length-reducing. Hence

Figure 4: (a) The tree T_3 projected onto a vertical plane; (b) The action of $\mathbf{v}$ on s_1 and s_2 , as described in Case 1.

we can restrict our consideration to perturbations that keep the Steiner points in $\mathcal{P}$. We can also assume that a length-reducing perturbation $\mathbf{v}$ moves s_2 along the line L_2 through ts_2 since the optimal way for t to join onto the backbone is via an edge of gradient m . Similarly, we can assume that $\mathbf{v}$ moves s_1 along L_1 and s_3 along L_3 .

Since s_2 is extreme, it follows that there are no length-reducing 1-point perturbations on s_2 . Hence, when looking for a 1 or 2-point length-reducing perturbation $\mathbf{v}$, we only need to consider perturbations that act non-trivially on either s_1 or s_3 . This leads to 2 cases.

CASE 1: $\mathbf{v}$ acts non-trivially on s_1 and fixes s_3 .

First note that, if $\mathbf{v}$ is 1-point perturbation on s_1 along L_1 , then it is not length-reducing since $\dot{L}(u_1s_1)_{\mathbf{v}} = \dot{L}(u_2s_1)_{\mathbf{v}} = 0$, and $\dot{L}(s_2s_1)_{\mathbf{v}} = 0$ if $\mathbf{v}$ is a perturbation to the right (since s_2s_1 becomes a b-edge), whereas $\dot{L}(s_2s_1)_{\mathbf{v}} > 0$ if $\mathbf{v}$ is a perturbation to the left.

If s_2 is not fixed, then it is easy to see that it must move towards t , and we can assume that, under the action of $\mathbf{v}$, s_1s_2 remains an m-edge (otherwise $\mathbf{v}$ could be decomposed). This is illustrated in Figure 4 (b). It follows that $\dot{L}(T_3)_{\mathbf{v}} = \dot{L}(s_3s_2)_{\mathbf{v}} > 0$ (since $m < 1$). Hence, there are no length-reducing perturbations for Case 1.

CASE 2: $\mathbf{v}$ acts non-trivially on s_3 and fixes s_1 .

We first consider 1-point perturbations; i.e., we assume that $\mathbf{v}$ fixes s_2 . If $\mathbf{v}$ perturbs s_3 along L_3 to the left, then s_2s_3 becomes a b-edge and $\dot{L}(v_1s_3 \cup s_2s_3)_{\mathbf{v}} = 0$; hence $\dot{L}(T_3)_{\mathbf{v}} = \dot{L}(v_2s_3)_{\mathbf{v}} > 0$. If $\mathbf{v}$ perturbs s_3 along L_3 to the right, say to a point s'_3 , then we have the situation illustrated in Figure 5(a). By choosing suitable units, we can assume the increase in height of s_3 under the perturbation be $1+\varepsilon$. Let p be the projection of s_3 onto $s_2s'_3$. Since the gradient of $s_2s'_3$ is arbitrarily close to m , we can assume that $L(s_2s_3) = L(s_2p)$.

Figure 5: Two actions of $\mathbf{v}$ on: (a) s_3 ; and (b) s_3 and s_2 , as described in Case 2.

Also, using Equation (3), it is straightforward to show that the vertical distance between p and s_3' is $2 + \cos(2\alpha) + \varepsilon$, where $\alpha = \arctan(m)$. Hence the increase in length of s_2s_3 under the action of $\mathbf{v}$ is $(2 + \cos(2\alpha) + \varepsilon) \csc \alpha$. On the other hand, v_1s_3 and v_2s_3 each decrease in length by $(1 + \varepsilon) \csc \alpha$. Hence the overall increase is $(\cos(2\alpha) - \varepsilon) \csc \alpha > 0$ since $\varepsilon < \cos(2\alpha)$. Hence $\dot{L}(T_3)_{\mathbf{v}} > 0$.

We now consider 2-point perturbations that act non-trivially on s_2 and s_3 . As previously, we can assume that $\mathbf{v}$ moves s_2 along L_2 and s_3 along L_3 , and that s_2s_3 remains an m-edge under the action of $\mathbf{v}$. If $\mathbf{v}$ perturbs s_3 to the right, then s_2 also moves to the right, and it is easy to see that $L(T_3)$ increases at an even greater rate than under the previous 1-point perturbation. If $\mathbf{v}$ perturbs s_2 and s_3 to the left, as in Figure 5(b), then, if the decrease in height of s_3 under $\mathbf{v}$ is $1 + \varepsilon$ (via a suitable choice of units), it follows from Equation (3) that the increase in height of s_2 is 1. Hence, $L(s_2s_3)$ decreases under the action of $\mathbf{v}$ by $(2 + \varepsilon) \csc \alpha$. Under the action of $\mathbf{v}$, s_1s_2 becomes a b-edge, and hence the combined length of s_1s_2 and ts_2 does not change. The remaining edges, v_1s_3 and v_2s_3 , each increase in length by $(1 + \varepsilon) \csc \alpha$, and thus there is an overall increase in length of $\varepsilon \csc \alpha$. So, for Case 2, there are no length-reducing perturbations.

Finally, we show that there is a length reducing 3-point backbone move $\mathbf{v}$ for T_3 . Let $\mathbf{v}$ perturb s_1, s_2 and s_3 to the right (along L_1, L_2 and L_3 , respectively) so that the internal edges remain m-edges. Note that $\dot{L}(v_1s_3 \cup s_2s_3 \cup s_1s_2 \cup u_1s_1 \cup u_2s_1)_{\mathbf{v}} = 0$, so we only need to consider the change in length of v_2s_3 and ts_2 . If the height of s_3 increases by $1 + \varepsilon$, then, as in the previous case (again, refer to Figure 5 (b)), the height of s_2 decreases by 1, giving an overall decrease in length of $v_2s_3 \cup ts_2$ by $\varepsilon \csc \alpha$. Hence $\dot{L}(T_3)_{\mathbf{v}} < 0$, completing the proof of the theorem for $k = 3$.

The above tree T_3 can be generalised to a tree T_k , for any odd k , by adding an even number of extra extreme points to the backbone. We claim that it is possible,

but tedious, to show that the resulting tree is minimal with respect to $(k - 1)$ -point perturbations, but not with respect to k -point backbone moves. Finally, a similar construction also applies to trees T_k , where k is even, by generalising suitable 2-point perturbations. $\square$

In relation to the previous Theorem, observe that, if the terminals are in general position, then no three terminals lie in a vertical plane, and hence no 5-point or higher backbone moves are required.

We now prove our main result — that for any non-minimal tree there exists either a 1-point or 2-point perturbation or a k -point backbone move that is length-reducing.

Theorem 6.2. *Let T be a gradient-constrained Steiner tree, minimal with respect to 1-point and 2-point perturbations. If T is not minimal, then there exists a length-reducing k -point backbone move on T , for some $k \geq 3$.*

Proof. By Lemma 5.4, any edge not incident with an extreme point is a decomposing edge. So, if there exists a length reducing perturbation on T , then we can find one that, for some component P of T_{ext} , fixes all nodes not in or adjacent to P . Let $\mathbf{v}$ be such a perturbation. Thus at most two non-extreme Steiner points move under the action of $\mathbf{v}$: the ones adjacent to the endpoints of P and collinear with P (one above and one below) — see Lemma 6.2.

Suppose s be an extreme point upon which $\mathbf{v}$ acts non-trivially. We may assume that s moves within the vertical plane π containing its incident edges. To see this, we make the following observations. Let e be an edge of T incident with s . Consider a perturbation on the endpoints of e . Without loss of generality, we can assume that one end of e is fixed under the perturbation $\mathbf{v}$. (If $\mathbf{u}_1$ and $\mathbf{u}_2$ are the actions of $\mathbf{v}$ on the two ends of e respectively, then apply $\mathbf{u}_2 - \mathbf{u}_1$ to the second end only.) Hence, $\dot{L}(e)_{\mathbf{v}}$ is the sum of $\dot{L}(e)_{\mathbf{v}_1}$ and $\dot{L}(e)_{\mathbf{v}_2}$ (both measured in the active metric of e under $\mathbf{v}$), where $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}$ and $\mathbf{v}_1$ is the component of $\mathbf{v}$ in π ; clearly, $\dot{L}(e)_{\mathbf{v}_2} = 0$. It now follows that it suffices to consider the projection of the perturbation on s in the plane π .

Now, consider the m -edges incident to nodes of P but not lying on the backbone. By Lemma 6.2, each of these edges lies between the backbone and a terminal. Since the backbone remains almost at gradient m and $m < 1$, it is easy to see that the optimal way for the terminals off the backbone to join onto it is with gradient m . So we may assume these edges remain ‘ m ’ in the perturbation $\mathbf{v}$.

From the above discussion, we can assume $\mathbf{v}$ fixes all extreme points except those in one component P of T_{ext} and does not move any nodes of P off P ’s vertical plane. We assume, furthermore, that $\mathbf{v}$ moves the minimum number of points of any length-reducing perturbation and that, subject to this, it has active metric m on the maximum number of edges. Let R denote the set of points moved by $\mathbf{v}$.

Now, suppose there be an edge $e = s_1 s_2$ joining two extreme points, such that the active

metric of $\mathbf{v}$ for that edge be either b or f . We will show that this leads to a contradiction. Define $\mathbf{v}_1$ to be the restriction of $\mathbf{v}$ to those points in the same component of $T_R - e$ as s_1 , and define $\mathbf{v}_2$ to be a similar restriction for s_2 . Then either $\mathbf{v}_1$ or $\mathbf{v}_2$, if acting alone, must give e the same active metric as $\mathbf{v}$. Without loss of generality, let it be $\mathbf{v}_1$, and assume the active metric be b (the same argument applies for f). If $\mathbf{v}_2$ also has active metric b on e , then e is decomposing by Lemma 4.1 — hence either $\mathbf{v}_1$ or $\mathbf{v}_1$ must be length-reducing (contradicting the minimality assumption for $\mathbf{v}$). So $\mathbf{v}_2$ has active metric f or m on e . Thus there is a perturbation $\theta\mathbf{v}_1 + \mathbf{v}_2$, for some $0 \leq \theta < 1$, for which the active metric on e is m . Call this $\mathbf{u}_1$. Then let $\mathbf{u}_2 = (1 - \theta)\mathbf{v}_1$. We have $\mathbf{u}_1 + \mathbf{u}_2 = \mathbf{v}$, which is length-reducing. Thus $\dot{L}(T)_{\mathbf{u}_1} + \dot{L}(T)_{\mathbf{u}_2} = \dot{L}(T)_{\mathbf{v}}$, since the only reason this would not be true is if the active metric for $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{v}$ on e were incompatible, but the active metric is m for $\mathbf{u}_1$ and b for both $\mathbf{u}_2$ and $\mathbf{v}$. Hence either $\mathbf{u}_1$ or $\mathbf{u}_2$ is length-reducing. But $\mathbf{u}_1$ cannot be length-reducing since it has active metric m for more edges than $\mathbf{v}$ (contradicting minimality), while $\mathbf{u}_2$ cannot be length-reducing because it moves fewer points than $\mathbf{v}$. The conclusion is that all edges in P have active metric m in the perturbation $\mathbf{v}$.

Thus, if the tree is minimal with respect to 1 and 2 point moves (and has no zero length edge), the only length reducing perturbations will move a component of T_{ext} , together with the two adjacent Steiner points, in a move that maintains the m slope of all backbone edges and edges to terminals and keeps all moved nodes in the one vertical plane. Each such perturbation is determined entirely by its effect on the bottom and top nodes in the backbone (and furthermore these must move to maintain slope m). $\square$

7 Conclusions

For the main theorems in our paper, we have restricted our results to networks which have no edges of length zero. There are, however, a number of reasons why it is sensible to permit zero length edges for a given topology: it ensures a minimum network exists for each topology; it is necessary to guarantee compactness of the configuration space for a fixed topology; and it means that all topologies can essentially be dealt with by considering only full topologies. For instance, some minimal Steiner trees have a Steiner vertex of degree 4, and this can be considered instead as a degenerate example of a Steiner tree whose topology has all Steiner points of degree 3, but two such points have an edge of length zero joining them.

Unfortunately, if we relax the restriction on no zero length edges, the main results of this paper no longer hold. This is illustrated by the following example.

Consider a gradient-constrained tree T in the vertical plane with terminals t_1, t_2, t_3, t_4 and t_5 , as shown in Figure 6. We assume T contains two Steiner points s_1 and s_2 such that $s_1 t_2$ be an f -edge sloping downwards from s_1 , $s_2 t_4$ be an f -edge sloping upwards from s_2 , and

Figure 6: A degenerate tree minimal with respect to 1-point and 2-point perturbations but not with respect to 3-point perturbations.

the remaining edges be m-edges. Now, suppose we view T as a full (but degenerate) tree containing a third Steiner point s adjacent to s_1 , s_2 and t_3 , such that st_3 be a zero-length edge. It is easy to see that T is minimal with respect to 1 and 2 point moves. However, there is a 3 point move, moving s_1 along the line through s_1t_1 away from t_1 , moving s_2 along the line through s_2t_5 away from t_5 , and moving s away from t_3 so that s_1s and s_2s remain m-edges that reduces the length of T as the f-edges come closer to being parallel. This move is indicated in the figure in broken lines.

This example does not rely on s being extreme. It is possible to move some of the terminals out of the vertical plane, without changing the edge labels, so that this problem persists.

The above example has the interesting property that although it is minimal with respect to 1 and 2 point moves, there nevertheless exists a perturbation $\mathbf{v}$ which moves only s that allows the tree to ‘escape’ from this configuration. In other words, applying $\mathbf{v}$ leads to new 1 or 2 point perturbations that make the (formerly) zero length edge even longer while moving towards the minimum tree. This is potentially a useful step for a descent algorithm. However, after applying $\mathbf{v}$, some length-reducing moves shorten the same edge (for example, applying a certain 1 point perturbation to s), so a descent algorithm using $\mathbf{v}$ to resolve this ‘deadlock’ could potentially cycle.

In a similar way, it is possible to construct examples with two Steiner points coinciding, which are minimal with respect to 1 and 2 point moves, but can be reduced in length by moving more than two Steiner points. For example, take the tree T in Figure 6 and its reflection

through a vertical axis, and join the two trees together by a zero length edge connecting t_3 and its counterpart in the reflected tree (both of which are now treated as Steiner points). Now adjust the terminals incident to the f -edges so that all f -edges are sloping upwards from their incident Steiner points, but with different slopes on the two sides of the tree. Under this fixed topology, the tree is minimal with respect to 1 and 2 point moves, but is not minimal – there is a move of all six Steiner points that reduces the length of the tree. It is interesting to note, however, that, for this example, there are 2 points moves that reduce the length of the tree, but change the topology of the tree (by splitting the merged Steiner points a different way).

Given the examples above, it appears that for the case of zero length edges, characterising the constraints on Steiner trees that are minimal with respect to the gradient metric, and having a given topology, becomes a much more difficult problem. However, this is not necessarily a major setback for a sensible algorithm that seeks to find the minimum gradient-constrained Steiner tree, such as UNO. The problem of finding the best topology is in general NP-complete, since this is true for the same problem without the gradient constraint. So, any practical algorithm requires a method of changing topologies, or at least considering many different topologies, that is unlikely to be highly efficient. UNO incorporates this aspect by including some random steps. By running for some time and repeatedly randomly encountering approximately minimal trees, the algorithm occasionally finds trees that are reasonably close to the minimal tree, even in the case of the first example above. A sensible descent method (using the results on this paper characterising minima with no zero length edges) then brings the algorithm even closer to the true global minimum. Thus, the characterisation of local minima described in this paper is enough for the algorithm to perform well in practical cases. The problem of finding a ‘good’ topology, for instance the topology of the minimum tree, is not the subject of this paper.

References

1. Brazil, M., Rubinstein, J.H., Thomas, D.A., Weng, J.F., Wormald, N.C.: Gradient-Constrained Minimum Networks, I. Fundamentals. *J. Global Optimisation* **21**, 139-155 (2001).
2. Brazil, M., Rubinstein, J.H., Thomas, D.A., Lee, D., Weng, J.F., Wormald, N.C.: Network optimization of underground mine design. *The Australasian Institute for Mining and Metallurgy Proc.* **305**, 57-65 (2000).
3. Brazil, M., Thomas, D.A., Weng, J.F.: Gradient-constrained minimum networks. II. Labelled or locally minimal Steiner points. *J. Global Optimization* **42**, 23-37 (2008).
4. Garey, M.R., Graham, R.L., Johnson, D.S.: The complexity of computing Steiner minimal trees. *SIAM Journal on Applied Mathematics* **32**, 835-859 (1977).

5. Rubinstein, J.H., Thomas, D.A., Weng, J.F.: Minimum networks for four points in space. *Geometriae Dedicata* **93**, 57-70 (2002).
6. Smith, W.D.: How to find Steiner minimal trees in Euclidean d -space. *Algorithmica* **7**, 137-177 (1992).
7. Thomas, D.A., Brazil, M., Lee D., Wormald, N.C.: Network modelling of underground mine layout: two case studies. *International Transactions in Operational Research* **14**, 143-158 (2007).
8. Brazil, M., Thomas, D.A., Weng, J.F.: Gradient-constrained minimal Steiner trees. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **40**, 23-38 (1998).

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Brazil, M;Rubinstein, JH;Thomas, DA;Weng, JF;Wormald, N

Title:

Gradient-Constrained Minimum Networks. III. Fixed Topology

Date:

2012-10-01

Citation:

Brazil, M., Rubinstein, J. H., Thomas, D. A., Weng, J. F. & Wormald, N. (2012). Gradient-Constrained Minimum Networks. III. Fixed Topology. JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS, 155 (1), pp.336-354. <https://doi.org/10.1007/s10957-012-0036-3>.

Persistent Link:

<http://hdl.handle.net/11343/283012>