Source Unfolding Convex Polytopes

Written Comprehensive Exam

Joseph Thomas

May 2, 2013

Contents

1 Introduction

The study of folding problems is quite old and is still actively researched. Much modern interest in this area is driven by industrial applications in which various structures must be constructed from flat, planar materials. The area also features theoretical problems of surprising difficulty. For example, the question of whether it is possible to cut an arbitrary convex polytope along its edges to obtain a non-overlapping unfolding into the plane has remained open for more than four centuries and is still studied today.

These notes discuss another type of unfolding called source-unfolding. Unlike edge-unfolding, we allow "cuts" to occur on any part of the polytope. Also, we designate a point v on the polytope to be the "source point" and insist that when the polyhedron is unfolded v should be the center of a star-shaped region, with the property that lines from v to other points in the region correspond to shortest paths on the polytope (that have v as an endpoint). Unlike many other types of unfolding, techniques for computing source unfoldings readily generalize to higher dimensions.

Source unfoldings lie at the intersection of several areas of mathematics, particularly Riemannian geometry, computational geometry, and combinatorial optimization. From the perspective of Riemannian geometry, one views the polytope's boundary as a flat Riemannian manifold S with a finite number of singularities. In this context, a source unfolding is a particular inverse of the exponential map. Such an inverse function Φ encodes useful information about shortest paths on S with endpoint v. Specifically, source unfoldings encode (in a transparent way) a one-to-one correspondence between shortest paths starting from v and line segments in the unfolded polytope starting from $\Phi(v)$. To find a shortest path on S between v and another point p, we simply calculate the Φ-preimage of the line segment from $\Phi(v)$ to $\Phi(p)$ in the unfolding. From the perspective of combinatorial optimization, then, the problem of computing a source unfolding may be viewed as a "single-source, all-destinations shortest path problem." This perspective is useful because well understood examples within this class of optimization problems suggest how to design an algorithm to compute a source unfolding. In this way, Riemannian geometry provides us with the language necessary to *analyze* source unfoldings, while computational geometry and combinatorial optimization suggest techniques to *synthesize* the desired unfolding.

Section 3 covers this analysis. Because the Riemannian manifolds we investigate are very specialized, geodesics on these manifolds admit a useful discrete characterization that is developed in Sections 3.1-3.3. This discrete characterization depends partly on the fact that the manifolds in question admit a notion of positive curvature at their singularities. This notion of curvature is made precise in Section 2. In Section 3.4, a result called "Mount's Lemma" explains how the desired inverse of the exponential map may be computed using certain Voronoi diagrams on the faces of the polytope. From this analysis, it will become clear that a source unfolding is completely determined by a certain finite collection of *source images* in the manifold's tangent bundle.

Sections 4 and 5 explain how the source images may be computed using a greedy optimization procedure. To specify the procedure, we develop the notion of a *jet frame* in Section 4, which gives a precise description of a "signal" or "wave front" that propagates from the source over the manifold. In Section 5, we explain how to use jet frames to compute the source images. Finally, Section 6 specifies the algorithm described in Section 5, proves its correctness, and briefly discusses its performance.

The idea of source unfolding was developed in the early 2000's by Ezra Miller and Igor Pak, and is described in their paper "Metric Combinatorics of Convex Polyhedra: Cut Loci and Nonoverlapping Unfoldings" [1]. Structurally speaking, our exposition follows their paper closely, though the author reworked many concepts and proofs in each section. As a simple example, Miller and Pak prefer to prove many properties of shortest paths by contradiction. This has the advantage of making arguments very short. However, if one argues the same results by proving the contrapositive (as we do) one gets a better sense of how paths on S may be shortened.

Since the subject of [1] lies in the intersection of several fields, different researchers may consider different definitions and results to be so fundamental that they may be omitted. We will assume the basics of differential geometry as a foundation, but in Section 2 we will present many of the basics of polytope theory. We also present arguments like Theorem 30, Lemma 44, and Lemma 50, in order to make our arguments for certain results from Miller and Pak's paper more concrete.

2 Background and Conventions

In this paper, we investigate manifolds that arise as the boundaries of convex polytopes. The theory of convex polytopes is quite extensive, so we will only list the few facts about polytopes that we require for later sections. As part of our investigation, we will also need to prove many theorems about paths on manifolds. Naturally, working efficiently with paths requires adopting some conventions regarding how paths should be parametrized and when two paths should be considered equivalent. These conventions are presented in Subsection 2.2.

2.1 Polytopes

Definition 1. Suppose $U \subset \mathbb{R}^n$ is convex. The **affine span** of U is the set:

$$\text{aff}(U) := \{\lambda \cdot (p - q) : p, q \in U, \lambda \in \mathbb{R}\}$$

Note that $\text{aff}(U)$ will always be an affine, k-dimensional plane in $\mathbb{R}^n$.

Readers familiar with smooth and Riemannian manifolds should note that this definition conflates the manifold $\mathbb{R}^n$ with the tangent spaces of various points in $\mathbb{R}^n$. This is typical in polytope theory. Occasionally, we will use the notation $T_F := \text{aff}(F)$ in order to emphasize that a d-face F may be identified with a subset of $\text{aff}(F)$ and that $\text{aff}(F)$ is the d-plane tangent to face F.

We have two equivalent definitions of a convex polytope:

- P is a n-dimensional polytope if P is a nonempty, bounded intersection of half-spaces in $\mathbb{R}^n$ such that the affine span of P (denoted $\text{aff}(P)$) is all of $\mathbb{R}^n$.
- P is the convex hull of a finite set of points in $\mathbb{R}^n$, such that the affine span of those points is all of $\mathbb{R}^n$.

The proof that these are equivalent definitions is given in Chapter 1 of [2]. Though the two definitions are mathematically equivalent, it is not algorithmically trivial to translate between the two presentations of a polytope.

Remark 2. Given a polytope P of dimension n, ∂P is a manifold of dimension $(n-1)$, embedded in $\mathbb{R}^n$. Since we will speak more frequently of this manifold than of P itself, our convention will be to let d denote the dimension of ∂P, so that P has dimension $(d+1)$.

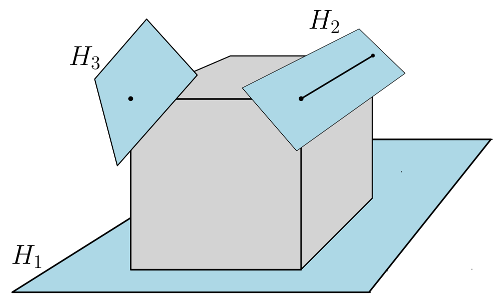

Figure 1: A unit cube in $\mathbb{R}^3$, with several planes.

Definition 3. A set $H \subseteq \mathbb{R}^n$ is a **hyperplane** if there exists $\varphi \in (\mathbb{R}^n)^*$ (the dual of the vectorspace $\mathbb{R}^n$) and a constant C so that

$$H = \{x \in \mathbb{R}^{d+1} : \varphi(x) = C\}$$

It is not hard to see that $\mathbb{R}^n \setminus H$ has two connected components, each of which is an open **half-space**. We denote these half-spaces as:

$$H^> = \{x \in \mathbb{R}^{d+1} : \varphi(x) > C\}$$
$$H^< = \{x \in \mathbb{R}^{d+1} : \varphi(x) < C\}$$

Definition 4. Suppose P is a convex polytope embedded in $\mathbb{R}^{d+1}$. A hyperplane H is said to **cut** P if both $H^>$ and $H^<$ have nonempty intersection with P. When $H^> \cap P = \varnothing$ and both $H \cap P$ and $H^< \cap P$ are nonempty, H **supports** p. If $P \subseteq H^< \cup H$, the inequality $\varphi(x) \leq c$ is **valid** for P and that $H \cap P$ is a **face**. The **dimension** of a face is the dimension of the affine span of $H \cap P$, if this set is nonempty, and -1 otherwise. When we want to succinctly denote that a face F has dimension k, we may speak of "the k-face F."

Example 5. Consider Figure 1. H_1 intersects the cube in a 2-face, H_2 intersects the cube in a 1-face, and plane H_3 intersects the cube in a 0-face.

The definition of a face implies P and $\varnothing$ are always faces of P, of dimensions $(d+1)$ and -1 respectively. The usual subset relation allows us to view the collection $\mathcal{F}(P)$ of all faces of P as a partially ordered set (or poset).

In our theory of polytopes, faces of dimensions d, $(d-1)$ and $(d-2)$ play a role similar to the role coordinate charts play in smooth manifold theory. This idea becomes more powerful if we introduce some ideas from topology.

Suppose F is a k-face of a $(d+1)$-dimensional polytope $P \subset \mathbb{R}^{d+1}$. Then $\text{aff}(F)$ is a k-dimensional affine subspace of $\mathbb{R}^{d+1}$, and hence is isometric to $\mathbb{R}^k$. Accordingly, we can speak of $B_F(x, \varepsilon)$, a k-dimensional Euclidean ball within $\text{aff}(F)$ of radius ε, centered at x.

Definition 6. If F is a k-face of a polytope P, the **relative interior** of F, denoted $\text{rint}(F)$, is the set of all $x \in F$ such that there exists an $\varepsilon > 0$ for which $B_F(x, \varepsilon) \cap \text{aff}(F) \subseteq F$. The relative boundary of F, denoted $\text{rbdy}(F)$, is defined to be $F \setminus \text{rint}(F)$.

Example 7. Consider the unit cube $C = [0,1] \times [0,1] \times [0,1] \subseteq \mathbb{R}^3$ in Figure 2. The relative interior of C is the interior of C in the usual topology on $\mathbb{R}^3$. Likewise, the relative boundary of C is ∂C. The relative interior of face F is isometric to the open square $(0,1) \times (0,1) \subseteq \mathbb{R}^2$. The relative boundary of F is the union of the 1-faces E_1, E_2, E_3 and E_4. The relative interior of E_1 is a line segment isometric to $(0,1) \subseteq \mathbb{R}$.

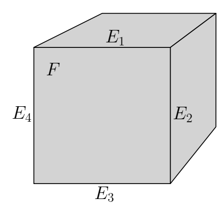

Figure 2: A Unit Cube in $\mathbb{R}^3$

The poset of faces has further combinatorial structure which we will describe with ideas from lattice theory.

Definition 8. A **lattice** is a poset (equipped with a partial ordering we will denote by $\leq$) in which any two elements have a unique supremum (or **join**) and a unique infimum (or **meet**). A **graded** lattice L is equipped with a function $r : L \to \mathbb{N}$, so that for any $x, y \in L$, $x < y$ implies $r(x) < r(y)$.

Suppose L is a lattice and 0 denotes an element of L such that $0 < x$ for any $x \in L$. An element $x \in L \setminus \{0\}$ is an **atom** if there does not exist a $y \in L$ so that $0 < y < x$. An **atomic** lattice is one in which every $z \in L$ may be uniquely identified as the join of a finite number of atoms.

Proposition 9. *Suppose P is a polytope. The poset consisting of the faces of P together with $\subseteq$ is a lattice with the following properties:*

- *The face lattice is graded and the grading is given by the dimensions of the faces.*

- *The face lattice is atomic and the atoms are 1-dimensional faces.*

- *(The Diamond Property) If we fix a k-dimensional face of F and a $(k-2)$-face V contained in F, there are exactly two $(k-1)$-faces E_1, E_2 such that $V \subseteq E_1, E_2 \subseteq F$.*

For the proof of these, and further properties of rint and rbdy see Chapter 2 of [2]. We will require the Diamond Property to define a notion of angle in higher dimensions. The second property, concerning atomicity, can be useful for proving two faces F_1, F_2 are equal — instead of trying to argue $F_1 = F_2$ as subsets of Euclidean space, we can instead argue they contain exactly the same 1-faces.

Notation 10. For the remainder of the paper, we will use S to denote ∂P. We will use S_{d-1} to denote the union of all $(d-1)$-faces of S and the notation S_{d-2} to denote the union of all $(d-2)$-faces of S. Notice $S_{d-2} \subset S_{d-1} \subset S$. These sets receive designations because they have special influence on which paths on the manifold can be shortest paths.

2.2 Paths and Distances on Polyhedral Manifolds

Definition 11. Suppose M is a smooth n-manifold and $I \subseteq \mathbb{R}$ is an interval. A **path** is a continuous function $\gamma : I \to M$.

Given t in the interior of I, we say that γ is C^m at t (or γ is C^m at $\gamma(t)$) if, for any coordinate chart (U, Φ) (where $U \subset M$ is open and $\Phi : U \to \widetilde{U} \subseteq \mathbb{R}^n$ is a homeomorphism), all n component functions of the mapping $\Phi \circ \gamma : I \to \mathbb{R}^n$ have at least m continuous derivatives at t. We understand $m = \infty$ to mean that the component functions have continuous derivatives of all orders. If γ is C^m at t for every t in the interior of I, we say that γ is C^m. If γ is C^∞, we say it is a **smooth** path.

We will often need to decompose paths into pieces by restriction and join them together "end to end" to create a longer path.

Definition 12. Given paths $\gamma_1 : [a, b] \to M$ and $\gamma_2 : [c, d] \to M$, with

$\gamma_1(b) = \gamma_2(c)$, their **concatenation** is the path $\gamma : [0, 2] \to M$ given by:

$$\gamma(t) = \begin{cases} \gamma_1(a + t(b - a)) & \text{if } t \leq 1 \\ \gamma_2(c + (t - 1)(d - c)) & \text{if } t > 1 \end{cases}$$

Definition 13. A path γ is **piecewise** C^m if it can be expressed as the concatenation of some finite number of paths $\gamma_1, \ldots, \gamma_k$, each of which is C^m.

Given a path $\gamma : I \to M$, we are sometimes more concerned with the image $\gamma(I) \subseteq M$ than with the function γ itself. For this reason, when the context is clear we will use γ to denote $\gamma(I)$. For instance, we might write $\gamma \cap F$ to indicate the set of all points within a face F that are traversed by γ.

In the course of proving theorems, we will frequently need to quantify over a space of paths. If we regard paths as functions, this can make some results (particularly uniqueness results) difficult to state. As an example, consider paths $\gamma : [0, 1] \to M$ and $\gamma' : [0, 2] \to M$, related by $\gamma(t) := \gamma'(2t)$. For our purposes, γ' and γ trace out the same shape in M, yet $\gamma' \neq \gamma$.

Definition 14. We will regard two paths $\gamma : I \to M$ and $\gamma' : I' \to M$ as equivalent if there exists an increasing continuous function $u : I \to I'$ so that $\gamma = \gamma' \circ u$.

For the remainder of the paper, our convention will be to consider paths to be piecewise C^∞, unless stated otherwise. For this class of paths, and our particular manifolds of interest, it is particularly simple to assign a notion of length to a path.

Suppose S is the boundary of a polytope embedded in $\mathbb{R}^{d+1}$. Then a smooth path in S may be regarded as a piecewise-smooth path in $\mathbb{R}^{d+1}$, via the embedding. The **arclength** of a smooth path $\gamma : [a, b] \to \mathbb{R}^n$ in a Euclidean space is given by:

$$L(\gamma) := \int_a^b \|\gamma'(t)\| \, dt$$

One can then extend L to piecewise smooth paths on M, by summing the lengths of the smooth pieces. Three theorems that typically follow this definition in an undergraduate analysis text (such as [3]) are that:

- $L(\gamma)$ does not depend on the parametrization of γ.

- Paths may be parametrized by arclength.
- In $\mathbb{R}^n$, a straight line between two points has a length weakly smaller than any other path between those points.

Though it is certainly possible to state an intrinsic definition for path length, in the interest of simplicity we will use the definition above. For notational convenience, we may write $L\gamma$ rather than $L(\gamma)$.

L induces a metric on S, $d_S : S \times S \to \mathbb{R}$, given by:

$$d_S(p, q) = \inf \{L\gamma : \gamma \text{ a piecewise } C^\infty \text{ path from } p \text{ to } q\}$$

Let us briefly check any two points $p, q \in S$ may be connected by a path that is not only continuous but actually piecewise smooth, so that the set above is nonempty. Since S is path connected, there exists a path $\gamma : [a, b] \to S$ from p to q. Since S is compact, there exists a finite subdivision of $[a, b]$ into a finite number of subintervals $[a_i, b_i]$ with the following property: For each i, there exists a coordinate neighborhood (U_i, Φ_i) such that for all $t \in [a_i, b_i]$, $\gamma(t) \in U$. On each interval, we may replace $\gamma|_{[a_i,b_i]}$ with a piecewise smooth path in coordinates, if need be, to produce a piecewise smooth path $\widetilde{\gamma}$ from p to q.

Proposition 15. *The function d_S is a metric, namely:*

- $d_S(p, q) \geq 0$ *for all* $p, q \in S$, *with equality if and only if* $p = q$.
- $d_S(p, r) \leq d_S(p, q) + d_S(q, r)$ *for all* $p, q, r \in S$.

Proof. That $d_S(p, q) \geq 0$ for all $p, q \in S$ is clear from the fact that L is a non-negative function. If $p = q$, then any constant path γ that remains at p is a path from p to q of length 0, hence $0 \leq d_S(p, q) \leq L\gamma = 0$. Finally, suppose $p \neq q \in S$. Recall that S is embedded in $\mathbb{R}^{d+1}$, so that:

$$\{\gamma : \gamma \text{ a path in } S \text{ from } p \text{ to } q\} \subset \left\{\gamma : \gamma \text{ a path in } \mathbb{R}^{d+1} \text{ from } p \text{ to } q\right\}$$
$$0 = \inf \{L\gamma : \gamma \text{ a path in } S \text{ from } p \text{ to } q\} \geq \inf \left\{L\gamma : \gamma \text{ a path in } \mathbb{R}^{d+1} \text{ from } p \text{ to } q\right\} \geq 0$$

But since we know that a line segment in $\mathbb{R}^{d+1}$ between p and q has length equal to the infimum on the right, we conclude that $p = q$.

For the second property, suppose $p, q, r \in S$. Suppose γ_1 is a piecewise smooth path from p to q and γ_2 is a piecewise smooth path from q to r.

Then their concatenation is a piecewise smooth path of length $L(\gamma_1) + L(\gamma_2)$ from p to r, and hence $d_S(p, r) \leq L(\gamma_1) + L(\gamma_2)$. Since we chose γ_1 and γ_2 generally from the classes of piecewise smooth paths from p to q and from q to r respectively, we may conclude that $d_S(p, r) \leq d_S(p, q) + d_S(q, r)$. ∎

2.3 Face Angles and Neighborhoods of S^{d-2}

When studying the boundary of a convex $(d+1)$-dimensional polytope, one is often interested in the d, $(d-1)$, and $(d-2)$ dimensional faces of the polytope. When $d = 2$, these are the faces, edges, and vertices of the polytope respectively, and there is a natural notion of the angle between two edges that share a vertex. Intuitively, in this case, the convexity of the polytope means the sum of the face angles around any vertex must be strictly less than 2π. We will describe how this idea may be generalized to higher dimensions. This, in turn, will provide a way to characterize certain neighborhoods of any point in the relative interior of a $(d-2)$-face. This characterization will be crucial to proving a result in the next section.

2.3.1 Face Angles for 3-Dimensional Polytopes

To prove results about face angles in higher dimensions, we will reduce the problem to the case of polytopes embedded in $\mathbb{R}^3$. In this subsection, we analyze this (very intuitive) situation.

Proposition 16. *Suppose T is a (non-degenerate) tetrahedron embedded in $\mathbb{R}^3$, with vertex v. Then among the three face angles adjacent to v, any one face angle is strictly smaller than the sum of the other two.*

Proof. Let w, x, y denote the other vertices of T. Assume that $\angle wvy$ has the greatest magnitude among the face angles adjacent to v, so that it suffices to show $\angle wvy < \angle xvy + \angle wvx$.

Without loss of generality, we may assume $\angle vwx = \angle vwy$, since if this is not the case, we may truncate T by a 2-plane to obtain another tetrahedron with this configuration. Because $\angle wvy \geq \angle wvx$, we may construct a line $\overline{vz}$ on face $\triangle wvy$, such that $\angle wvz = \angle wvx$.

By Angle-Side-Angle congruence, $\triangle vwz$ and $\triangle vwx$ are congruent. Let

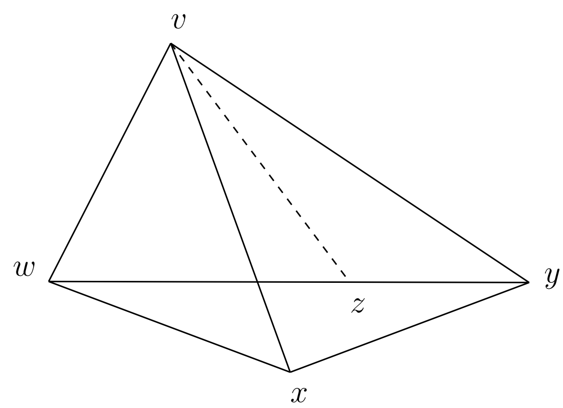

Figure 3

L denote the length function. Then by the triangle inequality:

$$\begin{aligned} L(\overline{wx}) + L(\overline{xy}) &> L(\overline{wy}) \\ L(\overline{wx}) + L(\overline{xy}) &> L(\overline{wz}) + L(\overline{zy}) \\ L(\overline{xy}) &> L(\overline{zy}) \end{aligned}$$

Because $L(\overline{vx}) = L(\overline{vz})$ (and $L(\overline{vy}) = L(\overline{vy})$), this last inequality implies $\angle xvy > \angle zvy$. Hence:

$$\angle wvy = \angle wvz + \angle zvy < \angle xvy + \angle wvx$$

■

Proposition 17. *Suppose P is a convex polytope embedded in $\mathbb{R}^3$, with vertex v. Then the sum of the face angles of v is a strictly less than 2π.*

Proof. Suppose n 2-faces of P meet at v. Since v is a 1-face of P, there exists a supporting hyperplane H such that $H \cap P = \{v\}$. Let n be the normal vector of H that is based at v and points toward P. Then there exists some small $\varepsilon > 0$ such that if we translate H a distance ε in the direction of n to obtain a new hyperplane $\widetilde{H}$, then we may use $\widetilde{H}$ to truncate P and obtain a new polytope $\widetilde{P}$ with $(n+1)$ faces, consisting of n triangles and one n-gon F, as depicted in Figure 4.

Choose a point p in the relative interior of F and draw lines from p to all vertices of $\widetilde{P}$. Because $\widetilde{P}$ is convex, these lines all lie within $\widetilde{P}$ and subdivide the polytope into n tetrahedra.

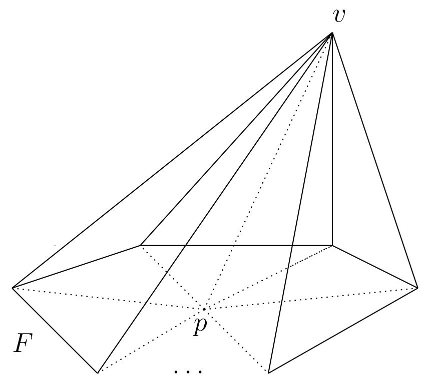

Figure 4: The remaining part of P after truncation.

Label the remaining vertices clockwise $v_1, v_2, \dots, v_n$. Let θ_j denote the face angle of F at v_j. Let φ_j denote the sum of the remaining face angles at v_j. By Proposition 16, $\theta_j < \varphi_j$ for all j. Let η_v, η_p denote the sums of the face angles (of the tetrahedra) at p and v respectively. Then:

$$\eta_v + \sum_{j=1}^{n} \varphi_j = n\pi$$

$$\eta_p + \sum_{j=1}^{n} \theta_j = n\pi$$

$$\eta_p - \eta_v + \sum_{j=1}^{n} \theta_j - \varphi_j = 0$$

Hence $\eta_v < \eta_p = 2\pi$ since p lies in a plane. ■

2.3.2 Face Angles in Higher Dimensions

Using the results for $d = 2$ and the definitions from Subsection 2.1, we are ready to define face angles in higher dimensions.

Definition 18. Suppose P is a polytope embedded in $\mathbb{R}^n$, F is a d-dimensional face of P and V is a $(d-2)$-face contained in F. By the Diamond Property of polytope lattices (Proposition 9), there exist exactly two $(d-1)$-faces E_1 and E_2 of F, that contain V. Pick a point $w \in V$. Then the affine span of V is a $(d-2)$-dimensional subspace which we identify with a subset of $T_w\mathbb{R}^n$, writing $\text{aff}(V) \subseteq T_w\mathbb{R}^n$. Likewise $\text{aff}(V) \subseteq \text{aff}(E_j) \subseteq T_w\mathbb{R}^n$ for $j = 1, 2$, with each $\text{aff}(E_j)$ a $(d-1)$-dimensional subspace. Hence, there exist vectors $a_1, a_2 \in \text{aff}(V)^\perp \cap E_j$ so that $\text{aff}(E_j) = \langle a_j \rangle \oplus \text{aff}(V)$. The face angle of F at v is defined to be

$$\arccos\left(\frac{\langle a_1, a_2 \rangle}{\|a_1\|\|a_2\|}\right)$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on $T_w\mathbb{R}^n$.

We should make sure this definition makes sense, because we chose a point $w \in V$ in the course of our definition. Because the polytope is embedded in $\mathbb{R}^n$, we may always parallel translate the vectors a_1, a_2 and a basis for $\text{aff}(V)$ along V to get a similar configuration at any other point in V. Thus our choice of a base point does not affect the definition.

2.3.3 Sums of Face Angles in Higher Dimensions

Theorem 19. *Suppose P is a $(d+1)$-dimensional polytope embedded in $\mathbb{R}^{d+1}$, and V is a $(d-2)$-face of P. Then the sum of the face angles at V is strictly less than 2π.*

Proof. Proposition 17 establishes the theorem in the case when $d = 2$, so we will assume $d > 2$ and reduce the problem to that case.

Fix a point $w \in V$, and observe that we may view $\text{aff}(V)$ as a subspace of $T_w\mathbb{R}^{d+1}$ and P as being embedded in $T_w\mathbb{R}^{d+1}$. We may decompose $T_w\mathbb{R}^{d+1}$ as $\text{aff}(V) \oplus \text{aff}(V)^\perp$, to obtain a projection map $\varphi : T_w\mathbb{R}^{d+1} \to \text{aff}(V)^\perp$.

Consider $\varphi(P)$. Since φ is linear and P is the convex hull of a finite set of points in $\mathbb{R}^{d+1}$, it is clear $\varphi(P)$ is a convex polytope embedded in $\text{aff}(V)^\perp \cong \mathbb{R}^3$. Because φ is a projection, $\text{aff}(\varphi(P))$ has dimension 3, because if $\text{aff}(\varphi(P))$ had dimension less than 3, this would imply $\text{aff}(P)$ could not have dimension $d+1$ in $\mathbb{R}^{d+1}$.

Since V is a $(d-2)$-face of P, there exists a d-dimensional plane H in $\mathbb{R}^{d+1}$ so that $H \cap P = V$. Since $\text{aff}(V)$ is a subspace of $T_w\mathbb{R}^{d+1}$, so is H, and hence $\varphi(H)$ is a 2-dimensional plane in $\text{aff}(V)^\perp$ that intersects $\varphi(P)$ at a single point, $\varphi(0)$. Thus, $\varphi(0)$ is a vertex on the 3-dimensional polytope $\varphi(P)$, and we may apply Proposition 2 to conclude the face angles about $\varphi(0)$ sum to strictly less than 2π.

Now consider a face angle at V. This angle is determined by two vectors, $a_1, a_2 \in \text{aff}(V)^\perp$ that specify the $(d-1)$-faces that form the face angle. But φ acts as the identity on $\text{aff}(V)^\perp$, so we conclude that the angle determined by a_1, a_2 is the same as the angle determined by $\varphi_* a_1$ and $\varphi_* a_2$. Moreover, the fact φ acts on $\operatorname{aff}(V)^\perp$ as the identity means there is a one-to-one correspondence between the face angles about V and the face angles about $\varphi(0)$. Hence, the two angle sums have the same magnitude. So, the sum of the face angles at V is strictly less than 2π. ■

We can re-use the details of this proof to gain a better understanding of neighborhoods of V, provided they are "small enough." Fix a point w in the relative interior of V. Then we may choose a neighborhood $\mathcal{O}$ of w so that points in $P \cap \mathcal{O}$ belong to no face of P of dimension less than $(d-2)$. As before, we view P as embedded in $T_w\mathbb{R}^{d+1}$, which we decompose as $\operatorname{aff}(V) \oplus \operatorname{aff}(V)^\perp$, with projections π_1, π_2. Then $\pi_1(\mathcal{O})$ is an open set in $\operatorname{aff}(V) \cong \mathbb{R}^{d-2}$ while $\pi_2(\mathcal{O})$, as we saw in the previous proof, is an open set about a vertex in the subspace topology on a polytope boundary embedded in $\mathbb{R}^3$. So we have the following corollary:

Corollary 20 (The Conical Neighborhood Corollary)**.** *Suppose V is a $(d-2)$-face of P and $w \in \operatorname{rint} V$. Any sufficiently small neighborhood of w is isometric to an open $\mathcal{O} \subset \widetilde{P} \times \mathbb{R}^{d-2}$ where:*

- *$\widetilde{P}$ is the boundary of a polyhedral cone embedded in $\mathbb{R}^3$ (a polyhedron with one vertex $v_{\widetilde{P}}$),*
- *$\widetilde{P}$ has as many faces adjacent to $v_{\widetilde{P}}$ as there are d-faces adjacent to V, with the same face angle structure, and*
- *$\mathcal{O}$ contains $(v_{\widetilde{P}}, 0)$.*

3 Unfoldings from Shortest Paths

The unfolding map we seek to construct is intimately connected with the problem of finding shortest paths on S. To have any hope of specifying an algorithm for computing this unfolding map, we must develop a discrete description of shortest paths suitable for use on a computer. In this section, we develop this description and analyze how an unfolding map is constructed from this data.

3.1 Discrete Characterizations of Geodesics and Shortest Paths

Recall from Subsection 2.2 that we have a functional L that maps a piecewise smooth path to its length, and this functional induces a metric d_S on S. The map d_S allows us to view S as a metric space.

Definition 21. A path γ from p to q is a **shortest path** if $L\gamma = d_S(p, q)$. A path $\gamma : [a, b] \to S$ is a geodesic if for every $t \in [a, b]$, there exists an $\varepsilon > 0$ and a closed metric space ball $B = \{p \in S : d_S(\gamma(t), p) \leq \varepsilon\}$ such that the intersection of γ and B is a subpath of γ and $B \cap \gamma$ is a shortest path.

In Riemannian geometry, the concepts above are usually defined very differently. Geodesics are characterized as paths which satisfy a particular differential equation (one that is ultimately derived from the Riemannian metric on the manifold), and then shortest paths are defined in terms of geodesics. The advantage of the preceding definition is that it may be understood without having to introduce the vector bundles, Riemannian metric, and differential equations needed to properly define geodesics on Riemannian manifolds.

We will frequently make use of two simple facts. First, geodesics in $\mathbb{R}^n$ are straight lines. Second, any subpath of a geodesic is again a geodesic.

Notation 22. Given points $p, q \in \mathbb{R}^k$, we write $[p, q]$ to denote the line segment from p to q. This line segment is understood to be a path, so we may apply the length functional L to $[p, q]$ and concatenate $[p, q]$ with other paths.

Proposition 23. *If $\gamma : [a, b] \to S$ is a shortest path, then $\gamma((a, b)) \cap S_{d-2} = \varnothing$.*

Proof. We will first consider the case when $d = 2$, then generalize. In each case, we will prove the contrapositive.

Suppose $d = 2$ and that γ passes through $w \in S_{d-2}$. Then there is a well defined first time $t \in [a, b]$ for which $\gamma(t) = w$. Let $\eta = \gamma|_{[a,t]}$ and $\eta' = \gamma|_{[t,b]}$. If η is not shortest, it may be replaced by a shorter path between its endpoints and concatenated with η' to produce a shorter path than γ. Hence η may be assumed to be a shortest path, and by a similar analysis η' may be assumed to be a shortest path as well.

On a sufficiently small neighborhood $\mathcal{O}$ of w, each of η, η' is a straight line. Consider $\mathcal{O} \setminus \eta$. Viewing η as a "cut" we obtain a local isometry φ from $\mathcal{O} \setminus \eta$ to $\mathbb{R}^2$ (see Figure 5) where $\theta \in (0, 2\pi)$, by Proposition 17. The path

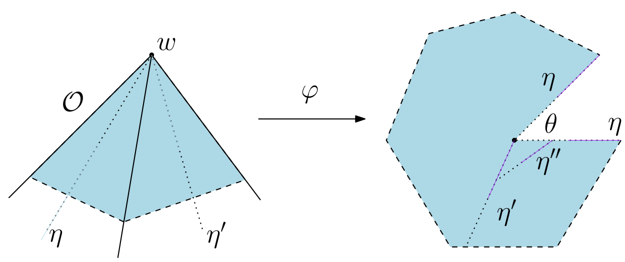

Figure 5

given by $\varphi \circ \gamma$ is not a straight line in $\mathbb{R}^2$, which means a shorter path in the plane (given by inserting η'' in the figure) may be pulled back via φ to obtain a shorter path on S, which is what we needed to show.

For the general case, suppose $d > 2$. We may take η and η' to be shortest paths from w as before. Then η, η' cannot both lie in the same d-dimensional face of S, because the restriction of γ to a single d-face is a shortest path in an affine plane isometric to $\mathbb{R}^d$. Consequently, the 2-plane $H \subset \mathbb{R}^{d+1}$ spanned by η, η' intersects Q, the lowest dimensional face of S containing w, in a single point, namely w. Since $\dim Q \leq d - 2$, the span of H and Q has dimension at most d. So, we may choose a line L not in the span of H and Q such that the affine 3-plane G spanned by L and H intersects S at w only. Then $P \cap G$ is a 3 dimensional polytope in $\mathbb{R}^3$ containing the paths η, η' on its surface, which reduces the problem to the base case. ■

Proposition 24. *Suppose γ is a shortest path on S. Then for any face F of S, γ intersects F in either a line segment (a single point is considered to be a line segment) or the empty set.*

Proof. We will argue the contrapositive. Suppose there exists a face F so that $\gamma \cap F$ is nonempty and is not a line segment. Then $\gamma \cap F$ contains points p and q so that the line segment $[p, q]$ is not contained in $\gamma \cap F$. Without loss of generality, suppose γ first visits p before q. Let γ be parametrized by arclength, so that $\gamma : [0, b] \to S$. By the continuity of γ, there are well defined first times t_p and t_q that γ visits p and q respectively. Define

$$\gamma_1 := \gamma|_{[0,t_p]} \qquad \gamma_2 := \gamma|_{[t_p,t_q]} \qquad \gamma_3 := \gamma|_{[t_q,b]}.$$

Consider the path γ' obtained by concatenating γ_1, the line segment $[p, q]$, and γ_3. Since γ_2 is not a line segment, γ and γ' are distinct paths with the same endpoints. Notice the line segment $[p, q]$ lies entirely within F, because F is convex. Since S is embedded in $\mathbb{R}^{d+1}$, both the line segment $[p, q]$ and γ_2 may be viewed as paths in $\mathbb{R}^{d+1}$. In $\mathbb{R}^{d+1}$, straight lines are shortest paths, hence the length of $[p, q]$ is strictly shorter than the length of γ_2. Hence γ' is strictly shorter than γ, so γ is not a shortest path. ■

This proof provides a simple example of a theme we will see again later. When we want to understand the structure of a geodesic or shortest path, we first find a way to move the path to a space we understand better than S, usually $\mathbb{R}^{d+1}$ or a tangent space (basically, $\mathbb{R}^d$), where we prove the estimate we need. Then, we explain why that estimate gives us information about the paths on S.

Corollary 25. *Suppose γ is a shortest path starting at a point in the relative interior of a d-face. Then for any $(d-1)$-face E, γ intersects E at a single point or the empty set.*

Proof. Suppose $\gamma : [a, b] \to S$ intersects E nontrivially. By Proposition 24, we know that if γ does not intersect E in either a point or nontrivial line segment, γ is not a shortest path. Thus, to argue the contrapositive, it suffices to show that if γ intersects E in a nontrivial line segment, then γ is not a shortest path.

Since γ is continuous and E is closed, $\gamma^{-1}(E)$ is compact. We may define $t_0 = \min \gamma^{-1}(E)$. Since γ does not intersect E in a point and does not start inside E, $a < t_0 < b$. By the Diamond Property of polytope lattices (Proposition 9), E is contained in exactly two d-faces of P, say F_1, F_2. By Proposition 23, we know that if γ passes through S_{d-2}, then it is not a shortest path. So we may assume γ does not pass through S_{d-2}, and conclude $\gamma(t_0)$

lies in $\text{rint}(E)$. By relabeling F_1 and F_2 if necessary, we may conclude that for some $\varepsilon > 0$ small enough, $\gamma(t) \in \text{rint}(F_1)$ for all $t \in [t_0 - \varepsilon, t_0)$.

If $\gamma_1 := \gamma|_{[t_0-\varepsilon,t_0)}$ is not a line segment, then γ is not a shortest path and we are done. Otherwise, γ_1 and $\gamma_2 := \gamma \cap E$ are two noncollinear line segments in F_1. If we replace this pair of line segments with a single line segment within F_1 from $\gamma(t_0 - \varepsilon)$ to the endpoint of $\gamma \cap E$, we obtain a strictly shorter path, which means γ is not a shortest path.

■

The previous results were phrased in terms of shortest paths, but they also show that a geodesic γ starting from a point in $S \setminus S_{d-1}$ will always cross $(d-1)$-faces transversely and does not pass through S_{d-2}. (This is simply because geodesics are locally shortest paths.) Since a $(d-1)$-face E is part of the relative boundary of exactly two distinct d-faces P, Q, this means we can determine the behavior of γ on E from its behavior on $\text{rint}(P)$ and $\text{rint}(Q)$.

Taking this idea further, we can see that since γ must be a disjoint union of straight lines on any given face, a natural way of describing γ is by the collection of faces it traverses. Our results show this collection will never contain any faces of dimension $(d-2)$ or lower and contains $(d-1)$-faces in a fairly trivial way. So in fact a geodesic may be specified by a starting point, an ending point, and a sequence of d-faces, provided one of the endpoints lies in the relative interior of a d-face. This discrete description of a geodesic comes up frequently enough that we will give it a formal definition.

Definition 26. A **face sequence** $\mathcal{F}$ **of length** n is an ordered n-tuple $\mathcal{F} = (F_1, \ldots, F_n)$ of d-faces $F_1, \ldots, F_n$ of S, such that for each i, F_i and F_{i+1} share a codimension 1 face.

Remark 27. In light of the arguments in this section, given a geodesic γ, the notion of the **face sequence of** γ is well defined. If γ is a shortest path, the fact γ may only traverse a face F in a single linear segment means each d-face F may only appear at most once in the face sequence for γ. This property can be quite useful both for formal arguments and for designing algorithms.

Unfortunately, this property is not immediately available on other piecewise flat manifolds. As a simple example, consider the flat torus obtained by gluing together two rectangles F_1 and F_2 of different widths, as in Figure 6. The shortest path from p to q in this example crosses from one rectangle to the other and back again, so that its face sequence contains F_1 more than once.

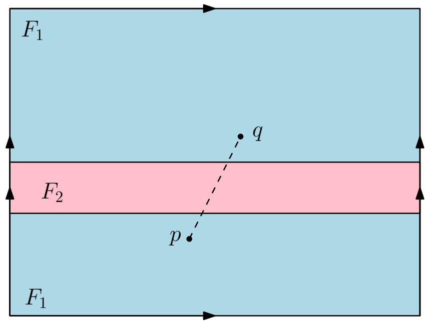

Figure 6: A flat torus obtained by gluing two rectangles.

Next, we argue that endpoints and a face sequence are actually enough to uniquely specify a geodesic.

Definition 28. Suppose $\mathcal{F} = (F_1, \ldots, F_n)$ is a sequence of d-faces such that F_i, F_{i+1} share a (necessarily unique) $(d-1)$-face for each i. Then we have a map

$$\Phi_{\mathcal{F}} = \Phi_{F_n, F_{n-1}} \circ \ldots \circ \Phi_{F_2, F_1}$$

where $\Phi_{F_{i+1}, F_i} : T_{F_i} \to T_{F_{i+1}}$ is a rotation of T_{F_i} into $T_{F_{i+1}}$ about the $(d-1)$-dimensional face between F_i and F_{i+1}. We call $\Phi_{\mathcal{L}}$ the **unfolding** of T_{F_1} onto T_{F_n} along $\mathcal{L}$. Given $\Gamma \subseteq \bigcup_{i=1}^{n} F_i$, the **unfolding** of Γ into T_{F_n} is the set $\bigcup_{i=1}^{n} \Phi_{\mathcal{F}_i}(\Gamma \cap F_i)$ where $\mathcal{F}_i = (F_i, F_{i+1}, \ldots, F_n)$. For convenience, we define the unfolding of a face sequence of length 1 to be the identity map.

Proposition 29. *Suppose $v, w \in S \setminus S_{d-2}$ and v, w are not both in some $(d-1)$-face. Then given a sequence of faces $\mathcal{F} = (F_1, \ldots, F_n)$, at most one shortest path*[1] *γ from v to w traverses $\mathcal{F}$.*

Proof. Suppose γ, γ' are two such shortest paths. Then both must unfold along $\mathcal{F}$ to lines from $\Phi_{\mathcal{F}}(v)$ to w in T_{F_n}. Since T_{F_n} is a d-hyperplane, the line from $\Phi_{\mathcal{F}}(v)$ to w is unique. So $\Phi_{\mathcal{F}} \circ \gamma$ and $\Phi_{\mathcal{F}} \circ \gamma'$ parametrize the same

[1] Recall from Section 2 our convention that paths are assumed to be piecewise smooth and two paths are considered equal if they differ by a re-parametrization.

line in T_{F_n}. On the other hand, on each face F_i, the map $\Phi_{\mathcal{F}_i}$ is an invertible isometry on $\Phi_{\mathcal{F}_i}(F_i)$. Hence for each i, $\Phi_{\mathcal{F}_i}(\gamma \cap F_i) = \Phi_{\mathcal{F}_i}(\gamma' \cap F_i)$ allows us to write $\gamma \cap F_i = \gamma' \cap F_i$ for each i. Hence $\gamma = \gamma'$. ■

Shortest paths play a crucial role in the theory of unfoldings that we will develop, so it is important to be certain that they exist. On a connected Riemannian manifold, the existence of a shortest path between two points is not guaranteed. For example, consider the Riemannian manifold $\mathbb{R}^2 \setminus \{(0,0)\}$, with the usual flat metric. If we consider points $(\pm 1, 0)$ in this manifold, it is clear that no path between them can achieve the distance between the two points, because such a path would cross through the deleted point. Unfortunately, if we want to view S as a flat Riemannian manifold we must delete S_{d-2}, which means the resulting manifold has much in common with the punctured plane. Fortunately, on S we have the following:

Theorem 30 (Shortest Path Existence Theorem)**.** *For any $p \in S$, there exists a shortest path from the source point $v \in S \setminus S_{d-1}$ to p.*

Proof. Let $\mathcal{P}$ denote the collection of all paths from v to p in S. Let $\mathcal{L}$ the subcollection of paths $\gamma \in \mathcal{P}$ with the additional properties that:

- γ intersects any d-face in a single line segment.
- γ intersects any $(d-1)$-face in at most a point.

Recall that $d_S(p, v) = \inf_{\gamma \in \mathcal{P}} L(\gamma)$. We will argue there exists a path that achieves the infimum in two steps.

1. For any path $\gamma \in \mathcal{P}$, there exists a $\widetilde{\gamma} \in \mathcal{L}$ such that $L(\widetilde{\gamma}) \leq L(\gamma)$.
2. Within $\mathcal{L}$ there is a path of minimal length.

Justification for 1: Fix a path $\gamma \in \mathcal{P}$, so that $\gamma : [0, b] \to S$ with $\gamma(0) = v$ and $\gamma(b) = p$. We will first argue that there exists a path γ' from v to p such that $L\gamma' \leq L\gamma$ and γ' is a concatenation of a finite number of line segments. Then, from γ' we may construct the desired $\widetilde{\gamma} \in \mathcal{L}$.

Notice that for any d-face F, the facts that γ is continuous and F is closed together imply $\gamma^{-1}(F)$ is a closed subset of $[0, b]$. Consequently, $\gamma^{-1}(F)$ is either empty or has a well defined minimum s_F and maximum t_F, corresponding to the time γ first enters F and the time γ last exits F, respectively. Let

$F_1, \ldots, F_N$ denote an enumeration of the d-faces of S and notice that from these faces we obtain a finite collection of closed intervals that cover $[0, b]$.

Without loss of generality, suppose $\gamma^{-1}(F_1)$ is nonempty and $[s_{F_1}, t_{F_1}] = [0, t_1]$. Form a new path γ_1 by concatenating the line segment $[\gamma(0), \gamma(t_1)]$ with $\gamma|_{[t_1,b]}$. Observe that by the arguments in Proposition 24, $L\gamma_1 \leq L\gamma$. If $t_1 = b$, we are done — take $\gamma' = \gamma_1$. Otherwise, $t_1 < b$, and $\gamma|_{(t_1,b]}$ has the property that it contains no points in F_1.

Now apply the preceding arguments to $\gamma|_{[t_1,b]}$. For each d-face F among $F_2, \ldots, F_N$, $\gamma|_{[t_1,b]}^{-1}(F)$ is either a closed subinterval of $[t_1, b]$, or the empty set, and these subintervals cover $[t_1, b]$. Without loss of generality, suppose $\gamma|_{[t_1,b]}^{-1}(F_2)$ is a nonempty interval of the form $[t_1, t_2]$, of positive length. By concatenating

- the line segment $[\gamma(0), \gamma(t_1)]$
- the line segment $[\gamma(t_1), \gamma(t_2)]$
- the path $\gamma|_{[t_2,b]}$

we obtain a new path γ_2 such that $L\gamma_2 \leq L\gamma_1$. As before, if $t_2 = b$ we take $\gamma' = \gamma_2$, otherwise, the remainder $\gamma|_{(t_2,b]}$ does not intersect F_1 and F_2.

Applying the above procedure at most $(N - 2)$ more times, we obtain a path γ' from v to p that is the concatenation of a finite number of line segments, where each d-face is traversed by at most one line segment.

The reason we cannot conclude $\gamma' \in \mathcal{L}$ is that γ' might intersect a $(d-1)$-face in more than just a point. Recall that the proof of Corollary 25 actually provides a procedure for shortening a path that intersects a $(d - 1)$-face in more than just a point, and this procedure works by replacing a piece of the path by a line segment. We know that γ' begins at v, within the relative interior of a face. Thus, some prefix of γ' is in $\mathcal{L}$. Iterating over the finitely many pairs of d-faces γ' traverses on its way from v to p, we may apply the procedure specified in Corollary 25 to replace any subpath of γ' that does not cross a $(d - 1)$-face transversely with a shorter line segment. After at most N such replacements, we obtain a path $\widetilde{\gamma} \in \mathcal{L}$ with $L\widetilde{\gamma} \leq L\gamma' \leq L\gamma$.

Justification for 2: Suppose $(F_1, \ldots, F_{k+1})$ is a face sequence and let $E_1, \ldots, E_k$ denote the $(d - 1)$-faces between each pair of faces F_j, F_{j+1}. Any tuple $(p_1, \ldots, p_k) \in \bigoplus_{j=1}^{k} E_j$ specifies a unique path γ in $\mathcal{L}$ with the face sequence $(F_1, \ldots, F_{k+1})$, in the following manner:

- Let $p_0 = v$, $p_{k+1} = p$.
- For each index $j = 1, \dots, k+1$, we have a line segment γ_j within F_j that travels from p_{j-1} to p_j.
- Form γ by concatenating the line segments $\gamma_1, \dots, \gamma_{k+1}$.

Moreover, any path in $\mathcal{L}$ with our chosen face sequence may be specified in this way.

Now consider the mapping $L : \bigoplus_{j=1}^{k} E_j \to \mathbb{R}$ that maps a tuple $(p_1, \dots, p_k)$ to the length of its corresponding path in S. Clearly, L is continuous. Each E_j is compact, so $\bigoplus_{j=1}^{k} E_j$ is compact. By the Extreme Value Theorem, then, there is a tuple $(p_1, \dots, p_k)$ which minimizes L. Among the paths traversing face sequence $(F_1, \dots, F_k)$ the path corresponding to $(p_1, \dots, p_k)$ has minimal length.

Let N denote the total number of d-faces of S. Paths in $\mathcal{L}$ are allowed to traverse each d-face at most once. Hence, each path in $\mathcal{L}$ has a face sequence of length at most N, and there are no more than $N!$ distinct face sequences. By choosing the shortest of at most $N!$ candidate paths, we may find a shortest path in $\mathcal{L}$. By the arguments above, this shortest path is also the shortest in $\mathcal{P}$. ■

3.2 Source Images and the Cut Locus

Our goal is to subdivide the faces of S in a way that naturally gives rise to a source unfolding. The main question we need to address is what points in S should be treated as "cuts" to define the subdivision. By emulating techniques from Riemannian geometry, one sees that a natural choice of cuts is the closure of all points with multiple shortest paths to the source point.

Definition 31. Fix a point $v \in S$. A point $w \in S$ is a **cut point** with respect to v if there exists more than one shortest path from v to w.

Example 32. In Figure 7, we can see that w is a cut point (in part due to the symmetries of S). In fact, any point in the relative interior of the edge containing w is a cut point. In contrast, x is not a cut point.

Definition 33. For a given $v \in S$, we denote the set of cut points for v by C_v. The **cut locus with respect to** v is defined to be the closure of C_v (that is, $\overline{C_v}$).

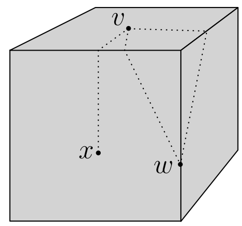

Figure 7: An example of a cut point on the boundary of a unit cube.

Example 34. Consider a polyhedral cone with vertex u in $\mathbb{R}^3$ and let S be its boundary. Fix a source point $v \neq u$ within a 2-face. We claim the cut locus C_v is a ray emanating from u.

Consider a ray R that starts at u and passes through v. Given any other ray R', R and R' divide the cone into two connected components. By Proposition 17 the sum of the face angles about u is $2\pi - \theta$ for some $\theta \in (0, 2\pi)$. So there exists another ray R', also emanating from u, such that the sums of the face angles within each component is $(2\pi - \theta)/2$. If we make a cut along R', we can unfold the cone into the plane, to obtain a configuration as in Figure 8. In this unfolded configuration, it is not hard to see that the only points with two distinct shortest paths to v are those along R'.

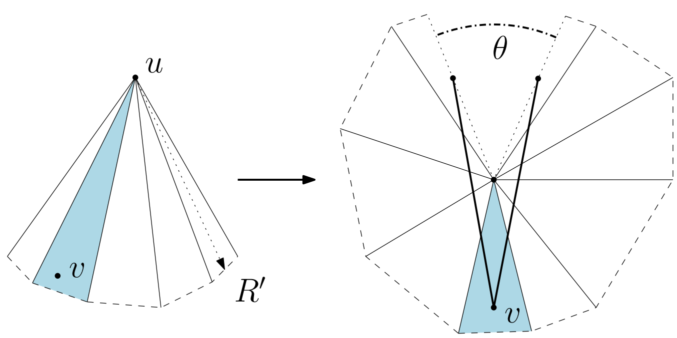

Figure 8: The unfolded traversable cone.

Proposition 35. *C_v is disjoint from the relative interior of any shortest path to v.*

Proof. We will show the contrapositive. Namely, if a path γ from v to w has a cut point c in its relative interior, then γ is not a shortest path. If any point in the relative interior of γ lies in S_{d-2}, then Proposition 23 already implies γ is not a shortest path. So we may further assume that γ avoids S_{d-2}.

Let γ_1 denote the part of γ from v to c and γ_2 denote the part of γ from c to w. We observed earlier that every sub-path of a shortest path is again a shortest path. Taking the contrapositive, we can see that if either γ_1 or γ_2 is not a shortest path, then γ is not a shortest path. So we may further assume both γ_1 and γ_2 are shortest paths.

Since $c \notin S_{d-2}$ we have a piecewise linear isometry Φ from a neighborhood $\mathcal{O}$ of c into $\mathbb{R}^d$. If $\Phi(\gamma \cap \mathcal{O})$ is not a line, we may shorten γ simply by replacing $\gamma \cap \mathcal{O}$ by the Φ^{-1}-image of a line segment. So we may now assume that $\Phi(\gamma \cap \mathcal{O})$ is a line, and that $\Phi(\gamma_1 \cap \mathcal{O})$ is the first part of the line.

Since c is a cut point, there exists another shortest path $\gamma_1' \neq \gamma_1$ from v to c. Due to the shortest path property of γ_1', we may (by making $\mathcal{O}$ smaller if necessary) assume that $\Phi(\gamma_1' \cap \mathcal{O})$ is a line segment.

We will now work only in $\Phi(\mathcal{O})$, and identify all the aforementioned objects with their Φ-images. By projecting onto the plane containing the distinct line segments corresponding to γ and γ_1, we obtain a configuration like the one in Figure 9. Necessarily, γ_1' forms an angle θ less than $\pi/2$ with either $\Phi(\gamma_1 \cap \mathcal{O})$ or $\Phi(\gamma_2 \cap \mathcal{O})$. Assume the former (a similar argument addresses the other case).

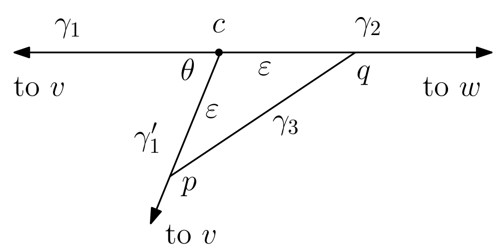

Figure 9: The path γ and a strictly shorter path obtained by avoiding c.

Let p and q be the points at some small arclength $\varepsilon > 0$ away from c along γ_1' and γ_2 respectively. By connecting p and q with a line segment γ_3 (which lies completely within $\mathcal{O}$) we can create a path γ' from v to w that travels from v to p along γ_1', then along γ_3 to q, and finally along γ_2 to w. By construction, γ' is a shorter path than the concatenation of γ_1' and γ_2, which has the same length as γ. ■

Corollary 36. *For any $v \in S$, C_v has a dense complement in S.*

Proof. We must argue that within every open metric-space ball on S there exists a point that is not a cut point. Suppose $B(x, \varepsilon)$ is a ball of radius $\varepsilon > 0$ centered at $x \in S$. By the Shortest Path Existence Theorem (Theorem 30), there exists a shortest path γ from v to x. By Proposition 35, no point in the relative interior of γ is a cut point. Since γ is a path from v to w, points from the relative interior of γ lie within B. ■

The cut locus actually has a great deal more structure than we just proved. In fact, the cut locus is made up of a finite number of $(d-1)$-dimensional polytopes. Together these polytopes divide the faces of P into convex regions such that points in the same region have similarities between their shortest paths to v. A result called Mount's Lemma makes this relationship precise, but to state the lemma we require more terminology.

Definition 37. A point $\nu \in T_F$ is a **source image** for a d-face F (or an F-source image) if there exists a point $x \in F \setminus S_{d-2}$ and a shortest path from v to x that unfolds to the line segment $[\nu, x]$ in T_F. We denote by src_F the set of all source images for F.

Proposition 38. *For any d-face F, src_F is finite.*

Proof. Recall that any shortest path may be uniquely described by its endpoints and the sequence of d-faces it traverses. The sequence of faces is further constrained in the sense that no face may be repeated in the sequence. Thus, there are at most $\sum_{n=1}^{N} n!$ face sequences that may be used to describe shortest paths in S, where N is the number of d-dimensional faces of S. If we take the (finitely many) of these sequences that describe valid sequences of adjacent faces from the face containing v to the face F, we conclude there are only finitely many possible unfoldings of v into T_F. Hence src_F is finite. ■

3.3 Converging Sequences of Geodesics

It often happens in the course of a proof that one wishes to take a limit of some geodesics or shortest paths in order to obtain a new shortest path with desirable properties. In this section, we will define what it means for sequences of paths to converge and analyze what properties of the paths in the sequence may be transferred to the limiting path.

Definition 39. Suppose γ_1 and γ_2 are piecewise smooth paths on S, of lengths l_1 and l_2 respectively, parametrized by arclength. Define the **distance** between γ_1 and γ_2 by:

$$d(\gamma_1, \gamma_2) = \sup_{t \in [0,1]} d_S(\gamma_1(tl_1), \gamma_2(tl_2))$$

Recall from section 2 that d_S denotes the distance on S.

Proposition 40. *The proposed map d is a metric on the space of piecewise smooth paths.*[2]

Proof. Based on the properties of the metric d_S, it is clear that for any paths γ_1, γ_2, $d(\gamma_1, \gamma_2) \geq 0$ and $d(\gamma_1, \gamma_1) = 0$. Now suppose $d(\gamma_1, \gamma_2) = 0$. Hence, for every $t \in [0, 1]$, we have $d_S(\gamma_1(tl_1), \gamma_2(tl_2)) = 0$. Since d_S is a metric, $\gamma_1(tl_1) = \gamma_2(tl_2)$ for all $t \in [0, 1]$. Thus, a reparametrization of γ_1 is equal to a reparametrization of γ_2. According to our convention of identifying reparametrized paths, we conclude $\gamma_1 = \gamma_2$.

Finally, we need to show d satisfies the triangle inequality. Suppose γ_1, γ_2, and γ_3 are piecewise smooth paths in S, with lengths l_1, l_2, and l_3, respectively. Since d_S is a metric, for every $t \in [0, 1]$ we may write

$$d_S(\gamma_1(tl_1), \gamma_3(tl_3)) \leq d_S(\gamma_1(tl_1), \gamma_2(tl_2)) + d_S(\gamma_2(tl_2), \gamma_3(tl_3)).$$

Hence

$$\begin{aligned}
\sup_{t \in [0,1]} d_S(\gamma_1(tl_1), \gamma_3(tl_3)) &\leq \sup_{t \in [0,1]} \left(d_S(\gamma_1(tl_1), \gamma_2(tl_2)) + d_S(\gamma_2(tl_2), \gamma_3(tl_3))\right) \\
&\leq \sup_{t \in [0,1]} d_S(\gamma_1(tl_1), \gamma_2(tl_2)) + \sup_{t \in [0,1]} d_S(\gamma_2(tl_2), \gamma_3(tl_3)) \\
d(\gamma_1, \gamma_3) &\leq d(\gamma_1, \gamma_2) + d(\gamma_2, \gamma_3).
\end{aligned}$$

■

Definition 41. A sequence $\{\gamma_k\}$ of piecewise smooth paths **converges** to a piecewise smooth path γ if $\lim_{k \to \infty} d(\gamma_k, \gamma) = 0$.

[2] Recall our convention that we consider two paths to be equal if they differ by a reparametrization. Each equivalence class has a unique representative that is parametrized by arclength.

Definition 42. Suppose $\mathcal{F} = (F_1, \ldots, F_n)$ is a face sequence. We say that $\mathcal{F}$ is **traversable with respect to** v if $v \in F_1$ and there exists a geodesic starting at v with face sequence $\mathcal{F}$. The existence of this geodesic implies the existence of an unfolding map U taking faces in $\mathcal{F}$ to a (likely nonconvex) polyhedral region $U(\mathcal{F})$ in $\mathbb{R}^d$. We may further insist U maps v to the origin. A geodesic starting at v and traversing $\mathcal{F}$ unfolds by U to a line. Such a line must pass through each of the $(d-1)$-faces $E_i = F_i \cap F_{i+1} \subseteq S_{d-1}$. It is not hard to see then that the $(d-2)$-faces in the E_i constrain geodesics that start at v and pass through $\mathcal{F}$. As such, there exists a maximal closed polyhedral cone with its vertex at the origin that is contained in $U(\mathcal{F})$. We call this polyhedral cone an **unfolded traversable cone**. We call the U-preimage of this set the **traversable cone** $C_{\mathcal{F},v}$.

Traversable cones are useful for estimating the lengths of certain paths on S. Suppose $p, q, r \in S$, and that we have geodesics γ_1 from p to q, γ_2 from q to r and γ_3 from r to p. In general, we cannot conclude that $L(\gamma_1) \leq L(\gamma_2) + L(\gamma_3)$. However, if we constrain these paths to lie within the same traversable cone, we do have such a triangle inequality — we can simply unfold the traversable cone into Euclidean space, so that in the unfolding each γ_i is a line (with length $L\gamma_i$), and apply the Euclidean triangle inequality.

Taking this idea further, suppose a sequence of geodesics γ_k are contained in the same traversable cone, and converge to some path γ. In the unfolded traversable cone, $U \circ \gamma_k$ and $U \circ \gamma$ are lines in a subset of Euclidean space. Each of these lines has one endpoint at the origin. Let p_k denote the other endpoint of $U \circ \gamma_k$ and p denote the other endpoint of $U \circ \gamma$. The lengths of these lines are exactly equal to the lengths of their corresponding paths. Moreover, if we consider the triangle formed by $U \circ \gamma$, $U \circ \gamma_k$ and the line segment between p_k and p, we may use the Euclidean triangle inequality to estimate the length of γ in terms of the length of γ_k (this is useful, since as $k \to 0$, the length of the line segment from p_k to p tends to 0).

One might guess that this limit path γ must be a geodesic. Unfortunately, we cannot claim the limit of a sequence of geodesics within the cone is again a geodesic because the geodesics might "converge into the boundary" of the cone, preventing γ from being a geodesic by forcing it to pass through a point in S_{d-2} (see Example 43). However, since we know how to estimate the length of γ in terms of the lengths of γ_k, with additional information we can sometimes argue this pathology does not occur. Lemma 44 elaborates on this idea.

Example 43. Figure 10 depicts a sequence of geodesics γ_k converging to a path γ within a given traversable cone. It also depicts those same paths under the unfolding map U. The limit path γ lies on the boundary of the

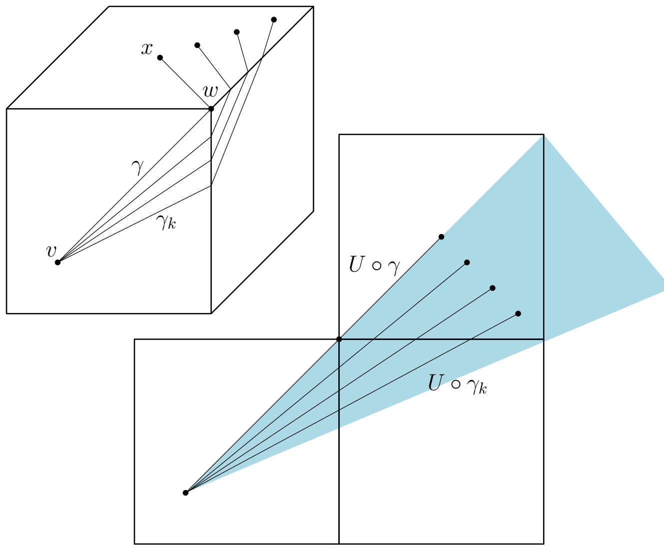

Figure 10: Geodesics "converging into the boundary" of an unfolded traversable cone (in blue). The limit path γ is not a geodesic, but we do know its length.

traversable cone. We can see that γ fails to be a geodesic because it passes through vertex w.

Lemma 44. *Suppose $\mathcal{F} = (F_1, \ldots, F_n)$ is a face sequence and $\{\gamma_k\}_{k=1}^{\infty}$ is a sequence of shortest paths that start at v and lie within $C_{\mathcal{F},v}$. Suppose the sequence converges to a path γ and that each of the γ_k has its endpoint x_k on a line segment $L \subseteq F_n \cap C_{\mathcal{F},v}$. Then γ is a shortest path.*

Proof. Let $x \in F_n$ be the endpoint of γ. Then $x_k \to x$ (in the sense of the subspace topology on $S \subset \mathbb{R}^{d+1}$). If we argue that $L\gamma \leq d_S(x, v)$, then γ is a shortest path.

Let U be an unfolding map for $C_{\mathcal{F},v}$. For each k, we have a triangle in the unfolded traversable cone with vertices $U(x)$, $U(x_k)$ and $U(v)$. This family of triangles share a common angle θ at vertex $U(x)$, because x and all the x_k lie on line segment L. Let δ_k denote the length of the line segment $[U(x), U(x_k)]$, so that as $k \to \infty$, $\delta_k \to 0$. Notice $L[U(x), U(v)] = L\gamma$ and

$L[U(x), U(x_k)] = L\gamma_k$. By the law of cosines:

$$d_S(x_k, v) = L[U(x_k), U(v)] = \sqrt{\delta_k^2 + (L\gamma)^2 - 2\delta_k(L\gamma)\cos(\theta)}$$

If we traverse a shortest path from v to x and then the line segment from x to x_k, we get a path from v to x_k. Hence, $d_S(x_k, v) \leq d_S(v, x) + \delta_k$. So for each k:

$$\sqrt{\delta_k^2 + (L\gamma)^2 - 2\delta_k(L\gamma)\cos(\theta)} \leq d_S(v, x) + \delta_k$$

Taking $k \to \infty$, we find $L\gamma \leq d_S(v, x)$ which completes the proof. ■

The proof above could be generalized to broader classes of convergent sequences of path, but it will suffice for our purposes. Throughout this section, we will use $v \in S$ to denote the source point. Consequently, though the set src_F depends on v we will not explicitly include this dependence in our notation.

3.4 Mount's Lemma

In this section, we present a lemma that is essential to the analysis of shortest paths on S, Mount's Lemma (Lemma 47). The fact S is the boundary of a convex polytope places strong constraints on which paths on S can be shortest paths. Even with these constraints, it is unclear what structure the cut locus must have. It is more unclear that one can compute this subset of S. Once we have proven Mount's Lemma, we will understand how the portion of the cut locus on a d-face F may be directly computed from src_F.

Given a shortest path γ from v to a point $w \in F \setminus S_{d-2}$, one can unfold γ to produce a line $[w, \nu]$ in T_F, with $\nu \in \mathrm{src}_F$ and $L[w, \nu] = L\gamma$. Lemma 47 considers the reverse situation — if $\nu \in \mathrm{src}_F$ is the closest source image in T_F to w, does there exist a shortest path from v to w that unfolds to yield the line segment $[w, \nu]$? Naturally, proving the existence of a shortest path is more difficult than characterizing the properties of such a path, so Lemma 47 requires some preliminaries. In Lemma 45 we prove 47 on a subset of points that are more amenable to analysis.

Lemma 45. *Suppose F is a d-face of S, $w \in F \setminus S_{d-2}$, $\nu \in \mathrm{src}_F$, and w does not lie in any hyperplane that is a perpendicular bisector of $[\nu_1, \nu_2]$, where $\nu_1, \nu_2 \in \mathrm{src}_F$. If $L[\nu, w] \leq L[\widetilde{\nu}, w]$ for all $\widetilde{\nu} \in \mathrm{src}_F$, then there is a shortest path from v to w that unfolds to $[\nu, w]$.*

Proof. We will prove the contrapositive: if no shortest path from v to w unfolds to $[\nu, w]$ then there must exist a source image ν' such that $L[\nu, w] > L[\nu', w]$.

Since $\nu \in \mathrm{src}_F$, there exists some $x \in F$ such that a shortest path from v to x unfolds to $[x, \nu]$. Consider the line segment $[x, w]$, which is contained in F by convexity. Let $\mathrm{src}_F([x, w])$ denote the collection of source images obtained by unfolding shortest paths from v to a point on $[x, w]$ into T_F.

For each $\widetilde{\nu} \in \mathrm{src}_F([x, w])$ we have a (necessarily nonempty) set $Y_{\widetilde{\nu}}$ of points $y \in [x, w]$ such that some shortest path from v to y unfolds to $[\widetilde{\nu}, y]$. We claim each $Y_{\widetilde{\nu}}$ is closed. To see this, suppose $\{x_i\}$ is a sequence of points in $Y_{\widetilde{\nu}}$ which converge to x. By the definition of $Y_{\widetilde{\nu}}$, for each i we have a shortest path γ_i from v to x_i that unfolds to produce $\widetilde{\nu}$. By going to a subsequence, if necessary, we may assume all of the shortest paths γ_i traverse a common face sequence $\mathcal{F}$.[3] Necessarily, the paths γ_i lie within the traversable cone $C_{\mathcal{F},v}$ and their endpoints lie on a line within $F \cap C_{\mathcal{F},v}$, namely $[x, w] \cap C_{\mathcal{F},v}$. By Lemma 44, the limit of the sequence $\{\gamma_i\}$ is a shortest path that unfolds along $\mathcal{F}$ to a line segment from x to $\widetilde{\nu}$. Hence $x \in Y_{\widetilde{\nu}}$ and so $Y_{\widetilde{\nu}}$ is closed. We may further observe that the sets $Y_{\widetilde{\nu}}$ cover $[w, x]$, that is

$$[w, x] = \bigcup_{\widetilde{\nu} \in \mathrm{src}_F([w,x])} Y_{\widetilde{\nu}}.$$

The compactness of Y_ν guarantees the existence of a point $x' \in Y_\nu$ that is closest to w. (Since $w \notin Y_\nu$, $w \neq x'$, but it is possible that $x = x'$.) Hence, in the relative topology on $[x, w]$, x' must lie on the boundary of Y_ν, and thus must also be a member of some other $Y_{\widetilde{\nu}}$. Consequently, x' is a cut point.

Consider a shortest path γ from v to x' that unfolds to yield ν. Then γ has a face sequence $\mathcal{F}$ ending with face F, and we may consider the open traversable cone $C_{\mathcal{F},v} \subseteq S$. Then $C_{\mathcal{F},v} \cap F$ is an open set containing x', and hence there exists a half-open subinterval $\mathcal{O}$ of the line segment $[x', w]$ such that $x' \in \mathcal{O}$ and $C_{\mathcal{F},v} \cap \mathcal{O}$ is nonempty. By the properties of the traversable cone, for any point $p \in \mathcal{O}$, there exists a geodesic from v to p that traverses the face sequence $\mathcal{F}$, and hence unfolds to the line segment $[\nu, p]$.

Now consider any sequence of points $\{x'_i\}$ in $\mathcal{O} \setminus \{x\}$ that converges to x. Each x'_i belongs to some $Y_{\widetilde{\nu}}$. By going to a subsequence if necessary, we may assume there exists some source image $\nu' \neq \nu$ and a face sequence $\mathcal{L}$ such

[3] We may do this because there are only finitely many possible face sequences.

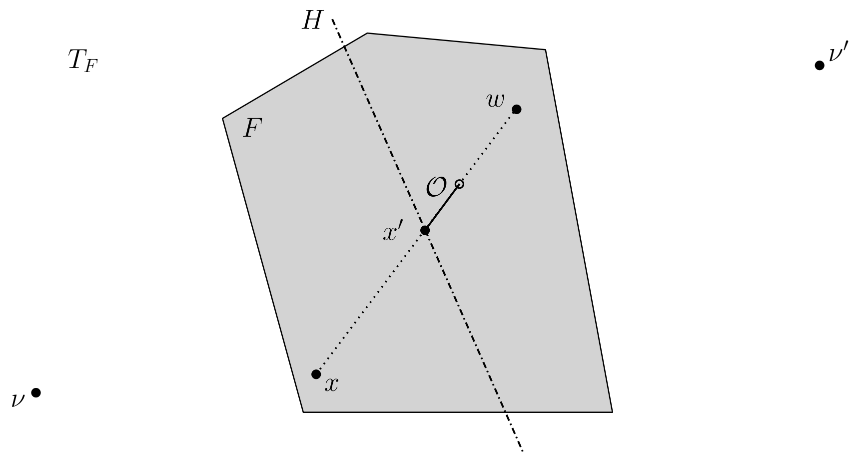

Figure 11: An example of Mount's Lemma in dimension 2.

that for all i, a shortest path from v to x_i traverses $\mathcal{L}$ and unfolds to yield ν'. Then $x_i \in Y_{\nu'}$ for all i, and since $Y_{\nu'}$ is closed, $x \in Y_{\nu'}$.

The shortest path from x' to v that unfolds to $[x', \nu']$ necessarily has length $d_S(x', v)$, so $L[x', \nu'] = d_S(x', v)$. Since $x' \in Y_\nu$, another shortest path from x' to v unfolds to $[x', \nu]$, hence $L[x', \nu] = d_S(x', v)$. So in T_F, x' is equidistant from ν and ν', which means x' lies on a hyperplane H that is the perpendicular bisector of $[\nu, \nu']$. If we can show w lies on the side of H closer to ν', we are done. (Recall our hypotheses guarantee $w \notin H$.)

Now, consider x'_1. By the properties of $\mathcal{O}$, the line segment $[\nu, x'_1]$ folds down to a geodesic η traveling through $C_{\mathcal{F},v}$. The optimality property of x' dictates $x'_1 \notin Y_\nu$, so η cannot be a shortest path. Hence η has length strictly greater than $d_S(x'_1, v)$ which means $L[\nu, x'_1] = L\eta > d_S(x'_1, v)$. So x'_1 is on the side of H that is closer to ν'. Since w is on the same side of H as x'_1, we conclude w is strictly closer to ν' than it is to ν. ■

Before we move on to Mount's Lemma, we consider the significance of the assumption "w does not lie on any perpendicular bisector of $[\nu_1, \nu_2]$, where $\nu_1, \nu_2 \in \mathrm{src}_F$" in the preceding argument.

Proposition 46. *Suppose F is a d-face, $w \in F \setminus S_{d-2}$ and $\nu \in \mathrm{src}_F$. If $L[\nu, w] \leq L[\nu', w]$ for all $\nu' \in \mathrm{src}_F$, then a shortest path from v to w unfolds to the line segment $[\nu, w]$ in T_F.*

Proof. [4] Since $\nu \in \mathrm{src}_F$, there exists an $x \in F \setminus S_{d-2}$ and a shortest path from v to x that unfolds to yield the line segment $[\nu, x]$. As in Lemma 45, the line segment $[x, w]$ is contained in the face F, because F is convex. Also as in that lemma, let us express $[x, w]$ as a union of closed sets $Y_{\widetilde{\nu}}$, where $Y_{\widetilde{\nu}}$ is defined to be the set of all $p \in [x, w]$ such that a shortest path from v to p unfolds to produce the line $[\widetilde{\nu}, p]$.

If $w \in Y_\nu$, we are done. So suppose by way of a contradiction that $w \notin Y_\nu$. We know that $x \in Y_\nu$, so since Y_ν is nonempty and compact, it contains a point x' closest to w, some positive distance ε away from w. As in Lemma 45, we conclude x' is a cutpoint, so that some shortest path from v to x' unfolds to a line segment $[\nu', x]$ with $\nu' \in \mathrm{src}_F \setminus \{\nu\}$. Also, x' lies on a hyperplane H separating T_F into two open half-spaces — one consisting of points in T_F closer to ν than to ν', the other consisting of points in T_F closer to ν than to ν'. Just as we did before, we may construct a half open interval of points $\mathcal{O}$ on $[x', w]$, and observe that for some $p \in \mathcal{O}$, a *geodesic* from v to p that is not a shortest path unfolds to the line segment $[\nu, p]$. In this way, we can determine that $[x', w]$, and hence w, lie on the side of H strictly closer to ν' than to ν. But the assertion that $L[\nu', w] < L[\nu, w]$ contradicts the assumption that $L[\nu, w] \leq L[\widetilde{\nu}, w]$ for all $\widetilde{\nu}$ for all $\widetilde{\nu} \in \mathrm{src}_F$. It must instead be the case that $w \in Y_\nu$. So a shortest path from v to w unfolds to the line segment $[\nu, w]$. ∎

We are now ready to prove the main result of this section.

Lemma 47 (Generalized Mount's Lemma)**.** *Suppose F is a d-face of S and $\nu \in \mathrm{src}_F$. Then for any $w \in F$, $d_S(v, w) \leq L[\nu, w]$. Moreover, if $w \in F \setminus S_{d-2}$, equality holds if and only if a shortest path from v to w unfolds to the line segment $[\nu, w]$.*

Proof. It suffices to prove the bound on a dense subset of F, since the maps $w \mapsto d_S(v, w)$ and $w \mapsto L[\nu, w]$ are both continuous. Let D be the relative interior of F, minus any $(d-1)$-hyperplane that is a perpendicular bisector of a line segment $[\nu_1, \nu_2]$, where $\nu_1, \nu_2 \in \mathrm{src}_F$, and minus any cut point in F. By Corollary 36 and Proposition 38, the set of cut points has a dense complement in F and src_F is finite, so D is dense in F.

[4] Hopefully the reader will forgive us for recapitulating many of the details from Lemma 45 in the proof below. By breaking up the argument in this way, we wish to illustrate that much of the argument can be proved without resorting to a proof by contradiction.

Suppose $w \in D$. Then w is closest to some $\nu \in \mathrm{src}_F$, so $L[\nu, w] \leq L[\nu', w]$ for all $\nu' \in \mathrm{src}_F$, and Lemma 45 implies the existence of a shortest path that unfolds to $[\nu, w]$. Hence

$$d_S(v, w) = L[\nu, w] \leq L[\nu', w]$$

for all $\nu' \in \mathrm{src}_F$, which establishes the desired bound.

For the second claim, suppose $w \in F \setminus S_{d-2}$. If there exists a shortest path from v to w that unfolds to the line segment $[\nu, w]$, then clearly $L[\nu, w] = d_S(v, w)$. The converse is the content of Proposition 46.

■

3.5 Voronoi Diagrams from Source Images

Definition 48. Suppose Y is a closed discrete set of points in $\mathbb{R}^d$. The **Voronoi diagram** $V(Y)$ is a covering of $\mathbb{R}^d$ by closed sets of the form:

$$V(Y, y) = \left\{ \zeta \in \mathbb{R}^d : |\zeta - y| \leq \inf_{y' \in Y} |\zeta - y'| \right\}$$

where $y \in Y$. We call each $y \in Y$ a **site** and $V(Y, y)$ the **cell** in the diagram belonging to y. The **Voronoi boundary** $\partial V(Y)$ is defined to be $\bigcup_{y \in Y} \partial V(Y, y)$.

When the set Y is finite, it is not hard to see that each cell $V(Y, y)$ must be polyhedral and convex. Specifically, in this case $V(Y, y)$ is an intersection of half-spaces

$$H_{y,y'} := \left\{ x \in \mathbb{R}^d : |x - y| \leq |x - y'| \right\}$$

for each $y' \in Y$. If it is unclear why the inequality $|x - y| \leq |x - y'|$ describes a half-space, consider that the following inequalities are all equivalent:

$$\begin{gathered} |x - y| \leq |x - y'| \\ |x - y|^2 \leq |x - y'|^2 \\ \langle x - y, x - y \rangle \leq \langle x - y', x - y' \rangle \\ \|x\|^2 - 2 \langle x, y \rangle + \|y\|^2 \leq \|x\|^2 - 2 \langle x, y' \rangle + \|y'\|^2 \\ -2 \langle x, y - y' \rangle \leq \|y'\|^2 - \|y\|^2 \end{gathered}$$

where the last inequality is clearly linear in x and $y - y'$ is a vector normal to the plane that bounds the half-space in question.

The Voronoi boundaries we have introduced are subsets of the affine spans of d-faces. We wish to use these boundaries to unfold S, so we need to create a corresponding set of points on S and understand its structure. To that end, let

$$V_{d-1} := \bigcup_{F \in \mathcal{F}} F \cap \partial V(\mathrm{src}_F).$$

where $\mathcal{F}$ denotes the collection of all d-faces of S. In words, V_{d-1} is made up of the parts of the Voronoi boundaries on each d-face F of S. Notice V_{d-1} is closed and is a union of a finite number of $(d-1)$-dimensional polytopes contained in the d-faces of S.

Theorem 49. *Suppose F is a d-face of S. Then* $\mathrm{rint}(F) \cap C_v = \mathrm{rint}(F) \cap \partial V(\mathrm{src}_F)$. *Moreover, if R is a $(d-1)$-face contained in F, then* $\mathrm{rint}(R) \cap C_v = \mathrm{rint}(R) \cap \partial V(\mathrm{src}_F)$.

Proof. Consider a point $w \in \mathrm{rint}\, R$ or $\mathrm{rint}\, F$. Any shortest path from v to w unfolds to a line segment $[\nu, w]$ for some $\nu \in \mathrm{src}_F$, and $L[\nu, w] = d_S(v, w)$. By Mount's Lemma (Lemma 47), $w \in V(\mathrm{src}_F, \nu)$ if and only if a shortest path from v to w unfolds to $[\nu, w]$. We know that $w \in \partial V(\mathrm{src}_F)$ if and only if w belongs to two or more Voronoi cells. Hence, $w \in \partial V(\mathrm{src}_F)$ if and only if there exist shortest paths from v to w that unfold to distinct source images ν and ν', which is if and only if w is a cut point. ■

We would now like to argue $S_{d-2} \subseteq \overline{C_v}$. To prove this, we will need the following lemma.

Lemma 50. *Suppose γ is a geodesic on S, with v as its starting point. Suppose $p \in S \setminus S_{d-2}$ is the point on γ where γ stops being a shortest path.*[5] *Then p is a cut point.*

Proof. Let γ_ε be the portion of γ that extends past p by length $\varepsilon > 0$. By choosing ε to be small enough, we may ensure γ_ε is contained in a single

[5] To be clear about our definition of p: Parametrize γ by arclength and let Ω be the set of all $t \geq 0$ for which γ is not a shortest path from v to $\gamma(t)$. Suppose Ω is nonempty, in other words γ stops being a shortest path *somewhere*. Then Ω is nonempty and bounded below by 0, hence it has an infimum. We take $p = \gamma(\inf \Omega)$.

d-face. Recall that any subpath of a shortest path is again a shortest path. Consequently any $q \in \gamma_\varepsilon$ is a point for which γ is *not* a shortest path from v to q.

Suppose q is a point in γ_ε. Then q has a shortest path γ_q from v to q. Necessarily, γ_q does not traverse the same face sequence as γ, because as we proved earlier, only one geodesic from v to q traverses the same face sequence as γ, namely γ itself, and γ is not a shortest path to q.

Since there are infinitely many points in γ_ε and finitely many face sequences, we may choose a sequence q_j in γ_ε, so that $q_j \to p$, each q_j has a shortest path γ_{q_j} from v to q_j, and all of these shortest paths traverse the same face sequence $\mathcal{F}$.

By Lemma 44, we conclude that the shortest paths γ_{q_j} converge to a shortest path $\widetilde{\gamma}$ from v to p, that traverses $\mathcal{F}$. By construction, $\mathcal{F}$ is a face sequence distinct from the face sequence of γ. Thus γ and $\widetilde{\gamma}$ are distinct shortest paths to p, which means p is a cut point. ■

This lemma illustrates that our intuitive definition of a cut point is compatible with the definition more commonly used in Riemannian geometry (namely, p is a cut point with respect to v if p is a point where the geodesic from v to p ceases to be a shortest path). Since we now know that only cut points and points in S_{d-2} stop a geodesic from being a shortest path, we have a mechanism for guaranteeing certain geodesics are shortest paths.

Proposition 51. $S_{d-2} \subseteq \overline{C_v}$.

Proof. Since the cut locus is closed, it will suffice to show any dense subset of S_{d-2} is contained in the cut locus. The collection $\mathrm{rint}(S_{d-2})$ (points in S_{d-2} that are in S_{d-2} but not any other face of higher codimension) are one such set. Suppose $p \in \mathrm{rint}(S_{d-2})$, so that $p \in \mathrm{rint}(V)$ for some $(d-2)$-face V. We will argue any (metric space) ball about p of radius ε, $B(p, \varepsilon)$, contains a cut point.

Take a shortest path γ from v to p, which traverses some face sequence $\mathcal{F}$. Let $C_{\mathcal{F}}$ be the traversable cone of γ. From Theorem 49, we know the Voronoi boundaries V_{d-1} (which contains the set of all cut points) on the faces of $\mathcal{F}$ are a union of precompact $(d-1)$-dimensional plane-segments. From earlier results, we know a shortest path cannot pass through a cut point. Hence, within $C_{\mathcal{F}}$, γ avoids the cut locus. Consequently, we may choose a smaller cone $\widetilde{C}_{\mathcal{F}}$ about γ such that the cone is free of cut points from the faces of $\mathcal{F}$.

Let F be the last face in $\mathcal{F}$, so F is one of several d-faces that contain V. Within $\widetilde{C}_{\mathcal{F}}$, we know that geodesics emanating from v remain shortest paths up until at least face F, by Lemma 50.

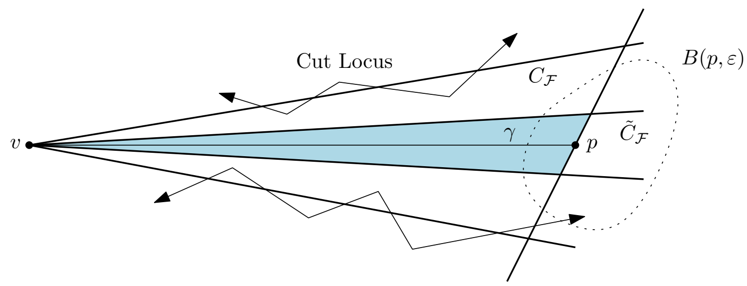

Figure 12: The cone $\widetilde{C}_{\mathcal{F}}$, free of cut points.

Pick a ball $B(p,\varepsilon)$ about p. Without loss of generality, we may insist ε be small enough that $B(p,\varepsilon)$ only intersects faces containing V. The sum of the face angles about V is less than 2π, so there is a well-defined "quasi-geodesic plane" H such that γ travels inside H near p and there are equal amounts of face angle on either side of H. Next, we may apply the Conical Neighborhood Corollary and the cut indicated by H to unfold $C_{\mathcal{F}} \cup B(p,\varepsilon)$ into T_F. We obtain a configuration like the one in Figure 13.

If we fix a point q within B and opposite γ, it is clear that within our cone we get two geodesics from v to q, one on either side of the cut. Call them φ and ψ, respectively. By construction, φ and ψ are shortest paths on $\mathcal{F}$. Hence, if one stops being a shortest path after traversing $\mathcal{F}$, it does so because it reached a cut point within B. In this case, we have found a cut point within $B(p,\varepsilon)$ and so we are done.

Otherwise, φ and ψ cross no cut points, and so they are shortest paths to q. But $\varphi \neq \psi$, so q is then a cut point in $B(p,\varepsilon)$. In either case, we have found a cut point in $B(p,\varepsilon)$, so we conclude p is a limit point of the set of cut points, and hence is in the cut locus. ■

Later on, we will need to know one additional detail about how different shortest paths from v to a face F can unfold to produce the same $\nu \in \mathrm{src}_F$. We record that fact here.

Proposition 52. *Suppose F is a d-face and $\nu \in \mathrm{src}_F$. For $i = 1, 2$, suppose $w_i \in F \setminus S_{d-2}$ and γ_i is a shortest path from v to w_i that unfolds to the line segment $[\nu, w_i]$. Then γ_1 and γ_2 enter F through the same $(d-1)$-face.*

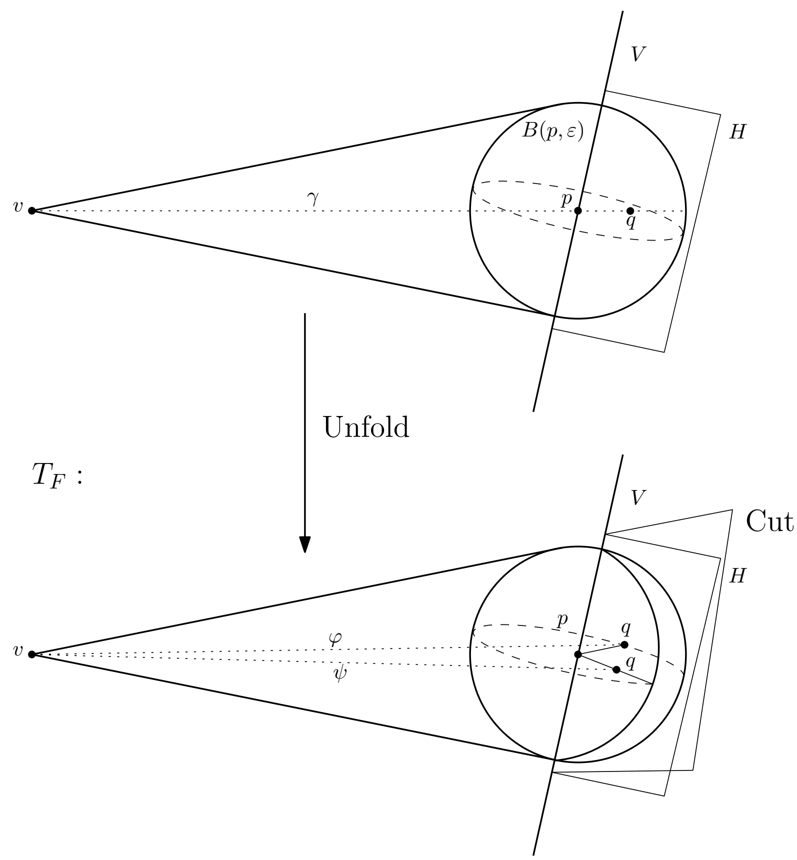

Figure 13: $C_{\mathcal{F}}$, unfolded into T_F.

Proof. Suppose, by way of a contradiction that γ_1 and γ_2 enter F through different ridges R_1 and R_2 respectively. Then the line segments $[\nu, w_1]$, $[w_1, w_2]$, and $[w_2, \nu]$ form a closed triangle in T_F that intersects the relative interior of at least two $(d-1)$-faces of F. As such, T must contain a point $p \in S_{d-2}$ in its relative interior.

Because $V(\mathrm{src}_F, \nu)$ is convex and contains $\{w_1, w_2, \nu\}$, $T \subseteq V(\mathrm{src}_F, \nu)$. But that implies p lies in the interior of $V(\mathrm{src}_F, \nu)$. By Proposition 51, there exists a sequence of cut points not in S_{d-2} that converge to p. Hence, a cut point $q \notin S_{d-2}$ lies in the interior of $V(\mathrm{src}_F, \nu)$. But that is a contradiction, because by Lemma 47 any such cut point $q \in V_{d-1}$ must lie on the boundary of the Voronoi cells. ■

Definition 53. A set $K \subseteq S$ is **polyhedral** if for every face F of S, $F \cap K$

is a finite union of convex polyhedra. Such a set K is **pure-polyhedral** of dimension k if for every d-face F of S, $F \cap K$ is a finite union of k-dimensional polyhedra.

Example 54. Consider the unit cube in $\mathbb{R}^3$, with subsets A, B, C of its boundary as in Figure 14. B and C are polyhedral, while A (a 2-disk) is

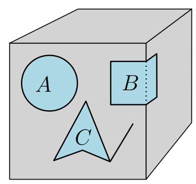

Figure 14: A unit cube with both pure-polyhedral and non-pure-polyhedral subsets.

not. B is pure-polyhedral of dimension 2, while C is not—though it is the union of a pure-polyhedral set of dimension 1 and a pure-polyhedral set of dimension 2.

By Theorem 49, we have that $\text{rint}(F) \cap C_v = \text{rint}(F) \cap V_{d-1}$ for any face F. Moreover, $F \cap V_{d-1}$ is a piece of a Voronoi boundary on a bounded subset of a d-dimensional Euclidean space. From our knowledge of general Voronoi boundaries in $\mathbb{R}^d$, we conclude V_{d-1} is a union of compact segments of $(d-1)$-dimensional hyperplanes. So we have the following Corollary to Theorem 49:

Corollary 55. V_{d-1} *is closed, polyhedral, and pure of dimension* $(d-1)$.

Next, we will use the results we have built up to characterize the cut locus.

Corollary 56. $\overline{C_v} = C_v \cup S_{d-2} = V_{d-1}$

Proof. We will argue two set equalities: first $\overline{C_v} = C_v \cup S_{d-2}$, then $\overline{C_v} = V_{d-1}$. Let $\mathcal{F}$ denote the collection of all d-faces and $(d-1)$-faces of S, and let $\mathcal{F}'$ denote the collection of all d-faces of S.

The Proposition 51 guarantees $C_v \cup S_{d-2} \subseteq \overline{C_v}$. In Theorem 49, we showed that for any face F of dimension d or $(d-1)$, $\text{rint}(F) \cap C_v = \text{rint}(F) \cap V_{d-1}$.

Thus:

$$\bigcup_{F\in\mathcal{F}} (\operatorname{rint}(F)\cap C_v) = \bigcup_{F\in\mathcal{F}} (\operatorname{rint}(F)\cap V_{d-1})$$
$$C_v \cap \left(\bigcup_{F\in\mathcal{F}} \operatorname{rint}(F)\right) = V_{d-1}\cap\left(\bigcup_{F\in\mathcal{F}} \operatorname{rint}(F)\right)$$

Notice that $\bigcup_{F\in\mathcal{F}} \operatorname{rint}(F) = S\setminus S_{d-2}$. Clearly, $C_v\cap S = C_v$ and $V_{d-1}\cap S = V_{d-1}$. So, we may conclude from the last equality above that

$$C_v \setminus S_{d-2} = V_{d-1}\setminus S_{d-2}$$
$$C_v \cup S_{d-2} = V_{d-1}\cup S_{d-2}$$

Since V_{d-1} and S_{d-2} are closed, their union is closed, so that by the previous equality $C_v\cup S_{d-2}$ is closed. But that means $\overline{C_v}\subseteq C_v\cup S_{d-2}$, since $\overline{C_v}$ is the smallest closed set containing C_v. So $\overline{C_v} = C_v\cup S_{d-2}$.

It remains to argue $\overline{C_v} = V_{d-1}$. We already know that $C_v\subseteq V_{d-1}$ and V_{d-1} is closed, so $\overline{C_v}\subseteq V_{d-1}$. Since $S_{d-2}\subseteq \overline{C_v}\subseteq V_{d-1}$ then, for the other containment it suffices to check $V_{d-1}\cap(S\setminus S_{d-2}) = \overline{C_v}\cap(S\setminus S_{d-2})$.

Now that we know $\overline{C_v} = C_v\cup S_{d-2}$, we may apply Theorem 49 again, using the fact $\operatorname{rint}(F)\cap C_v = \operatorname{rint}(F)\cap V_{d-1}$ for each d-face and $(d-1)$-face F to conclude:

$$\begin{aligned}\operatorname{rint}(F)\cap\overline{C_v} &= \operatorname{rint}(F)\cap(C_v\cup S_{d-2})\\ &= \operatorname{rint}(F)\cap C_v = \operatorname{rint}(F)\cap V_{d-1}\end{aligned}$$

for each d-face and $(d-1)$-face F. Taking a union over all such faces, we obtain:

$$\bigcup_{F\in\mathcal{F}'} (\operatorname{rint}(F)\cap\overline{C_v}) = \bigcup_{F\in\mathcal{F}'} (\operatorname{rint} F\cap V_{d-1})$$
$$\overline{C_v}\cap\left(\bigcup_{F\in\mathcal{F}'} \operatorname{rint}(F)\right) = V_{d-1}\cap\left(\bigcup_{F\in\mathcal{F}'} \operatorname{rint}(F)\right)$$
$$\overline{C_v}\cap(S\setminus S_{d-2}) = V_{d-1}\cap(S\setminus S_{d-2})$$

This completes the proof that $\overline{C_v} = V_{d-1}$. ■

In later sections, we will need to study nontrivial intersections of codimension 1 faces and Voronoi cells. Since both of these sets are polyhedral, their intersection is polyhedral as well, and so it will be useful to know its dimension.

Proposition 57. *Suppose F is a d-face, E is a $(d-1)$-face contained in F, and that $\nu \in \mathrm{src}_F$. If one of the shortest paths γ associated with ν unfolds to a line that intersects with the relative interior of R, then the dimension of $V(\mathrm{src}_F, \nu) \cap R$ is $(d-1)$.*

Proof. As a shortest path associated with ν, γ has the property that it is a shortest path from v to some $p \in \mathrm{rint}(F)$. Such a path intersects E transversely by Corollary 25; let q denote the point of intersection. Since γ is a shortest path through q, q is not a cut point. The set of cut points is closed, so we conclude that q lies in the exterior of the cut locus. That is, there exists an $\varepsilon > 0$ such that the d-dimensional ε ball $B(q, \varepsilon) \subseteq T_F$ is disjoint from the cut locus.

If we delete the Voronoi boundaries of $V(\mathrm{src}_F)$, T_F is separated into a finite number of connected components. $B(q, \varepsilon)$ contains no cut points, so it must be contained in one of these connected components, namely the one identified with ν. Hence $B(q, \varepsilon) \subseteq V(\mathrm{src}_F, \nu)$. By making ε smaller if necessary, we may assume the intersection of $B(q, \varepsilon)$ with R is an open $(d-1)$-ball B. So $B \subseteq R \cap V(\mathrm{src}_F, \nu) \subseteq R$, and hence $\mathrm{aff}(B) \subseteq \mathrm{aff}(R \cap V(\mathrm{src}_F, \nu) \subseteq \mathrm{aff}(R)$. But the affine span of a $(d-1)$-dimensional ε-ball is always a $(d-1)$-dimensional plane, hence $\mathrm{aff}(B) = \mathrm{aff}(R \cap V(\mathrm{src}_F, \nu)) = \mathrm{aff}(R)$. This implies $R \cap V(\mathrm{src}_F, \nu)$ has dimension $(d-1)$. ∎

3.6 Polyhedral Nonoverlapping Unfolding

In this subsection we will use the ideas developed in Sections 2 and 3 to explain why there must exist a nonoverlapping source unfolding for any convex polytope. After this section, our focus will shift toward devising a way to compute this unfolding.

Definition 58. $K \subseteq S$ is a **cut set** if all of the following hold:

1. $S_{d-2} \subseteq K$.

2. $S \setminus K$ is open and contractible.

Definition 59. Suppose (M_1, d_1) and (M_2, d_2) are two metric spaces and $\Phi : M_1 \to M_2$ is a continuous map. Suppose $p \in M_1$. Then Φ **preserves distances relative to** p if for every point $q \in M_1$, $d_1(p, q) = d_2(\Phi(p), \Phi(q))$.

Definition 60. A **polyhedral unfolding** of S is a pair (K, φ) where $K \subseteq S$ is a cut set and for some $p \in S$, $\varphi : S \setminus K \to \mathbb{R}^d$ preserves distances relative to p. (Here S is equipped with the metric d_S and $\mathbb{R}^d$ is equipped with the Euclidean metric.)

Definition 61. A **nonoverlapping foldout** of S is a continuous map $\Phi : \overline{U} \to S$ such that:

1. $U \subseteq \mathbb{R}^d$ is an open, star-shaped set. $\overline{U}$ is the closure of its interior, U.

2. Φ surjective.

3. $\Phi|_U$ is injective and preserves distances relative to some point $p \in U$. (Here $\overline{U}$ is equipped with the Euclidean metric and S is equipped with the metric d_S.)

The following lemma explains how this fact and the preceding definitions are related.

Lemma 62. *If $\Phi : \overline{U} \to S$ is a nonoverlapping foldout with a pure-polyhedral domain, $\Phi(\overline{U} \setminus U)$ is a cut set and $(\Phi|_U)^{-1}$ is a polyhedral unfolding of S.*

Proof. Since $S \setminus K = \Phi(U)$, Φ is a homeomorphism from U to $S \setminus K$, and U is open and contractible, we conclude that $S \setminus K$ is open and contractible in S. Since U is star shaped, every point in $\overline{U}$ can be connected to some central $v \in U$ by a line segment. By the distance-preserving property of Φ, the image of any line segment starting from v is a geodesic in S. Since a geodesic in S may not pass through S_{d-2}, we conclude $\Phi(U) \cap S_{d-2} = \varnothing$, and thus $S_{d-2} \subseteq \Phi(\overline{U} \setminus U)$. Thus, K is a cut set. Since $\Phi|_U$ preserves distances relative to some point $p \in U$, we can see that $(\Phi|_U)^{-1}$ preserves distances relative to $\Phi(p)$. Thus, $(K, (\Phi|_U)^{-1})$ is a polyhedral unfolding. ■

In light of this lemma, we will use the term **polyhedral nonoverlapping foldout** to indicate a nonoverlapping foldout with a pure-polyhedral domain. The particular nonoverlapping foldouts that interest us arise from exponential maps. We now require some definitions that deal specifically with the exponential map. For the remainder of this section, we will fix a point $v \in S \setminus S_{d-1}$ — in other words, v lies in the relative interior of a d-face. We will refer to v as the **source point**. Recall that $S^\circ := S \setminus S_{d-2}$ is a flat Riemannian manifold. For this reason, we may speak of the tangent space at v, $T_v S^\circ$.

Definition 63. A vector $\zeta \in T_v S^\circ$ **can be exponentiated** if there exists a (necessarily unique) shortest path γ such that $\gamma : [0,1] \to S^\circ$ is parametrized to have constant speed (i.e. proportional to arclength), such that $\gamma(0) = v$, $\gamma'(0) = \zeta$, and finally that γ may be extended to a longer shortest path some distance $\varepsilon > 0$ past $\gamma(1)$. (Thus $L\gamma = \|\zeta\| = d_S(\gamma(0), \gamma(1))$.)[6]

We define U_v, the **source interior**, to be the set of all vectors in $T_v S^\circ$ that can be exponentiated. We call $\overline{U_v}$ the **source foldout**. The **exponential map**, $\exp : U_v \to S^\circ$ assigns each $\zeta \in U_v$ the endpoint $\gamma(1)$, where γ is the shortest path associated with ζ as described above.

Clearly, if $\zeta \in U_v$ can be exponentiated, then $\lambda\zeta$ can be exponentiated for any $\lambda \in [0,1]$. Thus U_v is star shaped, and the condition that γ may be extended slightly past $\gamma(0)$ while remaining a shortest path guarantees U_v is open. As we have defined it, a shortest path from v to a point $p \in \overline{C_v}$ does not correspond to a vector in $T_v S^\circ$ that may be exponentiated. It is relatively clear that exp is smooth; the full details proving this may be found in most texts on Riemannian geometry, such as [4]. We may extend exp to a continuous map on $\overline{U_v}$ by specifying that for $\zeta \in \overline{U_v}$,

$$\exp(\zeta) := \lim_{\lambda \to 1^-} \exp(\lambda\zeta)$$

so that the extension $\exp : \overline{U_v} \to S$ is a surjection.

Example 64. Consider the unit cube $P = [0,1] \times [0,1] \times [0,1]$ in $\mathbb{R}^3$. Suppose $v = (1/2, 0, 1/2)$, as depicted in Figure 15. The vectors $e_1, e_2 \in T_v S^\circ$ can be exponentiated. (In fact, using the broader definition of the exponential map from Riemannian geometry, λe_i can be exponentiated for arbitrary $\lambda \in \mathbb{R}$— the exponential map will "wrap around" the cube as depicted by the dotted lines.) The vector $\frac{1}{2}(e_1 + e_2)$ does not lie in U_v, but does lie in $\overline{U_v}$, and may be exponentiated to yield p using the extension of exp. In contrast, $3e_1 + 3e_2$ cannot be exponentiated even if we allow the use of the extension of exp, because the geodesic in the direction of this vector reaches a point in S_{d-2} strictly before it achieves length $3\sqrt{2}$.

[6]Readers familiar with Riemannian geometry will note that the definition of the exponential map can be extended to more vectors in $T_v S^\circ$ than we described here.

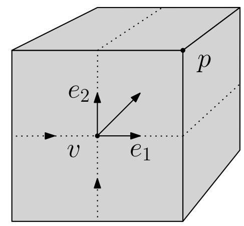

Figure 15: Exponentiated vectors on a unit cube in $\mathbb{R}^3$

As Example 64 illustrates, the exponential map identifies vectors in T_vS with geodesics in S. Namely, $\zeta \in T_vS^\circ$ yields a **geodesic flow** $\gamma_\zeta : [0,1] \to S$ given by $\gamma_\zeta(t) = \exp(t\zeta)$ (provided $\zeta \in \overline{U_v}$).

Theorem 65. *The map* $\exp : \overline{U_v} \to S$ *is a polyhedral nonoverlapping foldout and* $\overline{C_v} \subseteq \exp(\overline{U_v} \setminus U_v)$.

Proof. We have already observed that U_v is open and star-shaped. Next, let us check exp is surjective. For any $p \in S$, Theorem 30 implies that there exists a shortest path γ from v to p. If we parametrize $\gamma : [0,1] \to S$ proportional to arclength, then $\zeta := \gamma'(0)$ is a nonzero vector in T_vS°. If $p \in S \setminus \overline{C_v}$, then $\zeta \in U_v$. Otherwise $p \in \overline{C_v}$, and by shortening ζ to produce vectors ζ_ε, we obtain a sequence of vectors in U_v approaching ζ that allow us to conclude $\zeta \in \overline{U_v}$ and $\exp(\zeta) = p$.

Next we must argue $\exp|_{U_v}$ is injective. Suppose $\zeta \in U_v$. By the definition of U_v, there exists a shortest path $\gamma : [0,1] \to S$ parametrized proportional to arclength so that $\gamma'(0) = \zeta$ and γ may extended some small positive distance past $\gamma(1) = \exp(\zeta)$ while remaining a shortest path. Thus, a shortest path has the point $\exp(\zeta)$ in its relative interior, which implies $\exp(\zeta)$ is not a cut point. Since $\exp(\zeta) \notin C_v$, there can exist only one shortest path from v to $\exp(\zeta)$, namely the one that corresponds to ζ. Hence ζ is the only element in U_v that exp maps to p. Similarly, to see that $\exp|_{U_v}$ preserves distances relative to v, simply consider that $\exp|_{U_v}$ maps a vector ζ to a point p using a shortest path from v to p of length $\|\zeta\|$—hence $d_S(v,p) = \|\zeta\| = d_{\mathbb{R}^d}(0,\zeta)$.

To see that $\overline{C_v} \subseteq \exp(\overline{U_v} \setminus U_v)$, recall that in Corollary 56, we proved $\overline{C_v} = C_v \cup S_{d-2}$. So, if $x \in \overline{C_v}$ either $x \in S_{d-2}$ or x is a cutpoint. In either case, Lemma 50 states that a shortest path from v to x cannot be continued for any distance through x while remaining a shortest path. Hence, there cannot exist a $\zeta \in U_v$ for which $\exp(\zeta) = x$. On the other hand, $\exp : \overline{U_v} \to S$ is surjective, so $\exp(\zeta) = x$ for some $\zeta \in \overline{U_v} \setminus U_v$. ■

4 Jet Frames

In the preceding sections, we learned that all of the information needed to determine a source unfolding of S is encoded in the source images of the d-faces of S. The theory of affine transformations and the Voronoi diagram calculations needed to determine the unfolding from the source images are well understood algorithmically. Consequently, specifying an algorithm for unfolding S really amounts to specifying an algorithm for computing the source images.

The algorithm Miller and Pak present to solve this problem is based on the idea of a "signal" or "wave" propagating from the source point through S.[7] Informally speaking, the signal encounters different features of the manifold at different times, producing a discrete sequence of so-called "events" from which the source images may be easily recovered. Jet frames provide the mathematical abstraction used to specify, calculate, and order events.

In previous sections, we focused on the topological/metric properties of *the boundary* of some polytope P. In contrast, this section develops tools for studying the geometry *of polytopes themselves*, as they are embedded in Euclidean space. The polytopes that we are particularly interested in applying these concepts to are the Voronoi cells induced by the source images.

Definition 66. Suppose $x \in \mathbb{R}^n$. A **frame of order r at x** is an ordered r-tuple of orthonormal vectors in $T_x\mathbb{R}^n$. Given a frame $\zeta = (\zeta_1, \ldots, \zeta_r)$, the **$\zeta$-trajectory** is the path $\gamma_\zeta : [0, \infty) \to \mathbb{R}^n$ given by[8]

$$\gamma_\zeta(t) = x + \sum_{j=1}^{r} t^j \cdot \zeta_j.$$

The displacement vector in the equation for γ_ζ arises frequently, so we will denote this vector-valued function by

$$J_\zeta(t) = \sum_{j=1}^{r} t^j \cdot \zeta_j.$$

[7] "Signal propagation" is a common idiom in computational geometry. Dijkstra's algorithm (a shortest path algorithm for weighted graphs) is another example of a shortest path algorithm that can be explained in terms signal propagation.

[8] Notice that in this formula, t is just a scalar parameter and we are forming a linear combination of the ζ_j — Einstein summation notation is not being used.

and refer to it as the **ζ-displacement**.[9]

Jet frames record local geometric information about a polytope embedded in Euclidean space. For this reason, we are interested in $\gamma_\zeta(t)$ for small, positive values of t. We will frequently need to qualify which t we are considering in this way, which becomes quite cumbersome, so we introduce more succinct notation.

Notation 67. Suppose $P(t)$ is a proposition that depends on a real parameter t. We say that $P(t)$ holds for **positive, small enough** t (abbreviated p.s.e. t) if there exists a $\delta > 0$ such that for all $t \in (0, \delta)$, $P(t)$ holds. Likewise, we say $P(t)$ holds for **non-negative, small enough** t (abbreviated nn.s.e. t) if there exists a $\delta > 0$ such that for all $t \in [0, \delta)$, $P(t)$ holds.

Because we are developing an algorithm, we will typically only deal with a finite number of propositions depending on the same real parameter. It is not hard to see that if we have propositions $P_1(t), \ldots, P_n(t)$, and each individually holds for p.s.e. t, then $P_1(t) \wedge \ldots \wedge P_n(t)$ holds for p.s.e. t. Likewise, if one of $P_1(t), \ldots, P_n(t)$ individually holds for p.s.e. t, then $P_1(t) \vee \ldots \vee P_n(t)$ holds for p.s.e. t. Similar observations hold for nn.s.e..

Given a d-polytope $V \subseteq \mathbb{R}^d$ and a point $x \in V$, only certain frames at x give us information about the geometry of V. We identify them as follows.

Definition 68. Suppose V is a d-dimensional polytope and $x \in V$, and ζ is a frame at x. If $\gamma_\zeta(t) \in V$ for p.s.e. t, we say ζ is a **weak jet frame along** V. If $\gamma_\zeta(t) \in \operatorname{rint} V$ for p.s.e. t, we say ζ is a **strong jet frame along** V, or just a **jet frame along** V.

4.1 Calculating with Jet Frames

Miller and Pak's algorithm for calculating source unfoldings depends on constructing many jet frames. In this section, we specify an algorithm for constructing jet frames and explain how the collection of all jet frames at a point x may be ordered in a geometrically meaningful way.

[9] Remark: Miller and Pak prefer to normalize the quantity J_ζ to obtain a unit-vector valued function. It can be shown that the two definitions are equivalent but for our calculations Miller and Pak's convention would add an inconvenient scalar normalization term (that depends on t), so we will not introduce this other definition.

A jet frame is a collection of orthonormal vectors which must satisfy an additional, nonlinear constraint. Despite this nonlinear constraint, many familiar ideas about orthonormal vectors may be applied to jet frames.

Lemma 69 (The Gram-Schmidt Procedure for Jet Frames)**.** *Suppose* $\zeta = (\zeta_1, \ldots, \zeta_r)$ *is a weak jet frame at* $x \in V \subseteq \mathbb{R}^d$ *and* $w \in \mathbb{R}^d \setminus span(\zeta)$ *such that* $x + \varepsilon w \in V$ *for all positive* ε *small enough. Let* $\widetilde{w} = w - \sum_{j=1}^{r} \langle w, \zeta_j \rangle \zeta_j$ *and* $\zeta_{r+1} = \widetilde{w}/\|\widetilde{w}\|$*. Then* ζ_{r+1} *extends* ζ *to a weak jet frame* $\zeta' = (\zeta_1, \ldots, \zeta_r, \zeta_{r+1})$*.*

Proof. We may assume we have chosen our coordinates so that $x = 0 \in \mathbb{R}^d$. The property that $x + J_{\zeta'}(\varepsilon) \in V$ for all positive ε small enough is a local one, so there is no loss of generality in assuming V is a polyhedral cone (with vertex $0 \in \mathbb{R}^d$). So V is the intersection of m half-spaces and may be described by a finite collection of m linear inequalities of the form $\langle u, n_i \rangle \geq 0$, where n_i denotes the unit normal vector perpendicular to the i-th half-space.

Since ζ is a weak jet frame, we have that $\gamma_\zeta(\varepsilon) \in V$ for p.s.e. ε. This means

$$\langle 0 + J_\zeta(\varepsilon), n_i \rangle = \sum_{j=1}^{r} \langle \zeta_j, n_i \rangle \varepsilon^j \geq 0 \tag{1}$$

for $i = 1, \ldots, m$ and p.s.e. ε. So, we have a system of m polynomial inequalities in ε.

Next, consider the system of m inequalities:

$$\langle 0 + J_{\zeta'}(\varepsilon), n_i \rangle = \sum_{j=1}^{r} \langle \zeta_j, n_i \rangle \varepsilon^j + \left\langle \frac{\widetilde{w}}{\|\widetilde{w}\|}, n_i \right\rangle \varepsilon^{r+1} \geq 0$$

which is equivalent to the system of inequalities

$$\left(\sum_{j=1}^{r} \langle \zeta_j, n_i \rangle \varepsilon^j \right) + \frac{\varepsilon^{r+1}}{\|\widetilde{w}\|} \left(\langle w, n_i \rangle - \sum_{j=1}^{r} \langle \zeta_j, w \rangle \langle \zeta_j, n_i \rangle \right) \geq 0 \tag{2}$$

Fix an index $i = 1, \ldots, m$. Recall that a polynomial $p(x) \in \mathbb{R}[x]$ is non-negative on some interval $[0, \widehat{\varepsilon})$ if and only if

- $p(0) \geq 0$ and
- the term of lowest degree term of $p(x)$ has a positive coefficient (or p is identically 0).

If we view the sum in inequality (2) as a polynomial function p of ε, we see that $p(0) = 0$. Since inequality (1) holds for p.s.e. ε, we conclude that either the first nonzero coefficient of p is positive or else $\langle \zeta_j, n_i \rangle = 0$ for all j.

In the former case, we conclude that the first nonzero term in the polynomial in inequality (2) has a positive coefficient, so inequality (2) holds for p.s.e. ε.

In the latter case, the terms $\sum_{j=1}^{r} \langle \zeta_j, n_i \rangle \varepsilon^j$ and $\sum_{j=1}^{r} \langle \zeta_j, w \rangle \langle \zeta_j, n_i \rangle$ in inequality (2) vanish, leaving:

$$\frac{\varepsilon^{r+1}}{\|\widetilde{w}\|} \langle w, n_i \rangle \geq 0$$

Hence, the fact $0 + \varepsilon w \in V$ for all positive ε small enough implies inequality (2) holds for p.s.e. ε. ∎

The ideas from the preceding proof clarify the relationship between jets of different orders. Within the class of all jet frames based at a point x, along a polytope V, we can say that

- If one takes a weak jet frame ζ of order r and discards the last k vectors from it, the result is a weak jet frame of order $(r - k)$.
- Any weak jet frame of order $(r+1)$ can be constructed from some weak jet frame ζ of order r and some vector $w \notin \text{span}(\zeta)$ by applying the Gram-Schmidt procedure.

When x lies in the relative interior of V, any frame at x built from vectors in the affine span of V will be a jet frame along V. For this reason, jet frames are mainly of interest in the case when $x \in \partial V$.

4.2 Ordering Jet Frames

Henceforth, we assume $x \in \partial(V)$. We are particularly interested in the case when x belongs to several faces of V. To further explore how jet frames record the way in which faces of V meet, we require two more definitions.

Definition 70. Suppose V is a d-dimensional polytope and $x \in V$. We say that a unit vector $\nu \in T_x\mathbb{R}^d$ is an **outer support vector** at x if $\langle \nu, u - x \rangle \leq 0$ for all $u \in V$.[10]

[10] Note that, like Definition 1, this definition makes an implicit identification between the manifold $\mathbb{R}^d$ and the tangent space of x.

We will use outer support vectors to induce an ordering on jet frames.

Definition 71. Suppose $x \in \partial V$ and ν is an outer support vector at x. We define an ordering $\preceq_\nu$ (or just $\preceq$) on the space of jet frames at x as follows: Given jet frames ζ, ζ' of orders r, r' respectively, $\zeta \preceq \zeta'$ if the first entry of the real-valued sequence

$$\{\langle -\nu, \zeta_i' \rangle - \langle -\nu, \zeta_i \rangle\}_{i=1}^{\infty}$$

is non-negative.[11] Here, we view ζ and ζ' as vector valued sequences with finite support (so $\zeta_i = 0$ for $i > r$ and similarly for ζ'). We call this the **lexicographic ordering** induced by ν. The relation $\prec$ is defined similarly.

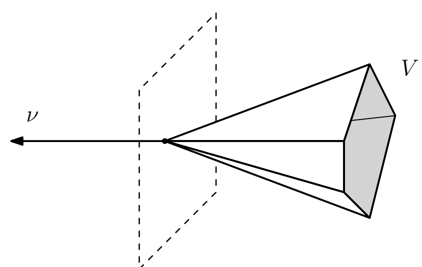

Figure 16: An example of an outer support vector illustrated with the corresponding supporting hyperplane and polytope.

An outer support vector ν should be understood to be a normal vector to a plane that determines a half-space containing V, such that ν points "away" from V (see Figure 16). This means that for any vector $v \in T_x\mathbb{R}^d \cap V$, the angle formed by v and ν will be at least π. In the next few subsections, we will often need to consider the sequence of real numbers $\{\langle -\nu, \zeta_i \rangle\}_{i=1}^{\infty}$ obtained from a partial-jet frame $\zeta = (\zeta_1, \ldots, \zeta_r)$ and outer support vector ν, again following the convention that ζ is really a vector sequence with finite support. For convenience, we will denote this sequence $\langle -\nu, \zeta \rangle$, and allow ourselves to compare two such sequences lexicographically with the notation $\langle -\nu, \zeta \rangle < \langle -\omega, \xi \rangle$.

Proposition 72. *Suppose ζ and ζ' are vectors of equal length in $\mathbb{R}^d$ and $\nu \in \mathbb{R}^d$. Then $|\nu - \zeta| < |\nu - \zeta'|$ if and only if $\langle \nu, \zeta \rangle > \langle \nu, \zeta' \rangle$.*

[11]Notice that the outer support vector ν and all of the vectors ζ_i and ζ_i' belong to the tangent space at x. Thus this sequence of inner products gives us information about how ζ_i and ζ_i' are oriented relative to ν.

Proof. Just notice that:

$$\begin{aligned}0 \leq |\nu - \zeta| < |\nu - \zeta'| &\Leftrightarrow |\nu - \zeta|^2 < |\nu - \zeta'|^2 \\ &\Leftrightarrow \langle \nu - \zeta, \nu - \zeta \rangle < \langle \nu - \zeta', \nu - \zeta' \rangle \\ &\Leftrightarrow -2\langle \nu, \zeta \rangle < -2\langle \nu, \zeta' \rangle \\ &\Leftrightarrow \langle \nu, \zeta \rangle > \langle \nu, \zeta' \rangle\end{aligned}$$

■

Proposition 73. *Fix polyhedra V, W with outer support vectors ν, ω of equal magnitude, at points $p \in V, q \in W$ respectively. Suppose ξ, ζ are partial jet frames at p along V and q along W respectively. Then $-\langle \nu, \xi \rangle < -\langle \omega, \zeta \rangle$ lexicographically if and only if*

$$d(p + \nu, \gamma_\xi(t)) \leq d(q + \omega, \gamma_\zeta(t)) \quad \textit{for p.s.e. } t$$

where d denotes the Euclidean metric.

Proof. Suppose $-\langle \nu, \xi \rangle < -\langle \omega, \zeta \rangle$ lexicographically, so that for some i, $-\langle \nu, \xi_i \rangle < -\langle \omega, \zeta_i \rangle$ and $-\langle \nu, \xi_j \rangle = -\langle \omega, \zeta_j \rangle$ for all $j < i$. So then

$$\begin{aligned}d(p + \nu, \gamma_\xi(t)) &= |(p + \nu) - (p + J_\xi(t))| = |\nu - tJ_\xi(t)|, \\ d(q + \omega, \gamma_\zeta(t)) &= |(q + \omega) - (q + J_\zeta(t))| = |\omega - tJ_\zeta(t)|\end{aligned}$$

and so for any fixed t

$$|\nu - J_\xi(t)| < |\omega - J_\zeta(t)| \Leftrightarrow \langle \nu, J_\xi(t) \rangle > \langle \omega, J_\zeta(t) \rangle$$

by Proposition 72. By linearity and the definitions of J_ξ and J_ζ,

$$\langle \nu, J_\xi(t) \rangle - \langle \omega, J_\zeta(t) \rangle = \sum_{j=1}^{r} t^j (\langle \nu, \xi_j \rangle - \langle \omega, \zeta_j \rangle) = \sum_{j=i}^{r} t^j (\langle \nu, \xi_j \rangle - \langle \omega, \zeta_j \rangle)$$

where r is the maximum of the ranks of ξ and ζ. The last sum is a polynomial in t, and the fact that $-\langle \nu, \xi_i \rangle < -\langle \omega, \zeta_i \rangle$ guarantees the lowest order term of the polynomial is positive. Hence, we may find a δ so that for all $t \in (0, \delta)$,

$$\sum_{j=i}^{r} t^j (\langle \nu, \xi_j \rangle - \langle \omega, \zeta_j \rangle) > 0$$

which means that

$$d(p+\nu,\gamma_\xi(t)) \le d(q+\omega,\gamma_\zeta(t)) \quad \text{for p.s.e.}\, t$$

This completes one direction of the proof. A similar argument shows that $-\langle \nu,\xi\rangle \ge -\langle \omega,\zeta\rangle$ implies $q+\zeta$ is no further from $q+tJ_\zeta(t)$ than $p+\nu$ is from $p+tJ_\xi(t)$. ■

We can use the proposition above in the case when $V = W$ and $p = q$, to understand geometrically what it means for a jet frame at p along V to be "lexicographically minimal." The following corollary makes this precise.

Corollary 74. *Fix an outer support vector ν at a point p in a polyhedron V. A jet frame ζ at x along V is minimal if and only if, for every jet frame ξ at p along V, $p+\nu$ is closer to $p+\varepsilon J_\zeta(\varepsilon)$ than $p+\varepsilon J_\xi(\varepsilon)$ for p.s.e. ε.*

A jet frame ζ for V at x that is minimal with respect to an outer support vector $\nu \in T_xV$ records data about how V is oriented with respect to ν. Miller and Pak explain that among all jet frames for V at x, ζ is the one "tilted as much toward ν as possible." To put it another way, $\gamma_\zeta(t)$ is the trajectory that passes closest to ν, while still traveling within V for p.s.e. t. It is precisely these kind of trajectories that we need to compute in order to model a signal propagating through M along shortest paths from the source point.

4.3 Computing Minimal Jet Frames

Specifying an algorithm for computing a minimal jet frame given an outer support vector ν requires some analysis, since we have a continuum of jet frames from which to choose. Fortunately, the functions governing a jet frame's minimality are simple enough that we may compute minimal jet frames with an explicit algorithm.

Since a jet frame at x depends only on the properties of V in a small neighborhood of x, there is no loss of generality (mathematically or algorithmically) in assuming V is a polyhedral cone with its sole vertex at x. (Algorithmically speaking, if we represent V by a system of inequalities, we simply disregard those inequalities that are not equalities at x to obtain the necessary cone.) The face lattice of a polyhedral cone has properties which help us to analyze the the functions we need to optimize. We outline these properties in the following proposition.

Proposition 75. *Suppose $V \subseteq \mathbb{R}^d$ is a polyhedral cone based at 0. Then V is the cone of its one dimensional faces. That is, if $v_1, \ldots, v_n$ are unit vectors, exactly one from each 1-face of V, then $V = \{t_1 v_1 + \ldots + t_n v_n : t_i \in [0, \infty)\}$. Moreover, any lower-dimensional face can be uniquely identified by which of the vectors v_i it contains.*

Proof. Let $V' = \{t_1 v_1 + \ldots + t_n v_n : t_i \in [0, \infty)\}$. Then $V' \subseteq V$, since V is closed under scaling by non-negative real numbers and addition. As the convex hull of some rays, V' is a polytope with the same 1-faces as V. But now recall Proposition 9, which states that the face lattice of V is atomic, where 1-faces are the atoms. Since V and V' contain exactly the same atoms (namely, *all* of the atoms), they must be equal, because they are both described as the join of all 1-faces of V. The second claim likewise follows from the fact that the face lattice of V is atomic with respect to the one dimensional faces. ■

Since polyhedral cones have some of the properties of vector spaces, our convention will be to identify a cone $V \subseteq \mathbb{R}^n$ (with its vertex at $x \in \mathbb{R}^n$) with a subset of $T_x\mathbb{R}^n$. One configuration of a cone with an outer support vector is particularly important:

Definition 76. Let $V \subseteq T_x\mathbb{R}^n$ be a cone and ν an outer support vector of V at x. We say V is **ν-sharp** (or just **sharp**) if for all nonzero $v \in V$, $\langle v, \nu \rangle < 0$.

We are now ready to analyze which unit vectors within a polyhedral cone have the smallest angle with an outer support vector. In the case of a ν-sharp cone, only finitely many unit vectors may have minimal angle with ν. In the case of a non-ν-sharp cone, we will have a continuum of choices for a unit vector in V having minimal angle with V.

Proposition 77. *Let ν be an outer support vector for a polyhedral cone C, and suppose C is ν-sharp. The minimum angle between ν and a unit vector $\zeta \in C$ occurs when ζ lies on one of the (finitely many) one dimensional faces of C.*

Proof. We will induct on the dimension d of the cone. For the base case, the claim is trivial when $d = 1$, since there is exactly one angle between ν and the single unit vector $\zeta \in C$.

For the induction step, suppose that $d > 1$. Without loss of generality, take ν to be a unit vector. The angle formed by ν and some other unit vector $\zeta \in C$ is given by $\arccos(\langle \zeta, \nu \rangle)$, so minimizing this angle over all unit vectors in C is equivalent to maximizing $\langle \cdot, \nu \rangle$, due to the monotonicity of arccos.

We may suppose $C \subseteq \mathbb{R}^{d+1}$, so that the domain we wish to optimize over is C intersected with the unit sphere $S^d \subseteq \mathbb{R}^{d+1}$. (This is a compact set, so by the Extreme Value Theorem, such a maximizer exists.) Suppose $p, q \in C \cap S^d$. By the convexity of C, the family of vectors $\widetilde{\gamma}(t) = tp + (1-t)q$ (where $t \in [0,1]$) is contained in C, and since C is closed under scaling by positive real numbers, we may conclude the path

$$\gamma(t) = \frac{tp + (1-t)q}{\|tp + (1-t)q\|}$$

is contained in $C \cap S^d$. Now consider the function $\langle \cdot, \nu \rangle$, restricted to γ. For the purposes of determining what point along γ maximizes $\langle \cdot, \nu \rangle$, we may restrict our attention to $\widetilde{\nu}$, the projection of ν into the p, q plane. Using an orthogonal transformation, we may choose coordinates x, y on the p, q-plane such that orthogonal transformation takes $\widetilde{\nu}$ to the negative y axis and γ to a segment of the unit-semicircle in the upper half x, y-plane. In these coordinates, it is clear that $\langle \gamma(t), \nu \rangle$ is just -1 times the y-coordinate of $\gamma(t)$, hence, on this semicircle $\langle \gamma(t), \nu \rangle$ has local maxima only at $t = 0$ and $t = 1$.

Since any point in the relative interior of $C \cap S^d$ is in the relative interior of an arc in $C \cap S^d$, the observation above implies no point in the relative interior of $C \cap S^d$ can be the global maximizer of $\langle \cdot, \nu \rangle$ on $C \cap S^d$. But that means any global maximizer must be in the relative boundary of $C \cap S^d$, which is the union of S^d intersected with each $(d-1)$-dimensional face of C. Applying the induction hypothesis to these finitely many $(d-1)$-dimensional, faces, we conclude any global maximizer must lie in the 1-faces of C.

■

Jet frames consist of orthonormal vectors, so to construct a jet frame for a cone V it is natural to project V onto the orthogonal complement of the partial jet frame we have found so far. In certain scenarios this preserves sharpness.

Proposition 78. *Suppose V is a ν-sharp cone. Suppose w is a unit vector in V, and that among the unit vectors in V, w has the smallest angle with ν. Let π denote the projection of V onto $\text{span}(w)^{\perp}$. Then $\pi(V)$ is also ν-sharp.*

Proof. By Proposition 75, a polyhedral cone consists of linear combinations of vectors with positive coefficients. By the linearity of π, then $\pi(V)$ is again a polyhedral cone. Suppose $v \in \pi(V)$. Then for some $\widetilde{v} \in V$, we have $\widetilde{v} = v + \lambda w$ where $\lambda \in \mathbb{R}$.

Now recall that V is closed under addition and positive-scalar multiplication. Thus, if $\lambda < 0$, we may write $v = \widetilde{v} - \lambda w$, and conclude $v \in V$, so that $\langle v, \nu \rangle < 0$ by the ν-sharpness of V.

Otherwise, $\lambda > 0$. Minimizing the angle with ν is equivalent to maximizing the function $\langle \cdot, \nu \rangle$. Among unit vectors in V, w has the property that it maximizes $\langle \cdot, \nu \rangle$. This allows us to write:

$$\begin{aligned}
\langle v, \nu \rangle &= \langle \widetilde{v} - \lambda w, \nu \rangle = \langle \widetilde{v}, \nu \rangle - \lambda \langle w, \nu \rangle \\
&= \|v\| \left\langle \frac{\widetilde{v}}{\|v\|}, \nu \right\rangle - \lambda \langle w, \nu \rangle \\
&\leq \|v\| \langle w, \nu \rangle - \lambda \langle w, \nu \rangle = (\sqrt{\|w\|^2 + \lambda^2} - \lambda) \langle w, \nu \rangle
\end{aligned}$$

Since $\sqrt{\|w\|^2 + \lambda^2} - \lambda > 0$ for $\lambda > 0$, we conclude that $\langle v, \nu \rangle < 0$. We chose v generally from $\pi(V)$, so we conclude $\pi(V)$ is ν-sharp. ■

Suppose V is a polyhedral cone with outer support vector ν, and $\zeta = (\zeta_1, \ldots, \zeta_n)$ denotes a jet frame for V that we plan to construct algorithmically. Intuitively, the lexicographical ordering on jet frames means that if we wish to construct ζ to be ν-minimal, it is more important to maximize $\langle \nu, \zeta_i \rangle$ for small i than for large i. The fact ν is an outer support vector means $\langle \nu, \zeta_i \rangle$ cannot exceed 0, so it is important that vectors in $V \cap \langle \nu \rangle^{\perp}$ appear first in the jet frame to ensure ν-minimality. Given these facts and the proposition above, one might guess that an algorithm for computing minimal jet frames must have two phases:

1. Determine a maximal linearly independent set of vectors in $V \cap \langle \nu \rangle^{\perp}$. Apply Gram-Schmidt to these vectors to obtain $\zeta_1, \ldots, \zeta_k$.

2. Let $\pi(V)$ be the projection of V onto $\operatorname{span}(\zeta_1, \ldots, \zeta_k)^{\perp}$. $\pi(V)$ is then a ν-sharp cone. Determine the remaining vectors $\zeta_{k+1}, \ldots, \zeta_n$ by examining the one-dimensional faces of projections of $\pi(V)$.

The following algorithm and its proof of correctness makes this intuition precise.

Algorithm 1: Minimal Jet Frame Finding

Input: A d-dimensional polytope V, a point $x \in \partial V$, and an outer support vector ν for V based at x. (All vectors, including ν are assumed to be in $T_x\mathbb{R}^{d+1}$.)

Output: A minimal jet frame $(v_1, \ldots, v_d)$ for V at x.

Extend V to a polyhedral cone $\widetilde{V}$ with x as its vertex.
Adjust coordinates of $\widetilde{V}$ so that x is the origin.
Initialize $P, F := \varnothing$.[a]

—Phase 1 of Jet Frame Construction—
Calculate[b] a maximal lin. indep. set of vectors $\{z_k\} \subset C \cap \operatorname{span}(\nu)^\perp$.
Apply Gram-Schmidt to $\{z_k\}$, producing orthonormal vectors $\{\zeta_k\}$.
Initialize a partial jet frame $(\zeta_1, \ldots, \zeta_K)$, add it to P.

—Phase 2 of Jet Frame Construction—

```
while P ≠ ∅ do
    Remove a partial jet frame (v_1, ..., v_{j-1}) from P
    Let π be the projection of ℝ^n onto span(v_1, ..., v_{j-1})^⊥.
    —Note π(C) is a ν-sharp cone. We apply Proposition 77.—
    Calculate^c the one-dimensional faces F_1, ..., F_m of π(C).
    Determine vectors w_1, ..., w_m spanning F_1, ..., F_m respectively.
    α := min_{k=1,...,m} Angle(w_k, ν)
    for F among F_1, ..., F_m do
        Let v be a unit vector spanning F.
        if Angle(w_k, ν) = α then
            Let ζ denote the partial jet frame obtained by extending
            (v_1, ..., v_{j-1}) with w_k.
            if k = d then
                Add ζ to F.
            else
                Add ζ to P.
```

— Now P is empty and F is a set of finitely many jet frames.—
return argmin(F, $\preceq_\nu$)

[a] The set F will store jet frames, while P stores partial jet frames. Here the "sets" we speak of are data structures, rather than mathematical objects. For this reason, we may speak of adding and removing elements from P and F.

[b] Via standard polytope-theoretic techniques.

[c] Also by standard polytope-theoretic techniques.

Correctness of Algorithm 1. That this algorithm will produce a d-dimensional orthonormal basis is clear. To see the algorithm produces a jet frame, we will check that at each step of the calculation, we possess a partial jet frame. Certainly this is true initially, when we have no vectors. Suppose now $\{v_1, \dots, v_{j-1}\}$ is a partial jet frame, and the algorithm's newly chosen vector is v_j.

By construction, $v_j = \pi_{j-1}(v)$ for some $v \in C$. Now, recall that because V is a cone, the fact that $v \in C$ means that $\lambda v \in C$ for all $\lambda \in [0, \infty)$. Lemma 69 allows us to use v to extend the partial jet frame $(v_1, \dots, v_{j-1})$, and the vector we obtain by applying the Gram-Schmidt procedure to v is precisely v_j.

Next, let us argue the algorithm actually finds a minimal jet frame. Suppose the jet frame $(w_1, \dots, w_d)$ is strictly smaller (with respect to $\prec_\nu$), than any jet frame in F. Then there exists some jet frame $(v_1, \dots, v_d) \in F$ that agrees with $(w_1, \dots, w_d)$ up to some greatest position j (possibly $j = 0$), so that in the lexicographical ordering we have $\langle \nu, w_{j+1} \rangle < \langle \nu, v_{j+1} \rangle$. But that implies we did not make an optimal choice of v_{j+1}, in contradiction to the specifications of our algorithm.

■

In later arguments, we will determine a minimal jet frame ζ and use its trajectory γ_ζ to obtain certain useful estimates. To prove these estimates, it will be important to know that the path γ_ζ is not a particularly "exotic" path, like the topologist's sine curve. Specifically, we will need to know that given any hyperplane H, $\gamma_\zeta(t)$ remains on only one side of H for p.s.e. t, rather than oscillating between the two half spaces.

Proposition 79. *Suppose ζ is a jet frame at $x \in \mathbb{R}^d$, and that p and q are distinct points in $\mathbb{R}^d$. Then*

- $d(p, \gamma_\zeta(t)) \le d(q, \gamma_\zeta(t))$ *for p.s.e.* t *or*
- $d(q, \gamma_\zeta(t)) < d(p, \gamma_\zeta(t))$ *for p.s.e.* t.

Moreover, exactly one of the two alternatives holds.[12]

Proof. Consider the $(d-1)$-dimensional hyperplane H of points that are equidistant between p and q. Let $m = \frac{1}{2}p + \frac{1}{2}q$, so that m lies on H. Then

[12] Note that the negation of the proposition "$P(t)$ for p.s.e. t" is not "$(\neg P(t))$ for p.s.e. t", so the claim is not trivial.

the vector $n = p - m$, based at m is a vector normal to H, and so we can say for any $u \in \mathbb{R}^d$, if $\langle u - m, n\rangle \geq 0$ then $d(u,p) \leq d(u,q)$, otherwise $\langle u - m, n\rangle < 0$ and $d(u,p) < d(u,q)$.

Now consider $\gamma_\zeta(t) = x + \sum_{j=1}^{k} t^j \zeta_j$. We may write

$$\langle \gamma_\zeta(t) - m, n\rangle = \langle x - m, n\rangle + \sum_{j=1}^{k} \langle \zeta_j, n\rangle\, t^j$$

so that we may view $\langle \gamma_\zeta(t) - m, n\rangle$ as a polynomial $p(t)$ with real coefficients, not all of which are zero. Since the polynomial is nonzero, there exists an $\varepsilon > 0$ so that on the interval $(0, \varepsilon)$, either $p(t) \geq 0$ for all $t \in \varepsilon$ or $p(t) < 0$ for all $t \in (0, \varepsilon)$. Hence, exactly one of the two desired alternatives holds. ■

5 Constructing Source Images

In Section 3.5, we learned source images induce Voronoi diagrams on the faces of the polytope. In fact, by Theorem 49, we know that when faces F and F' share a $(d-1)$-face R and we rotate $V(\mathrm{src}_F)$ into $T_{F'}$, the rotated diagram will agree with $V(\mathrm{src}_{F'})$ along R.

Taken together, these observations suggest we can specify an algorithm for finding source images by analyzing the relationship between $V(\mathrm{src}_F)$ and $V(\mathrm{src}_{F'})$ for neighboring faces F, F'. In this section, we will develop the results necessary to conduct this analysis precisely.

5.1 Describing Voronoi Diagrams and Unfoldings

We require terminology that describes configurations of a point, a $(d-1)$-face, a d-face, and a Voronoi diagram formed by the source images in the affine span of a d-face. Recall from Definition 3 that if $H \subseteq \mathbb{R}^n$ is a hyperplane, then $\mathbb{R}^n \setminus H$ has two connected components, which are open half-spaces.

Definition 80. Suppose $H \subseteq \mathbb{R}^n$ is a hyperplane. Suppose A and B are subsets of $\mathbb{R}^n$, such that:

- A is contained in the closure of exactly one of the two connected components of $\mathbb{R}^n \setminus H$, call it H_A,
- B is contained in the closure of exactly one of the two connected components of $\mathbb{R}^n \setminus H$, call it H_B.

If $H_A = H_B$, we say that A and B are **on the same side** of H. Otherwise, we say they are **on opposite sides** of H and that H **lies between** A and B. When $A = \{a\}$ and $B = \{b\}$ are singleton sets, we will suppress the set notation (for example, "a and b are on the same side of H").

If $P \subseteq \mathbb{R}^n$ is a $(n-1)$-dimensional polytope we say that sets A, B are on the same side (resp. on opposite sides) of P if they are on the same side (resp. on opposite sides) of $\mathrm{aff}(P)$.

Definition 81. Suppose F is a d-face, $R \subset F$ is a $(d-1)$-face, and $Y \subset T_F$ is finite. Suppose $\omega \in Y$.

- ω **sees** F **through** R **in** $V(Y)$ if R lies between ω and F and $V(Y, \omega)$ contains a point in $\mathrm{rint}(R)$.

- ω **sees** R **through** F **in** $V(Y)$ if ω and F lie on the same side of R and $V(Y, \omega)$ contains a point in $\text{rint}(R)$.

In either of these two configurations, we say the pair (R, ω) has:

- A **closest point**, $\rho(R, \omega)$, that is the unique point in $R \cap V(Y, \omega)$ that is closest to ω.

- A **radius**, $r(R, \omega)$, that is the (Euclidean) distance from ω to $\rho(R, \omega)$ in T_F.

- An **outer support vector**, $\omega - \rho(R, w)$.

- An **angle sequence** $\angle(R, \omega)$, given by

$$\langle -\omega - \rho(R, \omega), \zeta \rangle$$

for any minimal jet frame ζ at $\rho(R, \omega)$ along $R \cap V(Y, \omega)$.

We should briefly justify why the notion of a closest point is well defined.

Proposition 82. *The closest point specified in Definition 81 is unique.*

Proof. Suppose, by way of a contradiction, that $R \cap V(Y, \omega)$ contains distinct points p, q that minimize the distance to ω. Then by convexity of $R \cap V(Y, \omega)$, the line segment $[p, q]$ is contained in $R \cap V(Y, \omega)$. By basic trigonometry, the midpoint of this line segment must be strictly closer to ω, contradicting the minimality of p and q. ∎

Remark 83. Due to Proposition 57, the jet frame ζ in the definition must have dimension $(d - 1)$, and thus any angle sequence has length $(d - 1)$.

Example 84. Miller and Pak present the following useful example in their paper. Consider the following configuration in T_F. Voronoi boundaries are indicated with dotted lines. The point ω can see F through R in the Voronoi diagram. It can also see R'' through F in the Voronoi diagram. However, ω' cannot see R'' through F in the Voronoi diagram. This means every point in R'' is strictly closer to some other point in Y than it is to ω'. The dashed arrow depicts the outer support vector for (R, ω), based at $\rho(R, w)$.

Notice that $\rho(R, w) \in S_{d-2}$. By the definition of a source image, this means we cannot unfold a shortest path from the source to $\rho(R, w)$ to produce ω. Minimal jet frames will help us to overcome this difficulty.

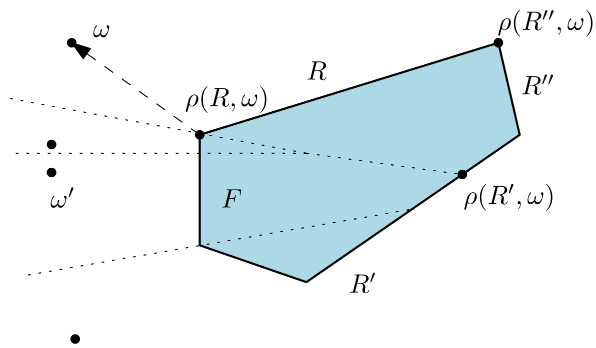

Figure 17

Definition 81 allows us to introduce the objects we will use for calculating source images. These are **events** and **potential events**.

Definition 85. An **event** is a pair (ν, F) with $\nu \in \mathrm{src}_F$. Each event has:

- A **radius** $r(\nu, F)$, which is the radius $r(R, \nu)$ from ν to the $(d-1)$-face R through which ν can see F in $V(\mathrm{src}_F)$.[13]

- An **event point** $\rho(\nu, F) = \rho(R, \nu)$.

- An angle sequence $\angle(\nu, F) = \angle(R, \nu)$.

The **source event** consists of the source point and the source face; its event point is also the source point and its angle sequence consists only of zeros.

Events, we recall, are supposed to provide a discrete, precise way of describing a signal propagating from the source point out through the manifold. Naturally, events are only a useful abstraction if we can order them.

Definition 86. Suppose (ν, F) and (ν', F') are events. We write $(\nu, F) \prec (\nu', F')$, and say (ν, F) **precedes** (ν', F')

- if $r(\nu, F) < r(\nu', F')$ or

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

[13] Recall that Proposition 52 explains why it makes sense to speak of *the* $(d-1)$-face R through which ν sees F.

We write $(\nu, F) \preceq (\nu', F')$, if $(\nu, F) \prec (\nu', F')$ or $(\nu, F) = (\nu', F')$.

Clearly, the relation $\preceq$ is transitive. It is possible to have distinct events with the same radius and angle sequence — consider for example the unit cube embedded in $\mathbb{R}^3$ with the source point centered in one of the faces.

Remark 87. Some elementary observations about the relation $\preceq$ are:

- $\preceq$ is a partial ordering.
- The source event precedes any other event.
- Two events are $\preceq$-incomparable if and only if their radii and angle sequences match exactly.

The third observation suggests that it is better to think of $\preceq$ like the relation $\leq$ on $\mathbb{R}$, with certain "singularities," rather than like the relation $\subseteq$ defined on sets. It is particularly important to recognize that virtually all events are comparable. In everyday language, we often use the word "precedes" to mean "immediately precedes" as in "James Buchanan preceded Abraham Lincoln as President, and Lincoln preceded Andrew Johnson." As we have defined it, "precedes" should be understood in the same manner as "happened before" so that event A may precede event B without the two events occurring in succession.

Definition 88. The **source poset** for v and S, denoted $\text{src}(v, S)$, is the set of all events for the source point v together with the partial ordering $\preceq$. Thus $\text{src}(v, S)$ is the set of pairs:

$$\text{src}(v, S) := \{(\nu, F) : F \text{ a } d\text{-face of } S,\ \nu \in \text{src}_F(v)\}$$

We say $\mathcal{I} \subseteq \text{src}(v, S)$ is an **order ideal** if $\mathcal{I}$ has the property that for any $e \in \text{src}(v, S)$, $e \preceq i$ for some $i \in \mathcal{I}$ implies $e \in \mathcal{I}$. To describe this property concisely, we say $\mathcal{I}$ is **closed under decreasing** (with respect to $\preceq$). Clearly, both $\varnothing$ and $\text{src}(v, S)$ are order ideals.

The radii and angle sequences of the events in an ideal carry important information about which events will be in that ideal.

Definition 89. Given an order ideal $\mathcal{I}$ the **radius** of the ideal is the quantity $r(\mathcal{I}) := \max_{(\nu,F)\in\mathcal{I}} r(\nu, F)$.[14]

[14] Recall that since src_F is finite for any d-face F, and we have finitely many d-faces, the source poset is finite. Thus, it makes sense to write a maximum rather than a supremum here.

Remark 90. Rephrasing the information in Remark 87 in terms of order ideals, we can say that if (ω, F) is an event with $r(\omega, F) < r(\mathcal{I})$, then $(\omega, F) \in I$. Put still another way, if $(\omega, F) \notin I$, then $r(\mathcal{I}) \leq r(\omega, F)$. We will require this last characterization in later arguments.

5.2 Understanding Order Ideals

When constructing an algorithm, a common strategy is to identify a recursive structure in the problem of interest. Once this recursive structure is identified, the algorithm may be specified by an inductive procedure in which one begins with a trivial sub-solution that is upgraded on each iteration to a larger sub-solution until a full solution is found.

We will specify an algorithm for calculating $\mathrm{src}(v, S)$ in which order ideals play the role of sub-solutions. The algorithm begins with the trivial sub-solution, $\varnothing$. In each step, it uses the current order ideal I to select an event E with a minimality property given by $\prec$. This gives us a new subsolution $I' = I \cup \{E\}$. If we prove I' is again an order ideal, we are guaranteed that once we can no longer find any more events, we have found $\mathrm{src}(v, S)$, the maximal order ideal. In this sense, the algorithm is analogous to a greedy optimization procedure—we need to argue that at no point do we "get stuck" with a proper subset of $\mathrm{src}(v, S)$ as a result of our greedy choice of events.

Order ideals represent, in a precise combinatorial sense, a "signal" or "wave front" propagating from the source point across S. Events within an ideal $\mathcal{I}$ correspond to regions of S that the wave front has passed through, and the radius of $\mathcal{I}$ records how far this wave front has moved away from the source. The definition of $\prec$ is useful for calculations, but it does not capture the "dynamic" interpretation we have just described in an obvious way. The following definition and proof clarify this situation by explaining how $\prec$ is related to shortest paths.

Definition 91. Suppose γ is a shortest path with face sequence $(F_1, \ldots, F_n)$ and that $v \in F_1$. Let $\nu_i \in T_{F_i}$ denote the point obtained by unfolding v into T_{F_i} via $(F_1, \ldots, F_i)$. Then for any ν_i, ν_j, with $i < j$, we say (ν_i, F_i) **geodesically precedes** (ν_j, F_j).

Remark 92. An event A may geodesically precede another event B another without the two events arising from adjacent faces. Thus, as in Remark 87, "precedes" should not be confused with "immediately precedes."

Proposition 93. *Suppose (ν_1, F_1) geodesically precedes (ν_2, F_2). Then $(\nu_1, F_1) \prec (\nu_2, F_2)$.*

Proof. Let γ be a shortest path. Suppose F_1 and F_2 appear somewhere in the face sequence of γ, with F_1 appearing before F_2. If we can prove the claim in the case of faces that share a $(d-1)$-face, we may use transitivity of $\prec$ and induct over the face sequence of γ to prove the claim in general. So we may assume the F_1 and F_2 share a $(d-1)$-face.

If F_1 is the first face in the face sequence, then F_1 is the source face and (ν_1, F_1) is the source event. In this case, the source event trivially satisfies $(\nu_1, F_1) \prec (\nu_2, F_2)$. We may now assume F_1 is not the first face in the face sequence, so that there exist $(d-1)$-faces $R_1 \subset F_1$ and $R_2 \subseteq F_1 \cap F_2$ such that γ passes through R_1 into F_1, then through R_2 into F_2.

We compare the events (ν_1, F_1) and (ν_2, F_2) by comparing their radii and angle sequences. By definition, these quantities are determined by the lower dimensional polytopes $V_j := R_j \cap V(\mathrm{src}_{F_j}, \nu_j)$, $j = 1, 2$. Let $\Phi : T_{F_1} \to T_{F_2}$ be the rotation about R_2 that unfolds T_{F_1} into T_{F_2}. Notice $\nu_2 = \Phi(\nu_1)$ and the geometry of F_1 relative to ν_1 is the same as the geometry of $\Phi(F_1)$ relative to ν_2. For this reason, we will now work entirely in T_{F_2}, identifying V_1 with $\Phi(V_1)$ and letting $\nu = \nu_2$ without further discussion of unfoldings.

V_1 and V_2 lie within the same traversable cone emanating from v. By unfolding this traversable cone into T_{F_2} we see that any line from ν to V_2 must pass through V_1. Hence, $r(R_1, \nu_1) \leq r(R_2, \nu_2)$. If $V_1 \cap V_2 = \varnothing$, then the inequality is strict and $(\nu_1, F_1) \prec (\nu_2, F_2)$. So we may assume $V_1 \cap V_2 \neq \varnothing$.

By Proposition 82, the closest point in V_1 to ν is unique and the closest point in V_2 to ν is unique. Since $V_1 \cap V_2 \neq \varnothing$, we see that in fact these two points must be the same. So the angle sequences of (ν_1, F_1) and (ν_2, F_2) are based at the same point p and are determined by the same outer support vector. Change coordinates so that p is the origin in T_{F_2} and the outer support vector is ν. We obtain a configuration like the one in Figure 18.

We have reduced the problem to a very specific configuration and now must compare the minimal jet frames of (ν_1, F_1) and (ν_2, F_2). It is difficult to compare minimal jet frames directly, so instead we will produce a jet frame J between them and use the transitivity of $\prec$. By Proposition 57 both V_1 and V_2 are $(d-1)$-dimensional, so their minimal jet frames have the same order. Let ω denote the minimal jet frame associated with (ν_1, F_1) and $\eta = (\eta_1, \ldots, \eta_{d-1})$ denote the minimal jet frame associated with (ν_2, F_2).

Since V_1 and V_2 are $(d-1)$-dimensional, $\mathrm{aff}(V_j) = \mathrm{aff}(R_j)$ for $j = 1, 2$.

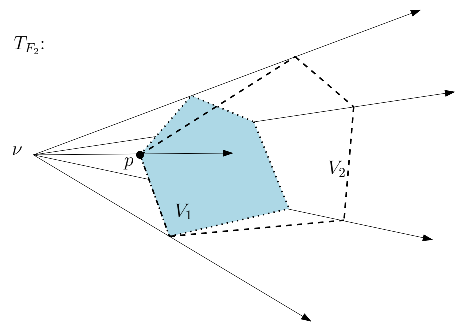

Figure 18: An example (in dimension 3), of the unfolded traversable cone (indicated by rays), with V_1 (the blue plane segment) in the background and V_2 (the transparent plane segment with dotted line boundary) in the foreground. They share the point p (and, in this example, an entire edge). The line segment from p to v denotes the outer support vector for both V_1 and V_2, and the minimal jet frames with respect to this outer support vector record the fact V_1 is tilted closer to ν than V_2.

Since R_1 and R_2 are distinct codimension 1 faces of a polytope, $\operatorname{aff}(R_1)$ and $\operatorname{aff}(R_2)$ are not equal and neither contains the other. The vectors in a minimal jet frame along a polytope V form a basis for $\operatorname{aff}(V)$. So the fact $\operatorname{aff}(V_1) = \operatorname{aff}(R_1)$ and $\operatorname{aff}(V_2) = \operatorname{aff}(R_2)$ have nonempty symmetric difference implies the existence of an index $j < d-1$ for which $(\eta_1, \ldots, \eta_j)$ is a weak jet frame for V_1, but $(\eta_1, \ldots, \eta_{j+1})$ is not.

We have two cases to consider: Either η_{j+1} and ν are collinear or else they are not. Suppose the former, so the angle between η_{j+1} and ν is π. Notice that there can only be one unit vector that is both collinear with ν and extends $(\eta_1, \ldots, \eta_j)$. Consequently, any vector $\widetilde{\eta}_{j+1}$ that extends $(\eta_1, \ldots, \eta_j)$ to a partial $(j+1)$-jet frame for V_1 forms an angle with ν strictly smaller than the angle between η_{j+1} and ν. If we complete $(\eta_1, \ldots, \eta_j, \widetilde{\eta}_{j+1})$ to a jet frame J for V_1, we know that $\omega \preceq_\nu J$ by minimality of ω, and that $J \prec_\nu \eta$ by construction, so that by transitivity $\omega \prec_\nu \eta$.

Next consider the case where η_{j+1} and ν are not collinear. For some $\varepsilon > 0$ small enough, $q_\varepsilon = p + \varepsilon\eta_1 + \ldots + \varepsilon^{j+1}\eta_{j+1} \in V_2$. The line connecting q_ε and ν contains a point in V_1, which means there exists $\lambda \in (0,1)$ for which $\lambda\nu + (1-\lambda)q_\varepsilon \in V_1$. Let $\pi : \mathbb{R}^d \to \mathbb{R}^d$ denote the projection that annihilates exactly $\operatorname{span}(\eta_1, \ldots, \eta_j)$. Then

$$\pi(\lambda\nu + (1-\lambda)q_\varepsilon) = \lambda\pi(\nu) + (1-\lambda)\varepsilon^{j+1}\eta_{j+1} \neq 0$$

since ν and η_{j+1} are not collinear. Let $\widetilde{\eta}_{j+1}$ be the unit vector in the direction of $\pi(\lambda\nu+(1-\lambda)q_\varepsilon)$. Then, by an application of the Gram-Schmidt Procedure for Jet Frames (Lemma 69), $\widetilde{\eta}_{j+1}$ extends $(\eta_1,\ldots,\eta_j)$ to a partial $(j+1)$-jet frame for V_1. Now consider the triangle in the plane spanned by $\pi(\nu)$ and η_{j+1} (Figure 19). By construction the angle between $\widetilde{\eta}_{j+1}$ and ν is strictly smaller

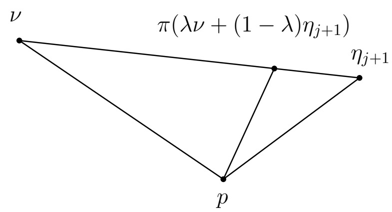

Figure 19

than the angle between η_{j+1} and ν. As before, we may extend $(\eta_1,\ldots,\eta_{j+1})$ to a jet frame J for V_1, to obtain the desired inequality $\omega \preceq_\nu J \prec_\nu \eta$. ■

Definition 94. Suppose $\mathcal{I} \subseteq \operatorname{src}(v,S)$ is an order ideal. For each d-face F, define $Y_F := \{\omega \in \operatorname{src}_F : (\omega,F) \in \mathcal{I}\}$.[15] The **potential events** for $\mathcal{I}$ is the set of triples (ω,F,R') such that

- ω sees R' through F in the Voronoi diagram $V(Y_F)$.
- Another face F' also contains R, and folding ω into $T_{F'}$ over R' yields a point ω' such that $(\omega',F') \notin \mathcal{I}$.

If $(\omega',F') \in \operatorname{src}(v,S)\setminus I$, then (ω',F') comes from **processing** (ω,F,R'). As with events, each potential event (ω,F,R') is equipped with

- A **radius**, $r(\omega,F,R')$, given by the distance from ω to $V(Y_F,\omega)\cap R'$.
- An **event point**, $\rho(\omega,F,R')$, the closest point in $V(Y_F,\omega)\cap R'$ to ω.
- An **angle sequence** $\angle(\omega,F,R')$, given by the minimal jet frame ζ along $V(Y_F,\omega)\cap R'$, at the event point.

[15]Note that Y_F depends on $\mathcal{I}$. In the arguments that follow, we will have only one order ideal to consider so for convenience we do not explicitly record the dependence on $\mathcal{I}$ in the notation. However, we will need to take this dependence into account when specifying our algorithms.

The relation $\preceq$ extends naturally to $\mathcal{E}_{\mathcal{I}}$, so that we may compare potential events and events. Naturally, if we consider the collection of all events and potential events together with $\preceq$, the result is not a poset. With this understanding, we will simplify the notation in our arguments by adopting the convention that when comparing potential events A, B (or a potential event and an event), we allow ourselves to write $A \preceq B$ even if the radius and angle sequences of A and B are identical. We say a potential event $E \in \mathcal{E}_{\mathcal{I}}$ is **minimal** if for any other $E' \in \mathcal{E}_{\mathcal{I}}$, $E \preceq E'$.

Our goal throughout this section is to argue that if (ω, F, R') is a minimal potential event in $\mathcal{E}_I$, then $\Phi_{F,F'}(\omega) \in \mathrm{src}_{F'}$ and $I \cup \{(\Phi_{F,F'}(\omega), F')\}$ is an order ideal. In other words, we need to prove that by processing a minimal potential event, one can create a larger order ideal.

Earlier, we said an order ideal $\mathcal{I}$ represents combinatorially how far a signal has propagated over the manifold. The following definition extends this metaphor:

Definition 95. Suppose $\mathcal{I}$ is an order ideal. The **horizon** (for $\mathcal{I}$), denoted $\mathcal{H}$ is the set of all events $(\nu, F) \in \mathrm{src}(v, S)$ obtained by processing a potential event. The **undiscovered events** (for $\mathcal{I}$), denoted $\mathcal{U}$ are defined to be $\mathrm{src}(v, S) \setminus (\mathcal{I} \cup \mathcal{H})$.

Clearly, for any ideal $\mathcal{I}$, we can write $\mathrm{src}(v, S) = \mathcal{I} \sqcup \mathcal{H} \sqcup \mathcal{U}$. The next proposition explains how this decomposition of $\mathrm{src}(v, S)$ relates to $\preceq$. Naturally, we are focused on characterizing $\mathrm{src}(v, S) \setminus \mathcal{I}$, because in the context of the algorithm we wish to construct, we need to calculate an event not in $\mathcal{I}$ based on information from $\mathcal{I}$.

Proposition 96. *Suppose $\mathcal{I}$ is a nonempty order ideal and (ω, F) is an event not in $\mathcal{I}$. Then either*

- $(\omega, F) \in \mathcal{H}$ *or else*
- $(\omega, F) \in \mathcal{U}$ *and there exists an event $(\omega', F') \in \mathcal{H}$ such that $(\omega', F') \preceq (\omega, F)$.*

Furthermore, in the second case, we may choose (ω', F') so that it geodesically precedes (ω, F).

Proof. By definition, since $\omega \in \mathrm{src}_F$, there exists a shortest path γ from v to some point $p \in F \setminus S_{d-2}$ such that unfolding v along the face sequence of γ yields ω. Let $(F_1, \ldots, F_n)$ denote the face sequence of γ, so that $v \in F_1$ and $F = F_n$.

Any subpath of γ is again a shortest path, so for each $j = 1, \ldots, n$, we may unfold the portion of γ with face sequence $(F_1, \ldots, F_j)$ into T_{F_j} to obtain a source image $\omega_j \in \mathrm{src}_{F_j}$ $(\omega_1 = v, \omega_n = \omega)$. Thus, we have a sequence of events $\{(\omega_j, F_j)\}_{j=1}^{n}$. By Proposition 93, this sequence is strictly increasing with respect to $\prec$.

The source event (ω_1, F_1) is trivially in $\mathcal{I}$, while (ω_n, F_n) is not. Using the definition of an order ideal, we conclude there must exists an index j for which $(\omega_k, F_k) \in \mathcal{I}$ if $k \leq j$ and $(\omega_k, F_k) \notin \mathcal{I}$ for $k > j$.

Let R' denote the $(d-1)$-face between F_j and F_{j+1}. We claim $(\omega_j, F_j, R') \in \mathcal{E}_{\mathcal{I}}$. By construction, it is clear that $\Phi_{F_j, F_{j+1}}(\omega_j) = \omega_{j+1}$, but we must check that ω_j can actually see R' through the Voronoi diagram $V(Y_{F_j})$ on F_j.

We know that γ passes through some point in $\mathrm{rint}(R')$. Since γ unfolds to produce ω_j, that same point must belong to the Voronoi cell $V(\mathrm{src}_{F_j}, \omega_j)$. So ω_j sees R' through $V(\mathrm{src}_{F_j})$. Since $Y_{F_j} \subseteq \mathrm{src}_{F_j}$, $V(\mathrm{src}_{F_j}, \omega_j,) \subseteq V(Y_{F_j}, \omega_j)$. So ω_j sees R' through $V(Y_{F_j})$ as well. Thus (ω_j, F_j, R') is indeed a potential event, and processing this event yields (ω_{j+1}, F_{j+1}).

If $j+1 = n$, then we have demonstrated that $(\omega, F) \in \mathcal{H}$. If $j+1 < n$, then by letting $(\omega', F') = (\omega_{j+1}, F_{j+1})$ we have $(\omega', F') \in \mathcal{H}$ with $(\omega', F') \prec (\omega, F)$ and moreover by our construction γ is exactly the shortest path needed to conclude (ω', F') geodesically precedes (ω, F). ■

There are a few noteworthy details in the preceding proof. Consider the relationship between the potential event (ω_j, F_j, R') and (ω_{j+1}, F_{j+1}). From Section 3.5, we know $R' \cap V(\mathrm{src}_{F_j}, \omega_j) = R' \cap V(\mathrm{src}_{F_{j+1}}, \omega_{j+1})$. Earlier, we observed that $V(\mathrm{src}_{F_j}, \omega_j) \subseteq V(Y_{F_j}, \omega_j)$. So, the radii and angle sequence of (ω_{j+1}, F_{j+1}) are determined by the position of ω_j relative to $R' \cap V(\mathrm{src}_{F_j}, \omega_j)$, while the radii and angle sequence of (ω_j, F_j, R') are determined by the position of ω_j relative to the larger set $R' \cap V(Y_{F_j}, \omega_j)$. Hence $r(\omega_j, F_j, R') \leq r(F_{j+1}, \omega_{j+1})$ with equality only if $\angle(\omega_j, F_j, R')$ weakly precedes $\angle(F_{j+1}, \omega_{j+1})$ lexicographically. Thus, we have

$$(\omega_j, F_j, R') \preceq (\omega_{j+1}, F_{j+1}) \prec (\omega_k, F_k) \text{ for all } k > j+1$$

In this way, we may use the preceding proposition to compare events in $\mathrm{src}(v, S) \setminus \mathcal{I}$ with potential events. We will particularly need the preceding observation in the case of processing a potential event, so we record it here.

Corollary 97. *Suppose (ω, F, R') is a potential event in $\mathcal{E}_I$. Suppose processing this potential event yields an event (ω', F'). Then $(\omega, F, R') \preceq (\omega', F')$.*

5.3 Expanding Order Ideals

In this subsection, we prove the main result of the paper, Theorem 98. Using this theorem, we will be able to easily prove the correctness of an algorithm (presented in Section 6) that determines $\mathrm{src}(v, S)$ for input v and S. Despite the large collection of abstractions we have already developed, the proof of Theorem 98 is somewhat complicated. For this reason, this section must be treated as one large argument that investigates the following situation:

- $\mathcal{I}$ is a nonempty order ideal in $\mathrm{src}(v, S)$.
- (ν, F, R') is a minimal potential event in $\mathcal{E}_{\mathcal{I}}$.

Unpacking the definition of our chosen minimal potential event (ν, F, R'), we may fix the following notation:

- x denotes the closest point of (ν, F, R').
- ζ denotes the minimal jet frame (at x, along $V(Y_F, \nu) \cap R'$) of (ν, F, R').
- $\nu' := \Phi_{F,F'}(\nu)$, where F' is the other d-face containing R'.

Finally, we will frequently need to calculate distances in the Euclidean spaces T_F and $T_{F'}$. We will use d to indicate the Euclidean distance on these spaces, rather than confuse the notation by introducing a name for each distance. Our ultimate goal is to argue:

Theorem 98. *Given the assumptions above*

1. $(\nu', F') \in \mathrm{src}(v, S)$ *and*
2. $\mathcal{I} \cup \{(\nu', F')\}$ *is an order ideal.*

However, it will be easier to prove this theorem if we begin by proving three bounds in Lemmas 99, 100, and 101. Using these bounds, we will argue the existence of a shortest path from v to a point in $F' \setminus S_{d-2}$ that unfolds to yield ν.

Lemma 99. *Suppose $\omega \in \mathrm{src}_F$ with $(\omega, F) \in \mathcal{I}$ and $\gamma_\zeta(t) \in V(\mathrm{src}_F, \omega)$ for p.s.e. t. Then:*

$$d(\nu, \gamma_\zeta(t)) \leq d(\omega, \gamma_\zeta(t)) \text{ for p.s.e. } t$$

Proof. Because $(\omega, F), (\nu, F) \in \mathcal{I}$, ω and ν are sites in the Voronoi diagram $V(Y_F)$. Then the fact $\gamma_\zeta(t) \in \mathrm{rint}(V(Y_F, \nu,) \cap R')$ for p.s.e. t means exactly that $d(\nu, \gamma_\zeta(t)) \leq d(\omega, \gamma_\zeta(t))$ for p.s.e. t. ■

Lemma 100. *Suppose $\omega \in \mathrm{src}_{F'}$ with $(\omega, F') \in \mathcal{I}$ and $(\Phi_{F',F}(\omega), F)$ is an event in $\mathcal{I}$. Suppose $\gamma_\zeta(t) \in V(\mathrm{src}_{F'}, \omega)$ for p.s.e. t. Then:*

$$d(\nu, \gamma_\zeta(t)) \leq d(\omega, \gamma_\zeta(t)) \text{ for p.s.e. } t$$

Proof. As in Lemma 99, the fact $(\omega, F') \in \mathcal{I}$ means ω is one of the sites in $V(Y_{F'})$. By our hypotheses, $\gamma_\zeta(t) \in Y(\mathrm{src}_{F'}, \omega) \subseteq V(Y_{F'}, \omega)$ for p.s.e. t. Let $\widehat{\omega} := \Phi_{F',F}(\omega)$. Then $\widehat{\omega}$ is one of the sites in $V(Y_F)$. But since R' is exactly the axis of rotation for $\Phi_{F',F}$, we have $d(\gamma_\zeta(t), \omega) = d(\gamma_\zeta(t), \widehat{\omega})$ for p.s.e. t. In the Voronoi diagram $V(Y_F)$ on F, we know that:

$$d(\gamma_\zeta(t), \nu) \leq d(\gamma_\zeta(t), \widehat{\omega}) \text{ for p.s.e. } t$$

So we conclude

$$d(\gamma_\zeta(t), \nu) \leq d(\gamma_\zeta(t), \omega) \text{ for p.s.e. } t$$

as we desired. ■

Lemma 101. *Suppose one of the following holds*

(i) $(\omega, F') \in \mathcal{I}$, $(\Phi_{F',F}(\omega), F) \notin \mathcal{I}$.

(ii) $\omega \in \mathrm{src}_F$ *and* $(\omega, F) \notin \mathcal{I}$, *or*

(iii) $\omega \in \mathrm{src}_{F'}$ *and* $(\omega, F') \notin \mathcal{I}$.

and $\gamma_\zeta(t)$ lies in the Voronoi cell of ω for p.s.e. t. Then:

$$d(\nu, \gamma_\zeta(t)) \leq d(\omega, \gamma_\zeta(t)) \text{ for p.s.e. } t$$

Proof. These scenarios are grouped together because they may be treated with the same calculation, though the calculation must be set up in slightly different ways.

First consider case (i). Then, $V(Y_{F'}, \omega)$ contains $\gamma_\zeta(t)$ for p.s.e. t, so ω sees $\mathrm{rint}(R')$ through $V(Y_{F'})$ and so (ω, F', R') is a potential event. Since (ν, F, R') is minimal, we conclude $(\nu, F, R') \preceq (\omega, F', R')$.

Now consider cases (ii) and (iii). For convenience, let $\widetilde{F}$ denote either F or F' so that $(\omega, \widetilde{F})$ is the event in question. By Proposition 96, since $(\omega, \widetilde{F}) \notin \mathcal{I}$ there exists a potential event P such that $P \preceq (\omega, \widetilde{F})$. Since (ν, F, R') is the minimal potential event, then, $(\nu, F, R') \preceq P \preceq (\omega, \widetilde{F})$.

Let E denote either the potential event (ω, F', R') or the event $(\omega, \widetilde{F})$, depending on the case. No matter the case, we have $(\nu, F, R') \preceq E$, which means exactly one of the following holds:

(a) $r(\nu, F, R') < r(E)$

(b) $r(\nu, F, R') = r(E)$ and the closest points of (ν, F, R') and E are different.

(c) $r(\nu, F, R') = r(E)$ and the closest points of (ν, F, R') and E are the same, namely x. In this case, $\angle(\nu, F, R')$ is weakly lexicographically smaller than $\angle(E)$.

We first consider (a) and (b). If (a) holds, we can immediately conclude $d(\nu, x) < d(\omega, x)$. If (b) holds, recall that closest points are unique, so that the distance from ω to x must be strictly greater than the distance from ω to its closest point. So in case (b) we also have $d(\nu, x) < d(\omega, x)$. Let $\varepsilon = d(\omega, x) - d(\nu, x) > 0$. Then there exists a metric space ball $B \subseteq S$ of radius $\varepsilon/2$ centered at x. For p.s.e. t, $\gamma_\zeta(t) \in B$, and for any $p \in B$ we may write:

$$\begin{aligned} d(\omega, p) &\geq |d(\omega, x) - d(x, p)| \qquad \text{(Reverse Triangle Inequality)} \\ &\geq d(\omega, x) - d(x, p) \\ &\geq (d(\nu, x) + \varepsilon) - \varepsilon/2 = d(\nu, x) + \varepsilon/2 \\ &\geq d(\nu, x) + d(x, p) \geq d(\nu, p) \end{aligned}$$

Thus $d(\nu, \gamma_\zeta(t)) \leq d(\omega, \gamma_\zeta(t))$ for p.s.e. t.

It remains to prove the bound when (c) holds. Because $(\nu, F, R') \preceq E$ and $r(\nu, F, R') = r(E)$, it must be the case that $\angle(\nu, F, R')$ is weakly lexicographically smaller than $\angle(E)$. Let ζ' denote the minimal jet frame associated with E and let

$$V_\omega = \begin{cases} V(Y_{F'}, \omega) & \text{in case (i)} \\ V(\mathrm{src}_{\widetilde{F}}, \omega) & \text{in cases (ii),(iii)} \end{cases} \tag{3}$$

Then ζ' is a minimal jet frame at x, along $R' \cap V_\omega$, with outer support vector $(\omega - x)$.

We assumed that $\gamma_\zeta(t)$ lies within both V_ω and $V(Y_F, \nu)$ for p.s.e. t. The angle sequence formed by ζ with outer support vector $(\nu - x)$ is weakly lexicographically smaller than the angle sequence formed by ζ' with the outer support vector $(\nu - x)$. By Corollary 74, this means:

$$d(\gamma_\zeta(t), \nu) \leq d(\gamma_{\zeta'}(t), \omega) \text{ for p.s.e. } t$$

On the other hand, ζ is also a jet frame along $V_\omega \cap R'$ and we said that ζ' is the minimal jet frame along $V_\omega \cap R'$ with respect to the outer support vector $(\omega - x)$. Hence, by another application of Corollary 74

$$d(\gamma_{\zeta'}(t), \omega) \leq d(\gamma_\zeta(t), \omega) \text{ for p.s.e. } t$$

By transitivity then:

$$d(\gamma_\zeta(t), \nu) \leq d(\gamma_\zeta(t), \omega) \text{ for p.s.e. } t$$

which is what we needed to show. ■

Proof of Theorem 98, Item 1. Since $\nu \in \mathrm{src}_F$ and $V(Y_F, \nu)$ contains a point in $\mathrm{rint}(R')$ (in other words, ν sees R' through $V(Y_F)$), it is clear that there exist geodesics from v to points in $\mathrm{rint}(R')$ that unfold to yield $\nu' \in T_{F'}$. We must justify why one of those geodesics is a shortest path.

Because ζ is a jet frame at $x \in R'$ along $V(Y_F, \nu) \cap R'$, $\gamma_\zeta(t) \in \mathrm{rint}(V(Y_F, \nu) \cap R')$ for p.s.e. t. For each such t, we have a geodesic from v to $\gamma_\zeta(t)$ that unfolds to produce ν'. The length of this geodesic is the same as the length of the line $[\nu, \gamma_\zeta(t)]$ in T_F, which is the same as the length of the line $[\nu', \gamma_\zeta(t)]$ in $T_{F'}$. Thus, if we could prove for p.s.e. t that

$$d(\nu', \gamma_\zeta(t)) = d(\nu, \gamma_\zeta(t)) \leq \min_{\omega \in \mathrm{src}_F \cup \mathrm{src}_{F'}} d(\omega, \gamma_\zeta(t))$$

we could conclude $\nu' \in \mathrm{src}_{F'}$ (and that ν' is obtained by unfolding a shortest path from v to $\gamma_\zeta(t)$ for any p.s.e. t).

Recall that by Proposition 79, for any two distinct $\omega, \omega' \in \mathrm{src}_F \cup \mathrm{src}_{F'}$, we have

- $d(\omega, \gamma_\zeta(t)) \le d(\omega', \gamma_\zeta(t))$ for p.s.e. t or
- $d(\omega', \gamma_\zeta(t)) < d(\omega, \gamma_\zeta(t))$ for p.s.e. t.

Consequently, it suffices to prove the bound

$$d(\nu, \gamma_\zeta(t)) \le d(\omega, \gamma_\zeta(t)) \text{ for p.s.e. } t$$

for only those source images ω such that $\gamma_\zeta(t)$ lies in the Voronoi cell of ω for p.s.e. t. Consider such an $\omega \in \mathrm{src}(F) \cup \mathrm{src}(F')$. Then either:

- $(\omega, F) \in \mathrm{src}_F$ with $(\omega, F) \in \mathcal{I}$,
- $(\omega, F') \in \mathrm{src}'_F$ with $(\omega, F') \in \mathcal{I}$ and $(\Phi_{F',F}(\omega), F) \in \mathcal{I}$, or
- One of the following holds:
  - (i) $(\omega, F') \in \mathcal{I}$, $(\Phi_{F',F}(\omega), F) \notin \mathcal{I}$.
  - (ii) $\omega \in \mathrm{src}_F$ and $(\omega, F) \notin \mathcal{I}$, or
  - (iii) $\omega \in \mathrm{src}_{F'}$ and $(\omega, F') \notin \mathcal{I}$.

These are exactly the cases covered by Lemmas 99, 100, and 101, so the desired bound is proved in each case. ■

Inspecting the details of the proof of Theorem 98, Item 1, we can extract several useful facts.

Corollary 102. *For p.s.e. t:*

- $\nu' \in \mathrm{src}_{F'}$*, unfolding from a shortest path from v to $\gamma_\zeta(t)$, and*
- $\gamma_\zeta(t) \in V(\mathrm{src}_F, \nu)$ *(note $V(\mathrm{src}_F, \nu)$ is a smaller Voronoi cell than $V(Y_F, \nu)$).*
- ζ *is not just a minimal jet frame at x along $V(Y_F, \nu) \cap R'$, but actually a minimal jet frame at x along $V(\mathrm{src}_F, \nu)$ as well. (In either case, the outer support vector is $\nu - x$.)*
- *The radii and angle sequence of (ν, F, R') are identical to those of (ν', F')*

With these facts recorded, it remains to prove Item 2 of the theorem.

Proof of Theorem 98, Item 2. To show $\mathcal{I} \cup \{(\nu', F')\}$ is an order ideal, it suffices to argue that given any event (ω, G) not in $\mathcal{I} \cup \{(\nu', F')\}$, either (ν', F') and (ω, G) are incomparable[16] or else $(\nu', F') \prec (\omega, G)$.

So suppose $(\omega, G) \notin \mathcal{I} \cup \{(\nu', F')\}$ and that (ω, G) and (ν', F') are comparable. Since $(\omega, G) \notin \mathcal{I}$, Proposition 96 implies that there exists a potential event (ω', G', R') so that $(\omega', G', R') \preceq (\omega, G)$. Since (ν, F, R') is a minimal potential event, we have $(\nu, F, R') \preceq (\omega', G', R')$. By Corollary 102, we know that the radius and angle sequence of (ν', F') and (ν, F, R') are identical. So we conclude that $(\nu', F') \preceq (\omega, G)$, which is enough to conclude $(\nu', F') \prec (\omega, G)$, because we know (ν', F') and (ω, G) are comparable and not equal. ■

[16] Recall that this means (ω, G) and (ν', F') are distinct but have identical radii and angle sequences.

6 The Source Unfolding Algorithm

In this section, we present an algorithm for calculating $\mathrm{src}(v, S)$, based upon Theorem 98. It will be much simpler to specify the algorithm if we establish the following conventions:

- A **set data structure** represents finite sets of points or other data. Such a data structure supports two operations:

  - Given datum d, we may **add** d to the set.
  - Given datum d, we may **query** whether d is in the set.

  We can also iterate over the data structure, provided we make no assumptions about the order of iteration.

- A **convex polytope data structure** is capable of representing any subset of $\mathbb{R}^d$ defined by intersecting finitely many half-spaces. It supports two operations.

  - Given a point $p \in \mathbb{R}^d$, we may **query** if p is in the set.
  - Given two convex polytope data structures, we may **intersect** them to create a new convex polytope.

  Internally, we may represent such a data structure as a finite collection of inner-product inequalities $\{\langle \cdot, n_j \rangle \leq c_j\}_{j=1}^{N}$.

- An **ordered data structure** provides a way of storing a finite set with an ordering. It supports all of the operations of a set data structure as well as an **extract minimum** operation which removes (and returns) from the data structure an element that is minimal with respect to the ordering. (If more than one element is minimal, we make no assumptions on which one is chosen.)

- A **Voronoi diagram** V is a data structure supporting two operations:

  - Given a point $p \in \mathbb{R}^d$, **add** p as a site in the Voronoi diagram.
  - Given a site p in the diagram, report the **Voronoi cell** associated with that site (thus, this operation reports a convex polytope data structure).

- Finally, we have a data structure that represents S. We assume the data structure records the face-lattice of S, so that given a d-face F and a $(d-1)$-face R with $R \subseteq F$, we may query what other d-face F' contains R. We also assume this data structure records how S is embedded in $\mathbb{R}^{d+1}$, so that we can easily access the rotation maps $\Phi_{F,F'} : T_F \to T_{F'}$.

All of the above data structures are standard and available in existing libraries. A description of how to implement most of these data structures can be found in [5].

6.1 The Source Image Construction Algorithm

In this section, we present Algorithm 2 for calculating $\mathrm{src}(v, S)$ and argue its correctness.

Algorithm 2: Source Image Construction

Input: A polytope boundary S and a source point v.
Output: The collection of source images $\mathrm{src}(v, S)$.

—Initialize all data structures.—
for *each d-face* $F \subseteq S$ **do**
    Construct an empty set data structure Y_F and an empty Voronoi diagram V_F.
$F_0 :=$ the source face containing v.
AddVoronoiSite(V_{F_0}, v)
Construct a set data structure $\mathcal{I} := \{(v, F_0)\}$.
Construct an empty, ordered data structure $\mathcal{E}$, ordered by $\preceq$.
for *each* $(d-1)$*-face* $R \subseteq F_0$ **do**
    $F' :=$ the d-face sharing R with F
    AddElement($\mathcal{E}$, MakePotentialEvent$(\Phi_{F_0,F'}(v), F_0, R)$)[a]

while $\mathcal{E} \neq \varnothing$ **do**
    —Find and process the minimal potential event.—
    $(\nu, F, R') :=$ ExtractMinimum$(\mathcal{E})$
    $F' :=$ the d-face sharing R' with F
    $\nu' := \Phi_{F,F'}(\nu)$
    AddElement$(Y_{F'}, \nu')$
    AddVoronoiSite$(V_{F'}, \nu')$

    —Update the collection of potential events.—
    for *each potential event of the form* $(\omega, F', R) \in \mathcal{E}$ **do**
        if VoronoiCell$(V_{F'}, \omega) \cap \mathrm{rint}(R) = \varnothing$ **then**
            RemoveElement($\mathcal{E}$, (ω, F', R))
    for *each* $(\omega, G, R) \in \mathcal{E}$*, where* $G \cap F' = R$ **do**
        if $\Phi_{G,F'}(\omega) = \nu'$ **then**
            RemoveElement($\mathcal{E}$, (ω, G, R))
    for *each* $(d-1)$*-face* $R \subseteq F'$ *such that* VoronoiCell$(V_{F'}, \nu')$ $\cap\, \mathrm{rint}(R) \neq 0$ **do**
        $G :=$ the d-face sharing R with F
        if $\Phi_{F',G}(\nu') \notin Y_G$ **then**
            AddElement($\mathcal{E}$, MakePotentialEvent(ν', F', R))

return Y

[a] Note the procedure MakePotentialEvent must calculate calculate a minimal jet frame using Algorithm 1.

Correctness of Algorithm 2. We must argue the Source Image Construction Algorithm terminates in finite time and constructs $\mathrm{src}(v, S)$. We claim that after any iteration of the while loop in the algorithm

- the data structure $\mathcal{I}$ holds an order ideal and
- the data structure $\mathcal{E}$ contains the potential events corresponding to $\mathcal{I}$.

We prove this by induction.

As a base case, before the first iteration through the loop, the set $\mathcal{I}$ is initialized to hold only the source event, so that $\mathcal{I}$ is trivially an order ideal. For any $(d-1)$-face R that belongs to the source face, the source point trivially sees R through the (necessarily one-celled) Voronoi diagram V_{F_0}. Moreover, since $\mathcal{I}$ contains only the source event, there is no possibility that v folds over R to produce a point ν' that already appears as part of an event in $\mathcal{I}$. Thus, the potential events are precisely those triples (v, F, R) where R is a $(d-1)$-face of F_0.

For the induction step, suppose $\mathcal{I}$ is an order ideal held in the data structure $\mathcal{I}$ at the beginning of the while-loop and $\mathcal{E}_{\mathcal{I}}$ is the set of potential events of $\mathcal{I}$, held in the data structure $\mathcal{E}$. Then by Theorem 98, we know that processing the minimal potential event (ν, F, R') yields an event (ν', F') and that $\mathcal{I}' := \mathcal{I} \cup \{(\nu', F')\}$ is an order ideal. It remains to analyze whether the changes we make to $\mathcal{E}$ yield the set of potential events $\mathcal{E}_{\mathcal{I}'}$ for $\mathcal{I}'$.

The only difference between $\mathcal{I}$ and $\mathcal{I}'$ is the event (ν', F'), so the only difference between the Voronoi diagrams induced by these ideals will be on face F'. This explains why we add the Voronoi site ν' to $V_{F'}$. Next, consider the difference between $\mathcal{E}_{\mathcal{I}}$ and $\mathcal{E}_{\mathcal{I}'}$. From the definition of a potential event, we see that $(\omega, G, R) \in \mathcal{E}_{\mathcal{I}} \setminus \mathcal{E}_{\mathcal{I}'}$ for one of the following reasons:

- R is a $(d-1)$-face that G shares with F', such that $\Phi_{G,F'}(\omega) = \nu'$. Informally speaking, since we have discovered the source image ν', we have no need for any more potential events that would give us ν' if processed.
- $(\omega, G, R) = (\omega, F', R)$, and ω can see R through the Voronoi diagram $\mathcal{I}$ induces on F' but it cannot see R through the Voronoi diagram $\mathcal{I}'$ induces on F'. By adding (ν', F') to the order ideal we have in some sense pushed the horizon further away from the source and reduced the size of the undiscovered region. (ω, G, R) represented a portion of the horizon that has moved, and so we no longer need it.

Since $\mathcal{I}$ and $\mathcal{I}'$ differ by a single element, the definition of a potential event dictates that the potential events in $\mathcal{E}_{\mathcal{I}'} \setminus \mathcal{E}_{\mathcal{I}}$ arise as a result of the event (ν', F'). In other words, $\mathcal{E}_{\mathcal{I}'} \setminus \mathcal{E}_{\mathcal{I}}$ consists of potential events of the form (ν', F', R), where R is a $(d-1)$-face of F' such that ν' sees R through the Voronoi diagram $\mathcal{I}'$ induces on F', and ν' does not fold over R to yield a point already belonging to an event in $\mathcal{I}'$.

With this characterization, we can see that we may update the data structure $\mathcal{E}$ to store the potential events of $\mathcal{I}'$ by removing the elements of $\mathcal{E}_{\mathcal{I}} \setminus \mathcal{E}_{\mathcal{I}'}$ and adding the elements of $\mathcal{E}_{\mathcal{I}'} \setminus \mathcal{E}_{\mathcal{I}}$, as appears in the pseudocode. Thus, at the end of each iteration of the while-loop, $\mathcal{E}$ holds the potential events that correspond to the order ideal held in the data structure $\mathcal{I}$, completing the induction step.

On each iteration of the while loop, $\mathcal{I}$ grows by one event. We know that $\mathcal{I} \subseteq \mathrm{src}(v, S)$, and that $\mathrm{src}(v, S)$ is finite, so after some finite number of iterations, $\mathcal{I} = \mathrm{src}(v, S)$. At this point, all of the source images have been discovered, so that any remaining potential events will be removed from $\mathcal{E}$. This ends the while loop and the algorithm.

■

Once one has used the above algorithm to calculate the source images for a given S and v, calculating the source unfolding is straightforward. First, one uses the source images to calculate the Voronoi diagram on each face (equivalently, one could reuse the Voronoi diagrams V_F calculated by the algorithm above). Then, one uses the face-lattice data structure describing S to determine how the Voronoi cells on different faces are connected along the $(d-1)$-faces. Finally, for each face-sequence of connected Voronoi cells, one calculates the sequence of unfolding maps for that face sequence and uses it to unfold the Voronoi cells into a star-shaped region in $\mathbb{R}^d$.

Remark 103. We have not analyzed the time complexity of the unfolding algorithm (which includes the calculation of $\mathrm{src}(v, S)$). Miller and Pak have shown the algorithm is polynomial in the number of source images. This is not a completely satisfying analysis, because the total number of source images cannot be easily estimated from the initial data, a polytope P and the source point.[17] Arguing that the number of source images is polynomial

[17]The size of $\mathrm{src}(v, S)$ can be trivially estimated as being at most $N \cdot N!$, where N is the number of d-faces of S, but this does not give us a useful estimate of the algorithm's true complexity.

in a meaningful measurement of the input data (such as, for example, the number of d-faces of S) is a major open problem.

7 Index of Notation

$\mathrm{aff}(F)$	affine span of F (Definition 1)
$\mathrm{rint}(F), \mathrm{rbdy}(F)$	relative interior (resp. relative boundary) of F (Definition 6)
$\overline{X}$	the closure of a set X in a topological space
S_{d-1}, S_{d-2}	union of the dimension $(d-1)$ (resp $(d-2)$) faces of S (Remark 10)
P	a convex, d-dimensional polytope contained in R^{d+1}
S	boundary of a convex, d-dimensional polytope
v	the source point
d_M	distance associated with a metric space M
L	length functional, defined on paths
$\langle \cdot, \cdot \rangle$	Euclidean inner product
T_pM	tangent space at point p on manifold M
T_F	another notation for $\mathrm{aff}(F)$
$\angle(X)$	angle sequence associated with X (Definition 81)
C_v	cut locus associated with source point v (Definition 33)
$\Phi_{F,F'}$	rotation map taking $\mathrm{aff}(F)$ to $\mathrm{aff}(F')$, where F and F' are d-faces sharing a $(d-1)$-face
$C_{\mathcal{F},v}$	traversable cone (Definition 42)
$[a, b]$	line segment between points a and b (Notation 22)
$V(Y)$	Voronoi diagram with sites Y (Definition 48)
$V(Y, y)$	Voronoi cell associated with a point y, in the Voronoi diagram $V(Y)$ (Definition 48)
$\partial V(Y)$	Voronoi boundary of Voronoi diagram $V(Y)$ (Definition 48)
src_F	Source images of face F (Definition 37)
ζ	jet frame (Definition 68)
$\gamma_\zeta(t)$	trajectory associated with ζ (Definition 66)
$J_\zeta(t)$	infinitesimal displacement associated with ζ (Definition 66)
p.s.e.	positive, small enough (Notation 67)
$r(R, \omega)$	radius (Definition 81)
$\rho(R, \omega)$	closest point(Definition 81)
$\mathcal{I}$	an order ideal (Definition 88)
$\mathcal{E}_\mathcal{I}$	the potential events of $\mathcal{I}$ (Definition 94)
$\preceq, \prec$	precedes relation (Definition 71, 86)
(ω, F)	an event (Definition 85)
(ω, F, R')	a potential event (Definition 94)

References

[1] Ezra Miller and Igor Pak. Metric combinatorics of convex polyhedra: Cut loci and nonoverlapping unfoldings. In *Twentieth Anniversary Volume: Discrete and Computational Geometry*, pages 1–50. Springer, 2009.

[2] Gunter M Ziegler. Lectures on polytopes (graduate texts in mathematics). 2001.

[3] Patrick Fitzpatrick. *Advanced Calculus*, volume 5. Amer. Mathematical Society, 2006.

[4] John M Lee. *Riemannian Manifolds: An Introduction to Curvature*, volume 176. Springer Verlag, 1997.

[5] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT press, 2001.

[6] Branko Grünbaum. *Convex Polytopes*. Springer, 2003.