

# Simplification of Polyline Bundles

Joachim Spoerhase 

Aalto University, Finland  
University of Würzburg, Germany  
joachim.spoerhase@aalto.fi

Sabine Storandt

University of Konstanz, Germany  
sabine.storandt@uni-konstanz.de

Johannes Zink 

University of Würzburg, Germany  
zink@informatik.uni-wuerzburg.de

---

## Abstract

We propose and study a generalization to the well-known problem of polyline simplification. Instead of a single polyline, we are given a set of  $\ell$  polylines possibly sharing some line segments and bend points. Our goal is to minimize the number of bend points in the simplified bundle with respect to some error tolerance  $\delta$  (measuring Fréchet distance) but under the additional constraint that shared parts have to be simplified consistently. We show that polyline bundle simplification is *NP*-hard to approximate within a factor  $n^{\frac{1}{3}-\varepsilon}$  for any  $\varepsilon > 0$  where  $n$  is the number of bend points in the polyline bundle. This inapproximability even applies to instances with only  $\ell = 2$  polylines. However, we identify the sensitivity of the solution to the choice of  $\delta$  as a reason for this strong inapproximability. In particular, we prove that if we allow  $\delta$  to be exceeded by a factor of 2 in our solution, we can find a simplified polyline bundle with no more than  $\mathcal{O}(\log(\ell + n)) \cdot OPT$  bend points in polytime, providing us with an efficient bi-criteria approximation. As a further result, we show fixed-parameter tractability in the number of shared bend points.

**2012 ACM Subject Classification** Theory of computation  $\rightarrow$  Approximation algorithms analysis; Theory of computation  $\rightarrow$  Computational geometry; Theory of computation  $\rightarrow$  Problems, reductions and completeness; Theory of computation  $\rightarrow$  Fixed parameter tractability

**Keywords and phrases** Polyline Simplification, Bi-criteria Approximation, Hardness of Approximation, Geometric Set Cover

**Digital Object Identifier** 10.4230/LIPIcs.SWAT.2020.35

**Funding** *Joachim Spoerhase*: Funded by European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 759557).

*Sabine Storandt*: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 50974019 – TRR 161.

## 1 Introduction

Visualization of geographical information is a task of high practical relevance, e.g., for the creation of online maps. Such maps are most helpful if the information is neatly displayed and can be grasped quickly and unambiguously. This means that the full data often needs to be filtered and abstracted. Many important elements in maps like borders, streets, rivers, or trajectories are displayed as polylines (also known as polygonal chains). For such a polyline, a simplification is supposed to be as sparse as possible and as close to the original as necessary.

A simplified polyline is usually constructed by a subset of bend points of the original polyline such that the (local) distance to the original polyline does not exceed a specifiable value according to a given distance measure, e.g., Fréchet distance or the Hausdorff distance.



© Joachim Spoerhase, Sabine Storandt, and Johannes Zink;  
licensed under Creative Commons License CC-BY

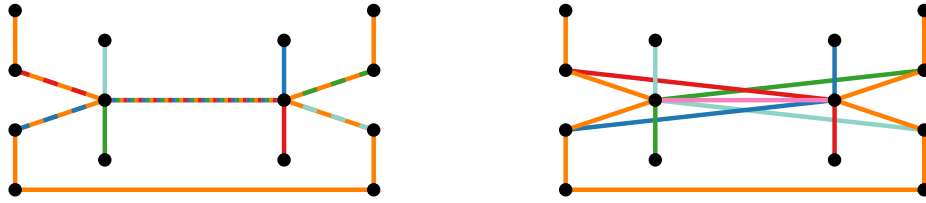
17th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2020).

Editor: Susanne Albers; Article No. 35; pp. 35:1–35:20

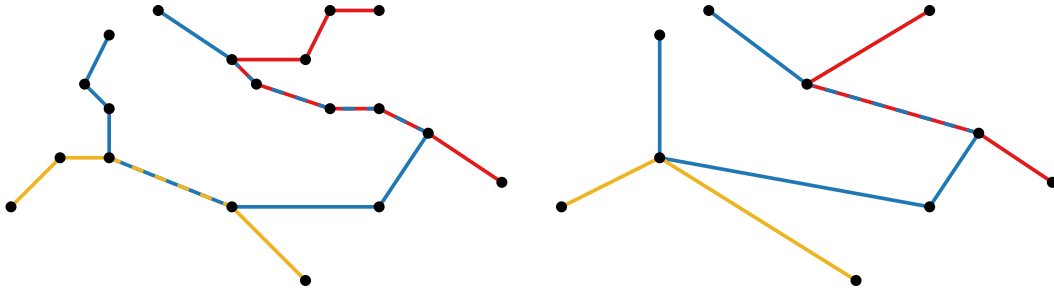
Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** Example where the total complexity increases if each polyline is simplified independently. Left: Initial bundle of polylines. Right: Bundle of independently simplified polylines.



■ **Figure 2** Example of a bundle of three polylines before and after consistent simplification.

The first such algorithm, which is still of high practical importance, was proposed by Ramer [16] and by Douglas and Peucker [7]. Hershberger and Snoeyink [13] proposed an implementation of this algorithm that runs in  $\mathcal{O}(n \log n)$  time, where  $n$  is the number of bend points in the polyline. It is a heuristic algorithm as it does not guarantee optimality (or something close to it) in terms of retained bend points. An optimal algorithm in this sense was first proposed by Imai and Iri [14]. Chan and Chin [5] improved the running time of this algorithm to  $\mathcal{O}(n^2)$  for the Hausdorff distance. For the Fréchet distance, the optimal solution can be determined in time  $\mathcal{O}(n^3)$  as described by Godau [10].

We remark that all of these algorithms consider the distance *segment-wise*. This is, the distance between each segment of the simplification and its corresponding sub-polyline of the input polyline does not exceed the given threshold. We adhere to this widespread approach. Intuitively and from an application point of view, it makes sense to map a point  $p$  of the input polyline only to a point of a segment of the simplification “spanning over”  $p$  with respect to the input polyline as this ensures a certain degree of locality. However, the general unrestricted approach has also received attention in the literature. Here, the Hausdorff or Fréchet distance between the input polyline and the simplification as a whole polyline is considered. For the (undirected) Hausdorff distance, this problem becomes *NP-hard* [17] and for the Fréchet distance, there is an  $\mathcal{O}(kn^5)$  time algorithm, where  $k$  is the output complexity of the simplification [17]. The problem variant where in addition the requirement is dropped that all bend points of the simplification must be bend points of the input polyline, is called a *weak* simplification. Agarwal et al. [1] show that an optimal simplification under the segment-wise Fréchet distance with distance threshold  $\delta$ , as computable using the algorithm by Imai and Iri, has no more bend points than an optimal weak simplification with distance threshold  $\delta/4$ . We note that computing the Fréchet distance between two polylines can be solved in polynomial time [2], but may become *NP-hard* when considering additional properties like allowing to take shortcuts, which replace outliers in one of the polylines [3].

## From a Single Polyline to a Bundle of Polylines

On a map, there are usually multiple polylines to display. Such polylines may share bend points (bends) and line segments between bends (segments) sectionwise. We call them a *bundle* of polylines. One example is a schematic map of a public transport network where bus lines are the polylines and these share some of the stations and legs.

One might consider simplifying the polylines of a bundle independently. This has some drawbacks, though. On the one hand, the total complexity might even increase when the shared parts are simplified in different ways; see Figure 1. On the other hand, it might suggest a misleading picture when we remove common segments and bends of some polylines, but not of all. Therefore, we require that a bend in a simplification of a bundle of polylines is either kept in all polylines containing it or discarded in all polylines. Our goal is then to minimize the total number of bend points that have to be kept. In Figure 2, we give an example of a simplification of a bundle of polylines.

## Related Work

Polyline bundles were studied before in different contexts. In [4], the goal is to interfere a concise graph which represents all trajectories in a given bundle sufficiently well. But this approach primarily aims at retrieving split and merge points of trajectories correctly and does not produce a simplification of each trajectory in the bundle. Methods for map generation based on movement trajectories [12] have a similar scope but explicitly allow to discard outliers and to unify sufficiently similar trajectories, which is not allowed in our setting.

Agarwal et al. [1] describe an  $\mathcal{O}(n \log n)$  time approximation algorithm for (classical) polyline simplification under the Fréchet distance. It is an approximation algorithm in the sense that the output simplification for distance threshold  $\delta$  has at most as many bends as an optimal solution with distance threshold  $\delta/2$ . In Theorem 10, we also relate the size of our approximate solution respecting a distance threshold of  $\delta$  to an optimal solution with distance threshold  $\delta/2$ .

There is also a multitude of polyline simplification problem variants for single polylines which involve additional constraints. One important variant is the computation of the smallest possible simplification of a single polyline which avoids self-intersection [6]. Another practically relevant variant is the consideration of topological constraints. For example, if the polyline represents a country border, important cities within the country should remain on the same side of the polyline after simplification. It was proven that those problem variants are hard to approximate within a factor  $n^{\frac{1}{5}-\epsilon}$  [8]. Hence, in practice, they are typically tackled with heuristic approaches [8, 9].

Note that the only allowed inputs to those problem variants are either a single polyline without self-intersections or a set of polylines without self-intersections and without common bends or segments (except for common start or end points). In contrast, we explicitly allow non-planar inputs and polyline bundles in which bends and segments may be shared among multiple polylines. We also remark that the known results on hardness of approximation of these problems heavily rely on the constraint that feasible solutions are still non-intersecting. Since we do not require this, we have to resort to different techniques.

### Contribution

We introduce the optimization problem of polyline bundle simplification, where we are given  $\ell$  polylines on an underlying set of  $n$  points as well as an error bound  $\delta$  and seek to find a simplified polyline bundle with the smallest possible number of remaining points, where each simplified polyline has a Fréchet distance of no more than  $\delta$  to the original polyline and the simplification is consistent for shared parts.

While the optimal simplification of a single polyline can be computed in polynomial time, we show that polyline bundle simplification is  $NP$ -hard to approximate within a factor  $n^{\frac{1}{3}-\varepsilon}$  for any  $\varepsilon > 0$ . This result applies already to bundles of two polylines, hence excluding an efficient FPT-algorithm depending on parameter  $\ell$ .

On the positive side, we show that this strong inapproximability can be overcome when relaxing the error bound  $\delta$  slightly. In particular, we design an efficient bi-criteria approximation algorithm. Here, we allow the simplified polylines in our solution to have a Fréchet distance of  $2\delta$  instead of only  $\delta$  to the original polylines. We can then approximate the optimal solution for the original choice of  $\delta$  within a factor logarithmic in the input size. As the choice of  $\delta$  for real-world problems often is made in a rather ad hoc fashion and uncertainties with respect to the precision of the input polylines have to be factored in as well, we deem our bi-criteria approximation to be of high practical relevance.

We furthermore show that, while the number of polylines in the bundles is not suitable to obtain an FPT-algorithm, the problem of polyline bundle simplification is indeed fixed-parameter tractable in the number of bend points that are shared among the polylines.

## 2 Formal Problem Definition

An instance of the *polyline bundle simplification* problem (from now on abbreviated by PBS) is specified by a triple  $(B, \mathcal{L}, \delta)$ , where  $B = \{b_1, \dots, b_n\}$  is a set of  $n$  points (*bends*) in the plane, a *polyline bundle*  $\mathcal{L}$ , which is a set  $\mathcal{L} = \{L_1, \dots, L_\ell\}$  of  $\ell$  polylines  $L_i = (s_i, \dots, t_i)$  represented as lists of points from  $B$ , and a distance parameter  $\delta$ , which specifies a threshold for the the maximum (segment-wise) Fréchet distance between original and simplified polyline bundle. Each polyline  $L_i$  ( $i \in \{1, \dots, \ell\}$ ) is simple in the sense that each bend of  $B$  appears at most once in its list.

► **Definition 1** (Polyline Bundle Simplification). *Given a triple  $(B, \mathcal{L}, \delta)$ , the goal is to obtain a minimum size subset  $B^* \subseteq B$  of points, such that for each polyline  $L_i \in \mathcal{L}$  its induced simplification  $S_i$  (which is  $L_i \cap B^*$  while preserving the order of points)*

- *contains the start and the end point of  $L_i$ , i.e.,  $s_i, t_i \in S_i$ , and*
- *has a segment-wise Fréchet distance of at most  $\delta$  to  $L_i$ , i.e., for each line segment  $(a, b)$  of  $S_i$  and the corresponding sub-polyline of  $L_i$  from  $a$  to  $b$ , abbreviated by  $L_i[a, \dots, b]$ , we have  $d_{\text{Fréchet}}((a, b), L_i[a, \dots, b]) \leq \delta$ .*

For the sake of self-containedness we restate the definition of the Fréchet distance below.

► **Definition 2** (Fréchet Distance). *Between two polylines  $L_1 = (b_{1,1}, b_{1,2}, \dots, b_{1,|L_1|})$  and  $L_2 = (b_{2,1}, b_{2,2}, \dots, b_{2,|L_2|})$  in the Euclidean plane, the Fréchet distance  $d_{\text{Fréchet}}(L_1, L_2)$  is*

$$d_{\text{Fréchet}}(L_1, L_2) := \inf_{\alpha, \beta} \max_{t \in [0,1]} \|c_{L_1}(\alpha(t)) - c_{L_2}(\beta(t))\|,$$

where  $\alpha: [0, 1] \rightarrow [1, |L_1|]$  and  $\beta: [0, 1] \rightarrow [1, |L_2|]$  are continuous and non-decreasing functions with  $\alpha(0) = \beta(0) = 1$ ,  $\alpha(1) = |L_1|$ ,  $\beta(1) = |L_2|$ ,

and  $c_{L_i}: [1, |L_i|] \rightarrow \mathbb{R}^2$  with  $c_{L_i}: x \mapsto (\lfloor x \rfloor + 1 - x)b_{i, \lfloor x \rfloor} + (x - \lfloor x \rfloor)b_{i, \lfloor x \rfloor + 1}$ .

### 3 Hardness of Polyline Bundle Simplification

In this section, we describe a polynomial-time reduction from Minimum Independent Dominating Set (MIDS) to PBS to show  $NP$ -hardness and hardness of approximation. In the MIDS problem, we are given a graph  $G = (V, E)$ , where  $V$  is the vertex set and  $E$  is the edge set of  $G$ . We define  $\hat{n} = |V|$  and  $\hat{m} = |E|$ . The goal is to find a set  $V^* \subseteq V$  of minimum cardinality that is a dominating set of  $G$  as well as an independent set in  $G$ . A dominating set contains for each vertex  $v$ ,  $v$  itself or at least one of  $v$ 's neighbors. An independent set contains for each edge at most one of its endpoints. Halldórsson [11] has shown that MIDS, which is also referred to as MINIMUM-MAXIMAL-INDEPENDENT-SET, is  $NP$ -hard to approximate within a factor of  $|V|^{1-\varepsilon}$  for any  $\varepsilon > 0$ . In his proof, he uses a reduction from SAT to MIDS: from a SAT formula  $\Phi$ , he constructs a graph such that an algorithm approximating MIDS would decide if  $\Phi$  is satisfiable. We observe that this reduction is still correct if  $\Phi$  is a 3-SAT formula. Moreover, we observe that the number of edges in the graph constructed in this reduction by a 3-SAT formula is linear in the number of vertices. Thus, we conclude the following corollary and assume henceforth that we reduce only from sparse graph instances of MIDS, in other words,  $\hat{m} \leq c\hat{n}$  for some sufficiently large constant  $c$ .

► **Corollary 3.** *MIDS on graphs of  $\hat{n}$  vertices and  $\mathcal{O}(\hat{n})$  edges, i.e., sparse graphs, is  $NP$ -hard to approximate within a factor of  $\hat{n}^{1-\varepsilon}$  for any  $\varepsilon > 0$ .*

In our reduction, we use three types of gadgets, which are in principle all lengthy zigzag pieces. We use *vertex gadgets* to indicate whether a vertex is in the set  $V^*$  or not, *edge gadgets* to enforce the independent set property, and *neighborhood gadgets* to enforce the dominating set property. See Fig. 3 for an overview. We define our gadgets in terms of an arbitrary  $\delta$  (threshold for the maximum Fréchet distance) and some  $\gamma \leq 2\delta/(10\hat{n}^2 + 5)$ . Note that our problem setting allows overlaps of different polylines without having a common bend or segment (non-planar input). In our reduction there can also be overlaps, which do not affect the involved polylines locally.

**Vertex Gadget.** For each vertex, we construct a *vertex gadget* (see Figure 3a), which we arrange vertically next to each other on a horizontal line in arbitrary order and with some distance  $x_{\text{spacing}} \geq (2\hat{n}^2 + 2)3\delta$  between one and the next vertex gadget.

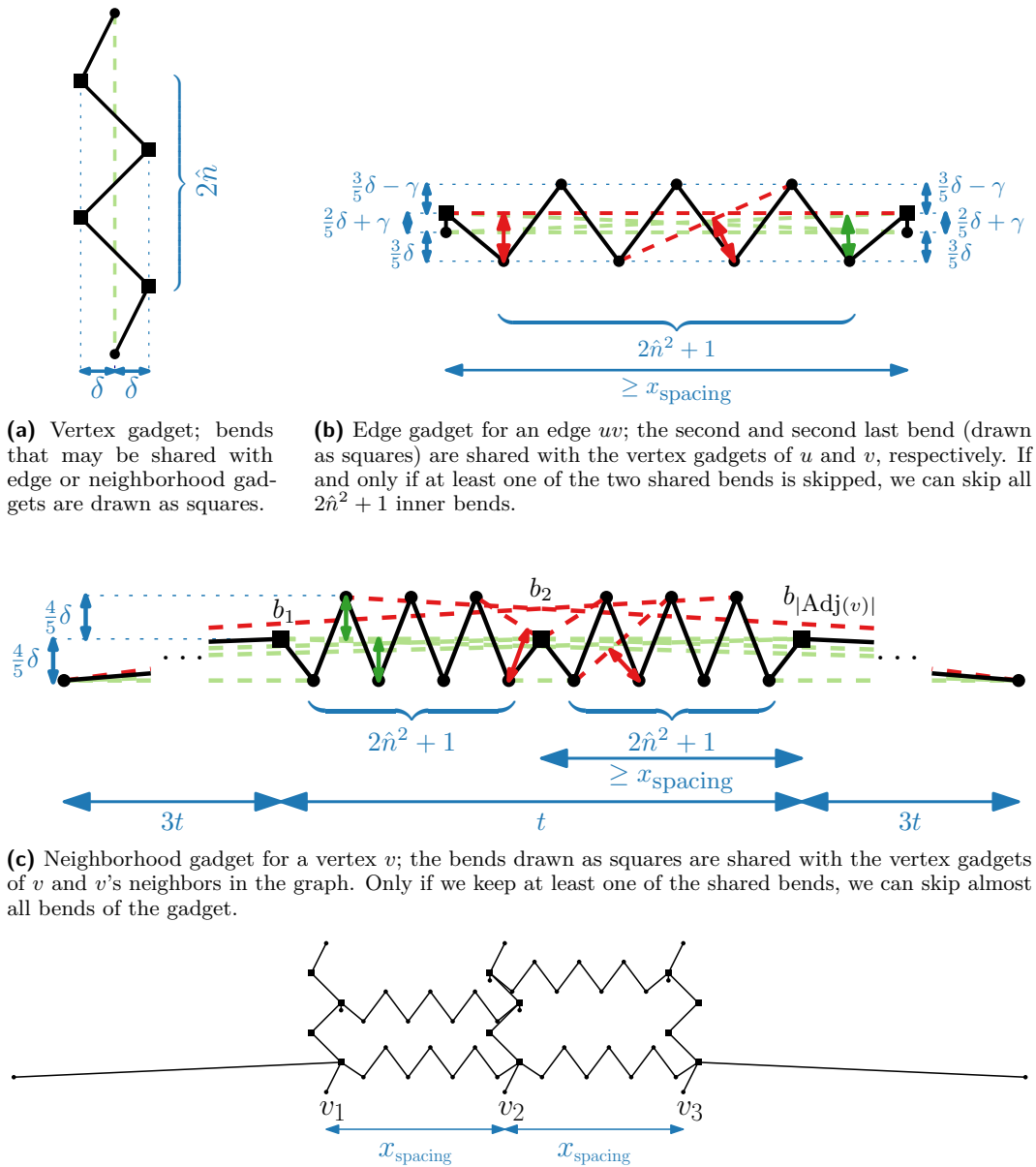
A vertex gadget has  $2\hat{n} + 2$  bends arranged in a zigzag course with x-distance  $2\delta$  ( $\delta$  for the first and the last segment) and y-distance  $3\delta$  between each two consecutive bends.

▷ **Claim 4.** In a vertex gadget, there is precisely one shortcut, which starts at the first and ends at the last bend.

Clearly, the line segment from the first to the last bend has Fréchet distance at most  $\delta$  to the other bends and segments of the vertex gadget. Moreover, observe that there is no shortcut starting or ending at any inner bend. Thus, either none or all inner bends are skipped. We say that the corresponding vertex is in  $V^*$  if and only if we do not skip the inner vertices.

**Edge Gadget.** For each edge  $\{u, v\}$ , we construct an *edge gadget* (see Figure 3b) being a zigzag course with  $2\hat{n}^2 + 5$  bends and sharing its second and second last bend with one of the two corresponding vertex gadgets – the vertex gadgets of  $u$  and  $v$ . All neighboring bends from the second to the second last are equidistant in x-dimension, while the first and second bend, and the second last and last bend have the same x-coordinate. In y-dimension, the first and the last bend are  $2/5\delta + \gamma$  below the second and second last bend, respectively. The other bends are  $3/5\delta - \gamma$  above the second bend or  $3/5\delta$  below the first bend.

35:6 Simplification of Polyline Bundles



(a) Vertex gadget; bends that may be shared with edge or neighborhood gadgets are drawn as squares. (b) Edge gadget for an edge  $uv$ ; the second and second last bend (drawn as squares) are shared with the vertex gadgets of  $u$  and  $v$ , respectively. If and only if at least one of the two shared bends is skipped, we can skip all  $2\hat{n}^2 + 1$  inner bends.

(c) Neighborhood gadget for a vertex  $v$ ; the bends drawn as squares are shared with the vertex gadgets of  $v$  and  $v$ 's neighbors in the graph. Only if we keep at least one of the shared bends, we can skip almost all bends of the gadget.

(d) Combination of three vertex gadgets (for the vertices  $v_1, v_2, v_3$ ) with two edge gadgets (for the edges  $v_1v_2$  and  $v_2v_3$ ) and a neighborhood gadget for the vertex  $v_2$ .

■ **Figure 3** Schematization of our reduction from MIDS to PBS. Shortcuts are indicated by dashed green line segments. Dashed red line segments between two bends indicate that there is no shortcut. The vertices in our minimum independent dominating set are precisely the ones for which we do not take the shortcut of the corresponding vertex gadgets.

▷ **Claim 5.** In an edge gadget, there are precisely three long shortcuts. These are (i) from the first to the last bend, (ii) from the first to the second last bend, and (iii) from the second to the last bend. Beside these three shortcuts, there are  $\leq 4$  more shortcuts, which skip only the second and the second last bend (and possibly also the third and third last bend). There is no shortcut not skipping one of the shared bends, i.e., the second or the second last bend.

In Appendix A, we argue that Claim 5 is correct. It follows that not skipping one of the two shared bends is a relatively expensive choice in terms of retained bends. Remember that not skipping one of the shared bends means not taking the shortcut in the corresponding vertex gadget, which means putting the corresponding vertex into  $V^*$ . So, skipping almost all bends in the edge gadget of  $\{u, v\}$  implies not having  $u$  or  $v$  in  $V^*$ , which means respecting the independent set property for the edge  $\{u, v\}$ .

**Neighborhood Gadget.** For each vertex  $v$ , we construct a *neighborhood gadget* (see Figure 3c). This gadget shares a bend with every vertex gadget corresponding to a vertex of  $\text{Adj}(v)$ , which is  $v$  and the vertices being adjacent to  $v$ . These shared bends are on the same height. The vertex gadgets of  $\text{Adj}(v)$  appear in some horizontal order in our construction. Say the corresponding vertices in order are  $u_1, \dots, u_{|\text{Adj}(v)|}$ . Let the shared bends with  $u_1$  and  $u_{|\text{Adj}(v)|}$  be  $b_1$  and  $b_{|\text{Adj}(v)|}$ , respectively, and define  $t$  as the distance between  $b_1$  and  $b_{|\text{Adj}(v)|}$ . We place the first bend (the starting point) of the neighborhood gadget  $4/5\delta$  below and  $3t$  to the left of  $b_1$ , where  $t$  is the distance between  $b_1$  and  $b_{|\text{Adj}(v)|}$ , and let the second bend be  $b_1$ . Symmetrically, we place the last bend (the end point) of the gadget  $4/5\delta$  below and  $3t$  to the right of  $b_{|\text{Adj}(v)|}$  and let the second last bend be  $b_{|\text{Adj}(v)|}$ . Between each two bends  $b_i$  and  $b_{i+1}$  shared with the vertex gadgets of  $u_i$  and  $u_{i+1}$  for each  $i \in \{1, \dots, |\text{Adj}(v)| - 1\}$ , we add a zigzag with  $2\hat{n}^2 + 1$  bends as in Figure 3c.

▷ **Claim 6.** In a neighborhood gadget, the only shortcuts are (i) the shortcuts skipping only  $b_i$  for  $i \in \{1, \dots, |\text{Adj}(v)|\}$  and (ii) the shortcuts starting at the first bend or  $b_i$  with  $i \in \{1, \dots, |\text{Adj}(v)|\}$  and ending at the last bend or  $b_j$  with  $i < j \in \{1, \dots, |\text{Adj}(v)|\}$  – except for the shortcut starting at the first and ending at the last bend.

In Appendix A, we argue that Claim 6 is correct. Consequently, we can skip almost all bends in a neighborhood gadget if we keep at least one bend of  $b_1, \dots, b_{|\text{Adj}(v)|}$ . If we skip all of them, we can skip no other bend. So, to avoid high costs, we must not take the shortcut of the vertex gadget of at least one vertex of  $\text{Adj}(v)$ . This means that we must, for each  $v \in V$ , add a vertex of  $\text{Adj}(v)$  to  $V^*$ , which enforces the dominating set property.

Observe that all shared bends are shared between only two polylines – by a vertex gadget and either an edge gadget or a neighborhood gadget. With  $2\hat{n}$  inner bends, a vertex gadget provides enough bends that are shared with the edge and neighborhood gadgets as a vertex is contained in at most  $\hat{n}$  neighborhoods and has at most  $\hat{n} - 1$  incident edges. In the following lemma, we analyze the size of the constructed PBS instance.

► **Lemma 7.** *By our reduction, we obtain from an instance  $G = (V, E)$  of MIDS an instance of PBS with  $n$  bends such that  $n \leq 10c\hat{n}^3$ , where  $\hat{n} = |V| \geq 2$ ,  $|E| \leq c\hat{n}$  ( $c \geq 1$  is constant).*

**Proof.** To count the bends of the vertex, edge, and neighborhood gadgets without double counting, we charge the shared bends to the vertex gadgets. All vertex gadgets together have  $\hat{n}(2\hat{n} + 2)$  bends, all edge gadgets have  $\hat{m}(2\hat{n}^2 + 3)$  bends without shared bends, and all neighborhood gadgets have  $2\hat{m} \cdot (2\hat{n}^2 + 1) + 2\hat{n}$  bends without shared bends. Summing these values up and using  $\hat{m} = |E| \leq c\hat{n}$  yields (for  $\hat{n} \geq 2$ )

$$n = 2\hat{n}^2 + 2\hat{n} + 2\hat{m}\hat{n}^2 + 3\hat{m} + 4\hat{m}\hat{n}^2 + 2\hat{m} + 2\hat{n} \leq 6c\hat{n}^3 + 2\hat{n}^2 + (4 + 5c)\hat{n} \leq 10c\hat{n}^3. \quad (1)$$

◀

We say a simplification of an instance of PBS obtained by this reduction *corresponds* to an independent and dominating set  $V'$  and vice versa if we take all “long” shortcuts in the vertex gadgets except for the ones corresponding to  $V'$  and we skip all inner unshared bends in all

edge and neighborhood gadgets, which is possible since  $V'$  is independent and dominating. Observe that for each independent and dominating set there is precisely one corresponding simplification (which is also valid acc. to  $\delta$ ).

► **Lemma 8.** *Let  $V'$  be a solution for an instance  $G = (V, E)$  of MIDS. In the instance  $(B, \mathcal{L}, \delta)$  of PBS obtained by our reduction, the size of the simplification corresponding to  $V'$  is  $2\hat{n}(|V'| + c + 2)$ , where  $\hat{n} = |V|$  and  $c \geq 1$  is constant.*

**Proof.** Only for all  $v \in V \setminus V'$ , we take the shortcuts in the corresponding vertex gadgets in  $(B, \mathcal{L}, \delta)$ . This gives us  $(\hat{n} - |V'|) \cdot 2 + |V'| \cdot (2 + 2\hat{n}) = 2\hat{n}(1 + |V'|)$  remaining bends in all vertex gadgets combined. In the following, we will count shared bends for the vertex gadgets. We take a “long” shortcut in all of the edge gadgets. This gives us two remaining unshared bends in all edges gadgets ( $c\hat{n} \cdot 2$  bends in total). Moreover, we skip all inner unshared bends in all of the neighborhood gadgets ( $2\hat{n}$  bends remaining). Altogether, this sums up to  $2\hat{n}(|V'| + 1 + c + 1)$ . ◀

By Lemma 8, we know that for an optimal solution  $V^*$  of an instance of MIDS, the corresponding simplification in the instance  $(B, \mathcal{L}, \delta)$  of PBS obtained by our reduction has size  $2\hat{n}(OPT_{\text{MIDS}} + c + 2)$ , where  $OPT_{\text{MIDS}} = |V^*|$  and which of course is at least the size  $OPT_{\text{PBS}}$  of the optimal solution of  $(B, \mathcal{L}, \delta)$ . We formalize this in the following corollary.

► **Corollary 9.** *For an instance  $G = (V, E)$  of MIDS and the instance  $(B, \mathcal{L}, \delta)$  of PBS obtained by our reduction from  $G$ ,  $OPT_{\text{PBS}} \leq 2\hat{n}(OPT_{\text{MIDS}} + c + 2)$ .*

► **Theorem 10.** *PBS is NP-hard to approximate within a factor of  $n^{\frac{1}{3}-\varepsilon}$  for any  $\varepsilon > 0$ , where  $n$  is the number of bend points in the polyline bundle.*

**Proof.** Assume that there is an approximation algorithm  $\mathcal{A}$  solving any instance of PBS within a factor of  $n^{\frac{1}{3}-\varepsilon}$  for some constant  $\varepsilon > 0$  relative to the optimal solution. We can transform any instance  $G = (V, E)$  of MIDS, where  $\hat{n} = |V|$ ,  $\hat{m} = |E|$ , and  $OPT_{\text{MIDS}} = |V^*|$ , this is the size of an optimal solution, to an instance  $(B, \mathcal{L}, \delta)$  of PBS using the reduction described above in this section, where  $|B| = n$  and the size of an optimal solution is  $OPT_{\text{PBS}}$ .

Employing  $\mathcal{A}$  to solve  $(B, \mathcal{L}, \delta)$  yields a (simplified) polyline bundle  $\mathcal{L}_{\mathcal{A}}$ . We denote the number of bends in  $\mathcal{L}_{\mathcal{A}}$  by  $n_{\mathcal{A}}$  and we know that  $n_{\mathcal{A}} \leq OPT_{\text{PBS}} \cdot n^{\frac{1}{3}-\varepsilon}$  for some  $\varepsilon > 0$ . If all  $(2\hat{n}^2 + 1)$ -bend-sequences in all edge and neighborhood gadgets are skipped, we can immediately read an independent dominating vertex set  $V' \subseteq V$  from the vertex gadgets where the shortcut is not taken. Otherwise, we replace  $\mathcal{L}_{\mathcal{A}}$  such that it corresponds to any maximal independent set  $V' \subseteq V$  (which is always an independent and dominating set and can be found greedily in polynomial time). Observe that this can only lower the number of bends compared to a solution not skipping all  $(2\hat{n}^2 + 1)$ -bend-sequences in the edge and neighborhood gadgets as in all vertex gadgets together we can skip at most  $\hat{n} \cdot 2\hat{n}$  bends.

Using Lemma 8 and Corollary 9, we can state that

$$n^{\frac{1}{3}-\varepsilon} \geq \frac{n_{\mathcal{A}}}{OPT_{\text{PBS}}} \geq \frac{2\hat{n}(|V'| + c + 2)}{2\hat{n}(OPT_{\text{MIDS}} + c + 2)} > \frac{|V'|}{OPT_{\text{MIDS}} + c + 2}, \quad (2)$$

which we can reformulate as  $|V'| < n^{\frac{1}{3}-\varepsilon}(OPT_{\text{MIDS}} + c + 2)$ . We can assume that  $OPT_{\text{MIDS}} > c + 2$  as otherwise we could check all subsets of  $V$  of size at most  $c + 2$  in polynomial time. Similarly, we can assume that  $\hat{n}$  is large enough so that  $\hat{n}^{2\varepsilon} > 20c$ . Beside this, we apply Lemma 7 and obtain

$$|V'| < 2n^{\frac{1}{3}-\varepsilon}OPT_{\text{MIDS}} \leq 2 \cdot (10c\hat{n}^3)^{\frac{1}{3}-\varepsilon}OPT_{\text{MIDS}} \quad (3)$$

$$< 20c \cdot \hat{n}^{1-3\varepsilon}OPT_{\text{MIDS}} < \hat{n}^{2\varepsilon} \cdot \hat{n}^{1-3\varepsilon}OPT_{\text{MIDS}} = \hat{n}^{1-\varepsilon}OPT_{\text{MIDS}}. \quad (4)$$



Since we know that it is  $NP$ -hard to approximate MIDS within a factor of  $\hat{n}^{1-\varepsilon}$  for any  $\varepsilon > 0$ , it follows that  $\mathcal{A}$  cannot be a polynomial time algorithm, unless  $P = NP$ . Or in other words, it is  $NP$ -hard to approximate PBS within a factor of  $n^{\frac{1}{3}-\varepsilon}$  for any  $\varepsilon > 0$ . ◀

Currently, we use one polyline per gadget. So, our reduction uses  $2\hat{n} + \hat{m}$  polylines. We can reduce the number of polylines to two by connecting all vertex gadgets – one after the other – in arbitrary order by two segments, which gives us the first polyline, and by connecting all edge and neighborhood gadgets in arbitrary order by two segments, which gives us the second polyline. The extra bend between each pair of new segments is placed far away from the construction, e.g. at  $(\infty, \infty)$ . This never creates new shortcuts for skipping a bend in a vertex gadget or in a neighborhood gadget. Yet, we might create new shortcuts that allow for additionally skipping the first and the last bend of an edge gadget. However, we cannot skip any further bend unless the second or second last bend is skipped, which preserves the functionality of our gadget. For the analysis, this gives us an additive constant of at most  $2\hat{n} + \hat{m}$  bends that cannot be skipped, which we can include to Inequalities (2)–(4) in Theorem 10 with the same result to obtain the following corollaries.

▶ **Corollary 11.** *Even for instances of two polylines, PBS is  $NP$ -hard to approximate within a factor of  $n^{\frac{1}{3}-\varepsilon}$  for any  $\varepsilon > 0$ , where  $n$  is the number of bend points in the polyline bundle.*

▶ **Corollary 12.** *PBS is not fixed-parameter tractable in the number of polylines  $\ell$ .*

## 4 Bi-criteria Approximation for Polyline Bundle Simplification

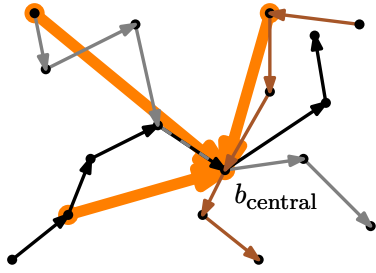
In this section, we describe a bi-criteria approximation algorithm for PBS. Conceptually, a bi-criteria approximation is a generalization of a (classical) approximation where it is allowed to violate a certain constraint by a specific factor. In particular, an algorithm is called a bi-criteria  $(\alpha, \beta)$ -approximation algorithm if it runs in polynomial time and produces a solution of size at most  $\alpha \cdot OPT$  while relaxing the constraint by a factor of  $\beta$ .

In our particular problem PBS, we relax the error bound  $\delta$ . In Section 3, we have shown that there is no bi-criteria  $(n^{\frac{1}{3}-\varepsilon}, 1)$ -approximation algorithm for PBS for any  $\varepsilon > 0$  unless  $P = NP$ . This strong inapproximability comes from the high sensitivity towards choices of keeping or discarding single bends, which is modulated by the given value of  $\delta$ . By making a bad choice we cannot take (helpful) shortcuts that have a distance just a little greater than the given distance threshold  $\delta$  to the original sub-polyline. This can be overcome by relaxing the constraint slightly. In particular, we show that allowing a constraint violation by a factor of  $\beta = 2$ , we can design an efficient algorithm with an approximation guarantee of  $\alpha \in \mathcal{O}(\log(\ell + n))$ . For an overview of our algorithm see Fig. 6.

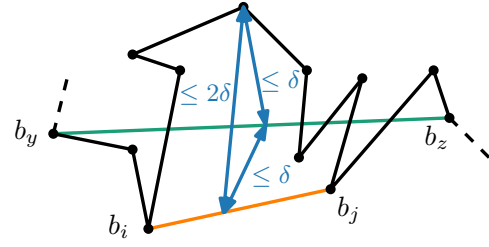
The key building block of our algorithm is a connection between PBS and a certain geometric set cover problem, which we call *star cover problem*. The star cover problem models the aspect of shortcutting polylines by few bend points but does not take into account consistency. We argue, however, that approximate solutions to the star cover problem can be post-processed to form consistent PBS solutions by slightly violating the error threshold  $\delta$ .

### Star Cover Problem

Next, we introduce the *star cover* problem, which is a special type of the set cover problem defined over instances of PBS. Informally spoken, a star is a bend together with some incident shortcut segments. These shortcut segments span sets of original segments of the polylines. To this end, we first direct each polyline  $L \in \mathcal{L}$  in a given PBS instance  $(B, \mathcal{L}, \delta)$  arbitrarily but ensuring that all (shortcut) segments of  $L$  are oriented in the same direction. Then a star consists of a set of incoming shortcuts of some bend; see Fig. 4 for an example.



■ **Figure 4** Example of a *star* (in orange) around a bend  $b_{\text{central}}$ , which lies on three polylines. Each polyline was assigned an arbitrary direction indicated by arrow heads.



■ **Figure 5** Example of the maximum Fréchet distance between a line segment  $(b_i, b_j)$  and its corresponding sub-polyline if there is a valid shortcut  $(b_y, b_z)$  going over  $b_i$  and  $b_j$ .

► **Definition 13 (Star).** A star is the combination of a bend  $b_{\text{central}} \in B$  and, for each polyline  $L \in \mathcal{L}$  that contains  $b_{\text{central}}$ , one or zero incoming shortcut segments (according to  $\delta$ ).

We say a star  $s$  covers a segment–polyline pair  $(e, L)$ , if  $s$  contains for  $L$  a shortcut  $(b_{\text{outer}}, b_{\text{central}})$  and  $e$  lies on  $L$  between  $b_{\text{outer}}$  and  $b_{\text{central}}$ . Our goal is to find a small set of stars that cover all segment–polyline pairs. We denote the set of all segment–polyline pairs in the input by  $\mathcal{U}$  and the subset of pairs covered by a particular star  $s$  by  $\mathcal{U}_s$ . Then the *star cover problem* is defined as follows.

► **Definition 14 (Star Cover).** A star cover  $C$  is a set of stars, such that  $\bigcup_{s \in C} \mathcal{U}_s = \mathcal{U}$ , i.e. all segment–polyline pairs are covered. The star cover problem (abbreviated by STCO) asks for a minimum size star cover.

### Relationship between Instances of Polyline Bundle Simplification and Star Cover

Next, we investigate the relationship between an instance of STCO and its corresponding instance of PBS. We argue that every (optimal) solution for PBS can be decomposed into a star cover. Hence an optimal STCO yields a lower bound for an optimal PBS solution.

► **Lemma 15.** The size  $OPT_{\text{STCO}}$  of an optimal solution of any instance of STCO obtained from an instance  $(B, \mathcal{L}, \delta)$  of PBS is bounded by  $OPT_{\text{STCO}} \leq OPT_{\text{PBS}}$ , where  $OPT_{\text{PBS}}$  is the size of an optimal solution of  $(B, \mathcal{L}, \delta)$ .

**Proof.** Consider an optimal solution  $B^*$  of  $(B, \mathcal{L}, \delta)$ . From the simplified polyline bundle induced by  $B^*$ , we can get a star cover for any instance of STCO obtained from  $(B, \mathcal{L}, \delta)$  by iteratively adding a star in the following way until there are only isolated bends. Get a star  $s$  by taking any connected bend  $b_{\text{central}} \subseteq B^*$  as a central bend and the bends that precede  $b_{\text{central}}$  on each of the simplified polylines as its outer bends. Remove the segment–polyline pairs covered by  $s$  from our simplified polyline bundle. Repeat this until there are no more segment–polyline pairs. The obtained star cover has at most  $|B^*|$  stars and at least as many stars as a minimum star cover. So,  $OPT_{\text{STCO}} \leq OPT_{\text{PBS}}$ . ◀

### Approximation for the Star Cover Problem

We can compute an approximate solution for STCO by employing the classical greedy algorithm [15] for set cover, which iteratively selects the set with the most uncovered elements until all elements are covered. However, if applied naively, the running time would be exponential in the size of the PBS instance as the number of stars might be in the order

of  $n \cdot 2^\ell$ . We observe, however, that it suffices to consider only maximal stars (containing on each polyline incident to the central bend the incoming shortcut that covers the largest number of segments). As there are only  $n$  maximal stars, this guarantees polynomial running time.

► **Lemma 16.** *We can compute an  $\mathcal{O}(\log(t+w))$ -approximation for an instance of STCO obtained from an instance  $(B, \mathcal{L}, \delta)$  in time  $\mathcal{O}(\ell n^3)$ , where  $t$  is the maximum number of polylines any bend point occurs in and  $w$  is the maximum number of segments any valid shortcut (according to  $\delta$ ) can skip.*

**Proof.** There is a polynomial time greedy algorithm that yields an  $\mathcal{O}(\log m)$  approximation for the set cover problem, where  $m$  is the size of the largest set in the given collection of subsets of the universe [15]. The greedy algorithm works as follows. While there are uncovered elements from the universe, add the set with the largest number of uncovered elements to the set cover. In an instance of STCO, this  $m$  is the maximum number of segment–polyline pairs  $\max_{\text{star } s} |\mathcal{U}_s|$  a single star can cover. If the central bend point of a star lies in at most  $t$  polylines, the star contains at most  $t$  shortcut segments, and each of which covers at most  $w$  segments, hence we have  $m = tw$ . Observe that  $\mathcal{O}(\log(tw)) = \mathcal{O}(\log(t+w))$ .

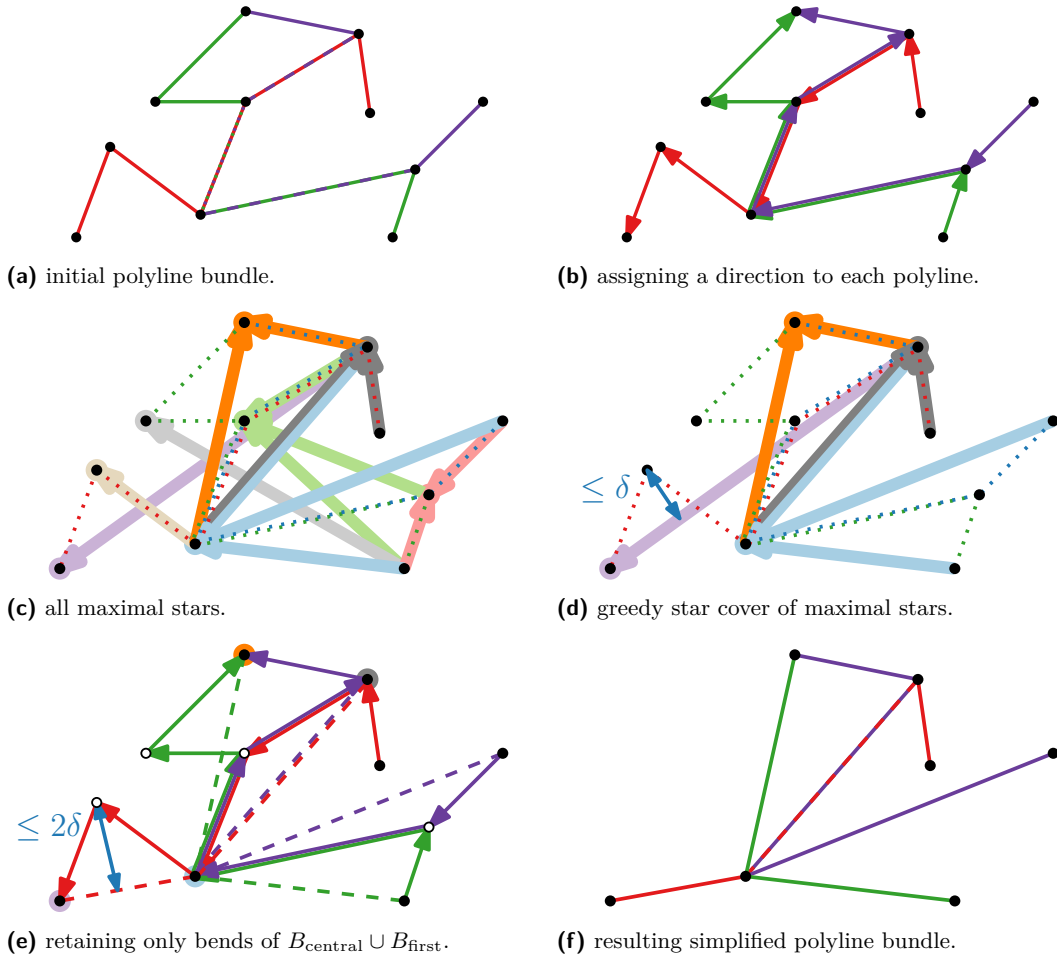
Having settled the  $\mathcal{O}(\log(t+w))$  approximation ratio, it remains to prove the polynomial running time. Using the algorithm by Imai and Iri [14] independently for each polyline, we can find all (maximal) shortcuts for every bend on every polyline in time  $\mathcal{O}(\ell n^3)$ . Combining these shortcuts at every bend gives us all  $n$  maximal stars in time  $\mathcal{O}(\ell n)$ . For each star, we also save the number of segment–polyline pairs it covers and, to each segment–polyline pair, we link all stars it appears in. Both can be done in time  $\mathcal{O}(\ell n^2)$ . As long as there are uncovered segments, we find the star with the most uncovered segments and then update the number of uncovered segments for the other stars. This can be done in  $\mathcal{O}(\ell n^2)$  time in total as well. ◀

### Relationship between Star Covers and Solutions of Polyline Bundle Simplification

While a solution for PBS can be directly converted into a star cover as argued above, the converse is more intricate. The shortcuts contained in the selected stars may be overlapping or nested along a polyline, that is, bends skipped by one shortcut may be end points of another shortcut in the set. Moreover, shared parts of different polylines may be shortcut differently. Therefore consistency is not guaranteed. We explain how to derive from a star cover solution a solution for its corresponding instance of PBS. Some of the shortcuts of the STCO solution are replaced by shorter shortcuts in order to integrate some intermediate point to the PBS solution. Lemma 17 states that those newly introduced shortcuts can be at most  $2\delta$  away from the original polyline. The situation described there is depicted in Fig. 5. It follows immediately from a lemma by Agarwal et al. ([1], Lemma 3.3).

► **Lemma 17.** *Given a polyline  $L = (b_1, b_2, \dots, b_{|L|})$  and a distance threshold  $\delta$ . If there are  $y, z \in \mathbb{N}$  with  $1 \leq y < z \leq |L|$  and  $d_{\text{Fréchet}}((b_y, b_z), L[b_y, \dots, b_z]) \leq \delta$  (i.e., segment  $(b_y, b_z)$  is a valid shortcut), then for any  $i, j \in \mathbb{N}$  with  $y \leq i < j \leq z$ ,  $d_{\text{Fréchet}}((b_i, b_j), L[b_i, \dots, b_j]) \leq 2\delta$ .*

Equipped with this lemma, we now discuss the actual transformation from a STCO solution to a PBS solution. The idea is to keep, beside the starting points of all polylines, only the central bend points of the selected stars while dropping their leaves. This is closely tied with the fact that we minimize the number of stars while ignoring their degree in the algorithm. The main insight here is that the shortcuts induced by this augmented point set still have a small distance to the original polylines.



■ **Figure 6** Example of our bi-criteria  $(\mathcal{O}(\log(\ell + n)), 2)$ -approximation algorithm for PBS.

► **Lemma 18.** *Let  $C$  be a star cover for an instance of STCO obtained from an instance  $(B, \mathcal{L}, \delta)$  of PBS. If  $C$  is an  $\alpha$ -approximation for its instance of STCO, a bi-criteria  $(\alpha + 1, 2)$ -approximation for  $(B, \mathcal{L}, \delta)$  can be computed in time  $\mathcal{O}(n)$  from  $C$ .*

**Proof.** Let  $B_{\text{central}}$  be the set of central bends of the stars in  $C$  and let  $B_{\text{first}}$  be the set of first bends of all polylines from  $\mathcal{L}$ . We return  $B_{\text{central}} \cup B_{\text{first}}$  as the bi-criteria approximate solution. Clearly, we can construct this set in time  $\mathcal{O}(n)$ . According to Lemma 15,  $\text{OPT}_{\text{STCO}} \leq \text{OPT}_{\text{PBS}}$ , where  $\text{OPT}_{\text{PBS}}$  is the size of the optimal solution of  $(B, \mathcal{L}, \delta)$  and  $\text{OPT}_{\text{STCO}}$  is the size of the optimal solution of the instance of STCO where  $C$  is an approximation for. We conclude

$$|B_{\text{central}} \cup B_{\text{first}}| \leq \alpha \text{OPT}_{\text{STCO}} + \text{OPT}_{\text{PBS}} \leq (\alpha + 1) \text{OPT}_{\text{PBS}}. \quad (5)$$

Let  $\mathcal{L}'$  be the polyline bundle induced by  $B_{\text{central}} \cup B_{\text{first}}$ . It remains to prove that the Fréchet distance between each induced segment of each polyline in  $\mathcal{L}'$  and its corresponding sub-polyline in  $\mathcal{L}$  is at most  $2\delta$ . Consider any segment  $(b_i, b_j)$  of any polyline  $L' \in \mathcal{L}'$  corresponding to a polyline  $L \in \mathcal{L}$  such that  $b_i$  precedes  $b_j$  in  $L$ . There is a star  $s$  in  $C$  that covers all segments of  $L[b_i, b_j]$ . Clearly, all segments of  $L[b_i, b_j]$  are covered by the stars of  $C$  and if there was no single star  $s$  covering all segments of  $L[b_i, b_j]$ , but multiple stars, there would be another central bend of a star between  $b_i$  and  $b_j$  on  $L$  and, in  $L'$ ,  $(b_i, b_j)$  would

not be a segment. The central bend  $b_{\text{central}}$  of  $s$  succeeds  $b_j$  or is equal to  $b_j$  as otherwise  $s$  would not cover all of  $L[b_i, b_j]$ . Accordingly, the outer bend  $b_{\text{outer}}$  of  $s$  on  $L$  precedes  $b_i$  or is equal to  $b_i$  as otherwise  $s$  would not cover all of  $L[b_i, b_j]$ . By the definition of a star, we know that  $d_{\text{Fréchet}}((b_{\text{outer}}, b_{\text{central}}), L[b_{\text{outer}}, b_{\text{central}}]) \leq \delta$ . By Lemma 17, it follows that  $d_{\text{Fréchet}}((b_i, b_j), L[b_i, b_j]) \leq 2\delta$ . ◀

### Bi-criteria Approximation for Polyline Bundle Simplification via Star Cover

Using the previous lemmas, we obtain the main theorem of this section. It is reasonable to assume that the number  $\ell$  of polylines is polynomial in  $n$  in practically relevant settings. Hence, we essentially obtain an exponential improvement over the complexity-theoretic lower bound  $n^{\frac{1}{3}-\varepsilon}$  if we allow the slight violation of the error bound.

► **Theorem 19.** *There is a bi-criteria  $(\mathcal{O}(\log(\ell + n)), 2)$ -approximation algorithm for PBS running in time  $\mathcal{O}(\ell n^3)$ , where  $\ell$  is the number of polylines and  $n$  is the number of bend points in the polyline bundle.*

**Proof.** We describe a (kind of) approximation-preserving reduction from PBS to STCO, which can be realized as a bi-criteria approximation algorithm. Its steps are depicted in Fig. 6. Given an instance  $(B, \mathcal{L}, \delta)$  of PBS, where we let the size of the optimal solution be  $\text{OPT}_{\text{PBS}}$ , we assign an arbitrary direction to each  $L \in \mathcal{L}$ . This yields our corresponding instance of STCO. For this corresponding instance of STCO, compute an  $\mathcal{O}(\log(t + w))$  approximation star cover  $C$ . We can do this in time  $\mathcal{O}(\ell n^3)$  according to Lemma 16. According to Lemma 18, we can compute a bi-criteria  $(\mathcal{O}(\log(t + w)), 2)$ -approximation for  $(B, \mathcal{L}, \delta)$  from  $C$  in  $\mathcal{O}(n)$  time. Since  $t \leq \ell$  and  $w \leq n$ , this is also a bi-criteria  $(\mathcal{O}(\log(\ell + n)), 2)$ -approximation. ◀

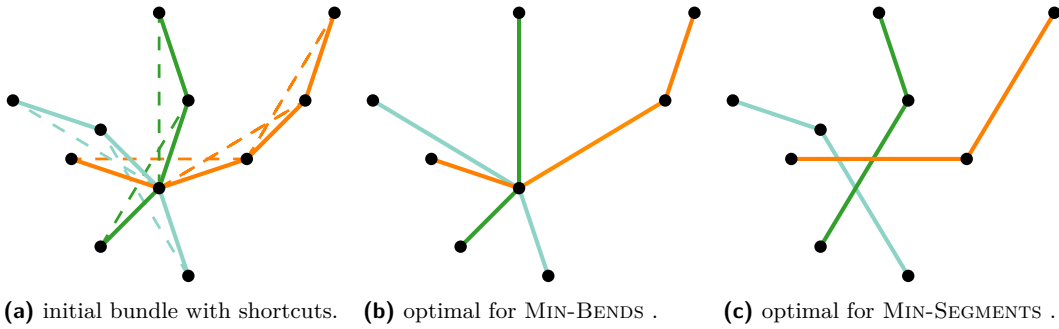
## 5 Fixed-Parameter Tractability

A brute force approach is checking for every subset of the bend set  $B$  in time  $\mathcal{O}(\ell \cdot n)$  whether it is a valid simplification and accepting the one with the smallest number of bends or segments. Consequently, the runtime of this approach is  $\mathcal{O}(2^n \cdot \ell \cdot n)$ . When considering fixed-parameter tractability, investigating parameters of the input is a natural choice. According to Corollary 12, PBS is not fixed-parameter tractable (FPT) in the number of polylines  $\ell$ . However, PBS is FPT in the number of shared bends, i.e., bends contained in more than one polyline. We denote the set of those bends by  $B_{\text{shared}}$  and we let  $k := |B_{\text{shared}}|$ .

► **Theorem 20.** *PBS is FPT in the number of shared bends  $k$ . There is an algorithm solving PBS in time  $\mathcal{O}(2^k \cdot \ell \cdot n^2 + \ell n^3)$ .*

**Proof.** We describe an algorithm that solves PBS in time  $\mathcal{O}(2^k \cdot \ell \cdot n^3)$ . Given an instance  $(B, \mathcal{L}, \delta)$  of PBS, the first step is to compute, for each  $L \in \mathcal{L}$ , its shortcut graph  $G_L$  using the algorithm by Imai and Iri [5]. This can be done in time  $\mathcal{O}(\ell \cdot n^3)$ . For a polyline  $L$  and a distance threshold  $\delta$ , the *shortcut graph* is the directed graph that has the bends of  $L$  as its vertices and has an edge from  $u$  to  $v$  if  $d_{\text{Fréchet}}((u, v), L[u, \dots, v]) \leq \delta$ , this is, if there is a shortcut from  $u$  to  $v$  in  $L$ . Given the shortcut graph  $G_L$  of  $L$ , the vertices of a shortest path in  $G_L$  from the first bend of  $L$  to the last bend of  $L$  define an optimal simplification of  $L$ .

The second step is to iterate over all subsets  $B' \subseteq B_{\text{shared}}$  and check if  $B'$  is part of an optimal solution. Before the first iteration, we initialize a variable  $n_{\text{min}} = \infty$  and we will save the current best solution by  $\mathcal{S}_{\text{min}}$ . Then, in each iteration, we temporarily remove from all shortcut graphs  $G_L$  all vertices  $B_{\text{not-contained}} = B_{\text{shared}} - B'$  and all edges that correspond



■ **Figure 7** Example of three polylines, where the goals MIN-SEGMENTS and MIN-BENDS differ.

to a shortcut skipping a bend in  $B'$ . Clearly, removing  $B_{\text{not-contained}}$  can be performed in  $\mathcal{O}(n^2)$  time for each  $G_L$ . For the removal of the edges in  $G_L$ , note that we can sort the list of bends  $B_{\text{not-contained}}$  and the list of all edges (defined by their endpoints) alphanumerically by the occurrence of the bends within the polyline  $L$ . If we traverse both lists simultaneously in ascending order, we remove an edge if and only if its endpoint-bends come before and after the currently considered bend from  $B_{\text{not-contained}}$ . Therefore, the removal operations can be performed in  $\mathcal{O}(n^2)$  time per  $G_L$ .

If some shortcut graph becomes disconnected by these removal operations, we continue with the next iteration. Otherwise, we take the bends of a shortest path from the first to the last bend in each reduced version of  $G_L$ . Together they define a simplification  $\mathcal{S}$  of our PBS instance. If the number  $n_{\mathcal{S}}$  of bends in  $\mathcal{S}$  is less than  $n_{\min}$ , we set  $n_{\min} = n_{\mathcal{S}}$  and  $\mathcal{S}_{\min} = \mathcal{S}$ . After the iteration process, we return  $\mathcal{S}_{\min}$ . Since we have  $2^k$  subsets of  $B_{\text{shared}}$  and each iteration can be performed in  $\mathcal{O}(\ell \cdot n^2)$  time, the running time of the algorithm is in  $\mathcal{O}(2^k \cdot \ell \cdot n^2 + \ell n^3)$ .

It remains to prove that  $\mathcal{S}_{\min}$  is in the end an optimal solution of our input instance of PBS. First note that our algorithm always returns some polyline simplification because for  $B' = B_{\text{shared}}$ , we do not get a disconnected  $G_L$  after the removal operations.

The returned solution is valid because the shared bends of  $B'$  are taken in all simplified polylines (they cannot be skipped) and the other shared bends are skipped in all simplified polylines. Our algorithm finds the minimum size solution because in one iteration it considers  $B' = B^* \cap B_{\text{shared}}$ , where  $B^*$  is the set of retained bends of an optimal solution. Moreover, an optimal solution cannot have fewer bends occurring in only one polyline  $L$  than our algorithm since this would imply a shorter shortest path within the reduced version of  $G_L$ . ◀

## 6 Conclusion and Outlook

We have generalized the well-known problem of polyline simplification from a single polyline to polyline bundles. Although in the case of one polyline, efficient algorithms have long been known, it turned out that simplifying two or more polylines is a problem that is indeed hard to approximate within a factor of  $n^{\frac{1}{3}-\varepsilon}$  for any  $\varepsilon > 0$ . However, if we relax the constraint on the maximum Fréchet distance between original and simplified polyline by a factor of 2, we can overcome this strong inapproximability bound. Moreover, we can find an optimal simplification quickly if we have only a small number of shared bends since the problem of polyline bundle simplification is fixed-parameter tractable (FPT) in this parameter.

Based on our results, there are many possible directions for future research.

- Our current bi-criteria approximation guarantee is logarithmic in the number of polylines  $\ell$  plus the number of bend points  $n$ . In most practical application,  $\ell$  is smaller than  $n$  or at most polynomial in  $n$ . From a theoretical perspective, however, it might be interesting to get rid off the dependency on  $\ell$  in the bi-criteria approximation in order to get improvements for the case where  $\ell$  is significantly larger than  $n$ .
- As a distance measure, we employed the Fréchet distance, which we consider to be more natural and intuitive than the Hausdorff distance when comparing polylines. However, the Hausdorff distance is sometimes used in classical polyline simplification as well. Our hardness results also apply to the Hausdorff distance, but our bi-criteria approximation algorithm fails since Lemma 17 is not true for the Hausdorff distance. One might consider PBS using the Hausdorff distance or other (even non-segment-wise) distance measurements.
- In our generalization to bundles of polylines, we aim for a minimizing the number of retained bends (MIN-BENDS). However, minimizing the number of retained segments (MIN-SEGMENTS) is an alternative goal, which also generalizes the classical minimization problem for a single polyline. Optimal simplifications for both goals may differ; see Fig. 7. Our hardness and FPT results also apply for the goal MIN-SEGMENTS. However, it is not clear how to obtain a similar result for the bi-criteria approximability.
- For practical purposes, the scalability of the proposed bi-criteria approximation algorithm, the FPT algorithm, and possibly new heuristics should be investigated on real-world data.

---

## References

- 1 Pankaj K. Agarwal, Sarel Har-Peled, Nabil H. Mustafa, and Yusu Wang. Near-linear time approximation algorithms for curve simplification. *Algorithmica*, 42(3-4):203–219, 2005. doi:10.1007/s00453-005-1165-y.
- 2 Helmut Alt and Michael Godau. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry and Applications*, 5:75–91, 1995. doi:10.1142/S0218195995000064.
- 3 Maike Buchin, Anne Driemel, and Bettina Speckmann. Computing the Fréchet distance with shortcuts is NP-hard. In *Proceedings of the 30th Annual Symposium on Computational Geometry (SoCG'14)*, pages 367–376, 2014. doi:10.1145/2582112.2582144.
- 4 Maike Buchin, Bernhard Kilgus, and Andrea Kölzsch. Group diagrams for representing trajectories. In *Proceedings of the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, pages 1–10, 2018.
- 5 W. S. Chan and F. Chin. Approximation of polygonal curves with minimum number of line segments or minimum error. *International Journal of Computational Geometry and Applications*, 6(1):59–77, 1996. doi:10.1142/S0218195996000058.
- 6 Marc de Berg, Marc J. van Kreveld, and Stefan Schirra. Topologically correct subdivision simplification using the bandwidth criterion. *Cartography and Geographic Information Systems*, 25(1):243–257, 1998. doi:10.1559/152304098782383007.
- 7 David H. Douglas and Thomas K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica*, 10(2):112–122, 1973.
- 8 Regina Estkowski and Joseph SB Mitchell. Simplifying a polygonal subdivision while keeping it simple. In *Proceedings of the 17th Annual Symposium on Computational Geometry (SoCG'01)*, pages 40–49. ACM, 2001. doi:10.1145/378583.378612.
- 9 Stefan Funke, Thomas Mendel, Alexander Miller, Sabine Storandt, and Maria Wiebe. Map simplification with topology constraints: Exactly and in practice. In *Proceedings of the 19th Workshop on Algorithm Engineering and Experiments (ALENEX'17)*, pages 185–196. SIAM, 2017.

- 10 Michael Godau. A natural metric for curves – computing the distance for polygonal chains and approximation algorithms. In *Proceedings of the 8th Annual Symposium on Theoretical Aspects of Computer Science (STACS'91)*, pages 127–136, 1991. doi:10.1007/BFb0020793.
- 11 Magnús M. Halldórsson. Approximating the minimum maximal independence number. *Information Processing Letters*, 46(4):169–172, 1993. doi:10.1016/0020-0190(93)90022-2.
- 12 Songtao He, Favyen Bastani, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, and Sam Madden. Roadrunner: improving the precision of road network inference from GPS trajectories. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 3–12, 2018.
- 13 John Hershberger and Jack Snoeyink. Speeding up the Douglas-Peucker line-simplification algorithm. In *Proceedings of the 5th International Symposium on Spatial Data Handling (SDH'92)*, pages 134–143, 1992.
- 14 Hiroshi Imai and Masao Iri. Polygonal approximations of a curve – formulations and algorithms. In Godfried T. Toussaint, editor, *Computational Morphology*, volume 6 of *Machine Intelligence and Pattern Recognition*, pages 71–86. North-Holland, 1988. doi:10.1016/B978-0-444-70467-2.50011-4.
- 15 David S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3):256–278, 1974. doi:10.1016/S0022-0000(74)80044-9.
- 16 Urs Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1(3):244–256, 1972. doi:10.1016/S0146-664X(72)80017-0.
- 17 Marc J. van Kreveld, Maarten Löffler, and Lionov Wiratma. On optimal polyline simplification using the Hausdorff and Fréchet distance. *Journal of Computational Geometry*, 11(1):1–25, 2020. URL: <https://journals.carleton.ca/jocg/index.php/jocg/article/view/415>.

## A Omitted Content of Section 3

It remains to show the correctness of Claim 5 and Claim 6, which we use in our reduction from MIDS to PBS. Our gadgets are depicted in Fig. 3. For convenience, we provide by Fig. 8 a copy of them with some additional details, to which we will refer in this appendix. For example  $r$  is the x-distance between two consecutive (inner) vertices in an edge and a neighborhood gadget (if a gadget is rotated, the distance is measured along the corresponding rotated axis). We know that  $r \geq x_{\text{spacing}}/(2\hat{n}^2 + 2)$ .

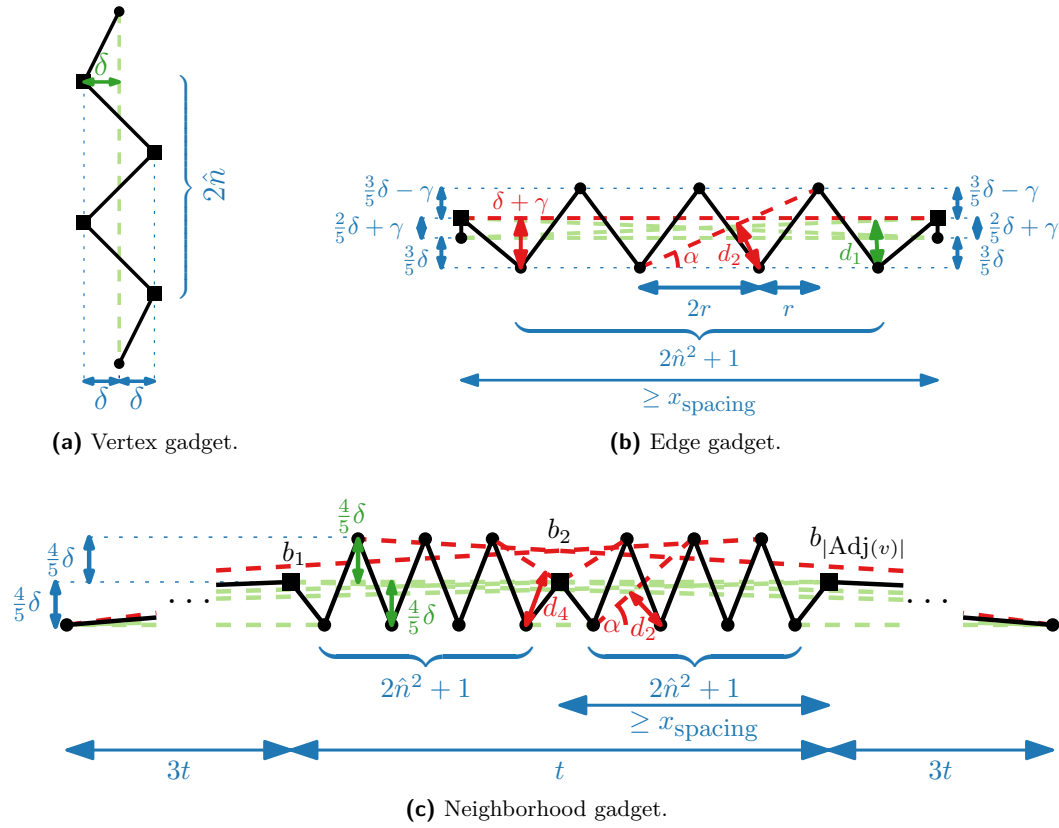
▷ **Claim 5.** In an edge gadget, there are precisely three long shortcuts. These are (i) from the first to the last bend, (ii) from the first to the second last bend, and (iii) from the second to the last bend. Beside these three shortcuts, there are  $\leq 4$  more shortcuts, which skip only the second and the second last bend (and possibly also the third and third last bend). There is no shortcut not skipping one of the shared bends, i.e., the second or the second last bend.

In (i), both of the shared bends, these are the second and the second last, are skipped and we can take the “long” shortcut from the first to the last bend because the line segment between them is horizontal and has y-distance  $3/5\delta$  or  $2/5\delta + \gamma$  or  $\delta$  to all inner bends. In (ii), the most critical part is the distance  $d_1$  between the third last bend and the straight-line segment from the first to the second last bend (see Figure 3b). It is

$$d_1 \leq \frac{3}{5}\delta + \left(\frac{2}{5}\delta + \gamma\right) - \frac{\frac{2}{5}\delta + \gamma}{2\hat{n}^2 + 2} \leq \delta + \gamma - \frac{\frac{2}{5} \cdot \frac{10\hat{n}^2 + 5}{2}\gamma + \gamma}{2\hat{n}^2 + 2} = \delta. \quad (6)$$

Observe that (iii) is the same as (ii) but mirrored. If neither the second nor the second last bend is skipped, i.e., if  $u$  and  $v$  are in the set  $V^*$ , then we cannot cut short anything in this gadget. Clearly, we cannot take a “long” shortcut from the second to the second last bend because the lower row of inner bends has distance  $\delta + \gamma$  from the potential shortcut segment.





■ **Figure 8** Some additional details to Fig. 3.

Moreover, we cannot take a “short” shortcut from a bend of the lower row to a bend of the upper row or the other way around. If we would aim to skip two inner bends, the distance  $d_2$  (see Figure 3b) from an inner bend to the shortcut segment would have to be at most  $\delta$ . However, it is

$$d_2 = \sin \alpha \cdot 2r = \sin \arctan \frac{\frac{8}{5}\delta}{3r} \cdot 2r = \frac{\frac{8\delta}{15r}}{\sqrt{\left(\frac{8\delta}{15r}\right)^2 + 1}} \cdot 2r = \frac{16\delta r}{\sqrt{(8\delta)^2 + (15r)^2}}, \quad (7)$$

where

$$r \geq \frac{x_{\text{spacing}}}{2\hat{n}^2 + 2} \geq \frac{(2\hat{n}^2 + 2)3\delta}{2\hat{n}^2 + 2} = 3\delta, \quad (8)$$

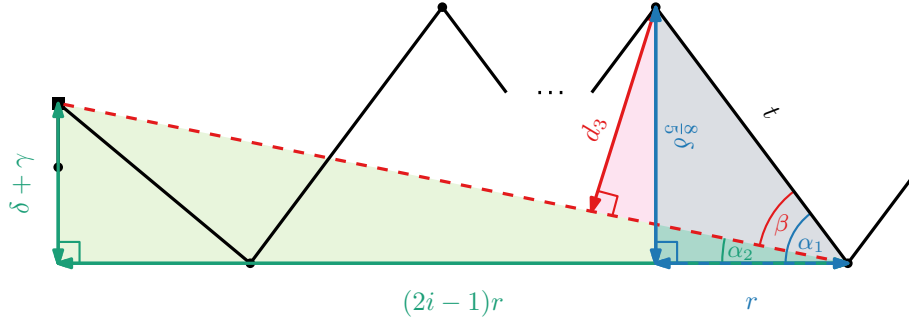
and hence,

$$d_2 \geq \frac{48\delta^2}{\sqrt{64\delta^2 + 2025\delta^2}} = \frac{48}{\sqrt{2089}}\delta = 1.0502 \dots \delta. \quad (9)$$

Observe that this becomes even greater if we aim for skipping four or more bends or if we start or end at one of the two shared bends. To make this clearer, we explicitly consider the latter case where a potential shortcut would start at the second bend and end at the  $(2i + 1)$ -th bend. This situation is depicted in Fig. 9. If it was a valid shortcut,  $d_3$  would be less than or equal to  $\delta$ . Since  $d_3$  is inside a rectangular triangle, its length is

$$d_3 = t \cdot \sin \beta, \quad (10)$$

### 35:18 Simplification of Polyline Bundles



■ **Figure 9** The potential shortcut segment in an edge gadget from the second bend to an inner bend (the  $(2i + 1)$ -th bend) is dashed in red. However,  $d_3 > \delta$  makes it no valid shortcut segment.

where  $t$  is inside another rectangular triangle with legs of length  $r$  and  $8/5\delta$ , so

$$t = \sqrt{r^2 + \left(\frac{8}{5}\delta\right)^2}. \quad (11)$$

We can determine  $\beta$  via the angles  $\alpha_1$  and  $\alpha_2$  as

$$\beta = \alpha_1 - \alpha_2 = \arctan \frac{\frac{8}{5}\delta}{r} - \arctan \frac{\delta + \gamma}{(2i-1)r}. \quad (12)$$

In the arctan-functions, all parameters are positive, so they live in the range  $[0, \pi/2)$ . Hence,  $\beta$  lives in the range  $(-\pi/2, \pi/2)$ . In this range, the sin-function is monotonously increasing. Therefore, to give a lower bound on  $d_3$ , we can use a lower bound on  $\sin \beta$  by specifying a lower bound on  $\beta$ . Since  $i \geq 2$ ,  $\gamma < \delta/(5\hat{n}^2)$  and  $\hat{n} \geq 1$ , we state that

$$\beta = \alpha_1 - \alpha_2 > \arctan \frac{\frac{8}{5}\delta}{r} - \arctan \frac{\frac{6}{5}\delta}{3r} = \arctan \frac{8}{5r'} - \arctan \frac{2}{5r'}, \quad (13)$$

where  $r' = r/\delta$ . A lower bound on  $t$  is

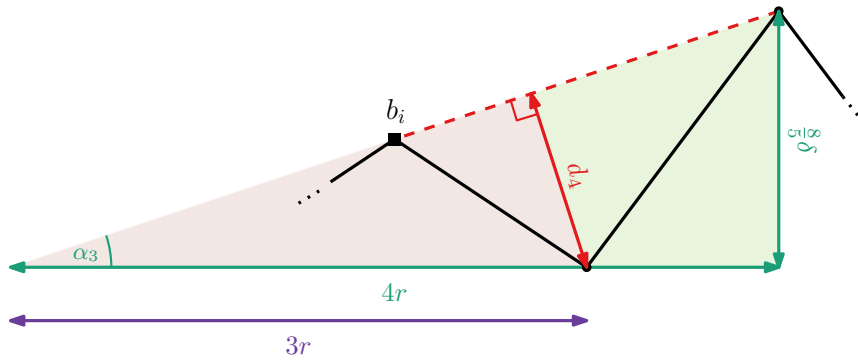
$$t = \sqrt{(r'\delta)^2 + \left(\frac{8}{5}\delta\right)^2} > r'\delta. \quad (14)$$

So, we can get a lower bound on  $d_3$  by

$$d_3 = t \cdot \sin \beta > r' \underbrace{\sin \left( \arctan \frac{8}{5r'} - \arctan \frac{2}{5r'} \right)}_{c(r')} \cdot \delta. \quad (15)$$

To prove that  $d_3$  is always greater than  $\delta$ , it suffices to show that the prefactor  $c(r')$  is equal to or greater than 1 for all possible values of  $r'$ . We reformulate  $c(r')$  using well-known trigonometric identities:

$$\begin{aligned} c(r') &= r' \sin \left( \arctan \frac{8}{5r'} - \arctan \frac{2}{5r'} \right) \\ &= \frac{6}{\sqrt{25 + \frac{68}{r'^2} + \frac{256}{25r'^4}}} \end{aligned} \quad (16)$$



■ **Figure 10** The potential shortcut segment in a neighborhood gadget from a bend  $b_i$  to an inner bend is dashed in red. However,  $d_4 > \delta$  makes it no valid shortcut segment.

For  $r' = 3$ , this is  $c(r') = 1.0495\dots$  and, from Equation (16), it is easy to see that  $c(r')$  is even greater for  $r' > 3$ . Thus, we conclude that  $d_3 > \delta$  always holds.

It remains to consider potential shortcuts starting or ending at the first or the last bend. Clearly, skipping only the second or second last bend is always possible. Skipping the second and the third bend or skipping the second last and the third last bend may sometimes be possible depending on how much the edge gadget is stretched horizontally. However, according to the previous analysis, skipping more bends is not possible since the distance between the potential shortcut segment and the bend before the end point of the potential shortcut is at least  $d_3$ .

▷ **Claim 6.** In a neighborhood gadget, the only shortcuts are (i) the shortcuts skipping only  $b_i$  for  $i \in \{1, \dots, |\text{Adj}(v)|\}$  and (ii) the shortcuts starting at the first bend or  $b_i$  with  $i \in \{1, \dots, |\text{Adj}(v)|\}$  and ending at the last bend or  $b_j$  with  $i < j \in \{1, \dots, |\text{Adj}(v)|\}$  – except for the shortcut starting at the first and ending at the last bend.

Clearly, the shortcuts (i) for skipping any  $b_i$  (or exactly one neighbor of  $b_i$ ) are valid and there is no shortcut from the first to the last bend since the potential shortcut segment has distance  $8/5\delta$  to the upper row of bends. In (ii), there clearly is a shortcut if we start at any  $b_i$  and end at any  $b_j$ . If we start at some  $b_i$  and end at the last bend, observe that, in the most extreme case, the segment from  $b_1$  to the last bend has a y-distance to the upper row of

$$\frac{4}{5}\delta + \frac{t}{4t} \cdot \frac{4}{5}\delta = \delta \tag{17}$$

when it passes  $b_{|\text{Adj}(v)|}$  in x-dimension. Thus, this shortcut is valid and the same holds for the shortcuts from the first bend to some  $b_j$ .

It remains to argue that there are no more shortcuts. A shortcut starting and ending at a bend on the upper or lower row is not possible because it would either be a horizontal segment, which has distance  $8/5\delta$  to the other row, or the distance to some bend in between would be at least  $d_2$ , which we have shown to be greater than  $\delta$  in Equations (7)–(9). It is easy to see that there is no shortcut starting at the first bend and ending at some inner bend of the upper or lower row. The same holds true for shortcuts starting at some inner bend of the upper or lower row and ending at the last bend.

Moreover, a shortcut segment starting (ending) at some  $b_i$  for  $i \in \{1, \dots, |\text{Adj}(v)|\}$  and skipping one bend would have a distance of  $d_4$  to this bend as depicted in Fig. 10. Since  $d_4$

## 35:20 Simplification of Polyline Bundles

is inside a rectangular triangle, we can determine  $d_4$  by

$$d_4 = 3r \cdot \sin \alpha_3, \quad (18)$$

where  $\alpha_3$  is in another rectangular triangle and thus can be determined by

$$\alpha_3 = \arctan \frac{\frac{8}{5}\delta}{4r} = \arctan \frac{2}{5r'}. \quad (19)$$

Putting them together, we get

$$d_4 = 3r'\delta \cdot \sin \arctan \frac{2}{5r'} = 3r'\delta \frac{\frac{2}{5r'}}{\sqrt{1 - \frac{4}{25r'^2}}} = \frac{6}{\sqrt{25 - \frac{4}{r'^2}}}\delta. \quad (20)$$

For  $r' = 3$ , this is  $1.2108 \dots \delta$  and again, for  $r' > 3$ ,  $d_4$  is even greater.

If we skip more than one inner bend, the distance to the last skipped bend becomes only greater. Hence, we conclude that Claim 6 is correct.