

METRIC COMBINATORICS OF CONVEX POLYHEDRA: CUT LOCI AND NONOVERLAPPING UNFOLDINGS

EZRA MILLER* AND IGOR PAK†

ABSTRACT. Let S be the boundary of a convex polytope of dimension $d + 1$, or more generally let S be a *convex polyhedral pseudomanifold*. We prove that S has a polyhedral nonoverlapping unfolding into \mathbb{R}^d , so the metric space S is obtained from a closed (usually nonconvex) polyhedral ball in \mathbb{R}^d by identifying pairs of boundary faces isometrically. Our existence proof exploits geodesic flow away from a source point $v \in S$, which is the exponential map to S from the tangent space at v . We characterize the *cut locus* (the closure of the set of points in S with more than one shortest path to v) as a polyhedral complex in terms of Voronoi diagrams on facets. Analyzing infinitesimal expansion of the wavefront consisting of points at constant distance from v on S produces an algorithmic method for constructing Voronoi diagrams in each facet, and hence the unfolding of S . The algorithm, for which we provide pseudocode, solves the discrete geodesic problem. Its main construction generalizes the source unfolding for boundaries of 3-polytopes into \mathbb{R}^2 . We present conjectures concerning the number of shortest paths on the boundaries of convex polyhedra, and concerning continuous unfolding of convex polyhedra. We also comment on the intrinsic non-polynomial complexity of nonconvex manifolds.

Introduction

The past several decades have seen intense development in the combinatorics and geometry of convex polytopes [Zie95]. Besides their intrinsic interest, the advances have been driven by applications to areas ranging as widely as combinatorial optimization, commutative algebra, symplectic geometry, theoretical physics, representation theory, statistics, and enumerative combinatorics. As a result, there is currently available a wealth of insight into (for example) algebraic invariants of the face posets of polytopes; arithmetic information connected to sets of lattice points inside polytopes; and geometric constructions associated with linear functionals, such as Morse-like decompositions and methods for locating extrema.

On the topological side, there are metric theories for polyhedral spaces, primarily motivated by differential geometry. In addition, there is a vast literature on general convexity. Nonetheless, there seems to be lacking a study of the interaction between the combinatorics of the boundaries of convex polytopes and their metric geometry in

Date: 22 February 2006.

Key words and phrases. Convex polyhedra, nonoverlapping unfolding, polyhedral metric, cut locus, discrete geodesic problem, geodesic Voronoi diagram, convex polyhedral pseudomanifold.

*School of Mathematics, University of Minnesota, Minneapolis, MN; ezra@math.umn.edu.

†Department of Mathematics, MIT, Cambridge, MA; pak@math.mit.edu.

arbitrary dimension. This remains the case despite relations to a number of classical algorithmic problems in discrete and computational geometry.

The realization here is that convexity and polyhedrality together impose rich combinatorial structures on the collection of shortest paths in a metrized sphere. We initiate a systematic investigation of this *metric combinatorics* of convex polyhedra by proving the existence of polyhedral nonoverlapping unfoldings and analyzing the structure of the cut locus. The algorithmic aspect, which we include together with its complexity analysis, was for us a motivating feature of these results. That being said, we also show that our general methods are robust enough so that—with a few minor modifications—they extend to the abstract spaces we call ‘convex polyhedral pseudomanifolds’, whose sectional curvatures along low-dimensional faces are all positive. To conclude, we propose some directions for future research, including a series of precise conjectures on the number of combinatorial types of shortest paths, and on the geometry of unfolding boundaries of polyhedra.

Overview

Broadly speaking, the metric geometry of boundaries of 3-dimensional polytopes is quite well understood, due in large part to work of A. D. Aleksandrov [Ale48, Ale50] and his school. For higher dimensions, however, less theory appears in the literature, partly because Aleksandrov’s strongest methods do not extend to higher dimension. Although there do exist general frameworks for dealing with metric geometry in spaces general enough to include boundaries of convex polyhedra, such as [BGP92], the special nature of polyhedral spaces usually plays no role.

The existing theory that does appear for polyhedral spaces is motivated from the perspective of Riemannian geometry, via metric geometry on simplicial complexes, and seems mainly due to D. A. Stone; see [Sto76], for example. In contrast, our original motivation comes from two classical problems in discrete and computational geometry: the ‘discrete geodesic problem’ [Mit00] of finding shortest paths between points on polyhedral surfaces, and the problem of constructing nonoverlapping unfoldings of convex polytopes [O’R00]. Both problems are well understood for the 2-dimensional boundaries of 3-polytopes, but have not been attempted in higher dimensions. We resolve them here in arbitrary dimension by a unified construction generalizing the ‘source unfolding’ of 3-dimensional convex polyhedra [VP71, SS86].

Previous methods for source unfoldings have been specific to low dimension, relying for example on the fact that arcs of circles in the plane intersect polygons in finite sets of points. We instead use techniques based on differential geometry to obtain general results concerning cut loci on boundaries of polytopes in arbitrary dimension, namely Theorem 2.9 and Corollary 2.11, thereby producing polyhedral foldouts in Theorem 3.5. In more precise terms, our two main goals in this paper are to:

1. describe how the set of points on the boundary S of a convex polyhedron at given radius from a fixed *source point* changes as the radius increases continuously;
2. use this description of ‘wavefront expansion’ to construct a polyhedral nonoverlapping unfolding of the d -dimensional polyhedral complex S into \mathbb{R}^d .

By ‘describe’ and ‘construct’ we mean to achieve these goals not just abstractly and combinatorially, but effectively, in a manner amenable to algorithmic computation. References such as [AAOS97, AO92, CH96, Kap99, MMP87, Mou85, SS06, SS86], which have their roots and applications in computational geometry, carry this out in the $d = 2$ case of boundaries of 3-polytopes (and for the first goal, on any polyhedral surface of dimension $d = 2$). Here, in arbitrary dimension d , our Theorem 5.2 says precisely how past wavefront evolution determines the location in time and space of its next qualitative change. The combinatorial nature of Theorem 5.2 leads immediately to Algorithm 6.1 for effectively unfolding boundaries of polyhedra.

The results and proofs in Sections 1–6 for boundaries of convex polyhedra almost all hold verbatim in the more abstract setting of what we call d -dimensional *convex polyhedral pseudomanifolds*. The study of such spaces is suggested both by Stone’s point of view in [Sto76] and by the more general methods in [BGP92]. Our Corollary 7.12 says that all convex polyhedral pseudomanifolds can be represented as quotients of Euclidean (usually nonconvex) polyhedral balls by identifying pairs of boundary components isometrically. The reader interested solely in this level of generality is urged to begin with Section 7, which gives a guide to Sections 1–6 from that perspective, and provides the slight requisite modifications where necessary. Hence the reader can avoid checking the proofs in the earlier sections twice.

The results in Section 7 on convex polyhedral pseudomanifolds are in many senses sharp, in that considering more general spaces would falsify certain conclusions. We substantiate this claim in Section 8, where we also discuss extensions of our methods that are nonetheless possible. For example, we present an algorithm to construct geodesic Voronoi diagrams on boundaries of convex polyhedra in Section 8.9.

The methods of this paper suggest a number of fundamental open questions about the metric combinatorics of convex polyhedra in arbitrary dimension, and we present these in Section 9. Most of them concern the notion of *vistal tree* in Definition 9.1, which encodes all of the combinatorial types of shortest paths (or equivalently, all bifurcations of the wavefront) emanating from a source point. The first two questions, Conjectures 9.2 and 9.4, concern the complexity of our unfolding algorithm and the behavior of geodesics in boundaries of polyhedra. Along these lines, we remark also on the complexity of nonconvex polyhedral manifolds, in Proposition 9.8. Our third question is about the canonical subdivision of the boundary of any convex polyhedron determined by the sets of source points having isomorphic vistal trees (Definition 9.5 and Conjecture 9.6); it asks whether this *vistal subdivision* is polyhedral, and how many faces it has. Our final question asks how to realize unfoldings of polyhedral boundaries by embedded homotopies (Conjecture 9.12).

As a guide for the reader navigating this paper, the list of sections is as follows:

0. Methods
1. Geodesics in polyhedral boundaries
2. Cut loci
3. Polyhedral nonoverlapping unfolding
4. The source poset
5. Constructing source images

6. Algorithm for source unfolding
7. Convex polyhedral pseudomanifolds
8. Limitations, generalizations, and history
9. Open problems and complexity issues

0. Methods

This section contains an extended overview of the paper, including background and somewhat informal descriptions of the geometric concepts involved.

Unfolding polyhedra. While unfolding convex polytopes is easy [Ale48], constructing a *nonoverlapping* unfolding is in fact a difficult task with a long history going back to Dürer in 1528 [Sch97]. When cuts are restricted to ridges (faces of dimension $d - 1$ in a polyhedron of dimension $d + 1$), the existence of such unfoldings is open even for polytopes in \mathbb{R}^3 [Sch97, O'R00]. It is known that nonconvex polyhedral surfaces need not admit such nonoverlapping unfoldings [Tar99, B+03].

In this paper we consider unfoldings of a more general nature: cuts are allowed to slice the interiors of facets. Nonoverlapping such unfoldings are known, but only for 3-dimensional polytopes [SS86, AO92, CH96, AAOS97]. In fact, two different (although strongly related) unfoldings appear in these and other references in the literature: the *Aleksandrov unfolding* (also known as the *star unfolding*) [Ale48, AO92], and the *source unfolding* [VP71, SS86]. Unfortunately, the construction of Aleksandrov unfoldings fails in principle in higher dimension (Section 8.4). As we mentioned earlier, we generalize the source unfolding construction to prove that the boundary S of any convex polyhedron of dimension $d + 1$, and more abstractly any convex polyhedral pseudomanifold S , has a nonoverlapping polyhedral unfolding \bar{U} in \mathbb{R}^d . The second of the following two foldouts of the cube is a $d = 2$ example of a source unfolding. For

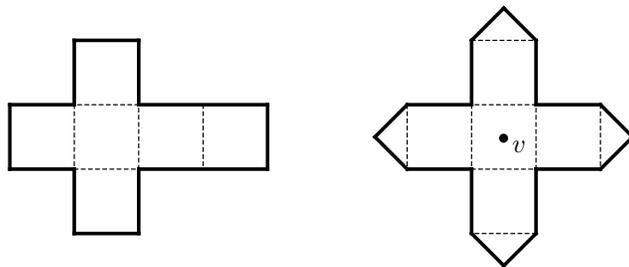


FIGURE 1. An edge-unfolding and a source unfolding of a cube into \mathbb{R}^2

clarity, we present the discussion below in the context of boundaries of polyhedra.

Cut loci. The idea of the source unfolding in arbitrary dimension d is unchanged from the case $d = 2$ of convex polyhedral surfaces. Pick a *source point* v interior to some facet (d -dimensional face) of S , so the tangent space T_v is well-defined. Then, treating S like a Riemannian manifold, define the *exponential map* from T_v to S by flowing along geodesics emanating from v . Our main unfolding result, Theorem 3.5,

says that exponentiation takes a certain open polyhedral ball $U_v \subset T_v$ isometrically to a dense open subset of S consisting of points possessing a unique shortest path (length-minimizing geodesic) to v . The image of the closure \overline{U}_v of the open ball U_v is all of S . The boundary $\overline{U}_v \setminus U_v$ maps onto the *cut locus* \overline{K}_v , which by definition is the closure of the set of points in S with more than one shortest path to the source point v . These properties characterize U_v .

In Riemannian geometry, when the manifold and the metric are both smooth, describing the cut locus for a source point is already an important and interesting problem (see [Kob89] for an excellent introduction and numerous references), although of course the exponential map can only be an isometry, even locally, if the metric is flat. Extending the notion of cut locus from Riemannian geometry to the polyhedral context is just as easy as extending it to arbitrary metric spaces. But showing that the open ball U_v is a polyhedral foldout requires strong conditions on the complement of the cut locus, such as metric flatness and polyhedrality. We prove these results in Sections 1 and 2 using methods based on the foundations of polyhedral geometry, and on Voronoi diagrams, culminating in Theorem 2.9 and Corollary 2.11. These conclusions depend crucially on convexity and do not hold in the nonconvex case.

Geometry of wavefront expansion. Our existence proof for polyhedral nonoverlapping source foldouts, even given their Voronoi characterization in Theorem 2.9, does not by itself provide a satisfactory combinatorial picture of the dynamics of wavefront expansion on polyhedra. For this, we must gain control over how the exponential map behaves as it interacts with *warped* points in S , namely those of nontrivial curvature,¹ or equivalently points on faces of dimension $d - 2$ or less.

Imagine the picture kinetically: the source point v emits a signal, whose wavefront proceeds as a $(d - 1)$ -sphere of increasing radius—at least until the sphere hits the boundary of the facet containing v . At that stage, the wavefront folds over a *ridge*, or face of dimension $d - 1$. Metrically, nothing has happened: points interior to ridges look to the wavefront just as flat as points interior to facets. But later, as the wavefront encounters faces of lower dimension, it is forced to bifurcate around warped points and interfere with itself, as signals emitted originally in different directions from v curl around the nontrivial curvature and converge toward the cut locus.

The question becomes: What discrete structure governs evolution of the wavefront on polyhedra? The most obvious first step is to define a finite collection of ‘events’, representing the points in time and space where the wavefront changes in some nontrivial way. If this is done properly, then it remains only to order the events according to the times at which they occur. However, in reality, the definition of an event is rather simple, while the geometry dictating time order of events is more complex.

Starting from scratch, one might be tempted (and we were) to mark an event every time the wavefront encounters a new warped face. Indeed, this works in dimension $d = 2$ [MMP87]: since the wavefront is a curve, its intersection with the set of edges

¹In differential geometry, when polyhedra are expressed as limits of (sequences of) smooth Riemannian manifolds, all of the curvature is forced into decreasing neighborhoods of the $(d-2)$ -skeleton. Consequently, the curvature actually tends to infinity near faces of dimension $d - 2$ or less, even though in a polyhedral sense the curvature is finite.

is a finite set, and it is easy to detect when one of these points hits a vertex of S before another. But because the geometry is substantially more complicated in higher dimensions, in the end we found it more natural to say an *event* has occurred every time the wavefront encounters a new facet through the relative interior of a *ridge* (see Definitions 2.3 and 4.14). This may seem counterintuitive, since the wavefront only interacts with and curls around faces of smaller dimension. However, wavefront collisions with warped points lead to intersections with ridge interiors infinitesimally afterward. In other words, the closest point (*event point*) on a facet to the source point v need not lie interior to a ridge, but can just as easily be warped.

Again think kinetically: once the wavefront has hit a new face (of small dimension, say), it begins to creep up each of the ridges containing that face. Although in a macroscopic sense the wavefront hits all of these ridges simultaneously, it creeps up their interiors at varying rates. Therefore the wavefront hits some of these ridges before others in an infinitesimal sense. The moral is that if one wants to detect curling of the wavefront around warped faces, it is simpler to detect the wake of this interaction infinitesimally on the interiors of neighboring ridges. Sufficiently refined tangent data along ridges then discretizes the finite set of events, thereby producing the desired ‘metric combinatorics’ of wavefront expansion.

Source poset. Making the above moral precise occupies Section 4. To single out a ridge whose interior is engulfed by the wavefront at a maximal rate (thereby making it *closer* to the source point) essentially is to find a ridge whose angle with the corresponding signal ray emitted from the source is minimal. When $d = 2$, this means that we do not simply observe two signals hitting vertices simultaneously, but we notice also the angles at which they hit the edges containing those vertices. The edge forming the smallest angle with its signal ray is the earlier event, infinitesimally beating out other potential events. (That each angle must be measured inside some ambient facet is just one of the subtleties that we gloss over for now.)

To distinguish events in time macroscopically, only radii (distances from the source) are required. When $d = 2$, as we have just seen, a first derivative is enough to distinguish events infinitesimally. Generally, in dimension $d \geq 2$, one needs derivatives of order less than d , or more precisely, a directional derivative successively along each of $d - 1$ orthogonal directions inside a ridge. In Section 4, these derivatives are encoded not in single angles, but in *angle sequences* (Definition 4.2), which provide quantitative information about the goniometry of intersections between signal rays and the faces of varying dimension they encounter. More qualitative—and much more refined—data is carried by *minimal jet frames* (Definition 4.1), which record not just the sizes of the angles, but their directions as well.

The totality of the (finite amount of) radius and angle sequence data induces a partial order on events. The resulting *source poset* (Definition 4.14), which owes its existence to the finiteness result in Theorem 4.11, describes precisely which events occur before others—both macroscopically and infinitesimally. Since wavefront bifurcation is a local phenomenon at an event point, incomparable events can occur simultaneously, or can be viewed as occurring in any desired order. Thus as time progresses, wavefront expansion builds the source poset by adding one event at a time.

The algorithm. It is one thing to order the set of events after having been given all of them; but it is quite another to predict the “next” event having been given only past events. That the appropriate event to add can be detected locally, and *without knowing future events*, is the content of Theorem 5.2. Its importance is augmented by it being the essential tool in making our algorithm for constructing the source poset, and hence also the source unfolding (Algorithm 6.1). Surprisingly, our geometric analysis of infinitesimal wavefront expansion in Sections 4 and 5 allows us to remove all calculus from Theorem 5.2 and hence Algorithm 6.1: detecting the next event requires only standard tools from linear algebra.

As we mentioned earlier, our original motivation for this paper was its algorithmic applications. Using the theoretical definitions and results in earlier sections, we present pseudocode for our procedure constructing source unfoldings in Algorithm 6.1. That our algorithm provides an *efficient* method to compute source unfoldings is formalized in Theorem 6.5.

There are several arguments in favor of presenting pseudocode. First, it underscores the explicit effective nature of our combinatorial description of the source poset in Theorem 5.2. Second, it emphasizes the simplicity of the algorithm that results from the apparently complicated analysis in Sections 1–5; in particular, the reader interested only in the computational aspects of this paper can start with Section 6 and proceed backwards to read only those earlier parts of the paper addressed in the algorithm. Finally, the pseudocode makes Algorithm 6.1 amenable to actual implementation, which would be of interest but lies outside the scope of this work.²

A note on the exposition. Proofs of statements that may seem obvious based on intuition drawn from polyhedral surfaces, or even solids of dimension 3, demand surprising precision in the general case. Occasionally, the required adjustments in definitions and lemmas, and even in statements of theorems, were borne out only after considering configurations in dimension 5 or more. The definition of source image is an example, about which we remark in Section 8, in the course of analyzing where various hypotheses (convexity, pseudomanifold, and so on) become essential. Fortunately, once the appropriate notions have been properly identified, the definitions become transparent, and the proofs remain intuitive in low dimension.

1. Geodesics in polyhedral boundaries

In this paper, a *convex polyhedron* F of dimension d is a finite intersection of closed half-spaces in some Euclidean space \mathbb{R}^d , such that F does not lie in a proper affine subspace of \mathbb{R}^d . The polyhedron F need not be bounded, and comes with an induced Euclidean metric. Gluing a finite collection of convex polyhedra by given isometries on pairs of codimension 1 faces yields a (*finite*) *polyhedral cell complex* S . More precisely, S is a regular cell complex endowed with a metric that is piecewise Euclidean, in which every face (closed cell) is isometric to a convex polyhedron.

²We refer the reader to a recent efficient implementation [S+05] of the classical $d = 2$ algorithm in [MMP87].

The case of primary interest is when the polyhedral cell complex S equals the boundary ∂P of a convex polyhedron P of dimension $d + 1$ in \mathbb{R}^{d+1} .

Convention 1.1. We assume that $S = \partial P$ is a polyhedral boundary in all theorems, proofs, and algorithms from here through Section 6.

We do not require P to be bounded, though the reader interested in polytopes will lose very little of the flavor by restricting to that case. Moreover, with the exception of Lemma 1.3, Proposition 2.10, Corollary 2.11, and Theorem 3.5, the statements of all results from here through Section 6 are worded to hold verbatim for the more abstract class of *convex polyhedral pseudomanifolds*, as we shall see in Section 7.

Denote by μ the metric on S , so $\mu(a, b)$ denotes the distance between points $a, b \in S$. A path $\gamma \subset S$ with endpoints a and b is a *shortest path* if its length equals $\mu(a, b)$. Since we assume S has finitely many *facets* (maximal faces), such length-minimizing paths exist, and are piecewise linear. A path $\eta \subset S$ is a *geodesic* if η is locally a shortest path; i.e. for every $z \in \eta$ that is not an endpoint of η , there exist points $a, b \in \eta \setminus \{z\}$ such that $z \in \gamma \subset \eta$ for some shortest path γ connecting a to b .

Henceforth, as S has dimension d , a face of dimension $d - 1$ will be called a *ridge*. For convenience, we say that a point x is *warped* if x lies in the union S_{d-2} of all faces in S of dimension at most $d - 2$, and call x *flat* otherwise. Every flat point has a neighborhood isometric to an open subset of \mathbb{R}^d .

Proposition 1.2. *If γ is a shortest path in S between its endpoints, then γ has no warped points in its relative interior.*

Proof. For any point w lying in the relative interior of γ , the intersection of γ with some neighborhood of w consists of two line segments η and η' that are each straight with one endpoint at w , when viewed as paths in \mathbb{R}^{d+1} . This is a consequence of local length-minimization and the fact that each facet of P is isometric to a polytope in \mathbb{R}^d . Moreover, if w happens to lie on a ridge while η intersects the relative interior of some facet containing w , then local length-minimization implies that η' is not contained in the facet containing η . Lemma 1.3 shows that w does not lie in S_{d-2} , so the point w is not warped. \square

Lemma 1.3. *Let $\eta, \eta' \subset S$ be two paths that (i) are straight in \mathbb{R}^{d+1} , (ii) share a common warped endpoint $w \in S_{d-2}$, and (iii) do not both lie in a single facet. There exists a neighborhood \mathcal{O} of w in S such that for every $a \in \eta \cap \mathcal{O}$ and $b \in \eta' \cap \mathcal{O}$, the path η_{ab} from a to w to b along η and η' is not a shortest path in S between a and b .*

Proof. Translate P so that w equals the origin $\mathbf{0} \in \mathbb{R}^{d+1}$, and let Q be the unique minimal face of P that contains w . Since η and η' do not lie in a single facet, the 2-plane E spanned by η and η' meets Q at exactly one point, namely $\mathbf{0}$. Since $\dim(Q) \leq d - 2$, the span of Q and E has dimension at most d . Choose a line L whose direction is linearly independent from the span of Q and E . Then the 3-plane $H = L + E$ intersects Q only at $\mathbf{0}$. Replacing P by $P \cap H$, we can assume that $\dim(P) = 3$, so that $d = 2$; note that $\mathbf{0}$ is a vertex of $H \cap P$ by construction.

Although the case $d = 2$ was proved in [Ale48, Theorem 4.3.5] (see also [SS86, Lemma 4.1]), we provide a simple argument here, for completeness. Let $\mathcal{O} \subset S$ be

the neighborhood of w consisting of all points at some fixed small distance from the vertex w . Then \mathcal{O} can be laid flat on the plane \mathbb{R}^2 by slicing along η . One of the two points in this unfolding that glue to $a \in \mathcal{O}$ connects by a straight segment in the unfolding to the unique point corresponding to b . This straight segment shortcuts η_{ab} after gluing back to S . \square

An illustration of Lemma 1.3 and its proof is given in Figure 2 below.

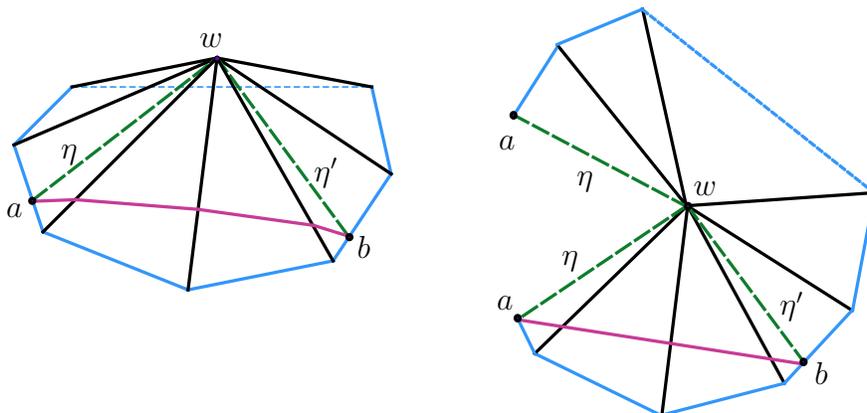


FIGURE 2. Neighborhood of a vertex and its foldout after slicing along the segment η . The points a and b are connected by a shortest path.

Corollary 1.4. *Let η be a bounded geodesic in S starting at a point z not on any ridge. Then η intersects each ridge in a discrete set, so η traverses (in order) the interiors of a well-defined sequence \mathcal{L}_η of facets (the facet sequence of η).*

Proof. Since η is locally length minimizing, Proposition 1.2 implies that every intersection of η with a ridge takes place at a flat point. Such points have neighborhoods isometric to open subsets of \mathbb{R}^d , and these intersect η in paths isometric to straight segments. It follows that η intersects every ridge transversely. \square

For each facet F of $S = \partial P$, let T_F be the affine span of F in \mathbb{R}^{d+1} .

Definition 1.5. Suppose two facets F and F' share a ridge $R = F \cap F'$. The *folding map* $\Phi_{F,F'} : T_F \rightarrow T_{F'}$ is the isometry that identifies the copy of R in T_F with the one in $T_{F'}$ in such a way that the image of F does not intersect the interior of F' .

In other words, the folding map $\Phi_{F,F'}$ is the rotation of T_F with $(d-1)$ -dimensional axis $R = F \cap F'$ so that F becomes coplanar with F' and lies on the other side of R from F' . It can be convenient to view $\Phi_{F,F'}$ as rotating all of \mathbb{R}^{d+1} instead of only rotating T_F onto $T_{F'}$. Informally, we say $\Phi_{F,F'}$ *folds* T_F *along* R to lie in the same affine hyperplane as F' .

Definition 1.6. Given an ordered list $\mathcal{L} = (F_1, F_2, \dots, F_\ell)$ of facets such that F_i shares a (unique) ridge with F_{i+1} whenever $1 \leq i < \ell$, we write

$$\Phi_{\mathcal{L}}^{-1} = \Phi_{F_1, F_2}^{-1} \circ \Phi_{F_2, F_3}^{-1} \circ \dots \circ \Phi_{F_{\ell-1}, F_\ell}^{-1}$$

for the *unfolding of T_{F_ℓ} onto T_{F_1}* , noting that indeed $\Phi_{\mathcal{L}}^{-1}(T_{F_\ell}) = T_{F_1}$. Setting $\mathcal{L}_i = (F_1, \dots, F_i)$, the *sequential unfolding* of a subset $\Gamma \subseteq F_1 \cup \dots \cup F_\ell$ *along \mathcal{L}* is the set

$$(\Gamma \cap F_1) \cup \Phi_{\mathcal{L}_2}^{-1}(\Gamma \cap F_2) \cup \dots \cup \Phi_{\mathcal{L}_\ell}^{-1}(\Gamma \cap F_\ell) \subset T_{F_1},$$

By Corollary 1.4, we can sequentially unfold any geodesic. Next, we use this unfolding to show uniqueness of shortest paths traversing given facet sequences.

Lemma 1.7. *Let v and w be flat points in S . Given a sequence \mathcal{L} of facets, there can be at most one shortest path γ connecting v to w such that γ traverses $\mathcal{L}_\gamma = \mathcal{L}$.*

Proof. Let γ be a shortest path from v to w traversing \mathcal{L} . Inside the union of facets appearing in \mathcal{L} , the relative interior of γ has a neighborhood isometric to an open subset of \mathbb{R}^d by Proposition 1.2 and the fact that the set of warped points is closed. Sequential unfolding of γ into T_F for the first facet F in \mathcal{L} thus yields a straight segment in T_F . This identifies γ uniquely as the path in S whose sequential folding along \mathcal{L} is the straight segment in T_F connecting v to $\Phi_{\mathcal{L}}^{-1}(w) \in T_F$. \square

In the proof of Lemma 1.7, we do not claim that the union of facets in the list \mathcal{L} unfolds sequentially without overlapping, even though some shortest path γ traverses \mathcal{L} . However, some neighborhood of γ in this union of facets unfolds without overlapping.

Example 1.8. Consider the unfolding of a $1 \times 1 \times 3$ rectangular box as in Figure 3. Denote by $F_{\text{bot}}, F_{\text{top}}, F_{\text{front}}, F_{\text{back}}, F_{\text{left}}, F_{\text{right}}$ the bottom, top, front, back, left, and right facets, respectively. Denote by \mathcal{L}_i the list of facets along which the points in the region marked by i have been sequentially unfolded to create the foldout $U \subset T_{F_{\text{bot}}}$ in Figure 3. Then:

$$\begin{aligned} \mathcal{L}_1 &= (F_{\text{bot}}), & \mathcal{L}_2 &= (F_{\text{bot}}, F_{\text{back}}), & \mathcal{L}_3 &= (F_{\text{bot}}, F_{\text{front}}), \\ \mathcal{L}_4 &= (F_{\text{bot}}, F_{\text{back}}, F_{\text{top}}), & \mathcal{L}_5 &= (F_{\text{bot}}, F_{\text{front}}, F_{\text{top}}), & \mathcal{L}_6 &= (F_{\text{bot}}, F_{\text{back}}, F_{\text{left}}), \\ \mathcal{L}_7 &= (F_{\text{bot}}, F_{\text{front}}, F_{\text{left}}), & \mathcal{L}_8 &= (F_{\text{bot}}, F_{\text{back}}, F_{\text{right}}), & \mathcal{L}_9 &= (F_{\text{bot}}, F_{\text{front}}, F_{\text{right}}). \end{aligned}$$

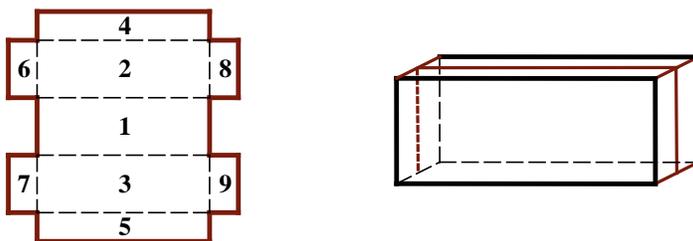


FIGURE 3. An unfolding of a $1 \times 1 \times 3$ box.

2. Cut loci

Most of this paper concerns the set of shortest paths with one endpoint fixed.

Definition 2.1. Fix a *source point* $v \in S$ lying interior to some facet. A point $x \in S$ is a *cut point*³ if x has more than one shortest path to v . Denote the set of cut points by K_v , and call its closure the *cut locus* $\overline{K}_v \subset S$.

Here is a consequence of Proposition 1.2.

Corollary 2.2. *No shortest path in S to the source point v has a cut point in its relative interior.*

Proof. Suppose c is a cut point in the relative interior of a shortest path from v to w . Replacing the path from v to c with another shortest path from v to c yields a new shortest path from v to w . These two paths to w meet at the flat point $c \in S$ by Proposition 1.2. The resulting Y-shaped intersection at c can be improved upon in a neighborhood of c isometric to an open set in \mathbb{R}^d (Fig. 4), a contradiction. \square

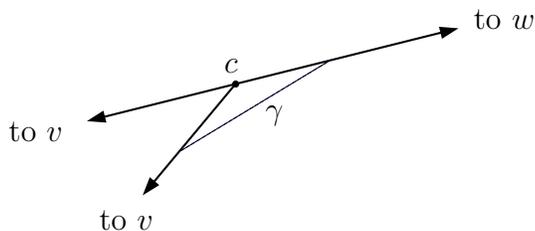


FIGURE 4. An intersection that is Y-shaped cannot locally minimize length in \mathbb{R}^d (segment γ is a shortcut).

Our study of polyhedrality of cut loci will use Voronoi diagrams applied to sets of points from the forthcoming definition, around which the rest of the paper revolves.

Definition 2.3. Suppose that the source point v connects by a shortest path γ to a point x that lies on a facet F or on one of its ridges $R \subset F$, but not on any face of S of dimension $d-2$ or less. If the sequential unfolding of γ into T_F is the segment $[\nu, x]$, then $\nu \in T_F$ is called a *source image* for F . Let src_F be the set of source images for F .

Lemma 2.4. *The set src_F of source images for any facet F of S is finite.*

Proof. The shortest path in \mathbb{R}^{d+1} between any pair of distinct points x and y in a facet F is the straight segment $[x, y]$. Since this segment is actually contained in S , any shortest path γ in S must contain $[x, y]$ whenever it contains both x and y . Taking x and y to be the first and last points of intersection between γ and the facet F , we find that F can appear at most once in the facet sequence of a shortest path starting at the source point v . Hence there are only finitely many possible facet sequences of shortest paths in S . Now apply Lemma 1.7. \square

³Our usage of the term ‘cut locus’ is standard in differential geometry, just as our usage of ‘ridge’ is standard in polyhedral geometry. However, these usages do not agree with terminology in computer science, such as in [SS86, AO92]: their ‘ridge points’ are what we call ‘cut points’. Furthermore, ‘cut points’ in [AO92] are what we would call ‘points on shortest paths to warped points’ (when $d = 2$).

Example 2.5. Consider a unit cube with a source point in its bottom face, as in Fig 5. Then the top face has 12 source images, shown in Fig 5. The four stars “ \star ” are sequential unfoldings of the source point (along three ridges each) that are not source images: each point in the top face is closer to some source image than to any of these stars.

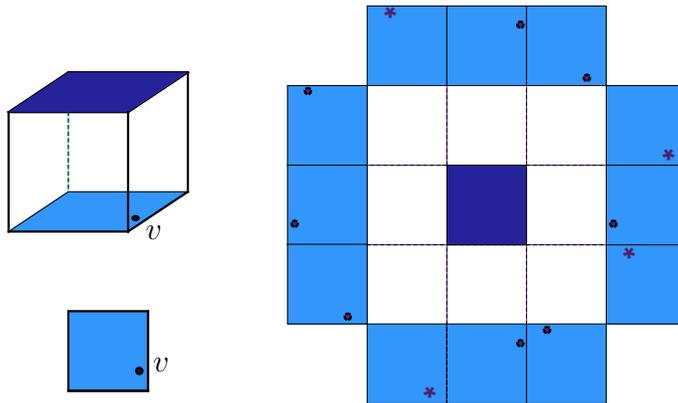


FIGURE 5. Source point v on the ‘bottom’ face, 12 source images for the ‘top’ face of a cube and 4 ‘false’ source images (view from the top).

Making Lemma 2.4 quantitative is one of our main open problems; see Section 9.

The next result on the way to Theorem 2.9 generalizes [Mou85, Lemma 3.1] to arbitrary dimension. Its proof is complicated somewhat by the fact (overlooked in the proof of [Mou85, Lemma 3.1]⁴) that straight segments can lie inside the cut locus, and our lack of *a priori* knowledge that the cut locus is polyhedral.

Proposition 2.6 (Generalized Mount’s lemma). *Let F be a facet of S , and suppose that $\nu \in \text{src}_F$ is a source image. If $w \in F$, then the straight segment $[\nu, w] \subset T_F$ has length at least $\mu(v, w)$.*

Example 2.7. In Fig. 6, the left figure is a typical illustration of Proposition 2.6 in dimension $d = 2$: any segment from a source image to $w \in F$ is weakly longer than the one contained in the region with w , and that one sequentially folds to a path in S of length $\mu(v, w)$. In contrast, the right figure will never occur: any point w interior to the shaded region is closest to the source image ν , but the straight segment connecting w to ν has not been sequentially unfolded along the correct facet sequence.

Proof. Since the two functions $F \rightarrow \mathbb{R}$ mapping w to $\mu(v, w)$ and to the length of $[\nu, w]$ are continuous, we can restrict our attention to those points w lying in any dense subset of F . In particular, the cut locus has dense complement in F (Corollary 2.2) as does the boundary of F , so we assume throughout that w lies in neither the cut locus nor the boundary of F .

⁴Much of [Mou85], but not Lemma 3.1 there, was later incorporated and published in [MMP87].

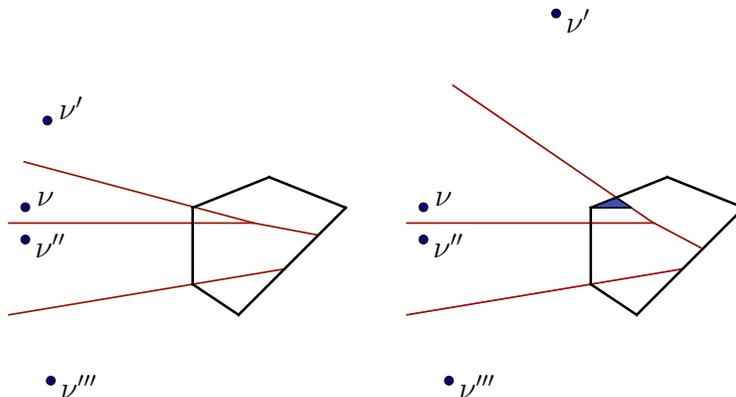


FIGURE 6. Generalized Mount's lemma (fails for the shaded region)

Having fixed $\nu \in \text{src}_F$, choose a point $x \in F$ as in Definition 2.3, so ν connects to x by a shortest path γ that sequentially unfolds to yield the segment $[\nu, x]$ in T_F . The set $\text{src}_F([x, w])$ of source images sequentially unfolded from shortest paths that end inside the segment $[x, w]$ is finite by Lemma 2.4. Hence we may furthermore assume that w does not lie on any hyperplane H that is equidistant from ν and a source image $\nu' \in \text{src}_F([x, w])$. In other words, we assume w does not lie inside the hyperplane perpendicularly bisecting any segment $[\nu, \nu']$.

Claim 2.8. *With these hypotheses, if $\nu \in \text{src}_F$ but no shortest path unfolds sequentially to the segment $[\nu, w]$, then w is closer to some point $\nu' \in \text{src}_F([x, w])$ than to ν .*

Assuming this claim for the moment, we may replace ν with ν' and x with another point x' on $[x, w]$. Repeating this process and again using that the set of source images sequentially unfolded from shortest paths ending in $[x, w]$ is finite, we eventually find that the unique source image $\omega \in \text{src}_F([x, w])$ closest to w is closer to w than ν is. Since $[\omega, w]$ has length $\mu(\omega, w)$, it suffices to prove Claim 2.8.

Consider the straight segment $[x, w]$, which is contained in F by convexity. Let Y be the set of points $y \in [x, w]$ having a shortest path γ_y from ν that sequentially unfolds to a segment in T_F with endpoint ν . Then Y is closed because any limit of shortest paths from ν traversing a fixed facet sequence \mathcal{L} is a shortest path that sequentially unfolds along \mathcal{L} to a straight segment from the corresponding source image. Thus, going from x to w , there is a last point $x' \in Y$. This point x' is by assumption not equal to w , so x' must be a cut point (possibly $x = x'$).

There is a facet sequence \mathcal{L} and a neighborhood \mathcal{O} of x' in $[x', w]$ such that every point in \mathcal{O} connects to ν by a shortest path traversing \mathcal{L} , and such that unfolding the source along \mathcal{L} yields a source image $\nu' \neq \nu$ in T_F . This point ν' connects to x' by a segment of length $\mu(\nu, x')$, so the hyperplane H perpendicularly bisecting $[\nu, \nu']$ intersects $[x, w]$ at x' . By hypothesis $w \notin H$, and it remains to show that w lies on the side of H closer to ν' .

The shortest path from ν to x' has a neighborhood in S disjoint from the set of warped points and hence isometric to an open subset of \mathbb{R}^d by Proposition 1.2, because

x' is itself not a warped point (we assumed x lies interior to F or to a ridge $R \subset F$). After shrinking \mathcal{O} if necessary, we can therefore ensure that each segment $[\nu, y]$ for $y \in \mathcal{O}$ is the sequential unfolding of a geodesic η_y in S . The geodesic η_y for $y \in \mathcal{O} \setminus x'$ cannot be a shortest path by definition of x' , so $[\nu, y]$ has length strictly greater than $\mu(v, y)$. We conclude that $\mathcal{O} \setminus x'$, and hence also w , lies strictly closer to ν' than to ν . This finishes the proof of Claim 2.8 and with it Proposition 2.6. \square

Before stating the first main result of the paper, we recall the standard notion of *Voronoi diagram* $\mathcal{V}(\Upsilon)$ for a closed discrete set $\Upsilon = \{\nu, \nu', \dots\}$ of points in \mathbb{R}^d . This is the subdivision of \mathbb{R}^d whose closed cells are the sets

$$V(\Upsilon, \nu) = \{\zeta \in \mathbb{R}^d \mid \text{every point } \nu' \in \Upsilon \text{ satisfies } |\zeta - \nu| \leq |\zeta - \nu'| \}.$$

Thus ζ lies in the interior of $V(\Upsilon, \nu)$ if ζ is closer to ν than to any other point in Υ .

Theorem 2.9. *Fix a facet F of S , and let $V_{d-1} \subseteq T_F$ be the union of the closed cells of dimension $d - 1$ in the Voronoi diagram $\mathcal{V}(\text{src}_F)$ for the set of source images in T_F . If F° is the relative interior of F , then the set $F^\circ \cap K_v$ of cut points in F° coincides with the intersection $F^\circ \cap V_{d-1}$. Moreover, if R° is the relative interior of a ridge $R \subset F$, then the set $R^\circ \cap K_v$ of cut points in R° coincides with $R^\circ \cap V_{d-1}$.*

Proof. Every shortest path from the source v to a point w in F° or R° unfolds to a straight segment in T_F of length $\mu(v, w)$ ending at a source image for F . Proposition 2.6 therefore says that w lies in the Voronoi cell $V(\Upsilon, \nu)$ if and only if the segment $[\nu, w]$ has length exactly $\mu(v, w)$. In particular, v has at least two shortest paths to w if and only if w lies in two such Voronoi cells—that is, $w \in V_{d-1}$. \square

To illustrate Theorem 2.9 consider Example 2.5. The Voronoi diagram of source images gives the cut locus in the top face of the cube.

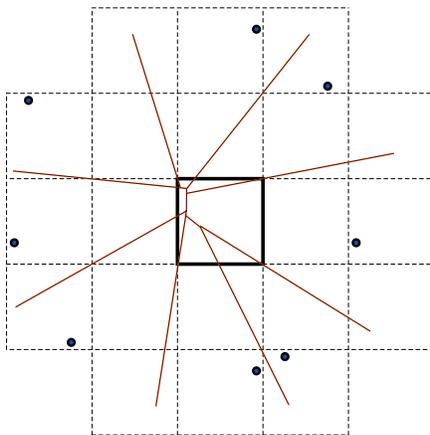


FIGURE 7. Cut locus of the ‘top’ face of the cube.

Theorem 2.9 characterizes the intersection of the cut locus with faces of dimension d or $d - 1$ in S . For faces of smaller dimension, we can make a blanket statement.

Proposition 2.10. *Every warped point lies in the cut locus \overline{K}_v ; that is, $S_{d-2} \subseteq \overline{K}_v$.*

Proof. It is enough to show that every point w in the relative interior of a warped face of dimension $d - 2$ is either a cut point or a limit of cut points, because the cut locus \overline{K}_v is closed by definition. Let γ be a shortest path from w to v .

First assume that every neighborhood of w contains a point having no shortest path to v that is a deformation of γ . Suppose that $(y_i)_{i \in \mathbb{N}}$ is a sequence of such points approaching w , with shortest paths $(\gamma_i)_{i \in \mathbb{N}}$ connecting the points y_i to v . Since there are only finitely many facets containing w and finitely many source images for each facet, we may assume (by choosing a subsequence if necessary) that for all i , the sequential unfolding of γ_i connects to the same source image for the same facet. The paths $(\gamma_i)_{i \in \mathbb{N}}$ then converge to a shortest path $\gamma' \neq \gamma$ to w from v , so $w \in K_v$.

Now assume that every point in some neighborhood of w has a shortest path to v that is a deformation of γ . Every point on γ other than w itself is flat in S by Proposition 1.2. Therefore some neighborhood of γ in S is isometric to an open subset of a product $\mathbb{R}^{d-2} \times C$, where C is a 2-dimensional surface that is flat everywhere except at one point $c \in C$ (so C is the boundary of a right circular cone with apex c). The set of points in $\mathbb{R}^{d-2} \times C$ having multiple geodesics to the image of v in $\mathbb{R}^{d-2} \times C$ is a relatively open half-space of dimension $d - 1$ whose boundary is $\mathbb{R}^{d-2} \times \{c\}$. Some sequence in this open half-space converges to the image of w . \square

Theorem 2.9 and Proposition 2.10 imply the following description of cut loci. For terminology, a subset K of a polyhedral cell complex is called *polyhedral* if its intersection with every facet is a union of convex polyhedra. Note that K need not be a polyhedral cell complex (as defined before Convention 1.1) because it might not come with a cell decomposition. Nonetheless, K can be made into a polyhedral cell complex by suitably subdividing. We call K *pure* of dimension k if it is the closure of a set whose dimension locally near every point is k .

Corollary 2.11. *If v is a source point in S , then*

1. *the cut locus \overline{K}_v is polyhedral and pure of dimension $d - 1$, and*
2. *the cut locus \overline{K}_v is the union $K_v \cup S_{d-2}$ of the cut points and warped points.*

Proof. Part 2 is a consequence of Theorem 2.9 and Proposition 2.10, the latter taking care of S_{d-2} , and the former showing that points in the cut locus but outside of S_{d-2} are in fact cut points. Since Voronoi diagrams in Euclidean spaces are polyhedral, Theorem 2.9 also implies the polyhedrality in part 1. For the purity, note that if P is any cut point, then the cut set divides a small neighborhood of P into finitely many regions (the regions being determined by the combinatorial types of shortest paths ending therein), with P lying in the closures of at least two of these regions. \square

3. Polyhedral nonoverlapping unfolding

In this section we again abide by Convention 1.1, so S is the boundary of convex polyhedron P of dimension $d + 1$ in \mathbb{R}^{d+1} .

Definition 3.1. A polyhedral subset $K \subset S$ of dimension $d - 1$ is a *cut set* if K contains the union S_{d-2} of all closed faces of dimension $d - 2$, and $S \setminus K$ is open and

contractible. A *polyhedral unfolding* of S into \mathbb{R}^d is a choice of cut set K and a map $S \setminus K \rightarrow \mathbb{R}^d$ that is an isometry locally on $S \setminus K$. A *nonoverlapping foldout* of S is a surjective piecewise linear map $\varphi : \overline{U} \rightarrow S$ such that

1. \overline{U} is the closure of its interior U , which is an open topological ball in \mathbb{R}^d , and
2. the restriction of φ to U is an isometry onto its image.

Note that K is not required to be a polyhedral subcomplex of S , but only a subset that happens to be a union of polyhedra; thus K can ‘slice through interiors of facets’. The open ball U in item 1 of the definition is usually nonconvex. The polyhedron P is a polytope if and only if \overline{U} is a closed ball—that is, bounded.

When the domain \overline{U} of a nonoverlapping unfolding happens to be polyhedral, so its boundary $\overline{U} \setminus U$ is also polyhedral, the image $K = \varphi(\overline{U} \setminus U)$ is automatically a cut set in S . Indeed, piecewise linearity of φ implies that K is polyhedral of dimension $d - 1$; while the isometry implies that K contains S_{d-2} , and that the open ball $U \cong S \setminus K$ is contractible. Therefore:

Lemma 3.2. *If \overline{U} is polyhedral, then a nonoverlapping foldout $\varphi : \overline{U} \rightarrow S$ yields an ordinary polyhedral unfolding by taking the inverse of the restriction of φ to U .*

This renders unambiguous the term *polyhedral nonoverlapping unfolding*.

The points in S outside of the $(d - 2)$ -skeleton S_{d-2} constitute a noncompact flat Riemannian manifold S° . When a point w lies relative interior to a facet F , the tangent space T_w is identified with the tangent hyperplane T_F of F , but when w lies on a ridge, there is no canonical model for T_w .

Most tangent vectors $\zeta \in T_w$ can be exponentiated to get a point $\exp(\zeta) \in S^\circ$ by the usual exponential map from the tangent space T_w to the Riemannian manifold S° . (One can show that the set of tangent vectors that cannot be exponentiated has measure zero in T_w ; we shall not use this fact.) In the present case, we have a partial compactification S of S° , which allows us to extend this exponential map slightly.

Definition 3.3. Fix a point $w \in S^\circ = S \setminus S_{d-2}$. A tangent vector $\zeta \in T_w$ can be exponentiated if the usual exponential of $t\zeta$ exists in S° for all real numbers t satisfying $0 \leq t < 1$. In this case, set $\exp(\zeta) = \lim_{t \rightarrow 1} \exp(t\zeta)$.

The exponential map $f_\zeta : t \rightarrow \exp(t\zeta)$ takes the interval $[0, 1]$ to a geodesic $\eta \subset S$, and should be thought of as ‘geodesic flow’ away from w with tangent ζ .

Henceforth fix a source point $v \in S$ not lying on any face of dimension less than d .

Definition 3.4. The *source interior* U_v consists of the tangent vectors $\zeta \in T_v$ at the source point v that can be exponentiated, and such that the exponentials $\exp(t\zeta)$ for $0 \leq t \leq 1$ do not lie in the cut locus \overline{K}_v . The closure of U_v is the *source foldout* \overline{U}_v .

Our next main result justifies the terminology for U_v and its closure \overline{U}_v .

Theorem 3.5. *Fix a source point v in S . The exponential map $\exp : \overline{U}_v \rightarrow S$ from the source foldout to S is a polyhedral nonoverlapping foldout, and the boundary $\overline{U}_v \setminus U_v$ maps onto the cut locus \overline{K}_v . Hence \overline{K}_v is a cut set inducing a polyhedral nonoverlapping unfolding $S \setminus \overline{K}_v \rightarrow U_v$ to the source interior.*

Proof. It suffices to show the following, in view of parts 1 and 2 from Corollary 2.11.

3. The metric space $S \setminus \overline{K}_v$ is homeomorphic to an open ball.
4. The exponential map $\exp : \overline{U}_v \rightarrow S$ is piecewise linear and surjective.
5. The exponential map $\exp : U_v \rightarrow S \setminus \overline{K}_v$ is an isometry.

Every shortest path is the exponential image of some ray in \overline{U}_v by Proposition 1.2, and the set of vectors $\zeta \in \overline{U}_v$ mapping to $S \setminus \overline{K}_v$ is star-shaped by part 2 along with Proposition 1.2 and Corollary 2.2. This implies part 3 and surjectivity in part 4. The space $S^\circ = S \setminus S_{d-2}$ is isometric to a flat Riemannian manifold. Hence the exponential map is a local isometry on any open set of tangent vectors where it is defined. The definition of \overline{U}_v implies that \exp is injective on the interior U_v , so the surjectivity in part 4 shows that $\exp : U_v \rightarrow S \setminus \overline{K}_v$ is an isomorphism of Riemannian manifolds, proving part 5. Every isometry between two open subsets of affine spaces is linear, so the piecewise linearity in part 4 is a consequence of part 5. \square

Example 3.6. Consider a cube P and a source point v located off-center on the bottom face of P , as in Example 2.5 and Figure 7. The cut locus \overline{K}_v and the corresponding source foldout \overline{U}_v are shown in Fig. 8. See Fig. 1 for the case when v is in the center of the bottom face.

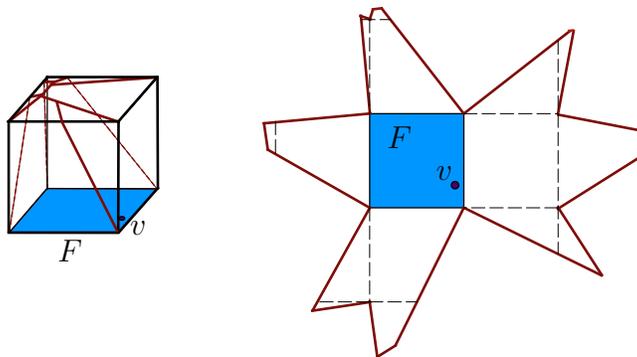


FIGURE 8. Cut locus \overline{K}_v and source foldout \overline{U}_v of the cube

Remark 3.7. Surjectivity of the exponential map does not follow from S° being a Riemannian manifold: convexity plays a crucial role (see Fig. 9 for the case of a non-

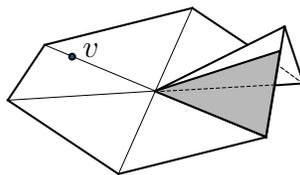


FIGURE 9. Shaded region lies outside of $\exp(T_v)$.

convex surface). In fact, surjectivity of \exp on a polyhedral manifold is equivalent—in

any dimension—to the manifold having positive curvature [Sto76, Lemma 5.1]. Theorem 3.5 extends to the class of *convex polyhedral pseudomanifolds*, but not quite verbatim; see Theorem 7.11 for the few requisite modifications.

4. The source poset

In this section we define the source poset (Definition 4.14), and in the next, we show how to build it step by step (Theorem 5.2). The reader should consider Definition 4.14 as the main result in this section, although it is the existence and finiteness properties for minimal jet frames⁵ in Theorem 4.11 that endow the source poset with its power to make continuous wavefront expansion combinatorially tractable.

Definition 4.1. Fix a polyhedron V in \mathbb{R}^d . Given a list $\bar{\zeta} = (\zeta_1, \dots, \zeta_r)$ of mutually orthogonal unit vectors in \mathbb{R}^d , define for $\varepsilon \in \mathbb{R}$ the unit vector

$$J_{\bar{\zeta}}(\varepsilon) = \frac{\varepsilon\zeta_1 + \dots + \varepsilon^r\zeta_r}{\sqrt{\varepsilon^2 + \varepsilon^4 + \dots + \varepsilon^{2r}}}.$$

If $x \in V$ and $x + \varepsilon J_{\bar{\zeta}}(\varepsilon)$ lies in V for all small $\varepsilon > 0$, then the vector-valued function $J_{\bar{\zeta}}$ is a *unit jet* of order r at x in V , and $\bar{\zeta}$ is a *partial jet frame* at x along V . If, in addition, $x + \varepsilon J_{\bar{\zeta}}(\varepsilon)$ lies relative interior to V for all small $\varepsilon > 0$, then $\bar{\zeta}$ is a *jet frame*.

The definition will be used later in the case where the convex polyhedron V is a closed Voronoi cell $R \cap V(\text{src}_F, \omega)$ for some ridge R of a facet F , and $\omega \in \text{src}_F$ is a source image. Think of the point $x \in V$ as the closest point in V to ω . It will be important later (but for now may help in understanding the next definition) to note that the relative interior of a polyhedron $V = R \cap V(\text{src}_F, \omega)$ is contained in the relative interior of the ridge R by Definition 2.3 and Theorem 2.9.

We do not assume the polyhedron V has dimension d . However, the order r of a unit jet in V , or equivalently the order of a jet frame along V , is bounded above by the dimension of V . In particular, we allow $\dim(V) = 0$, in which case the only jet frame is empty—that is, a list \emptyset of length zero—and $J_{\emptyset} \equiv 0$.

The *lexicographic order* on real vectors \bar{a} and \bar{b} of varying lengths is defined by

$$(a_1, \dots, a_r) < (b_1, \dots, b_s)$$

if the first nonzero coordinate of $\bar{a} - \bar{b}$ is negative, where by convention we set $a_i = 0$ for $i \geq r + 1$ and $b_j = 0$ for $j \geq s + 1$.

Definition 4.2. Fix a convex polyhedron V in \mathbb{R}^d , a point $x \in V$, and an *outer support vector* $\nu \in \mathbb{R}^d$ for V at x , meaning that $\nu \cdot y \leq \nu \cdot x$ for all points $y \in V$. A jet frame $\bar{\zeta}$ at x along V is *minimal* if the *angle sequence* $-(\nu \cdot \zeta_1, \dots, \nu \cdot \zeta_r)$ is lexicographically smaller than $-(\nu \cdot \zeta'_1, \dots, \nu \cdot \zeta'_r)$ for any jet frame $\bar{\zeta}'$ at x along V .

Again think of $V = R \cap V(\text{src}_F, \omega)$, with $\nu = \omega - x$ being the outer support vector.

In general, that ν is an outer support vector at x means equivalently that x is the closest point in V to $x + \nu$. Minimal jet frames $\bar{\zeta}$ can also be described more

⁵The notion of *jet frame* is new; it is motivated by constructions from differential and algebraic geometry, where a *jet* is to a higher-order derivative as a tangent vector is to a first derivative. Our goal is to measure infinitesimal expansion of the wavefront in the directions recorded by jet frames.

geometrically: the angle formed by ν and ζ_1 must be as small as possible, and then the angle formed by ν and ζ_2 must be as small as possible given the angle formed by ν and ζ_1 , and so on. It is worth bearing in mind that because ν is an outer support vector, the angle formed by ν and ζ_1 is at least $\pi/2$ (that is, obtuse or right).

Lemma 4.3. *If ζ and ζ' are vectors of equal length in \mathbb{R}^d , and $\nu \in \mathbb{R}^d$ is a vector satisfying $\nu \cdot \zeta \leq 0$ and $\nu \cdot \zeta' \leq 0$, then $|\nu - \zeta| < |\nu - \zeta'|$ if and only if $\nu \cdot \zeta > \nu \cdot \zeta'$.*

Proof. Draw ν pointing away from the center of the circle containing ζ and ζ' , with these vectors on the other side of the diameter perpendicular to ν . Then use the law of cosines: the radii ζ and ζ' have equal length, and ν has fixed length; only the distances from ν to ζ and ζ' change with the angles of ζ and ζ' with ν (see Fig. 10). \square

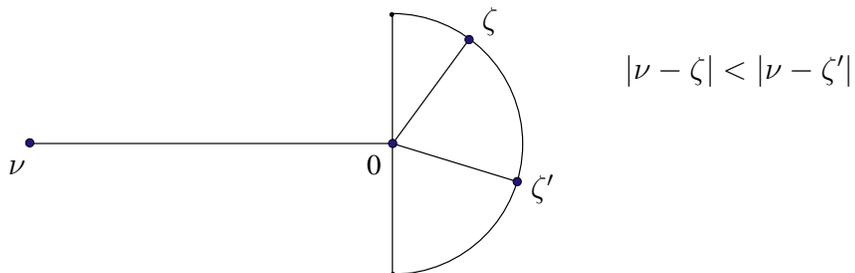


FIGURE 10. Dot product vs. length

Minimal jet frames admit a useful metric characterization as follows.

Proposition 4.4. *Fix two polyhedra V and V' with outer support vectors ν and ν' , of equal length, at points $x \in V$ and $x' \in V'$, respectively. Let $\bar{\zeta}$ and $\bar{\zeta}'$ be partial jet frames at x along V and x' along V' , respectively. The angle sequence $-\nu \cdot \bar{\zeta}$ is smaller than $-\nu' \cdot \bar{\zeta}'$ in lexicographic order if and only if there exists $\varepsilon_0 > 0$ such that $x + \nu$ is closer to $x + \varepsilon J_{\bar{\zeta}}(\varepsilon)$ than $x + \nu'$ is to $x' + \varepsilon J_{\bar{\zeta}'}(\varepsilon)$ for all positive $\varepsilon < \varepsilon_0$.*

Proof. Since the dot product of ν with each vector $J_{\bar{\zeta}}(\varepsilon)$ or $J_{\bar{\zeta}'}(\varepsilon)$ is negative, and these are unit vectors, it is enough by Lemma 4.3 to show that minimality is equivalent to

$$\nu \cdot J_{\bar{\zeta}}(\varepsilon) \geq \nu \cdot J_{\bar{\zeta}'}(\varepsilon) \quad \text{for all nonnegative values of } \varepsilon < \varepsilon_0.$$

If the first nonzero entry of $\nu \cdot \bar{\zeta} - \nu \cdot \bar{\zeta}'$ is $c = \nu \cdot (\zeta_i - \zeta'_i)$, then for nonnegative values of ε approaching 0, the difference $\nu \cdot J_{\bar{\zeta}}(\varepsilon) - \nu \cdot J_{\bar{\zeta}'}(\varepsilon)$ equals $c\varepsilon^{i-1}$ times a positive function approaching 1. The desired result follows easily. \square

Corollary 4.5. *Fix an outer support vector ν at a point x in a polyhedron V . A jet frame $\bar{\zeta}$ at x along V is minimal if and only if, for every jet frame $\bar{\zeta}'$ at x along V , $x + \nu$ is weakly closer to $x + \varepsilon J_{\bar{\zeta}}(\varepsilon)$ than to $x + \varepsilon J_{\bar{\zeta}'}(\varepsilon)$ for all small nonnegative ε .*

It is not immediately clear from the definition that minimal jet frames always exist: *a priori* there could be a continuum of choices for ζ_1 , and then a continuum of choices for ζ_2 in such a way that no minimum is attained. Although such continua of

choices can indeed occur, we shall see by constructing minimal jet frames explicitly in Theorem 4.11 that a minimum is always attained.

First we need to know more about how (partial) jet frames at x reflect the local geometry of V near x . The *tangent cone* to a polyhedron $V \subseteq \mathbb{R}^d$ at $x \in V$ is the cone

$$T_x V = \mathbb{R}_{\geq 0} \{ \zeta \in \mathbb{R}^d \mid x + \zeta \in V \}$$

generated by vectors that land inside V when added to x .

Definition 4.6. Fix a partial jet frame $\bar{\zeta}$ at x along a polyhedron V in \mathbb{R}^d . Let $\bar{\zeta}^\perp$ be the linear subspace of \mathbb{R}^d orthogonal to the vectors in $\bar{\zeta}$, and fix a sufficiently small positive real number ε . Then define the *iterated tangent cone*

$$T_x^{\bar{\zeta}} V = T_\xi((\xi + \bar{\zeta}^\perp) \cap T_x V)$$

as the tangent cone at $\xi = J_{\bar{\zeta}}(\varepsilon)$ to the intersection of $T_x V$ with the affine space $\xi + \bar{\zeta}^\perp$.

Just as the partial jet frames of order 1 generate the tangent cone $T_x V$, we have the following characterization of iterated tangent cones. We omit the easy proof.

Lemma 4.7. *The iterated tangent cone $T_x^{\bar{\zeta}} V$ is generated by all unit vectors ζ_{r+1} in \mathbb{R}^d extending the partial jet frame $\bar{\zeta} = (\zeta_1, \dots, \zeta_r)$ to a partial jet frame $(\zeta_1, \dots, \zeta_r, \zeta_{r+1})$ of order $r + 1$. In particular, iterated tangent cones do not depend on the small $\varepsilon > 0$.*

Now we set out to construct minimal jet frames inductively.

Lemma 4.8. *Fix a polyhedron V and an outer support vector ν at $x \in V$. If $\zeta \in T_x V$ is a unit vector with $\nu \cdot \zeta$ maximal, then ν is an outer support vector at $\mathbf{0} \in T_x^\zeta V$.*

Proof. If $\zeta' \in T_x^\zeta V$ is a unit vector satisfying $\nu \cdot \zeta' > 0$, then $\xi = (\zeta + \varepsilon \zeta') / \sqrt{1 + \varepsilon^2}$ for small $\varepsilon > 0$ is a unit vector in $T_x V$ satisfying $\nu \cdot \xi > \nu \cdot \zeta$, contradicting maximality. \square

In ‘generic’ cases, the functional $\zeta \mapsto \nu \cdot \zeta$ for an outer support vector ν on a cone will take on the maximum value zero uniquely at the origin. In this case, as we now show, there can be only finitely many unit vectors ζ in the cone having $\nu \cdot \zeta$ maximal, and these lie along the *rays*, meaning one-dimensional faces of the cone. Note that genericity forces the cone to be *sharp*, meaning that it contains no linear subspaces.

Proposition 4.9. *Let ν be an outer support vector for a sharp polyhedral cone C , and assume ν is maximized uniquely at the origin $\mathbf{0}$. The minimum angle between ν and a unit vector $\zeta \in C$ occurs when ζ lies on a ray of C .*

Proof. Let Z be the set of unit vectors in C . Suppose that L is a 2-dimensional subspace inside the span of C , and let $\bar{\nu}$ be the orthogonal projection of ν onto L . View ν and $\bar{\nu}$ as functionals on L via $\zeta \mapsto \nu \cdot \zeta$, and observe that $\nu \cdot \zeta = \bar{\nu} \cdot \zeta$ for all $\zeta \in L$. The circular arc $Z \cap L$ lies inside the unit circle in L , and $\bar{\nu}$ takes nonpositive values on $Z \cap L$ because ν is an outer support vector. Elementary geometry shows that $\bar{\nu}$ is therefore maximized on $Z \cap L$ only at one or both of the endpoints of the arc $Z \cap L$. This argument proves that ν cannot be maximized on Z at a point $\zeta \in Z$ unless ζ lies in the boundary of Z . The result now follows by induction on the dimension of the cone C . \square

In ‘nongeneric’ cases, including when the polyhedral cone C has nonzero *lineality*, which is by definition the largest vector space contained in C , the functional ν is maximized along a face of positive dimension. In this case, there is always a continuum of choices for unit vectors $\zeta \in C$ having $\nu \cdot \zeta = 0$. However, the sequences of iterated tangent cones to appear in Theorem 4.11 will not in any noticeable way depend on the continuum of choices, because of the next result.

Lemma 4.10. *Fix a polyhedron V , a point $x \in V$, and a face F of V containing x . The iterated tangent cone $T_x^{\bar{\zeta}}V$ is independent of the jet frame $\bar{\zeta}$ for F at x .*

Proof. Translate V so $x + \varepsilon J_{\bar{\zeta}}(\varepsilon)$ lies at the origin $\mathbf{0} \in \mathbb{R}^d$. Then F spans a dimension $\dim(F)$ linear subspace $\langle F \rangle \subseteq \mathbb{R}^d$, and the iterated tangent cone is $T_x^{\bar{\zeta}}V = \langle F \rangle^\perp \cap T_{\mathbf{0}}V$. Now use the fact that $T_{\mathbf{0}}V = T_\xi V$ for all vectors ξ relative interior to F . \square

The main theorem in this section says that given an outer support vector ν , there is a finite procedure using elementary linear algebra for producing a single jet frame that is, in a precise sense, tilted as much toward ν as possible.

Theorem 4.11. *Fix a polyhedron V and an outer support vector ν at $x \in V$. Inductively construct a finite set of jet frames for V at x by iterating the following procedure. For each of the finitely many partial jet frames $\bar{\zeta}$ already constructed:*

- *If ν is orthogonal to a nonzero vector in $T_x^{\bar{\zeta}}V$, then add any such vector to $\bar{\zeta}$.*
- *If $\nu \cdot \zeta < 0$ for all nonzero vectors ζ in $T_x^{\bar{\zeta}}V$, then create one new partial jet frame for each of the (finitely many) rays of $T_x^{\bar{\zeta}}V$ minimizing the angle with ν , by appending to $\bar{\zeta}$ the unit vector along that ray.*

At least one of the finitely many jet frames constructed in this way is minimal.

Proof. The sequences of vectors constructed by the iterated procedure are jet frames by Lemma 4.8. Given an arbitrary jet frame $\bar{\xi}$ for V at x , it is enough to show that the angle sequence of $\bar{\xi}$ satisfies $\nu \cdot \bar{\zeta} \geq \nu \cdot \bar{\xi}$ in lexicographic order for some constructed jet frame $\bar{\zeta}$. Indeed, then a jet frame whose angle sequence is lexicographically minimal among the constructed ones is minimal. Suppose that the first $i - 1$ entries $(\xi_1, \dots, \xi_{i-1})$ agree with a constructed jet frame, but that (ξ_1, \dots, ξ_i) do not.

If $\nu \cdot \xi_i < 0$ then $\nu \cdot \xi_i$ is less than $\nu \cdot \zeta_i$ for some constructed jet frame $\bar{\zeta}$ agreeing with $\bar{\xi}$ through the $(i - 1)^{\text{st}}$ entry, by Proposition 4.9.

If, on the other hand, $\nu \cdot \xi_i = 0$, then pick the index j maximal among those satisfying $\xi_j \neq 0$ and also $\nu \cdot \xi_i = \dots = \nu \cdot \xi_j = 0$. If there is a constructed jet frame $\bar{\zeta}$ that agrees with $\bar{\xi}$ through the j^{th} entry, but has $\nu \cdot \xi_{j+1} < \nu \cdot \zeta_{j+1} = 0$, then already we are done. Therefore we can assume that the constructed jet frame $\bar{\zeta}$ agrees with $\bar{\xi}$ through index $(i - 1)$, that $\bar{\zeta}$ has $\nu \cdot \zeta_i = \dots = \nu \cdot \zeta_j = 0$, and that either $\nu \cdot \zeta_{j+1} < 0$ or else $\bar{\zeta}$ has order j . Replacing the vectors ξ_i, \dots, ξ_j in $\bar{\xi}$ with ζ_i, \dots, ζ_j yields a new jet frame $\bar{\xi}'$, by Lemma 4.10 applied to the face F of the iterated tangent cone $T_x^{\bar{\zeta}'}V$ orthogonal to ν , where $\bar{\zeta}' = (\zeta_1, \dots, \zeta_{i-1})$. Downward induction on the number of entries of $\bar{\xi}'$ shared with a constructed jet frame completes the proof. \square

Our goal is to apply jets to define a poset structure on the set of source images. First, we need some terminology and preliminary concepts. The next definition is

made in slightly more generality than required for dealing only with complete sets of source images because we shall need it for Theorem 5.2.

Resume the notation from previous sections regarding the polyhedral complex S . Recall that $T_F \cong \mathbb{R}^d$ is the tangent hyperplane to the facet F . Removing from T_F the affine span T_R of any ridge $R \subset F$ leaves two connected components (open half-spaces). Thus it makes sense to say that a point $\nu \in T_F \setminus T_R$ lies either on the *same side* or on the *opposite side* of R as does F .

Definition 4.12. Fix a facet F , a ridge $R \subset F$, and a finite set $\Upsilon \subset T_F$.

1. A point $\omega \in \Upsilon$ can see F through R in $\mathcal{V}(\Upsilon)$ if ω lies on the opposite side of R as F does, and the closed Voronoi cell $V(\Upsilon, \omega)$ contains a point interior to R .
2. A point $\omega \in \Upsilon$ can see R through F in $\mathcal{V}(\Upsilon)$ if ω lies on the same side of R as F does, and the closed Voronoi cell $V(\Upsilon, \omega)$ contains a point interior to R .
3. In either of the above two cases, the ridge R lies at *radius* $r = r(R, \omega)$ from ω if r equals the smallest distance in T_F from ω to a point of $R \cap V(\Upsilon, \omega)$.
4. The unique *closest point* $\rho(R, \omega)$ to ω in $R \cap V(\Upsilon, \omega)$ has distance r from ω .
5. The *outer support vector* of the pair (R, ω) is $\omega - \rho(R, \omega)$.
6. The *angle sequence* $\angle(R, \omega)$ is the angle sequence $-(\omega - \rho(R, \omega)) \cdot \bar{\zeta}$ for any minimal jet frame $\bar{\zeta}$ at $\rho(R, \omega)$ along $R \cap V(\Upsilon, \omega)$.

Example 4.13. Fig. 11 depicts examples of the notions from Definition 4.12. The

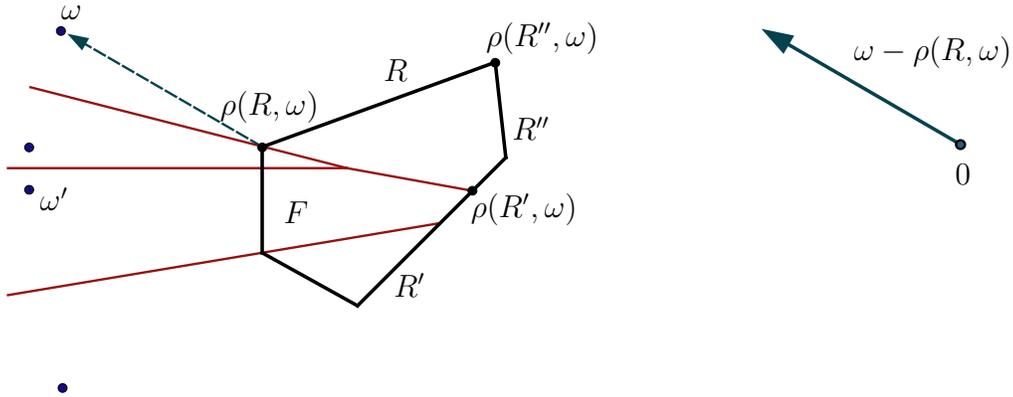


FIGURE 11. Illustrations for Definition 4.12

solid pentagon is the face F , while the set Υ contains four points. The point ω can see F through the ridge R , and can see the ridges R' as well as R'' through F . The three closest points for these are indicated, as is the outer support vector for (R, ω) . The point ω' can see the ridge R' through F , but ω' can *not* see R'' through F , because ω is closer to every point of R'' .

In our applications, the finite set Υ will always be a subset of source images in src_F , often a proper subset. Now we are ready for the main definition of this section. It may help to recall that each source image $\nu \in \text{src}_F$ can see F through a *unique* ridge R by Theorem 2.9, when $\mathbb{R}^d = T_F$ and the finite set Υ in Definition 4.12 equals src_F .

Definition 4.14. Fix a source point v in S . An *event* is a pair (ν, F) with $\nu \in \text{src}_F$ a source image for the facet F . The event (ν, F) has

1. *radius* $r(\nu, F)$ equal to the radius $r(R, \nu)$ from ν to the ridge R through which ν can see F in the Voronoi subdivision $\mathcal{V}(\text{src}_F)$ of T_F ;
2. *event point* $\rho(\nu, F)$ equal to the closest point $\rho(R, \nu)$ in $R \cap V(\text{src}_F, \nu)$ to ν ; and
3. *angle sequence* $\angle(\nu, F)$ equal to the angle sequence $\angle(R, \nu)$.

(The trivial event $(v, \text{facet}(v))$ has radius 0, event point v , and empty angle sequence.) The *source poset* $\text{src}(v, S)$ is the set of events, partially ordered with $(\nu, F) \prec (\nu', F')$ if

- $r(\nu, F) < r(\nu', F')$, or if
- $r(\nu, F) = r(\nu', F')$ and $\angle(\nu, F)$ is lexicographically smaller than $\angle(\nu', F')$.

Remark 4.15. Corollary 4.5 says that breaking ties by lexicographically comparing angle sequences at event points is the same as breaking ties by comparing distances from each source image to a minimal jet at its event point. This is the precise sense in which the source poset orders events by comparing infinitesimal expansion of the wavefront along the interiors of ridges containing event points.

5. Constructing source images

Aside from its abstract dynamical interpretation, the importance of the source poset here stems from its ability to be computed algorithmically, as we shall see here and in Section 6. Source images are built one by one, using only previously built source images as stepping stones. These stepping stones form an *order ideal* in $\text{src}(v, S)$, meaning a subset $\mathcal{I} \subset \text{src}(v, S)$ closed under going down: $E \in \mathcal{I}$ and $E' \prec E \Rightarrow E' \in \mathcal{I}$.

To make a precise statement in the main result, Theorem 5.2, we need one more dose of terminology, describing constructions in S determined by a choice of order ideal.

Definition 5.1. Fix an order ideal \mathcal{I} in the source poset $\text{src}(v, S)$. For each facet F , let $\Upsilon_F \subset T_F$ be the set of source images $\omega \in \text{src}_F$ with $(\omega, F) \in \mathcal{I}$. The set $\mathcal{E}_{\mathcal{I}}$ of *potential events* consists of triples (ω, F, R') such that

- ω can see the ridge R' through F in the Voronoi diagram $\mathcal{V}(\Upsilon_F)$, but
- a second facet F' contains R' , and the unfolding $\omega' = \Phi_{F, F'}(\omega)$ of ω onto the tangent space $T_{F'}$ results in a pair (ω', F') that does not lie in \mathcal{I} .

If (ω', F') is an event in $\text{src}(v, S) \setminus \mathcal{I}$, then we say it is obtained by *processing* (ω, F, R') . A potential event $E \in \mathcal{E}_{\mathcal{I}}$ is *minimal* if it has minimal radius r among potential events, and lexicographically minimal angle sequence among potential events with radius r .

Tracing back through notation, if $E = (\nu, F, R')$ is a minimal potential event, then the minimal radius is $r = r(R', \nu)$, and the minimal angle sequence is $\angle(R', \nu)$.

Theorem 5.2. *Given a nonempty order ideal \mathcal{I} in the source poset $\text{src}(v, S)$, pick a minimal potential event (ν, F, R') in $\mathcal{E}_{\mathcal{I}}$. If $\nu' = \Phi_{F, F'}(\nu)$ is the unfolding of ν to the other facet F' containing R' , then $\mathcal{I}' = \mathcal{I} \cup \{(\nu', F')\}$ is an order ideal in $\text{src}(v, S)$.*

The statement has two parts, really: first, $\nu' \in T_{F'}$ is indeed a source image; and second, \mathcal{I}' is an order ideal in the poset $\text{src}(v, S)$. To prove the theorem we need a

number of preliminaries. We state results requiring an order ideal inside the source poset $\text{src}(v, S)$ using language that assumes an order ideal \mathcal{I} has been fixed.

Recall from Section 1 the notion of facet sequence \mathcal{L}_γ for a shortest path γ . If, on the way to a facet F' , a shortest path γ from the source point v traverses a facet F , then the corresponding source images in F and F' have a special relationship. Precisely:

Definition 5.3. Let $(\nu, F) \prec (\nu', F')$ be events in the source poset. Suppose some shortest path γ has facet sequence $\mathcal{L}_\gamma = (F_1, \dots, F_\ell)$ with a consecutive subsequence

$$\mathcal{L} = (F_\ell, \dots, F_{\ell'}) \quad \text{in which} \quad F = F_\ell \text{ and } F' = F_{\ell'}.$$

If $\nu' = \Phi_{\mathcal{L}}(\nu) = \Phi_{\mathcal{L}_\gamma}(\nu)$ is the sequential unfolding of the source along γ , and also the sequential unfolding of $\nu \in T_F$ into $T_{F'}$, then (ν, F) *geodesically precedes* (ν', F') . We also say that the shortest path γ described above is *geodesically preceded* by (ν, F) .

Since the Voronoi cells in Theorem 2.9 come up so often, it will be convenient to have easy terminology and notation for them.

Definition 5.4. Given a source image $\omega \in \text{src}_F$, the *cut cell* of ω is $V_\omega = V(\text{src}_F, \omega)$.

Roughly speaking, our next result says that angle sequences increase at successive events along shortest paths, when the event point is pinned at a fixed point x .

Proposition 5.5. *If (ν, F) geodesically precedes (ν', F') then $(\nu, F) \prec (\nu', F')$.*

Proof. Because of the way partial order on $\text{src}(v, S)$ is defined, we may as well assume that F and F' share a ridge R' , and that $\nu' = \Phi_{F, F'}(\nu)$ is obtained by folding along this ridge. In addition, we may as well assume that both event points $\rho(\nu, F)$ and $\rho(\nu', F')$ equal the same point $x \in S$, since otherwise $r(\nu, F) < r(\nu', F')$. Translate to assume this point x equals the origin $\mathbf{0}$, to simplify notation. Let R be the ridge through which ν can see F , and set $V = R \cap V_\nu$ and $V' = R' \cap V_{\nu'}$; these are the cut cells through which the source images ν and ν' see their corresponding facets.

The angle geometry of ν' relative to V' in $T_{F'}$ is *exactly* the same as the geometry of ν relative to V in T_F , because ν' is obtained by rotation around an axis in \mathbb{R}^{d+1} containing V' . In other words, $\nu - \nu'$ is orthogonal to V' . Therefore we need only compare the angles with ν of jets along V and V' . All jet frames will be at x .

Suppose the finite sequence (ξ_1, ξ_2, \dots) is a jet frame along V' . Noting that V and V' have disjoint interiors, choose the index r so that $\bar{\xi} = (\xi_1, \dots, \xi_{r-1})$ is a partial jet frame along V , but $\bar{\xi}' = (\xi_1, \dots, \xi_r)$ is not. It is enough to demonstrate that some partial jet frame $(\xi_1, \dots, \xi_{r-1}, \zeta_r)$ along V has lexicographically smaller angle sequence than $\bar{\xi}'$. Equivalently, it is enough to produce a unit vector ζ_r in the iterated tangent cone $T_x^{\bar{\xi}} V$ satisfying $\nu \cdot \zeta_r > \nu \cdot \xi_r$.

Since $R' \cap V_{\nu'} = R' \cap V_\nu$ by Theorem 2.9, every line segment from ν to a point in V' passes through V . Therefore, since we have translated to make $x = \mathbf{0}$, every segment connecting ν to $T_x V'$ passes through $T_x V$. This observation will become crucial below; for now, note the resulting inequality $\dim(V) \geq \dim(V')$, which implies that the iterated tangent cone $T_x^{\bar{\xi}} V$ contains nonzero vectors. All such vectors by definition lie in the subspace $\bar{\xi}^\perp$ orthogonal to the space $\langle \bar{\xi} \rangle$ with basis ξ_1, \dots, ξ_{r-1} . The same holds

for ξ_r , so we may replace ν with a vector $\omega \in \bar{\xi}^\perp$ by adding a vector in $\langle \bar{\xi} \rangle$, since then

$$\omega \cdot \zeta = \nu \cdot \zeta \quad \text{for all vectors } \zeta \in \bar{\xi}^\perp.$$

Fix a small positive real number ε . The line segment $[\nu, J_{\bar{\xi}}(\varepsilon)]$ intersects $T_x V$ at a point near $J_{\bar{\xi}}(\varepsilon)$. The image segment in $\bar{\xi}^\perp$ by orthogonal projection modulo $\langle \bar{\xi} \rangle$ is $[\omega, \lambda \xi_r]$, for $\lambda = \varepsilon^r / \sqrt{\varepsilon^2 + \dots + \varepsilon^{2r}}$. This image segment passes through the cone $T_x^\xi V$ at some point ζ on its way from ω to $\lambda \xi_r$. Elementary geometry of the triangle with

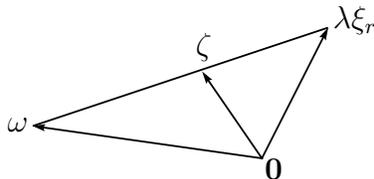


FIGURE 12. Geodesic precedence implies smaller angle sequence

vertices $\mathbf{0}$, ω , and $\lambda \xi_r$ (see Fig. 12) shows that the angle between ω and ζ is smaller than the angle between ω and $\lambda \xi_r$. Taking $\zeta_r = \zeta / |\zeta|$ completes the proof. \square

After choosing a minimal potential event E , we must make sure that when all is said and done, none of the other potential events end up below E in the source poset.

Lemma 5.6. *Suppose $(\omega, F, R') \in \mathcal{E}_{\mathcal{I}}$ is a potential event with angle sequence \angle and radius r . Let F' be the other facet containing R' and $\omega' = \Phi_{F, F'}(\omega)$ the unfolding of $\omega \in T_F$ onto $T_{F'}$. If (ω', F') is an actual event, then it either has radius strictly bigger than r , or else its angle sequence $\angle(\omega', F')$ is lexicographically larger than \angle .*

Proof. Assume that (ω', F') is an event. Quite simply, the result is a consequence of the fact that the cut cell $R' \cap V_{\omega'} = R' \cap V_\omega$ must be contained inside $R' \cap V(\Upsilon_F, \omega)$, which follows because $\Upsilon_F \subseteq \text{src}_F$. \square

In comparing a newly processed event (in source poset order) to other as yet unprocessed events, we will need to know approximately how those other events will eventually arise. This requires the forthcoming lemma, in which a *flat triangle* inside S is any subset of S isometric to a triangle in the Euclidean plane \mathbb{R}^2 .

Lemma 5.7. *Fix a point $x \in S$. There is an open neighborhood \mathcal{O}_x of x in S such that, given $y \in \mathcal{O}_x$ and a shortest path γ from the source point v to y , some shortest path γ' from v to x has the following property: the loop formed by traversing γ' and then the segment $[x, y]$ and finally the reverse of γ bounds a flat triangle in S .*

Proof. Choose \mathcal{O}_x so small that the only closed faces of the cut locus \bar{K}_v intersecting \mathcal{O}_x are those containing x . Every cut cell containing $y \in \mathcal{O}_x$ also contains x by construction. Convexity of cut cells (Theorem 2.9) implies that the segment $[x, y]$ lies inside every cut cell containing y (there may be more than one if y is itself a cut point). The source image obtained by sequentially unfolding γ therefore connects to every point of $[x, y]$ by a straight segment that sequentially folds to a shortest path. The union of these shortest paths is the flat triangle in question. \square

Conveniently, all of the shortest paths to x already yield events in \mathcal{I} :

Lemma 5.8. *Suppose some minimal potential event $E \in \mathcal{E}_{\mathcal{I}}$ has closest point x . Let G be the last facet whose interior is traversed by a shortest path γ from the source point to x . If $\omega \in \text{src}_G$ is the source image sequentially unfolded along γ , then $(\omega, G) \in \mathcal{I}$.*

Proof. As γ enters G , it crosses the relative interior of some ridge of G at a point w . The event point $\rho(\omega, G)$ can be no farther than w from ω . On the other hand, $\mu(v, w) < \mu(v, x)$, because γ traverses the interior of G . Therefore (ω, G) has radius less than $r(E) = \mu(v, x)$. \square

Proof of Theorem 5.2. Suppose the minimal potential event (ν, F, R') has closest point $x = \rho(R', \nu)$ to $R' \cap V(\Upsilon_F, \nu)$, of radius r , and a minimal jet frame $\bar{\zeta}$ at x with angle sequence \angle .

Let γ be a shortest path from the source that ends at a point in the neighborhood \mathcal{O}_x from Lemma 5.7. By that lemma and Lemma 5.8, γ unfolds to produce a source image whose event either lies in \mathcal{I} , or is obtained by processing a potential event in $\mathcal{E}_{\mathcal{I}}$, or is geodesically preceded by such a processed event. Applying Lemma 5.8 and then Proposition 5.5, we find that all events with event point x that are not in \mathcal{I} have angle sequences lexicographically larger than \angle .

For positive ε , set $y(\varepsilon) = x + \varepsilon J_{\bar{\zeta}}(\varepsilon)$. When ε is small enough, $y(\varepsilon)$ lies interior to R' , and close to x , in the neighborhood \mathcal{O}_x from Lemma 5.7. By the previous paragraph, every source image containing y in its cut cell is either in \mathcal{I} , or has angle sequence lexicographically larger than \angle .

Let us now compare, for all small positive ε , the distance to $y(\varepsilon)$ from ν with the distance to $y(\varepsilon)$ from any source image in src_F or $\text{src}_{F'}$. Clearly the distance from a source image ω is minimized when $y(\varepsilon)$ lies in the cut cell V_ω . Moreover, we may restrict our attention to those source images ω whose cut cells V_ω contain $y(\varepsilon)$ for all sufficiently small positive ε . Definition 4.1 says that $\bar{\zeta}$ is a jet frame at x along V_ω . Therefore, by Proposition 4.4, we conclude using the last sentence of the previous paragraph that $y(\varepsilon)$ is weakly closer to ν than to ω for all small positive ε . This argument shows that ν' is a source image, so (ν', F') is an event. Moreover, it shows:

Claim 5.9. *Any minimal jet frame $\bar{\zeta}$ at the event point $x = \rho(R', \nu)$ along the polyhedron $R' \cap V(\Upsilon_F, \nu)$ is a minimal jet frame at x along $R' \cap V_{\nu'}$.*

Every event in $\text{src}(v, S) \setminus \mathcal{I}$ is either obtained by processing a potential event in $\mathcal{E}_{\mathcal{I}}$, or is geodesically preceded by such a processed potential event. Using Claim 5.9, we conclude by Lemma 5.6 and Proposition 5.5 that \mathcal{I}' is an order ideal. \square

6. Algorithm for source unfolding

The primary application of the analysis up to this point is an algorithmic construction of nonoverlapping unfoldings of convex polyhedra, which we present in pseudocode followed by bounds on its running time. In particular, we show that the algorithm is polynomial in the number of source images, when the dimension d is fixed. (Later

we shall state Conjecture 9.2, which posits that the number of source images is polynomial in the number of facets.) Other applications, some of which are further discussed in Section 8, include the discrete geodesic problem (Corollary 6.6) and geodesic Voronoi diagrams (Algorithm 8.1 in Section 8.9).

Roughly, Algorithm 6.1 consists of a single loop that with every iteration constructs one new *event*. Each event is a pair consisting of a facet and a point that we have called a *source image* in the affine span of that facet. The loop is repeated exhaustively until all of the events are computed, so the affine span of every facet has its full complement of source images. The Voronoi diagram for the set of source images in each affine span induces a subdivision of the corresponding facet. For each maximal cell in this subdivision, the algorithm computes a Euclidean motion (composition of rotation and parallel translation) that moves it into the affine span of the facet containing the source point. The union of these moved images of Voronoi cells is the output foldout \bar{U}_v in the tangent space T_v to the source point v .

At each iteration of the loop, the algorithm must choose from a number of *potential events* that it could process into an actual event. Each potential event E consists of an already-computed event (ν, F) plus a ridge R in the facet F . Processing the event E applies a rotation to move the source image ν into the affine span of the other facet containing R . The potential event that gets chosen must lie as close to the source point as possible; this distance is the radius $r = r(R, \nu)$ at the beginning of the loop. The loop then calls Routine 6.2 to choose which event to process; although this routine is quite simple in structure, it is the part of the algorithm that most directly encounters the subtlety of working in higher dimensions. The end of the loop consists of updating the sets of source images and potential events; the latter requires Routine 6.3, which we have isolated because it is the only time-consuming part of the algorithm, due to its Voronoi computation.

Let us emphasize that once a source point ν is computed, it is never removed. This claim is part of Theorem 5.2, in which the correctness of Algorithm 6.1—and indeed the procedure of the algorithm itself—is more or less already implicit, as we shall see in the proof of Theorem 6.4.

We assume that the convex polyhedron P is presented in the input of the algorithm as an intersection of closed half-spaces. Within the algorithm, we omit the descriptions of standard geometric and linear algebraic operations, for which we refer to [GO97, PS85]. These operations include the determination of lower-dimensional faces (such as ridges) given the facets of P , and the computation of Voronoi diagrams.

Some additional notation will simplify our presentation of the algorithm. Denote by \mathcal{F} and \mathcal{R} the sets of facets and ridges of P , respectively. If a ridge $R \in \mathcal{R}$ lies in a facet $F \in \mathcal{F}$, denote by $\phi(F, R)$ the other facet containing R , so $F \cap \phi(F, R) = R$. Finally, for each facet $F \in \mathcal{F}$, denote by $\hat{\mathcal{E}}_F$ the set of all triples (ν, F, R) such that source point $\nu \in \Upsilon_F$ lies in the affine span T_F of F , and $R \in \mathcal{R}$ is a ridge contained in F .

Algorithm 6.1. (Computing source unfolding)

INPUT convex polyhedron $P \subset \mathbb{R}^{d+1}$ of dimension $d + 1$
point v lying in the relative interior of a facet F of P
OUTPUT source foldout of the boundary $S = \partial P$ into $T_v \cong \mathbb{R}^d$ (see Sec. 3)

DEFINE for each $F \in \mathcal{F}$: a finite set $\Upsilon_F \subset T_F$ of points
 for each pair (ν, F) satisfying $\nu \in \Upsilon_F$: an ordered list $\mathcal{L}_{\nu, F}$ of facets
 for each $F \in \mathcal{F}$: a set $\mathcal{E}_F \subset \widehat{\mathcal{E}}_F$ of *potential events*
 $\mathcal{E} = \bigcup_{F \in \mathcal{F}} \mathcal{E}_F$, the set of *all potential events*
INITIALIZE for $F \in \mathcal{F}$: if $v \notin F$ then $\Upsilon_F := \emptyset$ and $\mathcal{E}_F = \emptyset$;
 otherwise $\Upsilon_F := \{v\}$, $\mathcal{L}_{v, F} := (F)$, $\mathcal{E}_F := \{(v, F, R) \mid R \in \mathcal{R} \text{ and } R \subset F\}$
COMPUTE $\Phi_{F, F'}$ for all $F, F' \in \mathcal{F}$ such that $F \cap F' \in \mathcal{R}$ is a ridge (see Def. 1.5)
WHILE $\mathcal{E} \neq \emptyset$
 DO $r := \min\{r(R, \nu) \mid (\nu, F, R) \in \mathcal{E}\}$ (see Def. 4.12)
 CHOOSE A POTENTIAL EVENT $E = (\nu, F, R) \in \mathcal{E}$ **TO PROCESS**
 set $F' := \phi(F, R)$, $\nu' := \Phi_{F, F'}(\nu)$, $\mathcal{L}_{\nu', F'} := (\mathcal{L}_{\nu, F} F')$
 update $\Upsilon_{F'} \leftarrow \Upsilon_{F'} \cup \{\nu'\}$
 $\mathcal{E}_{F'} \leftarrow \{(\omega, F', R') \in \widehat{\mathcal{E}}_{F'} \text{ such that } \omega \in \Upsilon_{F'}, \text{ and}$
 POINT $\omega \in \Upsilon_{F'}$ **CAN SEE** R' **THROUGH** F' , and
 $\omega' \notin \Upsilon_G$, where $G = \phi(F', R')$, $\omega' = \Phi_{F', G}(\omega)\}$
 $\mathcal{E}_F \leftarrow \mathcal{E}_F \setminus \{E\}$, $\mathcal{E} \leftarrow \bigcup_{G \in \mathcal{F}} \mathcal{E}_G$
END WHILE-DO
COMPUTE for all facets $F \in \mathcal{F}$ and points $\nu \in \Upsilon_F$:
 $\Phi_{\mathcal{L}}$ for $\mathcal{L} = \mathcal{L}_{\nu, F}$ (see Def. 1.6), and then
 $\overline{U}_v(\nu, F) := \Phi_{\mathcal{L}}^{-1}(F \cap V(\Upsilon_F, \nu)) \subset T_v$ (see Thm. 2.9)
RETURN the foldout $\overline{U}_v = \bigcup_{(\nu, F)} \overline{U}_v(\nu, F)$, the union being over all $F \in \mathcal{F}$ and $\nu \in \Upsilon_F$

Routine 6.2. (CHOOSE A POTENTIAL EVENT TO PROCESS)

INPUT the set $\mathcal{E} = \bigcup_{F \in \mathcal{F}} \mathcal{E}_F$ of potential events, and the radius $r > 0$
OUTPUT an event $E \in \mathcal{E}$ (see Def. 5.1)
COMPUTE the closest potential events $\mathcal{E}_\circ := \{(\omega, F, R) \in \mathcal{E} \mid r(\omega, F) = r\}$
 angle sequence $\angle(R, \omega)$ for all $(\omega, F, R) \in \mathcal{E}_\circ$ (see Def. 4.12)
FIND a potential event $E = (\omega, F, R) \in \mathcal{E}_\circ$ with lexicographically
 minimal angle sequence $\angle(R, \omega)$ (see Sec. 4)
RETURN the event $E = (\omega, F, R)$

Routine 6.3. (POINT $\omega \in \Upsilon$ CAN SEE R THROUGH F)

INPUT facet $F \in \mathcal{F}$, ridge $R \in \mathcal{R}$, finite set of points $\Upsilon \subset T_F$, and $\omega \in \Upsilon$
OUTPUT boolean variable $\beth \in \{\text{True}, \text{False}\}$ (see Def. 4.12)
COMPUTE Voronoi diagram $\mathcal{V}(\Upsilon)$ (see Sec. 2)
IF Voronoi cell $V(\Upsilon, \omega) \subset \mathcal{V}(\Upsilon)$ contains a point interior to R
 and ω lies on the same side of R as F does in T_F
 then $\beth := \text{True}$;
 otherwise $\beth := \text{False}$
RETURN the variable \beth

In the pseudocode, we have used the two different symbols ‘ \leftarrow ’ and ‘ $:=$ ’ to distinguish between those variables that are being updated and those that are being

completely redefined at each iteration of the WHILE-DO loop. We hope this clarifies the structure of Algorithm 6.1.

Theorem 6.4. *For every convex polyhedron $P \subset \mathbb{R}^{d+1}$ with boundary $S = \partial P$, and any source point v in a facet of S , Algorithm 6.1 computes the source foldout $\overline{U}_v \subseteq T_v$.*

Proof. First, we claim by induction that after each iteration of the WHILE-DO loop, the set $\{(\nu, F) \mid F \in \mathcal{F} \text{ and } \nu \in \Upsilon_F\}$ is an order ideal in the source poset $\text{src}(v, S)$ from Definition 4.14. The claim is clear at the beginning of the algorithm. By construction, Routine 6.2 picks a minimal potential event E to process. The loop then adds an event by processing E , with the aid of Routine 6.3. Theorem 5.2 implies that what results after processing E is still an order ideal of events, proving our claim. Since the poset $\text{src}(v, S)$ is finite by Lemma 2.4, the algorithm halts after a finite number of loop iterations. Finally, by Theorem 2.9 the Voronoi cells in each facet coincide with the polyhedral subdivision of each facet by the cut locus \overline{K}_v , so Theorem 3.5 shows that the foldout in the output is the desired (nonoverlapping) *source foldout* \overline{U}_v . \square

For purposes of complexity, we assume throughout this paper that the dimension d is fixed. Thus, if the convex polyhedron $P \subset \mathbb{R}^{d+1}$ of dimension d has n facets, so P is presented as an intersection of n closed half-spaces, we can compute all of the vertices and ridges of P in polynomial time [GO97, Zie95]. For simplicity, we assume these are precomputed and appended to the input.

The timing of Algorithm 6.1 crucially depends on the number of source images. Let

$$\overline{\text{src}}_v := \max_{F \in \mathcal{F}} |\text{src}_F|$$

be the largest number of source images in a tangent plane T_F for a facet F . (This number can change if the source point v is moved. For example, $\overline{\text{src}}_v = 4$ if v is in the center of a face, while $\overline{\text{src}}_v = 12$ if v is off-center as in Figures 7 and 8.) Note that computing Voronoi diagrams for N points in \mathbb{R}^d can be done in $N^{O(d)}$ time [GO97, p.381]. See [Aur91, Cha91, For95] for details and further references on Voronoi diagrams, and [GO97, PS85] for other geometric and linear algebraic computations we use.

Theorem 6.5. *When the dimension d is fixed, the cost of Algorithm 6.1 is polynomial in the number n of facets and the maximal number $\overline{\text{src}}_v$ of source images for a facet.*

Proof. From the analysis in the proof of Theorem 6.4, the number of loop iterations is at most $|\text{src}(v, S)| \leq |\mathcal{F}| \overline{\text{src}}_v \leq n \overline{\text{src}}_v$. Within the main body of the algorithm, only standard geometric and linear algebraic operations are used, and these are all polynomial in n . Similarly, Routine 6.2 uses only linear algebraic operations for every potential event $E \in \mathcal{E}$. Note that the cardinality of the set of potential events \mathcal{E} during any iteration of the loop is bounded by $|\text{src}(v, S)| \cdot |\mathcal{F}|^2 \leq (n \overline{\text{src}}_v) \cdot n^2 = n^3 \overline{\text{src}}_v$.

Routine 6.3 constructs Voronoi diagrams $\mathcal{V}(\Upsilon)$ for finite sets $\Upsilon \subset \mathbb{R}^d$. This computation requires $|\Upsilon|^{O(d)} \leq (\overline{\text{src}}_v)^{O(d)}$ time, which is polynomial for our fixed dimension d . Therefore the total cost of the algorithm is also polynomial in n and $\overline{\text{src}}_v$. \square

Corollary 6.6. *Let v and w be two points on the boundary S of the convex $(d+1)$ -dimensional polyhedron $P \subset \mathbb{R}^{d+1}$, and suppose that v lies interior to a facet. Then the geodesic distance $\mu(v, w)$ on S can be computed in time polynomial in n and $\overline{\text{src}}_v$.*

The restriction that v lie interior to a facet is unnecessary, and in fact Algorithm 6.1 can be made to work for arbitrary points v ; see Sections 8.8 and 8.9.

Proof. Use Algorithm 6.1 to compute the foldout map $\varphi : \overline{U}_v \rightarrow S$. Find $w' \in T_v$ mapping to $w = \varphi(w') \in S$, and compute the distance $|v - w'|$. By the isometry of the exponential map in Theorem 3.5, we conclude that $\mu(v, w) = |v - w'|$. \square

Remark 6.7. The complexity of Algorithm 6.1 is exponential in d if the dimension is allowed to grow. For example, the number of vertices of P can be as large as $n^{\Omega(d)}$ [Zie95]. Similarly, the number of cells in Voronoi diagrams of N points in \mathbb{R}^d can be as large as $N^{\Omega(d)}$ [Aur91, For95].

On the other hand, for fixed dimension d Algorithm 6.1 can not be substantially improved, because the input and the output have costs bounded from below by (a polynomial in) n and \overline{src}_v , respectively. This is immediate for the input since P is defined by n hyperplanes. For the output, we claim that the foldout \overline{U}_v in the output of Algorithm 6.1 can not be presented at a smaller cost because it is a (usually nonconvex) polyhedron that has at least \overline{src}_v *boundary ridges*, meaning faces of dimension $d - 1$ in the boundary of \overline{U}_v . To see why, let F be a facet with \overline{src}_v source images, and for each $\nu \in \text{src}_F$ consider a shortest path γ_ν whose sequential unfolding into T_F has endpoint ν . If instead we sequentially unfold the paths γ_ν into T_v , we get $|\text{src}_F| = \overline{src}_v$ segments emanating from v . Extend each of these segments to an infinite ray. Some of these infinite rays might pierce the boundary of \overline{U}_v through faces of dimension less than $d - 1$, but adjusting their directions slightly ensures that each ray pierces the boundary of \overline{U}_v through a boundary ridge. These ridges are all distinct because their corresponding rays traverse different facet sequences.

Of course, the efficiency of Algorithm 6.1 does not necessarily imply that it yields an optimal solution to the discrete geodesic problem—or the unfolding problem, for that matter. (The problem of computing *any* nonoverlapping unfolding, not necessarily the source unfolding, is of independent interest in computational geometry [O'R00].) But although \overline{src}_v is not known to be polynomial in n , we conjecture in Section 9 that it is. See Section 8.10 for more history of the discrete geodesic problem.

Remark 6.8. Following traditions in computational geometry, we have not specified our model of computation. In most computational geometry problems the model is actually irrelevant, since the algorithms are oblivious to it. In our case, however, the situation is more delicate, due to the fact that during each iteration of the loop we make a number of *arithmetic operations* that increase the error. More importantly, we make *comparisons*, which potentially require sharp precision.

Theorem 6.5 and its proof hold as stated for the *complexity over \mathbb{R}* model [BCSS98], where there are no errors, and where all arithmetic operations and comparisons have unit cost. While it would be more natural to consider the (usual) *complexity over \mathbb{Z}_2* model [BCSS98], arithmetic over \mathbb{R} is unfortunately inherent in the problem: the cut locus, the source unfolding, and geodesic distances can all be irrational.

7. Convex polyhedral pseudomanifolds

Recall the notion of polyhedral complex from Section 1. The results in Sections 1–6 hold with relatively little extra work for polyhedral complexes S that are substantially more general than boundaries of polytopes. Since the generality is desirable from the point of view of topology, we shall complete this extra work here.

Suppose that x is a point in a polyhedral complex S . Denote by

$$S_x(\varepsilon) = \{y \in S \mid \mu(x, y) = \varepsilon\}$$

the *geodesic sphere* in S at radius ε from x . If $\langle x \rangle$ is the smallest face of S containing x , then for sufficiently small positive real numbers ε , the intersection $\langle x \rangle \cap S_x(\varepsilon)$ is an honest (Euclidean) sphere $\langle x \rangle_\varepsilon$ of radius ε around x . The set of points N_x in S near x and equidistant from all points on $\langle x \rangle_\varepsilon$ is the *normal space* at x orthogonal to $\langle x \rangle$ in every face containing x . The *spherical link* of x at radius ε is the set

$$N_x(\varepsilon) = \{y \in N_x \mid \mu(x, y) = \varepsilon\}$$

of points in the normal space at distance ε from x . When ε is sufficiently small, the intersection of $N_x(\varepsilon)$ with any k -dimensional face containing x is a sector inside a sphere of dimension $k - 1 - \dim\langle x \rangle$. The metric μ on S induces a subspace metric on the spherical link $N_x(\varepsilon)$. Always assume ε is sufficiently small when $N_x(\varepsilon)$ is written.

Definition 7.1. Let S be a connected finite polyhedral cell complex of dimension d whose facets all have dimension d . Given a point x inside the union S_{d-2} of all faces in S of dimension at most $d - 2$, we say that S is *positively curved* at x if the spherical link $N_x(\varepsilon)$ is connected and has diameter less than $\pi\varepsilon$. The space S is a *convex⁶ polyhedral complex* if S is positively curved at every point $x \in S_{d-2}$.

This definition of positive curvature is derived from the one appearing in [Sto76]. It includes as special cases all boundaries of convex polyhedra; this is essentially the content of Proposition 1.2.

Spherical links give local information about geodesics, as noticed by D. A. Stone (but see also [BGP92, Section 4.2.2]).

Lemma 7.2 ([Sto76, Lemma 2.2]). *Suppose S is a convex polyhedral complex. Then $\tilde{\gamma}$ is a shortest path of length $\alpha\varepsilon$ in the spherical link $N_x(\varepsilon)$ of a point $x \in S$ if and only if the union of all segments connecting points of $\tilde{\gamma}$ to x is isometric (with distances given by the metric on S) to a sector of angle α inside a disk in \mathbb{R}^2 of radius ε .*

Although Stone only uses simplicial complexes, we omit the straightforward generalization to polyhedral complexes. Stone’s lemma forces shortest paths to avoid low-dimensional faces in the presence of positive curvature.

Proposition 7.3. *Proposition 1.2 holds for convex polyhedral complexes S .*

⁶Using ‘convex’ instead of ‘positively curved’ allows usage of the term ‘nonconvex polyhedral complex’ without ambiguity: ‘nonpositively curved’ is already established in the context of CAT(0) spaces to mean (for polyhedral manifolds, at least) that no point has positive sectional curvature in any direction. In contrast, ‘nonconvex’ means that some point has a negative sectional curvature.

Proof. Using notation from Lemma 7.2, suppose that $\alpha < \pi$, and let γ be the segment connecting the endpoints of $\tilde{\gamma}$ through the sector of angle α . Then γ misses x . \square

The rest of Section 1 goes through without change for convex polyhedral complexes after we fix, once and for all, a *tangent hyperplane* $T_F \cong \mathbb{R}^d$ for each facet F . The choice of a tangent hyperplane is unique up to isometry. For convenience, we identify F with an isometric copy in T_F , so that (for instance) we may speak as if F is contained inside T_F . This makes Definition 1.5, in particular, work verbatim here.

The main difficulty to overcome in the remainder of Sections 1–6 is the finiteness in Lemma 2.4. In the context of convex polyhedral complexes, this finiteness is fundamental. It comes down to the fact that shortest paths never wind arbitrarily many times around a single face inside of a fixed small neighborhood of a point. The statement of the upcoming Proposition 7.4 would be false if we allowed infinitely many facets, though it could still be made to hold in that case if the sizes of the facets and their dihedral angles were forced to be uniformly bounded away from zero.

Proposition 7.4. *Fix a real number $r \geq 0$ and a convex polyhedral complex S . There is a fixed positive integer $N = N(r, S)$ such that the facet sequence \mathcal{L}_γ of each shortest path γ of length r in S has size at most N .*

Proof. Pick a real number $\varepsilon > 0$ small enough so that the following holds. First, the sphere $S_x(\varepsilon)$ of radius ε centered at each vertex x only intersects faces containing x . Then, for every point x on an edge but outside the union of the radius ε balls around vertices, the sphere $S_x(\varepsilon/2)$ only intersects faces containing x . Iterating, for every point x on a face of dimension i but outside the union of all the previously constructed neighborhoods of smaller-dimensional faces, the sphere $S_x(\varepsilon/2^i)$ only intersects faces containing x . The existence of such a number ε follows from the fact that every facet of S is convex, and that S has finitely many facets (Definition 7.1).

It suffices to prove the lemma with $r = \varepsilon/2^d$. Let y be the midpoint of γ . The closed ball $B_y(\varepsilon/2^{d+1})$ of radius $\varepsilon/2^{d+1}$ centered at y intersects some collection of faces, and among these there is a face of minimal dimension k . Fix a point x_k lying in the intersection of this face with the ball $B_y(\varepsilon/2^{d+1})$. The ball $B_{x_k}(\varepsilon/2^k)$ contains γ by the triangle inequality. However, $B_{x_k}(\varepsilon/2^k)$ might also contain a point x_j on a face of dimension $j < k$. If so, then choose j to be minimal. Iterating this procedure (at most d times) eventually results in a point x on a face of dimension i such that $B_x(\varepsilon/2^i)$ contains γ and only intersects faces containing x .

The metric geometry of S inside the ball $B_x(\varepsilon/2^i)$ is the same as in $B_{x'}(\varepsilon/2^i)$ for every point x' on the smallest face containing x , as long as $B_{x'}(\varepsilon/2^i)$ only intersects faces containing x' . Since S has finitely many faces by Definition 7.1, we reduce to proving the lemma for shortest paths γ after replacing S by the ball $B = B_x(\varepsilon/2^i)$. In fact, we shall uniformly bound the number of facets traversed by *any* shortest path in B . For simplicity, inflate the metric by a constant factor so that B has radius 2. By a *face of B* we mean the intersection of B with a face of S .

Note that B is isometric to a neighborhood of the apex on the boundary of a right circular cone when the dimension is $d = 2$. In this case, shortest paths in B can pass at most once through each ray emanating from x . We conclude that the lemma holds

in full (not just for B) when $d = 2$. Using induction on d , we shall assume that the lemma holds in full for convex polyhedral complexes of dimension at most $d - 1$.

First suppose that x is not a vertex of S , so the smallest face $\langle x \rangle$ containing x has positive dimension. Then B is isometric to a neighborhood of x in the product $\langle x \rangle \times N_x$ of the face $\langle x \rangle$ with the normal space N_x . Projecting γ onto N_x yields a shortest path $\tilde{\gamma}$ whose facet sequence in the convex polyhedral complex N_x has the same size as \mathcal{L}_γ . Induction on d completes the proof in this case.

Now assume that x is a vertex of S . If one of the endpoints of γ is x itself, then γ is contained in some face of B . Hence we may assume from now on that x does not lie on γ . Consider the radial projection from $B \setminus \{x\}$ to the unit sphere $S_x(1)$ centered at x in B . If the image of γ is a point, then again γ lies in a single face; hence we may assume that radial projection induces a bijection from γ to its image curve $\tilde{\gamma}$. Since the geometry of B is scale invariant, every path γ' in $B \setminus \{x\}$ mapping bijectively to $\tilde{\gamma}$ under radial projection has a well-defined facet sequence equal to \mathcal{L}_γ .

Choose another small real number ε as in the first paragraph of the proof, but with B in place of S . Assume in addition that $\varepsilon < 1/2\pi$. Subdivide $\tilde{\gamma}$ into at least $2^d/\varepsilon$ equal arcs, and use Lemma 7.2 to connect the endpoints of each arc by straight segments in (the cone over $\tilde{\gamma}$ in) B . Lemma 7.2 implies that $\tilde{\gamma}$ has length at most π , because γ is a shortest path. Therefore each of the at least $2^{d+1}\pi$ chords of $\tilde{\gamma}$ has length at most 2^d . The argument in the second paragraph of the proof now produces a new center x' for each chord, and we are assured that $x' \neq x$ because the ε -ball around x does not contain any of the chords. Hence the smallest face $\langle x' \rangle$ containing x' has positive dimension, and we are done by induction on d as before. \square

We shall see in Corollary 7.7 that Proposition 7.4 implies finiteness of the set of source images. But first, we need to introduce the class of polyhedral complexes for which the notion of source image—and hence the rest of Sections 1–6—makes sense.

Definition 7.5. A convex polyhedral complex S of dimension d is a *convex polyhedral pseudomanifold* if S satisfies two additional *pseudomanifold conditions*: (i) each facet is a bounded polytope of dimension d , and (ii) each ridge lies in at most two facets.

Remark 7.6. The “A.D. Aleksandrov spaces with curvature bounded below by 0” of [BGP92] include convex polyhedral pseudomanifolds; see Example 2.9(6) there. Some of our results here, such as surjectivity of exponential maps and non-branching of geodesics, are general—and essentially local—properties of spaces with curvature bounded below by 0. But our focus is on decidedly global issues pertaining to the combinatorial and polyhedral nature of convex polyhedral pseudomanifolds, rather than on a local analogy with Riemannian geometry. That being said, many of our results here can be extended to convex “polyhedral” pseudomanifolds with facets of constant positive curvature instead of curvature zero. We leave this extension to the reader.

A flat point in an arbitrary convex polyhedral complex need not have a neighborhood isometric to an open subset of \mathbb{R}^d , because more than two facets could meet there. In a convex polyhedral pseudomanifold, on the other hand, every flat point not lying on the topological boundary has a neighborhood isometric to an open subset of \mathbb{R}^d . This condition is necessary for even the most basic of our results to hold,

including Corollary 2.2 (whose proof works verbatim for convex polyhedral pseudomanifolds), and the definition of source image (which would require modification without it; see Section 8.3).

We would have preferred to avoid the boundedness condition on facets, but the finiteness of the set of source images in Lemma 2.4 can fail without it; see Section 8.6.

Corollary 7.7. *Lemma 2.4 holds for convex polyhedral pseudomanifolds S .*

Proof. Since every facet is bounded, the lengths of all shortest paths in S are uniformly bounded. Proposition 7.4 therefore implies that there are only finitely many possible facet sequences among all shortest paths in S from the source. \square

Corollary 7.7 yields the following consequences, with the same proofs.

Theorem 7.8. *Proposition 2.6 on the generalization of Mount's lemma and Theorem 2.9 on Voronoi diagrams hold verbatim for convex polyhedral pseudomanifolds S .*

The rest of Section 2 requires slight modification due to the fact that a convex polyhedral pseudomanifold S can have a nonempty topological boundary ∂S .

Proposition 7.9. *Fix a source point v in a convex polyhedral pseudomanifold S . Every warped point lies either in the topological boundary of S or in the cut locus \overline{K}_v .*

Proof. The same as Proposition 2.10, assuming w is not in the boundary of S . \square

In view of Proposition 7.9, the statement of Corollary 2.11 fails for convex polyhedral pseudomanifolds. Instead we get the following, with essentially the same proof.

Corollary 7.10. *If v is a source point in a convex polyhedral pseudomanifold S , then*

1. $\overline{K}_v \cup \partial S$ is polyhedral and pure of dimension $d - 1$, and
2. $\overline{K}_v \cup \partial S$ is the union $K_v \cup S_{d-2} \cup \partial S$ of the cut, warped, and boundary points.

The considerations in Section 3 go through with one small modification: the non-compact flat Riemannian manifold S° is the complement in S of not just the $(d - 2)$ -skeleton S_{d-2} , but also the topological boundary ∂S of S . The notion of what it means that a tangent vector at $w \in S$ can be exponentiated (Definition 3.3) remains unchanged, as long as w lies neither in S_{d-2} nor the boundary of S . Similarly, the notion of source interior (Definition 3.4) remains unchanged except that the exponentials $\exp(t\zeta)$ for $0 \leq t \leq 1$ must lie in neither the cut locus \overline{K}_v nor the boundary ∂S .

Theorem 7.11. *Fix a source point v in the convex polyhedral pseudomanifold S . The exponential map $\exp : \overline{U}_v \rightarrow S$ on the source foldout is a polyhedral nonoverlapping foldout, and the boundary $\overline{U}_v \setminus U_v$ maps onto $\overline{K}_v \cup \partial S$. Hence $\overline{K}_v \cup \partial S$ is a cut set inducing a polyhedral nonoverlapping unfolding $S \setminus (\overline{K}_v \cup \partial S) \rightarrow U_v$ to the source interior.*

Proof. Using Corollary 7.10 in place of Corollary 2.11, the proof is the same as that of Theorem 3.5, except that every occurrence of $S \setminus \overline{K}_v$ must be replaced by $S \setminus (\overline{K}_v \cup \partial S)$, and the open subspace S° must be defined as $S \setminus (S_{d-2} \cup \partial S)$ instead of $S \setminus S_{d-2}$. \square

Corollary 7.12. *Every convex polyhedral pseudomanifold of dimension d is, as a metric space, obtained from a closed, star-shaped, polyhedral ball in \mathbb{R}^d by identifying pairs of isometric boundary components.*

Section 4 concerns local geometry in the context of convex polyhedra, and therefore requires no modification for convex pseudomanifolds, given that all of the earlier results in the paper hold in this more general context.

In Section 5, the only passage that does not seem to work verbatim for convex polyhedral pseudomanifolds is the proof of Proposition 5.5. That proof is presented using language as if F and F' were embedded in the same Euclidean space \mathbb{R}^{d+1} , as they are in the case $S = \partial P$. This embedding can be arranged in the general case here by choosing identifications of T_F and $T_{F'}$ as subspaces of \mathbb{R}^{d+1} in such a way that the copies of F and F' intersect as they do in S .

Finally, the algorithm in Section 6 works just as well for convex polyhedral pseudomanifolds, as long as these spaces are presented in a manner that includes the structure of each facet as a polytope and the adjacency relations among facets. For example, folding maps along ridges shared by adjacent facets can be represented as linear transformations after assigning a vector space basis to each tangent hyperplane.

For the record, let us summarize the previous three paragraphs.

Theorem 7.13. *The results in Sections 4, 5, and 6 hold verbatim for convex polyhedral pseudomanifolds S in place of boundaries of convex polyhedra.*

8. Limitations, generalizations, and history

The main results in this paper are more or less sharp, in the sense that further extension would make certain aspects of them false. In this section we make this sharpness precise, and also point out some alternative generalizations of our results that might hold with requisite modifications. Along the way, we provide more history.

8.1. Polyhedral vs. Riemannian. The study of geodesics on convex surfaces, where $d = 2$, goes back to ancient times and has been revived by Newton and the Bernoulli brothers in modern times. The study of explicit constructions of geodesics on two dimensional polyhedral surfaces was initiated in [Lyu40], and is perhaps much older.

The idea of studying the exponential map on polyhedral surfaces goes back to Aleksandrov [Ale48, §9.5], who introduced it locally when $d = 2$. He referred to images of lines in the tangent space T_F to a facet F as *quasi-geodesic* lines on the surface, and proved some results on them specific to the dimension $d = 2$. Among his other results was the $d = 2$ case of Proposition 1.2.

A detailed analysis of the cut locus of 2-dimensional convex polyhedral surfaces was presented in [VP71]. This paper, seemingly overlooked in the West, gives a complete description of certain convex regions called ‘peels’ in [AO92], which can be used to construct source unfoldings. The approach in [VP71] is inherently 2-dimensional and nonalgorithmic.

The study of exponential maps on Riemannian manifolds is classical [Kob89]. F. Wolter [Wol85] proved properties of cut loci in the Riemannian context that are quite similar to our results describing the cut locus as the closure of the set of cut points. In fact, we could deduce part 2 of our Corollary 2.11 from [Wol85, Lemma 2]—in the manifold case, at least—using Proposition 1.2 (which has no analogue in Riemannian geometry). The method would be to “smooth out” the warped locus to

make a sequence of complete Riemannian manifolds converging (as metric spaces) to the polyhedral complex S , such that the complement of an ever decreasing neighborhood of the warped locus in S is isometric to the corresponding subset in the approximating manifold. Every shortest path to v in S is eventually contained in the bulk complement of the smoothed neighborhood.

This method does not extend to the polyhedral case where S is allowed to be nonconvex, because Proposition 1.2 fails: shortest paths (between flat points) can pass through warped points (Fig. 13). Moreover, the polyhedrality in the first part of Corollary 2.11 fails systematically when S is allowed to be nonconvex (see [MMP87]).

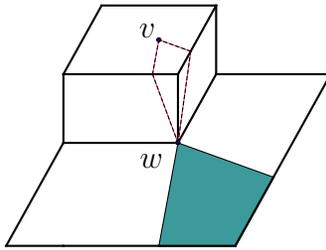


FIGURE 13. Points in the shaded region have exactly two shortest paths to v ; all of these paths go through the warped point w .

8.2. Low-dimensional flat faces. We assumed in Definition 7.1 that faces of dimension $d - 2$ or less in convex polyhedral complexes must be nontrivially curved. Allowing convex polyhedral complexes where low-dimensional faces can be flat would break the notion of facet sequence in Corollary 1.4, and would cause the set of warped points to differ from the union of all closed faces of dimension $d - 2$, in general. The resulting definitions of folding map and sequential unfolding would be cumbersome if not completely opaque. Nonetheless, the resulting definitions would be possible, because shortest paths would still enter facets (and in fact, all faces whose interiors are flat) at well-defined angles. The notion of exponential map would remain unchanged.

Definition 2.3 and Theorem 2.9 should hold verbatim for the modified notion of convex polyhedral pseudomanifold in which low-dimensional flat faces are allowed, because the generalized Mount Lemma (Proposition 2.6) should remain true. Note that Mount's lemma relies mainly on Proposition 1.2 and Corollary 2.2. The latter might be more difficult to verify in the presence of low-dimensional flat faces, because it needs every flat point to have a neighborhood isometric to an open subset of \mathbb{R}^d . Thus one might have to assume S is a *manifold*, and not just a pseudomanifold.

Observe that Fig. 6 depends on not having low-dimensional flat faces: it uses the fact that the vertex bordering the shaded region must lie in the cut locus.

8.3. Why the pseudomanifold conditions? Theorem 2.9 fails for convex polyhedral complexes that are not pseudomanifolds, even when there are no flat faces of small dimension. Indeed, with the notion of cut point set forth in Definition 2.1, entire facets could consist of cut points. To see why, suppose there is a cut point interior

to a ridge lying on the boundary of three or more facets, and note that the argument using Fig. 4 in the proof of Corollary 2.2 fails. For a more concrete construction in dimension $d = 2$, find a convex polyhedral pseudomanifold with a source point so that some edge in the cut locus connects two vertices (for example, take a unit cube with a source point in the center of a facet; see Fig. 1), and then attach a triangle along that edge of the cut locus. The attached triangle (“dorsal fin”) consists of cut points.

The proof of Theorem 2.9 fails for non-pseudomanifolds S when we use the thinness of the cut set in the proof of Proposition 2.6. The appropriate definition of cut point x for convex polyhedral complexes more general than pseudomanifolds should say that two shortest paths from x to the source leave x in different directions—that is, they pierce the geodesic sphere $S_x(\varepsilon)$ at different points. But Corollary 2.2 would still fail for shortest paths entering the “dorsal fin” constructed above.

8.4. Aleksandrov unfoldings. The dimension $d = 2$ foldouts called ‘star unfoldings’ in [AO92, CH96, AAOS97] were conceived of by Aleksandrov in [Ale48, §6.1]. Thus we propose here to use the term ‘Aleksandrov unfolding’ instead of ‘star unfolding’, since in any case these foldouts need not be star-shaped polygons. We remark that a footnote in the same section in [Ale48] indicates that Aleksandrov did not realize the nonoverlapping property, which was only established four decades later [AO92].

Aleksandrov unfoldings are defined for 3-dimensional polytopes P similarly to source unfoldings. The idea is again to fix a source point v , but then slice the boundary S of P open along each shortest path connecting v in S to a vertex. An example of the Aleksandrov unfolding of the cube is given in Fig. 14 (see also Fig. 5).

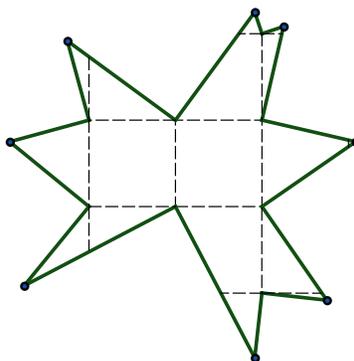


FIGURE 14. Aleksandrov unfolding of the cube (the source point v is on the front face, while the left and back faces have no cuts).

Note that when the source point is in the center of the face, the resulting Aleksandrov unfolding agrees with the source unfolding in Fig. 1.

There is a formal connection between source and Aleksandrov unfoldings. Starting from the source unfolding, cut the star-shaped polygon \bar{U}_v into sectors—these are ‘peels’ as in Section 8.1—by slicing along the shortest paths to images of vertices. Rearranging the peels so that the various copies of v lie on the exterior cycle yields

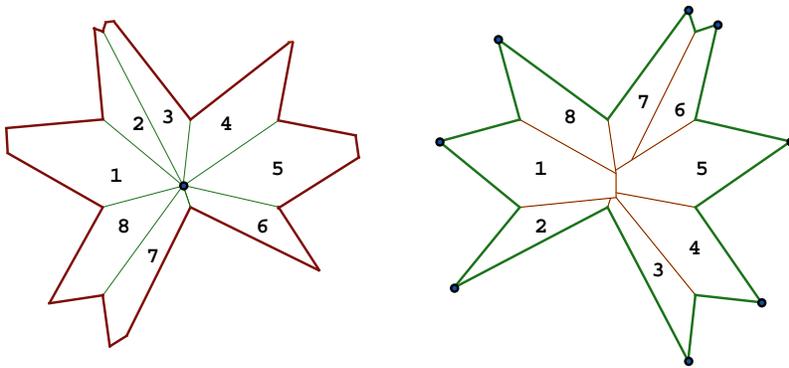


FIGURE 15. Source and Aleksandrov unfoldings of the cube, where the corresponding peels in both unfoldings are numbered from 1 to 8.

a nonoverlapping foldout [AO92] containing an isometric copy of the bulk of the cut locus. This rearrangement is illustrated in Fig. 15, which continues Example 3.6 (see [AAOS97] for further references).

No obvious higher-dimensional analogue of the Aleksandrov unfolding exists, because although the union of all shortest paths connecting the source point to warped points is polyhedral, this complex is not a cut set as per Definition 3.1. Indeed, thinking in terms of source foldouts again, the union of all rays passing from the origin through the images of warped points does not form the $(d - 1)$ -skeleton of a fan of polyhedral cones. Even when $d = 3$, edges of S closer to the source point can make edges farther away look disconnected, as seen from the source point. An example of how this phenomenon looks from v is illustrated in Fig. 16, where the picture is meant to look like the roof of a building as seen from above.

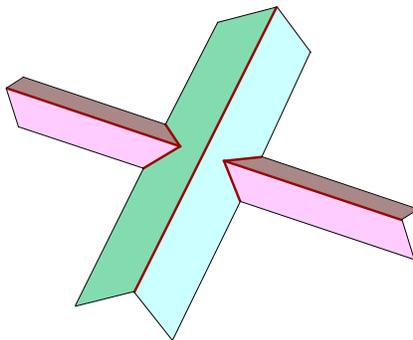


FIGURE 16. Shortest paths to warped points making edges look disconnected.

Circumventing the above failure of Aleksandrov unfoldings in high dimension would necessarily involve dealing with the fact that the set S_{d-2} of warped points generically intersects the cut locus \bar{K}_v in a polyhedral set of dimension $d - 3$. This “warped cut locus” usually contains points interior to maximal faces of \bar{K}_v , making it impossible for these interiors of maximal cut faces to have neighborhoods in S isometric to open

sets in \mathbb{R}^d , even locally. Thus the picture in Fig. 15, where most of the cut locus can lie intact in \mathbb{R}^2 , is impossible in dimension $d \geq 3$. The only remedy would be to make further slices across the interiors of the maximal faces of the cut locus \overline{K}_v before attempting to lay it flat in \mathbb{R}^d . Making these extra slices in a canonical way, to generalize Aleksandrov unfoldings to arbitrary dimension, remains an open problem.

8.5. Definition of source image. Some subtle geometry dictated our choice of definition of ‘source image’ (Definition 2.3). With no extra information available, we might alternatively have tried defining src_F as the (finite) set of endpoints of sequentially unfolded shortest paths

- ending at a point interior to F ; or
- ending anywhere on F , including at a warped point.

Both look reasonable enough; but the first fails to detect faces of dimension $d - 1$ in the cut locus that lie entirely within ridges of S , while the second causes problems with verifying the generalized Mount Lemma (Proposition 2.6) as well as Proposition 5.5 and Lemma 5.8. It is not that the generalized Mount Lemma would be false with these “bonus” source images included, but the already delicate proof would fail. In addition, having these extra source images would add unnecessary bulk to the source poset.

8.6. Finiteness of source images. As we saw in Lemma 2.4 for boundaries of polyhedra, or Proposition 7.4 and Corollary 7.7 for convex polyhedral pseudomanifolds, the number of source images is finite. The argument we gave in Lemma 2.4 relies on the embedding of S as a polyhedral complex inside \mathbb{R}^{d+1} in such a way that each face is part of an affine subspace (i.e. not bent or folded). This embedding can be substituted by the more general condition that the polyhedral metric on each facet is induced by the metric on S (so pairs of points on a single facet are the same distance apart in S as in the metric space consisting of the isolated facet). With this extra hypothesis, we would get finiteness of the set of source images even for convex polyhedral pseudomanifolds whose facets were allowed to be unbounded. However, allowing unbounded facets in arbitrary convex polyhedral pseudomanifolds can result in facets with infinitely many source images.

For example, consider an infinite strip in the plane, subdivided into three substrips (one wide and two narrow, to make the picture clearer). Fix a distance $\ell > 0$, and glue each point on one (infinite) boundary edge of the strip to the point ℓ units away from its closest neighbor on the opposite (infinite) edge of the strip. What results is the cylinder S in Fig. 17. This cylinder would be a convex polyhedral manifold if its facets were bounded. The source foldout \overline{U}_v determined by a source point v in the middle of the wide substrip is depicted beneath S , shrunk vertically by a factor of about 2. The cut locus \overline{K}_v , which is a straight line along the spine of S , divides each substrip into infinitely many regions, so each substrip has infinitely many source images.

8.7. Generic source points. For generic choices of source point v , the source poset will be a chain—that is, a total order on events. The reason is that moving v infinitesimally changes differently the angles from different source images to ridges containing the same event point, precisely because these source images are sequentially unfolded

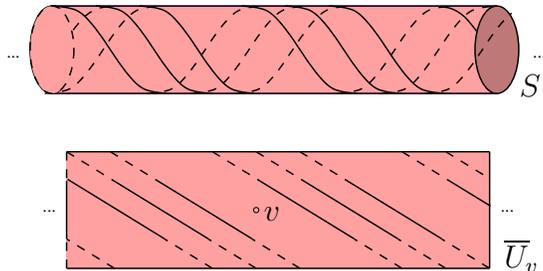


FIGURE 17. A source foldout with infinitely many source images

along shortest paths leaving v in different directions. Note, however, that the *distances* from these various source images to the same event point always remain equal.

8.8. Warped source points. We assumed that the source point $v \in S$ lies in the relative interior of some facet; however, nothing really changes when v lies in the relative interior of some ridge. This can be seen by viewing the exponential map as living on the interior flat points $S^\circ \subseteq S$, as in Section 3.

Moreover, simple modifications can generalize the exponential map to the case where v is warped. However, exponentiation on the complement of the cut locus cannot produce a nonoverlapping foldout in \mathbb{R}^d if $v \in S_{d-2}$, because the resulting cut locus would not be a cut set. Indeed, the cut locus would fail to contain all of S_{d-2} , so its complement could not possibly be isometric to an open subset of \mathbb{R}^d . On the other hand, exponentiation would instead produce a foldout of S onto the tangent cone to S at v . The main point is that the source point still connects to a dense set of points in S via shortest paths not passing through warped points, by Proposition 1.2.

8.9. Multiple source points. Let $\Upsilon = \{v_1, \dots, v_k\} \subset S$ be a finite set of points on the boundary $S = \partial P$ of a convex polyhedron P . Define the *geodesic Voronoi diagram* $\mathcal{V}_S(\Upsilon)$ to be the subdivision of S whose closed cells are the sets

$$V_S(\Upsilon, v_i) = \{w \in S \mid \mu(v_i, w) \leq \mu(v_j, w) \text{ for all } 1 \leq j \leq k\}.$$

Just like the (usual) Voronoi diagrams, computing geodesic Voronoi diagrams is an important problem in computational geometry, with both theoretical and practical applications [AGSS89, KWR97, PL98] (see [Mit00, GO97] for additional references).

Below we modify Algorithm 6.1 to compute the geodesic Voronoi diagrams in S when multiple source points are input. The modified algorithm outputs subdivisions of the facets of S that indicate which source point is closest. More importantly, it also computes which combinatorial type of geodesic gives a shortest path. In the code below, the WHILE-DO loop and the routines remain completely unchanged. The only differences are in the initial and final stages of the pseudocode.

Algorithm 8.1. (Computing the geodesic Voronoi diagram)

INPUT convex polyhedron $P \subset \mathbb{R}^{d+1}$ of dimension $d + 1$, and
flat points v_1, \dots, v_k in the boundary $S = \partial P$
OUTPUT geodesic Voronoi diagram $\mathcal{V}_S(\Upsilon)$ in S

INITIALIZE $\Upsilon_F := \{v_i \mid v_i \in F\}$, $\mathcal{L}_{v_i, F} := (F)$ for $v_i \in \Upsilon_F$, and
 $\mathcal{E}_F := \{(v_i, F, R) \in \widehat{\mathcal{E}}_F \mid \text{POINT } v_i \in \Upsilon_F \text{ CAN SEE } R \text{ THROUGH } F\}$
 $[\dots]$
 COMPUTE for each i : $\text{src}_i := \{(\nu, F) \mid F \in \mathcal{F}, \nu \in \Upsilon_F, \text{ and } \mathcal{L}_{\nu, F} \text{ begins with } F_i\}$, and
 for each i : the subset $V_S(\Upsilon, v_i) := \bigcup_{(\nu, F) \in \text{src}_i} V(\Upsilon_F, \nu) \cap F$ of S
 RETURN geodesic Voronoi diagram $\mathcal{V}_S(\Upsilon) = (V_S(\Upsilon, v_1), \dots, V_S(\Upsilon, v_k))$

That some of the source points v_1, \dots, v_k might lie in the same facet necessitates the call to Routine 6.3 in the initialization of \mathcal{E}_F . As we did before Theorem 6.5, define $\overline{\text{src}}$ to be the maximal number of source images for a single facet.

Theorem 8.2. *Let $P \subset \mathbb{R}^{d+1}$ be a convex polyhedron and $S = \partial P$, with source points v_1, \dots, v_k in $S \setminus S_{d-2}$. For fixed dimension d , Algorithm 8.1 computes the geodesic Voronoi diagram $\mathcal{V}_S(\Upsilon)$ in time polynomial in k , the number n of facets, and $\overline{\text{src}}$.*

The proof is a straightforward extension of the proof of Theorems 6.4 and 6.5; it is omitted. Using observations in Section 8.8, it is possible to modify Algorithm 8.1 to work for a set of arbitrary (that is, possibly warped) source points.

8.10. The discrete geodesic problem. One of our motivating applications for this paper was to the *discrete geodesic problem* of computing geodesic distances and the shortest paths between points v and w in S . The reduction of this problem to computing source unfoldings is easy: construct the source foldout \overline{U}_v in the tangent cone at v , and compute the Euclidian distance between the images.

We should mention here that for $d = 2$ essentially two methods are used in the literature to resolve the discrete geodesic problem: the construction of nonoverlapping unfoldings as above (see [AAOS97, CH96, SS86]), and the so-called ‘continuous Dijkstra’ method, generalizing Dijkstra’s classical algorithm [Dij59] for finding shortest paths in graphs. The second method originated in [MMP87] and is applicable to non-convex surfaces (see also [Kap99] and [SS06], where the appendix to the latter paper contains a critique of the former). Interestingly, this method constructs an explicit geodesic wavefront, and then selects and performs ‘events’ one at a time. However, the time-ordering of events is based on the $d = 2$ fact that the wavefront intersects the union of ridges (edges, in this case) in a finite set of points. Our approach is a combination of these two algorithmic methods, which have previously been separated in the literature. We refer the reader to [Mit00] for more references and results on the complexity of discrete geodesic problems. In general, computing geodesic distances on arbitrary polyhedral complexes remains a challenging problem of both theoretical and practical interest.

9. Open problems and complexity issues

The source poset succeeds at time-ordering the events during wavefront expansion, but it fails to describe accurately how the wavefront bifurcates during expansion, because every event of radius less than r occurs before the first event of radius r in the source poset. On the other hand, the notion of ‘geodesic precedence’ from Definition 5.3 implies a combinatorial structure recording bifurcation exactly.

Definition 9.1. Given a source point v on a convex polyhedral pseudomanifold S , the *vistal tree* $\mathcal{T}(v, S)$ is the set of events, partially ordered by geodesic precedence.

The definition of geodesic precedence immediately implies that $\mathcal{T}(v, S)$ is indeed a rooted tree. It records the facet adjacency graph of the polyhedral decomposition of the source foldout \overline{U}_v into cut cells of dimension d . Equivalently, this data describes the “vista” seen by an observer located at the source point—that is, how the visual field of the observer is locally subdivided by pieces of warped faces. Proposition 5.5 says precisely that the identity map on the set of events induces a poset map from the vistal tree to the source poset. In particular, when the source point is generic as in Section 8.7, the source poset is a linear extension of the vistal tree.

There are numerous interesting questions to ask about the vistal tree, owing to its geometric bearing on the nature of wavefront expansion on convex polyhedra. For example, its size, which is controlled by the extent of branching at each node, is important for reasons of computational complexity (Theorem 6.5).

Conjecture 9.2. *The cardinality $|\text{src}(v, S)|$ of the set of source images for a polyhedral boundary S is polynomial in the number of facets when the dimension d is fixed.*

Hence we conjecture that there is a fixed polynomial f_d , independent of both S and v , such that $|\text{src}(v, S)| < f_d(n)$ for all boundaries $S = \partial P$ of convex polyhedra P of dimension $d + 1$ with n facets, and all source points $v \in S$. Note that the cardinality in question is at most factorial in the number of facets: $|\text{src}(v, S)| < n \cdot (n - 1)! = n!$. Indeed, each source image yields a facet sequence, and each of these has length at most n , starts at with facet F containing v , and does not repeat any facet.

To demonstrate the strength of Conjecture 9.2, the following weaker (but perhaps more natural) claim is an immediate consequence.

Conjecture 9.3. *The number of shortest paths joining any pair of points in a polyhedral boundary is polynomial in the number of facets when the dimension is fixed.*

Conjecture 9.3 says that only polynomially many cut cells can meet at a single point, whereas Conjecture 9.2 says there are only polynomially many cut cells in total.

In contrast, we also believe a stronger statement than Conjecture 9.2 holds for boundaries of convex polyhedra. Given a shortest path γ , both of whose endpoints lie interior to facets, call the facet sequence \mathcal{L}_γ traversed by γ the *combinatorial type* of γ (this is called the *edge sequence* in [Mou85] for the $d = 2$ case).

Conjecture 9.4. *The cardinality of the set of combinatorial types of shortest paths in the boundary S of a convex polyhedron is polynomial in the number of facets of S , when the dimension is fixed.*

That is, we do not require one endpoint to be fixed at the source point. The statement is stronger than Conjecture 9.2 because source images are in bijection with combinatorial types of shortest paths in S with endpoint v . In all three of the previous conjectures, the degree of the polynomial will increase with d , even perhaps linearly. When $d = 2$ all three conjectures have been proved (see [AAOS97, CH96]).

The intuition for Conjecture 9.2 is that, as seen from the source point in a convex polyhedral boundary S , the faces of dimension $d - 2$ more or less subdivide the

horizon into regions. (The horizon is simply the boundary of the source foldout \overline{U}_v , as seen from v .) The phrase ‘more or less’ must be made precise, of course; and our inability to delete it altogether is a result of exactly the same phenomenon in Fig. 16 that breaks the notion of Aleksandrov unfoldings in higher dimension.

The reason we believe Conjecture 9.4 is that we believe Conjecture 9.2, and there should not be too many combinatorial types of vistal trees. More precisely, moving the source point a little bit should not alter the combinatorics of the vistal tree, and there should not be more than polynomially many possible vistal trees. In fact, we believe a stronger, more geometric statement. It requires a new notion.

Definition 9.5. Two source points are *equivistal* if their vistal trees are isomorphic, and corresponding nodes represent the same facet sequences.

Again, the facet sequence corresponding to a node of the vistal tree is the list of facets traversed by any shortest path whose sequential unfolding yields the corresponding source image. Hence two source points are equivistal when their views of the horizon look combinatorially the same.

Conjecture 9.6. *The equivalence relation induced by equivistality constitutes a convex polyhedral subdivision of the boundary S of any convex polyhedron. Moreover, the number of open regions in this subdivision is polynomial in the number of facets of S .*

Independent from the conjecture’s validity, the *vistal subdivision* it speaks of—whether convex polyhedral or not—is *completely canonical*: it relies only on the metric structure of S . In addition, lower-dimensional strata of the vistal subdivision should reflect combinatorial transitions between neighboring isomorphism classes of vistal trees. Thus Conjecture 9.6 gets at the heart of a number of issues surrounding the interaction of the metric and combinatorial structures of convex polyhedra.

Remark 9.7. An important motivation behind the above ideas lies in the computation of the geodesic diameter of the boundary of a convex polytope. This is a classical problem in computational geometry, not unlike computing diameters of finite graphs (for the $d = 2$ case see [AO92, AAOS97]). One possibility, for example, would be to compute the vistal subdivision in Conjecture 9.6, and use this data to list the combinatorial types of shortest paths. Each combinatorial type could then be checked to determine how long its corresponding shortest paths can be. Conjectures 9.4 and 9.6 give hope that the geodesic diameter problem can be solved in polynomial time.

Let us remark here that the polynomial complexity conjectures fail for nonconvex polyhedral manifolds of dimension $d \geq 2$. Note that this does not contradict the fact that when $d = 2$ there exists a polynomial time algorithm to solve the discrete geodesic problem (see section 8.10 above). Indeed, the number of source images gives only a lower bound for *our* algorithm, while the problem is resolved by a different kind of algorithm. On the other hand, we show below that for $d \geq 3$ the discrete geodesic problem is NP-hard. The following result further underscores the difference between the convex and nonconvex case.

Proposition 9.8. *On (nonconvex) polyhedral manifolds, the number of distinct combinatorial types of shortest paths can be exponential in the number of facets. In addition, finding a shortest path on a (nonconvex) polyhedral manifold is NP-hard.*

We present two proofs of the first part: one that is more explicit and works for all $d \geq 2$, and the other that is easy to modify to prove the second part. For the proof of the second part we construct a 3-dimensional polyhedral manifold, which is essentially due to Canny and Reif [CR87]. See Remark 9.9 for comments on how to doctor these manifolds to make them compact and without boundary.

Proof. To obtain a polyhedral domain with exponentially many shortest paths between two points x and y , we consider a dimension $d = 2$ example. Simply take a pyramid shape polyhedral surface as shown in Fig. 18 and observe that there exist 2^k shortest paths between top point v and bottom vertex w , where k is the number of terraces in the pyramid. The omitted details are straightforward.

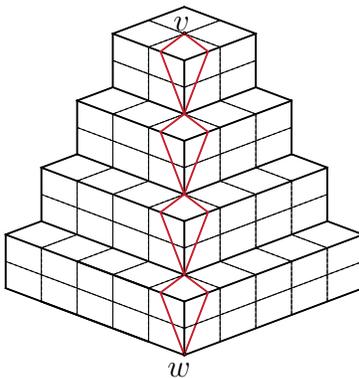


FIGURE 18. Nonconvex polyhedral surface in \mathbb{R}^3 and shortest paths between points v and w .

Now consider a dimension $d = 3$ example of a different type. Polyhedrally subdivide \mathbb{R}^3 by taking the product of a line $\ell = \mathbb{R}$ with the subdivision of \mathbb{R}^2 in Fig. 19. Observe that there are only finitely many cells. Now add $4n$ hyperplanes

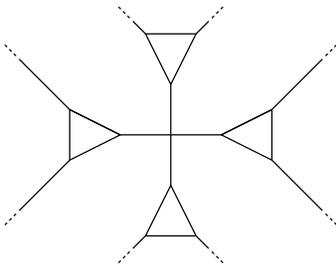


FIGURE 19. Polyhedral subdivision of a planar slice in \mathbb{R}^3

H_0, \dots, H_{4n-1} orthogonal to ℓ , and equally spaced along ℓ . This still leaves finitely

many convex cells. Between hyperplanes H_{4k} and H_{4k+1} , for all $k = 0 \dots n-1$, remove all cells *except* the prisms whose bases are the top and bottom triangles in Fig. 19. Similarly, between hyperplanes H_{4k+2} and H_{4k+3} , remove all cells *except* the prisms whose bases are the left and right triangles in Fig. 19.

Now choose x and y to be points on ℓ , with x being on one side of all the hyperplanes, and y being on the other side. Any shortest path connecting x to y must pass alternately through vertical and horizontal pairs of triangular prisms, and there is no preference for which of the two prisms in each pair the shortest path chooses. Thus the number of shortest paths is at least 4^n , while the number of cells is linear in n .

For the second part, a construction in [CR87] presents a polyhedral domain B where the shortest path solution is NP-hard. This domain B is obtained by removing a set of parallel equilateral triangles from \mathbb{R}^3 . To produce a manifold one has to thicken the triangles into nearly flat triangular prisms. We omit the details. \square

Remark 9.9. The polyhedral manifold S in the above proof is noncompact and has nonempty boundary; but with a little extra work, we could accomplish the same effect using a compact polyhedral manifold without boundary. The idea is to draw a large cube C around S in \mathbb{R}^3 , and place copies C_{top} and C_{bot} of C as the top and bottom facets of a hollow hypercube inside \mathbb{R}^4 . The remaining 6 facets of the hollow hypercube are to remain solid. The result is compact, but still has nonempty boundary in C_{top} and C_{bot} . This we fix by building tall 3-dimensional prisms in \mathbb{R}^4 on the boundary faces, orthogonal to C_{top} and C_{bot} , pointing away from the hypercube. Then we can cap off the prisms with copies of the cells originally excised from $C \subset \mathbb{R}^3$ to get a nonconvex polyhedral 3-sphere in \mathbb{R}^4 .

Remark 9.10. The reader should not be surprised by the fact that computing the geodesic distance is NP-hard for nonconvex manifolds. On the contrary: in most situations the problem of computing the shortest distance is intractable, and in general is not in NP. We refer to [MS04] for further hardness results in the geometric context. In a different, more traditional context, finding the shortest distance in a Cayley graph between two elements in a permutation group (presented by a list of generators in S_N) is known to be NP-hard even for abelian groups [EG81]. Furthermore, for directed Cayley graphs the problem is PSPACE-complete [Jer85].

Our final conjecture concerns the process of unfolding boundaries of convex polyhedra: if someone provides a polyhedral nonoverlapping foldout made of hinged wood, is it always possible to glue its corresponding edges together? Because wood is rigid, we need not only a nonoverlapping property on the foldout as it lies flat on the ground, but also a nonintersecting property as we continuously fold it up to be glued.

Viewing this process in reverse, can we continuously unfold the polyhedral boundary so that all dihedral angles monotonically increase, until the whole polyhedral boundary lies flat on a hyperplane? This idea was inspired by recent works [BC02, CDR03] and was suggested by Connelly.⁷ While the monotone increase of the dihedral angles may seem an unnecessary condition justified only by the aesthetics of the blooming, it is in fact crucial in the references above.

⁷Private communication.

As we have phrased things above, we asked for continuous unfolding of an arbitrary nonoverlapping foldout. But in fact, we only want to ask that there *exist* a foldout that can be continuously glued without self-intersection. Let us be more precise.

Definition 9.11. Let S be the boundary of a convex polyhedron of dimension $d + 1$ in \mathbb{R}^{d+1} . A *continuous blooming* of S is a choice of nonoverlapping foldout $\bar{U} \rightarrow S$, and a homotopy $\{\phi_t : \bar{U} \rightarrow \mathbb{R}^{d+1} \mid 0 \leq t \leq 1\}$ such that

1. ϕ_0 is the foldout map $\bar{U} \rightarrow S$;
2. ϕ_1 is the identity map on \bar{U} ;
3. ϕ_t is an isometry from the interior U of \bar{U} to its image, and ϕ_t is linear on each component of the complement of the cut set in each facet, for $0 < t < 1$; and
4. the dihedral angles between corresponding facets of $\phi_t(\bar{U})$ increase as t increases.

An example of a continuous blooming is given in Figure 20 below.

Conjecture 9.12. *Every convex polyhedral boundary has a continuous blooming.*

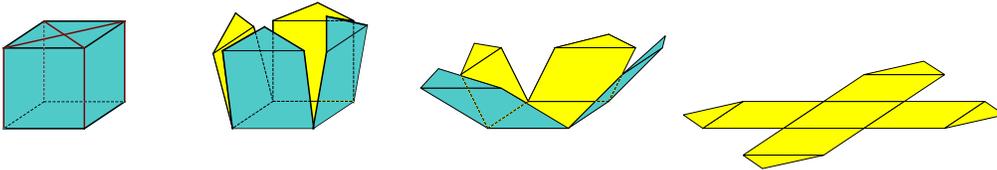


FIGURE 20. An example of a continuous blooming of the surface of the cube.

Even though we ask only for existence, we believe that in fact the source unfolding can be continuously bloomed. As far as we know, this is open even for $d = 2$. Interestingly, we know of no nonoverlapping unfolding that cannot be continuously bloomed, and remain in disagreement on their potential existence.

Acknowledgments. We are grateful to Boris Aronov, Yuri Burago, Bob Connelly, Erik Demaine, Maksym Fedorchuk, Robin Forman, Tracy Hall, Bob MacPherson, Jon McCammond, David Mount, János Pach, Micha Sharir, Alexey Tarasov, and Santosh Vempala for helpful conversations. We are indebted to Robin Forman, Joe Malkevitch and Frank Wolter for references, to Frank Sottile for finding a subtle error in an earlier version, and especially to Günter Rote and (non-anonymous) referees Joe Mitchell and Joseph O’Rourke for a careful reading and numerous helpful comments.

The first author was partially supported by the National Science Foundation, and he wishes to acknowledge the Mathematisches Forschungsinstitut Oberwolfach for a stimulating research environment during the week-long program on Topological and Geometric Combinatorics (April, 2003). Most of this work was completed while the first author was at the Massachusetts Institute of Technology (Cambridge, MA) and the Mathematical Sciences Research Institute (Berkeley, CA). The second author was partially supported by the National Security Agency and the National Science Foundation. He would like to thank the organizers of the “Second Geometry Meeting

dedicated to A.D. Aleksandrov” held at the Euler International Mathematical Institute in St. Petersburg (June, 2002), where these results were originally presented.

REFERENCES

- [AAOS97] P. K. Agarwal, B. Aronov, J. O’Rourke, and C. A. Schevon, Star unfolding of a polytope with applications, *SIAM J. Comput.* **26** (1997), no. 6, 1689–1713.
- [AGSS89] A. Aggarwal, L. J. Guibas, J. Saxe, and P. W. Shor, A linear-time algorithm for computing the Voronoi diagram of a convex polygon, *Discrete & Comput. Geom.* **4** (1989), no. 6, 591–604.
- [Ale48] A. D. Aleksandrov, *Vnutrennyaya geometriya vypuklykh poverkhnostey* (in Russian), M.–L.: Gostekhizdat, 1948; English translation: *Selected works. Intrinsic geometry of convex surfaces*, Vol. 2, Chapman & Hall/CRC, Boca Raton, FL, 2005.
- [Ale50] A. D. Aleksandrov, *Vypuklye mnogogranniki* (in Russian), M.: Gostekhizdat, 1950; English translation: *Convex polyhedra*, Springer, Berlin, 2005.
- [AO92] B. Aronov and J. O’Rourke, Nonoverlap of the star unfolding, *Discrete & Comput. Geom.* **8** (1992), no. 3, 219–250.
- [Aur91] F. Aurenhammer, Voronoi diagrams—a survey of a fundamental geometric data structure, *ACM Comput. Surv.* **23** (1991), 345–405.
- [BC02] K. Bezdek and R. Connelly, Pushing disks apart—the Kneser–Poulsen conjecture in the plane, *J. Reine Angew. Math.* **553** (2002), 221–236.
- [B+03] M. Bern, E. D. Demaine, D. Eppstein, E. Kuo, A. Mantler, and J. Snoeyink, Ununfoldable polyhedra with convex faces, *Comput. Geom.* **24** (2003), 51–62.
- [BCSS98] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer, New York, 1998.
- [BGP92] Yu. Burago, M. Gromov, and G. Perelman, A.D. Alexandrov spaces with curvature bounded below, *Russian Math. Surveys* **47** (1992), no. 2, 1–58.
- [CR87] J. F. Canny and J. H. Reif, New lower bound techniques for robot motion planning problems, *Proc. 28-th IEEE FOCS* (1987), 49–60.
- [Cha91] B. Chazelle, An optimal convex hull algorithm and new results on cuttings, *Proc. 32-nd IEEE FOCS* (1991), 29–38.
- [CH96] J. Chen and Y. Han, Shortest paths on a polyhedron. I. Computing shortest paths, *Internat. J. Comput. Geom. Appl.* **6** (1996), no. 2, 127–144.
- [CDR03] R. Connelly, E. D. Demaine and G. Rote, Straightening polygonal arcs and convexifying polygonal cycles, *Discrete & Comp. Geom.* **30** (2003), no. 2, 205–239.
- [Dij59] E. W. Dijkstra, A note on two problems in connexion with graphs, *Numer. Math.* **1** (1959), 269–271.
- [EG81] S. Even and O. Goldreich, The minimum-length generator sequence problem is NP-hard, *J. Algorithms* **2** (1981), no. 3, 311–313.
- [For95] S. Fortune, Voronoi diagrams and Delaunay triangulations, in *Computing in Euclidean geometry* (F. Hwang and D. Z. Du, eds.), World Scientific, Singapore, 1995, 225–265.
- [GO97] J. E. Goodman and J. O’Rourke (eds.), *Handbook of discrete and computational geometry*, CRC Press, Boca Raton, FL, 1997.
- [Jer85] M. R. Jerrum, The complexity of finding minimum-length generator sequences, *Theoret. Comput. Sci.* **36** (1985), no. 2-3, 265–289.
- [Kap99] S. Kapoor, An efficient computation of geodesic shortest paths, *Proc. of the 31-st ACM STOC* (1999), 770–779.
- [Kob89] S. Kobayashi, On conjugate and cut loci, in *Global differential geometry*, MAA, Washington, DC, 1989, pp. 140–169.
- [KWR97] R. Kunze, F. E. Wolter, T. Rausch, Geodesic Voronoi Diagrams on Parametric Surfaces, in *Proc. Comp. Graphics Int.* (1997), Hasselt-Diepenbeek, Belgium, 230–237.

- [Lyu40] L. A. Lyusternik, *Geodesic lines. The shortest paths on surfaces* (in Russian), M.–L.: Gostekhizdat, 1940.
- [Mit00] J. S. B. Mitchell, Geometric shortest paths and network optimization, in *Handbook of computational geometry*, 633–701, North-Holland, Amsterdam, 2000.
- [MMP87] J. S. B. Mitchell, D. M. Mount, and C. H. Papadimitriou, The discrete geodesic problem, *SIAM J. Comp.* **16** (1987), no. 4, 647–668.
- [MS04] J. S. B. Mitchell and M. Sharir, New Results on Shortest Paths in Three Dimensions, in *Proc. 20th ACM Sympos. Comput. Geom.*, New York, 2004, 124–133.
- [Mou85] D. M. Mount, *On finding shortest paths on convex polyhedra*, Technical Report 1495, Dept. of Computer Science, Univ. of Maryland, Baltimore, MD, 1985.
- [O’R00] J. O’Rourke, Folding and unfolding in computational geometry, in *Discrete and Computational Geometry* (Tokyo, 1998), Springer, Berlin, 2000, pp. 258–266.
- [PL98] E. Papadopoulou and D. T. Lee, A new approach for the geodesic Voronoi diagram of points in a simple polygon and other restricted polygonal domains, *Algorithmica* **20** (1998), no. 4, 319–352.
- [PS85] F. P. Preparata and M. I. Shamos, *Computational geometry. An introduction*, Texts and Monographs in Computer Science, Springer, New York, 1985.
- [Sch97] W. Schlickenrieder, *Nets of polyhedra*, Diplomarbeit, TU Berlin, 1997.
- [SS06] Y. Schreiber and M. Sharir, An efficient algorithm for shortest paths on a convex polytope in three dimensions, preliminary version (see also the extended abstract in *Proc. 22nd ACM Sympos. Comput. Geom.*, Sedona, Arizona, 2006.)
- [SS86] M. Sharir and A. Schorr, On shortest paths in polyhedral spaces, *SIAM J. Comp.* **15** (1986), no. 1, 193–215.
- [Sto76] D. A. Stone, Geodesics in piecewise linear manifolds, *Trans. AMS* **215** (1976), 1–44.
- [S+05] V. Surazhsky, T. Surazhsky, D. Kirsanov, S. J. Gortler, and H. Hoppe, Fast exact and approximate geodesics on meshes, in *Proc. ACM SIGGRAPH 2005*, 553–560.
- [Tar99] A. S. Tarasov, Polyhedra that do not admit natural unfoldings, *Russian Math. Surveys* **54** (1999), no. 3, 656–657.
- [VP71] Ju. A. Volkov and E. G. Podgornova, The cut locus of a polyhedral surface of positive curvature (in Russian), *Ukrainian Geometric Sbornik* **11** (1971), 15–25.
- [Wol85] F. E. Wolter, *Cut loci in bordered and unbordered Riemannian manifolds*, Ph.D. thesis, TU Berlin, FB Mathematik, Berlin, Germany, 1985.
- [Zie95] G. M. Ziegler, *Lectures on polytopes*, Springer, New York, 1995.