CONVEXITY IS OVERRATED: SHORTEST PATHS ON 4-POLYTOPES ARE HARD TO COMPUTE

IGOR PAK

ABSTRACT. We prove that computing geodesic distance on the boundary of a rational convex 4-polytope is NP-hard. We then construct rational 4-polytopes with two rational points joined by exponentially many shortest paths of equal rational length, disproving several conjectures along the way. For the same problem, we also prove SQRTSUM-hardness under deterministic polynomial-time Turing reductions. By a separate argument, we show that the square of the geodesic distance between distinct rational boundary points of a rational convex polytope is a totally positive multiquadratic number.

1. INTRODUCTION

1.1. **Foreword.** We start with two classical facts in discrete and computational geometry. First, on the boundaries of convex 3-polytopes, there is an elegant polynomial-time approach to computing geodesic distances via a nonoverlapping unfolding [SS86]. While the resulting algorithm extends to higher dimensions, its analysis does not [MP08]. Second, in $\mathbb{R}^3$ with open convex polyhedral obstacles, computing a shortest path is NP-hard [CR87].

This leaves a natural *conundrum*: are the boundaries of convex 4-polytopes structurally closer to those of 3-polytopes, or to $\mathbb{R}^3$ with obstacles removed? For over twenty years, we believed that the answer was the former. Only recently did we become universal contrarians [Pak20]. As it turns out, the right answer was the latter, just as the title suggests (cf. §11.1). The main result is our Theorem 1.1 below. Curiously, convexity still has a role to play, see Theorem 3.1 in the next section.

1.2. **Main theorem.** Our main result is NP-hardness of exact geodesic distance on the boundary of a rational convex 4-polytope. A *rational convex 4-polytope* is a bounded intersection of rational closed halfspaces with affine span $\mathbb{R}^4$. A supporting hyperplane meets the polytope with the polytope on one side, and its intersection with the polytope is a *face*. The 3-dimensional faces are *facets*. The *relative interior* of a face is its interior in its affine span. A *facet-interior point* belongs to the relative interior of a facet. For boundary points v, x, their *intrinsic distance* $\operatorname{dist}_{\partial P}(v, x)$ is the infimum of the Euclidean lengths of paths in ∂P joining v and x.

Theorem 1.1 (Main theorem)**.** *Given a rational convex 4-polytope $P \subset \mathbb{R}^4$, defined by rational linear inequalities, two rational boundary points $v, x \in \partial P$, and a rational threshold $\tau \in \mathbb{Q}_{>0}$, deciding whether*

$$\operatorname{dist}_{\partial P}(v, x) \ \leq \ \tau$$

is NP-hard, even when v and x lie in the relative interiors of two facets. All input numbers are encoded in binary.

More precisely, let PARTITION *be the decision problem that asks whether m positive integers can be split into two groups with equal sums. Such an instance is mapped in polynomial time to a polytope with exactly $18m + 8$ facets, two rational facet-interior points, and a rational threshold. The distance is at most the threshold precisely when the* PARTITION *instance has a solution.*

Date: November 18, 2026.

Recall that PARTITION is NP-complete [Kar72]; it is one of "six basic NP-complete problems" [GJ79, §3.1]. Curiously, the geodesic distance problem is also SQRTSUM-hard under deterministic polynomial-time Turing reductions (Theorem 3.5). We postpone this and the other arithmetic results, and the exact complexity bounds until Section 3. In this section, we present the geometric counting results and their relation to earlier work. The proof of the Main Theorem 1.1 is quite technical and is given in Sections 4–6. See §4.1 and §4.2 for an outline.

1.3. **Geometric counting result.** To emphasize how different the geometry becomes beyond dimension three, we record an exponential multiplicity result for shortest paths between fixed endpoints on the boundary of a convex polytope $P \subset \mathbb{R}^4$.

In dimension four, a *ridge* is a two-dimensional face, an *edge* is a one-dimensional face, and a *vertex* is a zero-dimensional face. Facets are *adjacent* when they share a ridge. This ridge is their hinge in an unfolding. It has codimension one in the boundary, whereas edges and vertices have codimension at least two.

A *globally shortest path* γ between points v, x attains the intrinsic (geodesic) distance in the boundary ∂P. Obviously, the length $|\gamma|$ is not necessarily rational even if both P and v, x are rational. Path γ is *regular* if it meets only facet interiors and relative interiors of ridges, and *simple* if it has no self-intersections. It is *strictly monotone* in a coordinate if that coordinate strictly increases along the path. Such paths are always simple.

A *facet gallery* is the ordered list of adjacent facets through which a path is to pass. To *unfold* a gallery, successively rotate the supporting hyperplanes around their common ridges until all pieces are represented in one Euclidean three-dimensional space. These are the folding and sequential unfolding maps of Miller and the author (MP from this point on), see [MP08, Defs 1.5–1.6] and discussion in §4.1.

A regular shortest path develops to a straight segment, and its initial point develops to an *unfolded copy* in the affine span of the last facet. A *formal unfolded image* is such a copy calculated from a prescribed gallery before validity or global minimality has been established. MP reserve the term *source image* for a copy arising from a globally shortest path [MP08, Def. 2.3]. We make the same distinction. A formal unfolded image need not correspond to a boundary path. We require the folded segment to cross the prescribed ridges, in order and in their relative interiors. Global minimality is a separate condition.

Theorem 1.2. *For every integer $m \geq 1$, there are a rational convex 4-polytope R_m with exactly $18m + 8$ facets and two rational facet-interior points $v_m, x_m \in \partial R_m$ joined by at least 2^m distinct globally shortest paths. These paths are regular and simple, strictly monotone in one coordinate, and visit each facet at most once. Their common length is rational. Their facet galleries are pairwise distinct, and a single facet of R_m has at least 2^m distinct source images of v_m, obtained by unfolding prefixes of these paths.*

In one big swoosh, the theorem disproves three Conjectures 9.2–9.4 in [MP08] and answers both versions of Rote's face-sequence question in the negative [Mit09, Problem 3]. See §1.4 for further details, historical context, and consequences for the MP algorithm. The theorem is proved in Section 8.

This result is in sharp contrast with the $O(n)$ bound on the number of shortest paths between two points on 3-polytopes, see e.g. [AAOS97]. Furthermore, it was shown by Mount [Mou90] that the number of possible facet galleries on 3-polytopes is $O(n^4)$. In both cases, n is the number of vertices of the polytope.

1.4. **Historical background.** The relevant earlier work concerns exact algorithms on polyhedral surfaces, hardness in obstacle domains, and source unfolding in higher dimension. We refer to Bose–Maheshwari–Shu–Wuhrer [BMSW11] for a short but detailed survey of algorithms on polyhedral surfaces.

Distance propagation. Dijkstra's algorithm settles graph distances in increasing order [Dij59]. Its geometric counterpart, *continuous Dijkstra*, propagates a *wavefront*, the points at a common distance from a source. A *shortest-path map* partitions the domain according to the combinatorial descriptions of shortest paths. For planar polygonal obstacles, Mitchell developed a subquadratic continuous Dijkstra algorithm [Mit96], and Hershberger–Suri obtained $O(n \log n)$ time and space, where n is the total number of obstacle vertices [HS99]. More recently, Wang gave an optimal $O(n + h \log h)$-time and $O(n)$ space algorithm, assuming a triangulation of the free space is given, where h is the number of obstacles [Wang23].

Polyhedral surfaces. Sharir–Schorr gave an early exact algorithm for shortest paths on a convex polyhedron [SS86]. Mitchell–Mount–Papadimitriou developed continuous Dijkstra for possibly nonconvex polyhedral surfaces in $\mathbb{R}^3$ [MMP87]. Chen–Han obtained a different algorithm with quadratic time and linear space [CH96], while Mount developed efficient representations of subdivisions bounded by geodesic arcs [Mou87]. For a convex polyhedron with n edges, Schreiber–Sharir constructed an implicit shortest-path map in $O(n \log n)$ time and space [SS08]. Schreiber extended this bound to certain classes of nonconvex polyhedra [Sch10]. In [Wang25], Wang gave recent results for two-point queries and edge-sequence enumeration on convex polyhedral surfaces. These algorithms work with surfaces of intrinsic dimension 2, meaning that the surface itself is 2-dimensional and is embedded in $\mathbb{R}^3$.

Hardness and convexity. Canny–Reif [CR87, Thm 2.3.2] proved NP-hardness of Euclidean shortest paths among polyhedral obstacles in $\mathbb{R}^3$. Their reduction explicitly controls coordinate encoding and geometric separation [CR87, §§2.1–2.5], and is a key to our work. In [MP08, §9], Miller and the author obtained hardness on nonconvex polyhedral manifolds of intrinsic dimension 3 and discussed compact examples without boundary in $\mathbb{R}^4$.

Exact arithmetic. Choi–Sellen–Yap gave a detailed bit-complexity analysis [CSY97], and Sellen–Choi–Yap studied the separation of competing lengths [SCY00] for the Euclidean shortest path problem in $\mathbb{R}^3$ with obstacles. By contrast, in the *real-arithmetic model* exact real-number operations are counted at unit cost, and such operation counts need not bound binary complexity, see e.g. [BCSS98]. See also [DHK20, §3] for a quick overview of models in the context of geodesics on convex polyhedra. The relation to sums of square roots and arithmetic-circuit sign tests is discussed in Section 3, together with the new restrictions on convex-boundary distance values. The exact surface algorithms above and the binary NP-hardness theorem address distinct computational models.

Source and star unfoldings. The *cut locus* is the closure of points with more than one shortest path from a fixed source. Cutting there gives the *source unfolding*. On a convex three-dimensional polyhedron, the *star unfolding*, first introduced by Alexandrov (see [MP08, §8.4]), instead cuts selected shortest paths from the source to the vertices. Aronov–O'Rourke proved its nonoverlap [AO92]. Agarwal–Aronov–O'Rourke–Schevon used it for edge-sequence enumeration, *geodesic diameter* (the maximum intrinsic distance over all endpoint pairs), and two-point distance queries [AAOS97].

In [MP08, Thm 3.5], Miller and the author extended the source unfolding to convex polyhedral boundaries in arbitrary dimension. The distinction from star unfolding in higher dimension is discussed in [DO07, §26.3.4] and [DO25, §7.7]. The set where the boundary is not locally Euclidean can then have positive dimension, and cutting shortest paths to this set need not give enough cuts. In a curious ad hoc result, Yamagishi studied the four-cube's *farthest-point map*, which assigns to a source its most distant boundary points [Yam25].

Counting conjectures. For convex 3-dimensional polyhedra, Sharir proved a polynomial bound on shortest-path edge sequences [Sha87], and Mount obtained the tight worst-case bound [Mou90]. An *edge sequence* records the edges crossed by a shortest path. This count ranges over all endpoint pairs. In fixed dimension, MP's geodesic Dijkstra algorithm is polynomial in both the facet count

and the maximum number of source images per facet [MP08, Thm 6.5]. Their Conjectures 9.2, 9.3, and 9.4 ask for polynomial bounds on, respectively, the total number of source images from a fixed source, the number of shortest paths between one fixed pair, and the number of shortest-path facet sequences over all endpoint pairs. Rote's question at Dagstuhl asks for polynomial face-sequence counts on a convex 4-polytope, over all endpoint pairs or with the source fixed [Mit09, Problem 3]. Theorem 1.2 disproves all three conjectures and answers both versions of Rote's question in the negative: the exponential facet-sequence count already occurs with both endpoints fixed.

Theorem 1.2 also gives an exponential lower bound of $2^{(n-8)/18}$ source-image insertions for the complete source-unfolding Algorithm 6.1 in [MP08], even in its exact real-arithmetic model. This consequence, discussed in Remark 8.2, is separate from the NP-hardness of computing one distance in the binary input model.

Relative approximation. Although computing the exact geodesic distance is computationally hard, getting an approximation can be done in polynomial time in any fixed dimension. Formally, a $(1+\eta)$-approximation returns a path whose length is at most $(1+\eta)$ times the shortest distance, where $\eta > 0$ is the relative error. Classical obstacle-avoidance approximations in $\mathbb{R}^3$ are due to Papadimitriou [Pap85] and Clarkson [Cla87]. Har-Peled constructs approximate single-source maps for polyhedral surfaces and for obstacle avoidance in $\mathbb{R}^3$ [HP99b]. See §11.6 and Appendix A for more on approximation of the geodesic distance.

1.5. **Paper structure.** Section 3 states the complexity and arithmetic results. This section has no proofs and can be viewed as the second introduction, for the reader interested in computational rather than geometric aspects of the problem. Sections 4–6 give the construction and prove NP-hardness (Theorem 1.1). Section 8 proves the counting theorem (Theorem 1.2), building on the unfolding lemma (Lemma 7.1) in Section 7. Section 9 proves the arithmetic results and exact upper bounds, and Section 10 proves SqrtSum-hardness. We conclude in Section 11 with assorted final remarks discussing further complexity background related to the geodesic distance. In the Appendix A, we give a concise self-contained proof of polynomial-time approximation of the geodesic distance; this is not a new result and is included for completeness. Appendix B gives the obstacle-distance example showing that Theorem 3.1 does not extend to this setting.

2. Basic definitions and notation

We use $\mathbb{N} = \{0, 1, 2, \ldots\}$, $\mathbb{Q}_{>0} = \{q \in \mathbb{Q} : q > 0\}$, and $[n] = \{1, \ldots, n\}$. Whenever convenient, we use bold letters for vectors and ordinary letters for scalars, e.g. we write $\boldsymbol{r} = (r_1, r_2)$ for a 2-dimensional vector. We use $\mathbf{0}$ to denote zero vector. We write $\boldsymbol{v} \cdot \boldsymbol{w}$ to denote the scalar product, so e.g. $\boldsymbol{v} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^2$ denotes the squared length of the vector.

We assume the reader is familiar with standard complexity classes, and include only an informal and very brief overview below. We refer to [AB09, Gol08] for a thorough introduction to computational complexity and context behind these complexity classes.

We use P for deterministic polynomial time, NP for decision problems with polynomial-length certificates verifiable in polynomial time, and PSPACE for polynomial space. The class coNP consists of the complements of problems in NP. A polynomial-time *many-one reduction* computes one target instance with the same answer. A problem is NP-*hard* if every problem in NP has such a reduction to it, and NP-*complete* if it is also in NP.

For a decision problem A, the notation P^A means deterministic polynomial time with an oracle answering instances of A. Both the number of queries and their total binary length are polynomially bounded. Later queries may depend on earlier answers. The notation NP^A additionally permits a nondeterministic polynomial-length guess before the oracle computation is carried out. By contrast, a deterministic polynomial-time Turing reduction from B to A uses no such guess:

the reduction algorithm itself makes polynomially many adaptive queries to the oracle for A. We therefore write this stronger deterministic statement as $B \in \mathsf{P}^A$.

The class PP consists of problems decided by a polynomial-time randomized machine whose acceptance probability is greater than $1/2$ on yes instances and at most $1/2$ on no instances. The *counting hierarchy* is $\mathsf{C}_0\mathsf{P} = \mathsf{P}$, $\mathsf{C}_{j+1}\mathsf{P} = \mathsf{P}\mathsf{P}^{\mathsf{C}_j\mathsf{P}}$, and $\mathsf{CH} = \bigcup_{j\geq 0} \mathsf{C}_j\mathsf{P}$. The class BPP uses bounded error, say at most $1/3$ on either answer. In the Merlin–Arthur class MA, a polynomial-length certificate is checked by a randomized polynomial-time verifier with bounded error.

3. Computational complexity and arithmetic properties

We state the arithmetic restrictions on geodesic distance values, the exact complexity bounds, and SqrtSum-hardness. The proofs appear in Sections 9 and 10.

3.1. Complexity setup.

For fixed $d \geq 2$, the decision problem $\textsc{GeoDist}_d$ asks whether

$$\operatorname{dist}_{\partial P}(s,t) \leq k,$$

where $P \subset \mathbb{R}^d$ is a bounded, full-dimensional convex polytope specified by rational linear inequalities, $s, t \in \partial P$ are rational, and k is rational. Numerators and denominators are written explicitly in binary, and running time counts bit operations. Note that in every fixed dimension, a rational vertex description can also be converted to a polynomial-size rational facet description.

Theorem 1.1 proves many-one NP-hardness through binary Partition: given positive integers $A_0, \ldots, A_{m-1}$, decide whether $\sum_{j=0}^{m-1} \sigma_j A_j = 0$ for signs $\sigma_j \in \{-1, 1\}$. The construction has polynomial binary size and a nonzero decision gap of polynomial encoding length, although the gap can be exponentially small. The reduction does not imply NP membership, or *strong* NP-hardness (hardness with the numerical data encoded in unary).

3.2. Arithmetic structure of the distance.

A real algebraic number is *multiquadratic* if it belongs to a field

$$\mathbb{Q}\big(\sqrt{q_1}, \ldots, \sqrt{q_t}\big), \qquad q_1, \ldots, q_t \in \mathbb{Q}_{>0}.$$

Equivalently, a multiquadratic number can be formed inside a field obtained from $\mathbb{Q}$ by adjoining finitely many square roots of positive rational numbers. The definition concerns the ambient field, not the particular shape of one displayed radical expression. For example, $\sqrt{2} + \sqrt{3}$ is multiquadratic, while $\sqrt[4]{2}$ and $\sqrt{2 + \sqrt{2}}$ are not. We will use only explicitly displayed multiquadratic fields, so no separate symbol for the field of all such numbers is needed.

The *conjugates* of an algebraic number are the roots of its minimal polynomial over $\mathbb{Q}$. An algebraic number is *totally positive* if all its conjugates are strictly positive real numbers. For example, $1 + \sqrt{2}$ is not totally positive, while $2 + \sqrt{2}$ is.

A *radical circuit* is a finite acyclic graph of arithmetic gates at vertices, with binary rational constants and operations $+, -, \times, \div, \sqrt{\ }$. Each gate uses values from earlier gates, and a computed value may be used more than once. Square roots are principal real roots. The circuit is *valid* when every divisor is nonzero and the input to every square-root gate is nonnegative. Its size is the binary length of its gate list and constants. These conventions apply to every radical circuit below.

Theorem 3.1 (Arithmetic properties of geodesic lengths)**.** *Let $d \geq 2$, let $P \subset \mathbb{R}^d$ be a full-dimensional rational convex polytope with m facets, and let $a_1, \ldots, a_m \in \mathbb{R}^d$ be rational outward facet normals. Fix distinct rational points $s, t \in \partial P$, and let $D := \operatorname{dist}_{\partial P}(s,t)$. Define the real multiquadratic field*

$$K_P := \mathbb{Q}\big(\sqrt{a_1 \cdot a_1}, \ldots, \sqrt{a_m \cdot a_m}\big).$$

Then $D^2 \in K_P$, so D^2 is a real multiquadratic number. Additionally, the number D^2 is totally positive. Furthermore, given a shortest facet gallery, a radical circuit for D can be constructed in polynomial time in the binary input length, for every fixed d.

For endpoints on lower-dimensional faces, the supplied gallery may be a limiting gallery with zero-length pieces, as constructed in §9.3. The algebraic conclusions hold in every ambient dimension $d \geq 2$. In particular, we conclude that

$$\sqrt[4]{2}, \qquad \sqrt{1+\sqrt{2}}, \qquad \sqrt{3+\sqrt{2}+\sqrt{3}}$$

cannot be geodesic distances between two rational boundary points on a rational convex polytope. We prove the theorem in §9.4. We note that Theorem 3.1 fails for the shortest-distance problem in $\mathbb{R}^3$ with obstacles studied in [CR87]; see Appendix B for an explicit two-obstacle counterexample.

Remark 3.2. In a somewhat related investigation, Bajaj [Baj85] proved that contact coordinates of spatial polyhedral shortest paths need not be expressible by radicals, already with two encountered obstacle edges. In his example, one contact coordinate has a degree-20 minimal polynomial with Galois group S_{20}. By contrast, for convex polyhedral boundaries, our unfolding argument instead gives $D^2 \in K_P$, where K_P is a multiquadratic extension of $\mathbb{Q}$. Recall that every such extension K_P is Galois, with the Galois group isomorphic to a finite direct product of copies of $\mathbb{Z}/2\mathbb{Z}$.

3.3. **Exact decision and arithmetic upper bounds.** A facet gallery is *nonrepeating* if no facet occurs more than once. A shortest boundary path has a nonrepeating gallery with at most one straight piece in each facet, see Lemma 9.1 and [MP08, proof of Lemma 2.4]. This is a polynomial-length combinatorial description. Its crossing points need not be rational, however, and a short gallery is not by itself a polynomial-time binary certificate for a length threshold, see also [MP08, Rem. 6.8].

The class $\exists\mathbb{R}$ consists of problems reducible to the existence of real solutions of polynomial equations and inequalities with integer coefficients. Canny proved $\exists\mathbb{R} \subseteq \mathsf{PSPACE}$ [Can88]. For a polytope with m facets, introduce m path segments with real vertices and nonnegative length variables. For each segment separately, a disjunction over the m facets requires both endpoints to lie in one facet; convexity then places the whole segment in that facet. Together with the quadratic length constraints, these m disjunctions give a polynomial-size existential-real formula. Zero-length segments pad any shorter path. The full construction is given in §9.1.

Proposition 3.3. *For every fixed $d \geq 2$, we have:*

$$\textsc{GeoDist}_d \ \in \ \exists\mathbb{R} \ \subseteq \ \mathsf{PSPACE}.$$

The problem PosSLP asks whether an arithmetic circuit over the integer constants $0, 1$, with operations $+, -, \times$, has positive output. Binary integer constants can also be supplied with only polynomial overhead, because a binary integer can be built by repeated doubling and addition. Consequently, an $\mathsf{NP}^{\text{PosSLP}}$ computation may guess a polynomial-size combinatorial object and then ask, in polynomial time, for the signs of polynomial-size integer arithmetic circuits arising from that guess. Allender, Bürgisser, Kjeldgaard-Pedersen and Miltersen proved that $\text{PosSLP} \in \mathsf{CH}$ [ABKM09].

Theorem 3.4. *For rational input in every fixed ambient dimension $d \geq 2$,*

$$\textsc{GeoDist}_d \in \mathsf{NP}^{\text{PosSLP}} \subseteq \mathsf{CH} \subseteq \mathsf{PSPACE}.$$

If the recent bound $\text{PosSLP} \in \mathsf{BPP}^{\mathsf{PP}}$ claimed in [Mei26] is used, this sharpens to

$$\textsc{GeoDist}_d \in \mathsf{NP}^{\text{PosSLP}} \subseteq \mathsf{MA}^{\mathsf{PP}} \subseteq \mathsf{C_2P}. \tag{3.1}$$

Section 9 supplies the arithmetic lemma and complete gallery verifier, including equality cases and endpoints lying on lower-dimensional faces. Its arithmetic uses rational data and square roots of rational squared facet-normal norms. The relevant sign methods are due to Tiwari [Tiw92], with more general algebraic-expression bounds in [B+09]. The second display is conditional on Meiburg's preprint. Ordinary NP membership remains open.

Concurrent and independent preprints [OAI26, Mei26, BCJS26] claim a fixed-level CH bound for $\exists\mathbb{R}$ and, more generally, for existential–universal real sentences with arithmetic-circuit input. Meiburg claims the sharper bound $\exists\mathbb{R} \subseteq \mathsf{BPP}^{\mathsf{C_3P}} \subseteq \mathsf{C_4P}$ [Mei26]. Assuming this claimed inclusion, Proposition 3.3 gives a fourth-level bound for GeoDist$_d$. The unfolding verifier instead gives the second-level bound in (3.1), conditional on the cited result for PosSLP.

3.4. **Square-root-sum hardness.** In the *Sums of Square Roots problem*, denoted SqrtSum, nonnegative integers $A_1, \dots, A_n, B$ are given in binary, and the question is whether

$$\sum_{i=1}^{n} \sqrt{A_i} \;\le\; B\,. \tag{3.2}$$

The integers A_i under the square roots are the *radicands*. This problem was introduced by O'Rourke [O'Rou81], and has been intensely studied in recent years.

The standard arithmetic route is

$$\textsc{SqrtSum} \in \mathsf{P}^{\textsc{PosSLP}} \subseteq \mathsf{CH}, \tag{3.3}$$

see [Tiw92, ABKM09]. Meiburg's recent preprint proposes the sharper bounds

$$\textsc{SqrtSum} \in \mathsf{BPP}^{\mathsf{PP}} \cap \mathsf{P}^{\mathsf{NP}^{\mathsf{PP}}} \subseteq \mathsf{C_2P}, \tag{3.4}$$

and the same two oracle upper bounds for PosSLP [Mei26]. Neither statement establishes membership in P or in ordinary NP. Separation estimates, meaning lower bounds on a nonzero difference between two algebraic values, for fixed radicands such as [EHS24], do not give a uniform polynomial-time comparison algorithm when the radicands are input.

Prescribed rational polygonal path lengths give the classical geometric connection to SqrtSum, including the verification of Euclidean tours [ABKM09, §1.4]. The reduction is given in §9.5. Our hardness theorem goes from the arithmetic comparison problem to shortest boundary distance.

Theorem 3.5 (Square-root-sum hardness)**.** *In the binary rational input model,*

$$\textsc{SqrtSum} \in \mathsf{P}^{\textsc{GeoDist}_4}. \tag{3.5}$$

The reduction uses polynomially many exact distance queries on rational convex 4-polytopes with rational facet-interior endpoints.

The geometric ingredient is the following exact distance identity. Its role is to turn each square root $\sqrt{A_i}$ into a prescribed additive contribution to one boundary distance, while keeping every supporting hyperplane and both endpoints rational.

Lemma 3.6 (Exact geometric encoding)**.** *Given positive integers $A_1, \dots, A_n$ and vectors $v_i \in \mathbb{Z}^4$ satisfying $v_i \cdot v_i = A_i$, let $A = \max_{1\le i\le n} A_i$ and $\lambda = 1/(nA)$. In deterministic polynomial time one can construct a rational convex 4-polytope P with exactly $4n+8$ facets and rational facet-interior points s, t such that*

$$\operatorname{dist}_{\partial P}(s,t) = 2n + 1 + \lambda \sum_{i=1}^{n} \sqrt{A_i}. \tag{3.6}$$

All supporting coefficients and endpoint coordinates have polynomial binary encoding length.

The lemma is proved in Section 10. The reduction first needs, for each radicand A_i, an integer vector $v_i \in \mathbb{Z}^4$ with $v_i \cdot v_i = A_i$. Section 10.2 shows how polynomially many GeoDist$_4$ oracle queries recover such vectors; these vectors are then fed into Lemma 3.6. The proof of Theorem 3.5 is given in §10.7. It does not assume that SqrtSum belongs to NP: the reduction is a deterministic polynomial-time Turing reduction.

The two arithmetic comparisons should not be conflated. A rational polygonal path gives an explicit square-root sum, whereas a general facet unfolding gives a circuit of radical expressions whose expansion can have exponentially many terms. Our upper bound is $\mathsf{NP}^{\text{PosSLP}}$, not $\mathsf{NP}^{\text{SqrtSum}}$. These two observations do not identify the two arithmetic problems. In particular, the argument gives neither a reduction from GeoDist$_d$ to SqrtSum nor any separation between the corresponding complexity classes; it only explains why the upper-bound verifier naturally uses PosSLP rather than SqrtSum.

4. The mirrored three-lens construction

The construction uses a chain of local two-route configurations, called lenses, that are defined below. We group three lenses into a block and compare route choices by a quadratic score. Reflection of the chain favors matching block choices in the two halves. A small rational perturbation makes the score depend on the corresponding Partition residual, namely the signed sum of the input integers. Section 5 relates this score to boundary-path length.

4.1. **Geometric outline.** We first describe the geometry and notation. The lower boundary of the polytope contains a chain of lenses. Their route choices determine the quadratic score introduced in §4.2 and formalized in §4.3. The same geometry is used for the hardness and counting constructions.

Throughout the geometric proofs, $|x|$ and $\|x\|$ denote the Euclidean norm, $x \cdot y$ the Euclidean inner product, and I the identity matrix of the indicated dimension. Matrix norms are the induced Euclidean operator norms. A *support* of a piecewise-affine convex graph is an affine function whose graph lies below it and meets it. A support is *active* at a point when equality holds there.

Write the ambient coordinates as $(\boldsymbol{r}, z, w) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}$. The coordinate z runs along the construction and is called *longitudinal*. The vector $\boldsymbol{r}$ records the two *transverse* coordinates. The coordinate w is the height of the lower boundary graph. The construction uses a small positive rational parameter $\varepsilon > 0$ to make the slopes of that graph small. The polytope is thin in the transverse directions, and the paths of interest remain in a narrower neighborhood of $\boldsymbol{r} = \mathbf{0}$.

For the hardness construction in this section and Sections 5–6, let m be the number of input integers. The complete mirrored chain has $k = 6m$ individual lenses, indexed by $i \in [k]$. The m three-lens blocks in each half are indexed by $j = 0, \dots, m-1$. Thus a complete lens word is $s = (s_1, \dots, s_k) \in \{-1, 1\}^k$. The counting construction in Section 8 reuses one half of this notation and explicitly resets $k = 3m$ there.

We call a particular four-facet neighborhood of an edge e_i a *lens*. Its facets occur around that edge in the cyclic order $D_{i-1}, A_i^+, D_i, A_i^-$. The facets D_{i-1} and D_i are the incoming and outgoing *baseline facets*, while A_i^+ and A_i^- are its two *side facets*. The two routes used in the construction have the galleries (D_{i-1}, A_i^+, D_i) and (D_{i-1}, A_i^-, D_i). We encode the choice of A_i^+ by the *lens sign* $s_i = +1$ and the choice of A_i^- by $s_i = -1$. Neighboring lenses share a baseline facet. Each lens has a longitudinal position t_i and a transverse *slope vector* $\boldsymbol{L}_i$. The side supports have transverse gradients $\varepsilon \boldsymbol{L}_i$ and $-\varepsilon \boldsymbol{L}_i$. The exact local supporting functions are given in Section 5.

A *block* is a group of three consecutive lenses, rather than a separate face or polytope. An ordered list of lens signs is a *sign word*. The two *uniform* triples are $(+,+,+)$ and $(-,-,-)$, and the remaining triples are *mixed*. When block j, $0 \le j \le m-1$, is uniform, we write $\sigma_j \in \{-1, 1\}$ for its common sign; the resulting vector $\sigma = (\sigma_0, \dots, \sigma_{m-1})$ is a *block word*. The slope vectors

within each block are chosen so that replacing a mixed triple by a uniform one improves the quadratic score by a definite amount. The interaction of a block with other blocks is recorded by two weighted sums of its slope vectors, called its *endpoint moments*; their formulas are given in §4.3. A rational fractional-linear change of the lens positions and slope vectors, defined explicitly in §4.4, makes interactions between different blocks small without changing the internal score comparisons. Figure 4.1 illustrates a lens and a block. In the figure, u and ζ are schematic local base coordinates; the free edge direction and the graph height are suppressed.

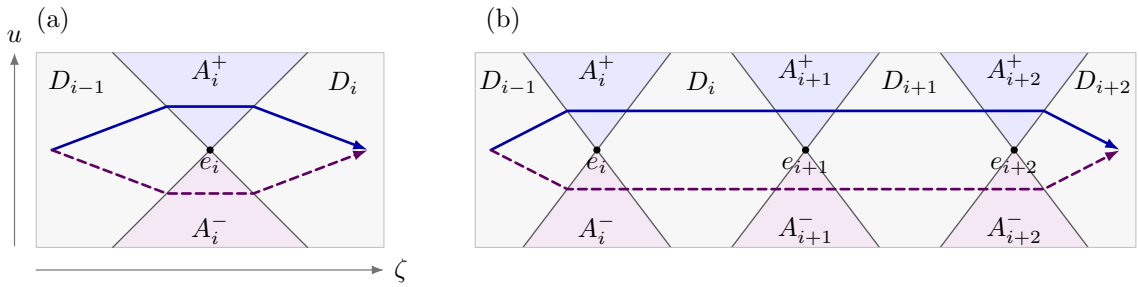


FIGURE 4.1. A lens and a three-lens block. (a) A lens has two possible galleries, through A_i^+ or A_i^-. (b) A block consists of three consecutive lenses; the solid and dashed routes show the two uniform choices.

One chain assigns a distinct scalar label to every block word. A longitudinally reflected second chain imposes a quadratic loss determined by the difference of the two labels, so the score is maximal only when the labels, and hence the words, agree. The labels and the loss are made explicit in §4.2. To encode binary PARTITION, we then insert the input integers with opposite signs in the two chains. On agreeing words, all linear terms cancel and the remaining word-dependent penalty is a negative multiple of the square of the signed sum specified by that block word.

The realization step controls *all* competing shortest paths. Shortest-path regularity permits unfolding across successive ridges. For fixed endpoints and a fixed gallery, the unfolded straight segment determines at most one path. A shortest path cannot leave and later return to a convex facet [MP08, Cor. 1.4, Lemma 1.7, and proof of Lemma 2.4]. These facts reduce the comparison to finitely many galleries. Small slopes give finite control of direction changes under ridge unfolding, and a Pythagorean estimate compares boundary lengths with the finite score.

For the shortest paths we use the *Cayley parametrization of rotations* [HLW06, §IV.8.3], originally due to Cayley [Cay46], to make an exact rational adjustment. The selected unfolded copies and reflected target images lie on two parallel lines. Perpendicular joining segments then have exactly equal lengths, while the score comparison excludes other shorter galleries.

Remark 4.1. The hardness construction already has exponentially many maximum-score words for suitable PARTITION instances. This does not imply exponentially many shortest paths, since the realization in Section 5 controls individual lengths only through an $O(\varepsilon^6)$ error. The purpose of Section 8 is to remove this error at the relevant words by an exact unfolding construction.

The square-root-sum construction uses a different local piece: a strip of three successive facets whose two ridges are parallel. The incoming and outgoing facets admit rational orthonormal frames, meaning ordered orthonormal bases of their direction spaces, while the middle facet has a rational normal of prescribed squared norm. Its development is an exact translation by a rational vector times that square root. A reflected pair cancels transverse components, and rational assembly makes these translations add along one developed line. Ordered supports and a common length-preserving development prove global shortestness. Unlike the PARTITION comparison, this distance identity has no remainder.

4.2. **Proof idea.** We now use the notation from the preceding subsection. The starting point is to replace a discrete binary choice by the two routes through one lens. When the lower graph is thin, the length of a path has a *small-slope expansion*, namely the Taylor expansion of its Euclidean length in powers of the transverse slope. After the side choices $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$ are fixed, the first word-dependent term is the quadratic score

$$h(s) \,:=\, \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} G(t_i, t_j)\, s_i s_j \, (\boldsymbol{L}_i \cdot \boldsymbol{L}_j)\,, \quad \text{where} \quad G(u,v) \,:=\, \min(u,v) \,-\, uv\,. \tag{4.1}$$

Here G is the *Dirichlet Green kernel* for $-d^2/dz^2$ on $[0,1]$, see e.g. [SH11]. Namely, for fixed v, the function $G(\,\cdot\,, v)$ is piecewise linear, vanishes at 0 and 1, and its slope drops by one at v. This jump property is exactly what will later convert the side choice at a lens into a jump of the slope of a polygonal transverse chain.

After integration by parts, the total length correction is therefore governed by the finite quadratic form $h(s)$ displayed above. In this sense the geometric problem is reduced to designing a quadratic energy on the lens signs. Three consecutive lenses are chosen so that the two uniform triples are separated from the six mixed triples; a block then behaves as one binary variable. The fractional-linear relocation defined in §4.4 preserves this internal comparison while making interactions between different blocks small.

Reflection supplies the second ingredient. The first half labels a block word σ by

$$b_\sigma \,:=\, \sum_{j=0}^{m-1} \sigma_j b_j\,,$$

where the positive weights $b_0, \ldots, b_{m-1}$ are *superincreasing*, meaning that

$$b_j \,>\, \sum_{i=0}^{j-1} b_i \qquad \text{for all} \qquad 1 \le j \le m-1.$$

If two block words first differ at their largest differing index j, the contribution $2b_j$ at that index is larger than the total possible contribution of all earlier indices. Hence the two signed sums cannot agree, and b_σ uniquely determines σ. For two block words σ, τ, let $H_0(\sigma, \tau)$ denote the unperturbed score of the reflected pair. The construction gives

$$H_0(\sigma,\tau) \;=\; 2F_* \,-\, \frac{1}{2}(b_\sigma - b_\tau)^2,$$

where F_* is independent of σ and τ. Thus a maximizing pair must carry the same word in the two halves.

Let $A_0, \ldots, A_{m-1}$ be the positive integers of the Partition instance and define its signed residual at a block word by

$$A_\sigma \,:=\, \sum_{j=0}^{m-1} \sigma_j A_j\,.$$

Opposite small perturbations turn the surviving word-dependent term into a negative multiple of A_σ^2. This is the mechanism behind the reduction.

This quadratic sign optimization has a long history in computational complexity, and famously appears in the standard Ising model interpretation. The preference for the two aligned triples is analogous to a ferromagnetic coupling, while the squared residual A_σ^2 is exactly the usual Ising encoding of number partitioning, see [Luc14, §2.1]. For classical complexity results on Ising ground-state problems, see e.g. [Bar82].

At the end, the construction produces exponentially many candidate geodesic paths between two fixed points whose lengths are nearly identical. Their small length differences are engineered

to encode a PARTITION instance: the perturbation associated with a path depends on the corresponding partition residual A_σ. The shortest candidate is therefore the one whose encoded residual gives the best perturbation term. Deciding whether a zero residual occurs is exactly the original PARTITION decision problem. Hence an algorithm that determines the exact shortest path length for these instances would decide PARTITION, which proves the claimed NP-hardness.

Finally, let us mention that the sequential-unfolding viewpoint comes from [MP08], while the use of local route choices to encode discrete data is familiar from geometric shortest-path hardness constructions in [CR87]. The mirrored three-lens gadget we employ below is specifically tailored to the convex setting here.

4.3. **Sign words and finite scores.** We use three levels of notation. The block index is $j = 0, \dots, m-1$, while the full mirrored chain uses the lens index $i \in [k]$, where $k = 6m$. A sign $s_i \in \{-1, 1\}$ records which side of lens i is used. Once a block is forced to be uniform, its three lens signs are represented by one block sign σ_j. The geometric data of lens i are its longitudinal position t_i and transverse slope vector $\boldsymbol{L}_i$. The score h measures the leading change in path length, while the endpoint moments $\boldsymbol{B}, \boldsymbol{C}$ summarize how one group of lenses interacts with another.

Let $A_0, \dots, A_{m-1}$ be positive integers, $m \geq 1$, and let

$$S := \sum_{j=0}^{m-1} A_j\,, \qquad T := 10000\,, \qquad a := \frac{16}{5}\,,$$

$$q_j := T^{j-m} \quad (0 \leq j \leq m-1), \qquad q := T^{-m}\,, \qquad \kappa := \frac{q^6}{10^6 S}\,.$$

Here S is the total input weight and T is a fixed large base. The number $q_j = T^{j-m}$ is the geometric scale assigned to block j. As j increases by one, q_j is multiplied by the large factor T, so $q_0 < q_1 < \ldots < q_{m-1}$. We also write $q = T^{-m}$ for a single convenient lower scale that is no larger than any q_j and can therefore be used in uniform estimates. The constant $a = 16/5$ normalizes the endpoint moments, and the much smaller parameter κ carries the actual PARTITION integers as a perturbation.

For each input integer A_j, $0 \leq j \leq m-1$, we place one three-lens block in the first half and a reflected copy of that same block in the second half. Each input integer therefore contributes six lenses in total: three before the midpoint and three after it. With m input integers, the complete construction has $2m$ blocks and $k = 6m$ individual lenses. The scales q_j determine where successive blocks sit and how strongly they interact, while κ controls the much smaller perturbation that carries the integers A_j. All parameters below are rational. We write $\boldsymbol{e}_1, \boldsymbol{e}_2 \in \mathbb{R}^2$ for the standard basis vectors of the transverse plane.

For $i = [k]$, let $t_i \in (0,1)$ be the longitudinal lens positions and $\boldsymbol{L}_i \in \mathbb{R}^2$ the transverse slope vectors. Define $G(u,v)$ and $h(s)$ as in (4.1), where the *lens sign* $s_i \in \{-1,1\}$ selects one of the two side facets at lens i. A *sign word* is the ordered list of these signs. For fixed t, the function $z \mapsto G(z,t)$ is a tent: it is $(1-t)z$ for $z \leq t$ and $t(1-z)$ for $z \geq t$. It vanishes at 0 and 1, and its slope decreases by one at t. The quantity $h(s)$ is the *score* of the word. Its diagonal terms are independent of the signs, while its pair terms compare the combined choices at distinct lenses. Larger scores will give shorter paths to leading order, with an error bounded in Section 5.

For later comparisons, it is important that a score gap is defined from *distinct score values*, not from distinct words. If $\mathcal{W}$ is a nonempty finite set of sign words, list its distinct score values as

$$v_1 > v_2 > \ldots > v_r, \qquad \{v_1, \dots, v_r\} = \{h(s) : s \in \mathcal{W}\}.$$

In particular, $H_{\mathcal{W}} := \max_{s \in \mathcal{W}} h(s) = v_1$. We define the *score gap* on $\mathcal{W}$ by

$$\operatorname{gap}_{\mathcal{W}}(h) := \begin{cases} \min_{1 \leq j < r}(v_j - v_{j+1}), & r \geq 2, \\ +\infty, & r = 1, \end{cases}$$

and the *top score gap* by

$$\operatorname{gap}_{\mathcal{W}}^{\mathrm{top}}(h) := \begin{cases} v_1 - v_2, & r \geq 2, \\ +\infty, & r = 1. \end{cases}$$

Hence tied words share one score value and do not create a zero gap; in particular, the "second-best" score means the second-largest *distinct* score value. The convention $+\infty$ in the one-value case only records that there is no competing distinct score value. In the proof below, whenever a finite positive lower bound is needed we name the relevant comparison explicitly: the mixed-block repair gap, the unequal-copy matching gap, or the arithmetic PARTITION gap.

For any consecutive group of lenses, let $I \subseteq [k]$ be its lens-index set and define its *endpoint moments* by

$$\boldsymbol{B}(s) := \sum_{i \in I} t_i s_i \, \boldsymbol{L}_i \,, \qquad \boldsymbol{C}(s) := \sum_{i \in I} (1 - t_i) s_i \, \boldsymbol{L}_i \,.$$

These weighted vector sums have a geometric interpretation. The polygonal comparison chain for the same group is

$$\boldsymbol{R}_s(z) := \sum_{i \in I} G(z, t_i) \, s_i \, \boldsymbol{L}_i$$

has initial slope $\boldsymbol{C}(s)$ and final slope $-\boldsymbol{B}(s)$. The same sums also determine interactions between groups: if every lens of one group precedes every lens of another, then $G(t_i, t_j) = t_i(1 - t_j)$, so their complete cross contribution to the score is exactly $\boldsymbol{B}_{\text{earlier}} \cdot \boldsymbol{C}_{\text{later}}$.

We record a uniform triple in block j, $0 \leq j \leq m - 1$, by one *block sign* σ_j. A block word therefore records one sign for each three-lens block, while the corresponding lens word repeats that sign three times inside each block. Consequently, a word of m block signs determines the $3m$ lens signs in that half of the construction.

4.4. **Three-lens blocks in the first half.** We specify one three-lens pattern and relocate a copy for each block. The relocation preserves every internal score but rescales the two endpoint moments in opposite directions. This lets us retain the preference for uniform triples while controlling interactions between different blocks.

The reference positions are

$$(u_1, u_2, u_3) := \left(\frac{1}{10}, \frac{1}{2}, \frac{9}{10} \right).$$

For block j, $0 \leq j \leq m - 1$, let

$$\eta_j := \frac{9}{8} \, a q_j^2 = \frac{18}{5} \, q_j^2 \,,$$

and use reference slope vectors

$$\boldsymbol{l}_{j,1} := (-1 - \eta_j \,, 3), \quad \boldsymbol{l}_{j,2} := \left(\frac{6}{5} \,, \frac{6}{5} \right), \quad \boldsymbol{l}_{j,3} = \left(3 + \frac{\eta_j}{9} \,, -1 \right).$$

Define also scalar reference perturbation coefficients

$$(p_{j,1}, p_{j,2}, p_{j,3}) := \left(-\frac{A_j}{8q_j} \,, 0 \,, \frac{9A_j}{8q_j} \right).$$

For the local lens index $\nu \in \{1, 2, 3\}$, set

$$t_{j,\nu} := \frac{q_j^2 u_\nu}{1 + (q_j^2 - 1) u_\nu} \,, \qquad \boldsymbol{L}^0_{j,\nu} := \frac{1 + (q_j^2 - 1) u_\nu}{q_j} \, \boldsymbol{l}_{j,\nu} \,, \qquad p^0_{j,\nu} := \frac{1 + (q_j^2 - 1) u_\nu}{q_j} \, p_{j,\nu} \,.$$

The first formula is the new longitudinal position of lens ν in block j; the second rescales its two-dimensional slope vector by the factor paired with that change of position; the third applies the same rescaling to the scalar perturbation coefficient. The superscript 0 means that these are

the relocated data before the small PARTITION perturbation is added. Writing the common denominator explicitly avoids introducing another symbol that would later conflict with the baseline supports D_j. The first-half slope vector is

$$\boldsymbol{L}^{+}_{j,\nu} \;=\; \boldsymbol{L}^{0}_{j,\nu} \;+\; \kappa p^{0}_{j,\nu}\, \boldsymbol{e}_2 .$$

In the second half put a lens at $1 - t_{j,\nu}$ with slope vector

$$\boldsymbol{L}^{-}_{j,\nu} \;=\; \boldsymbol{L}^{0}_{j,\nu} \;-\; \kappa p^{0}_{j,\nu}\, \boldsymbol{e}_2 .$$

The second-half lenses are listed in the reverse of first-half order. All first-half positions are less than $1/2$, and the blocks occur in increasing order of j. The superscripts $+$ and $-$ here distinguish the two perturbed halves. Each individual lens still has its own two side choices, specified by a lens sign.

The relocation above is a *fractional-linear* change of position together with a rescaling of the slope vector. The parameter $c > 0$ controls the relocation: τ_c moves the longitudinal position and α_c rescales the corresponding slope vector. Explicitly,

$$\tau_c(u) = \frac{cu}{1 + (c-1)u}, \qquad \alpha_c(u) = \frac{1 + (c-1)u}{\sqrt{c}}.$$

These two changes are paired so that

$$G\big(\tau_c(u), \tau_c(v)\big) \cdot \alpha_c(u)\,\alpha_c(v) \;=\; G(u, v).$$

The displayed identity says that each pairwise Green-kernel term in the quadratic score is unchanged when the position u is replaced by $\tau_c(u)$ and its slope vector is simultaneously multiplied by $\alpha_c(u)$. Therefore replacing $(u, \boldsymbol{L})$ by $(\tau_c(u), \alpha_c(u)\boldsymbol{L})$ preserves every within-block score. The same substitution changes the two endpoint moments by $\boldsymbol{B} \mapsto \sqrt{c}\,\boldsymbol{B}$, $\boldsymbol{C} \mapsto \boldsymbol{C}/\sqrt{c}$. This identity is an exact algebraic identity: multiplying the displayed rational expressions verifies it directly. For block j we choose $c := q_j^2$. Since $q_j > 0$, we have $\sqrt{c} = q_j$. Substituting this value of c into the two relocation formulas gives $\tau_c(u_\nu) = t_{j,\nu}$ and $\alpha_c(u_\nu)\boldsymbol{l}_{j,\nu} = \boldsymbol{L}^{0}_{j,\nu}$. Thus the previously defined quantities $t_{j,\nu}$ and $\boldsymbol{L}^{0}_{j,\nu}$ are exactly the relocated position and rescaled slope of the νth reference lens in block j; they are not independent parameters. No limiting argument is involved. This operation changes the data of one lens block; it is not a projective transformation of the whole polytope and is not asserted to preserve arbitrary lengths.

We call the endpoint moments obtained after setting $\kappa = 0$ the *unperturbed endpoint moments*. For the positive uniform triple in block j, they are

$$\boldsymbol{B}_j \;=\; b_j \boldsymbol{e}_1\,, \qquad \boldsymbol{C}_j \;=\; -b_j \boldsymbol{e}_1 + d_j \boldsymbol{e}_2\,, \qquad b_j \;=\; a q_j\,, \quad d_j \;=\; a/q_j\,.$$

The negative uniform triple reverses both moments because all three lens signs change sign. Before the fractional-linear relocation, the positive reference triple has endpoint moments $(a, 0)$ and $(-8\eta_j/9, a)$. After relocation these become the displayed vectors $\boldsymbol{B}_j$ and $\boldsymbol{C}_j$. The scalar perturbation coefficients have moments

$$B_{p,j} \;=\; A_j, \qquad C_{p,j} \;=\; 0\,.$$

Here the subscript p refers to the scalar perturbation coefficients $p_{j,\nu}$. The quantities $B_{p,j}$ and $C_{p,j}$ are therefore one-dimensional weighted sums of those coefficients; they are not the two-dimensional vector moments $\boldsymbol{B}_j$ and $\boldsymbol{C}_j$. Their purpose is to insert the input integer A_j into the second transverse coordinate without altering the first-coordinate label that is used to match the two halves. The unperturbed within-block score of either uniform triple is

$$h_j \;=\; \frac{36}{25} + \frac{8}{225}\eta_j + \frac{2}{45}\eta_j^2 \;=\; \frac{36}{25} + \frac{16}{125}q_j^2 + \frac{72}{125}q_j^4\,. \tag{4.2}$$

The corresponding scalar quadratic contribution is

$$h_{p,j} \;=\; \frac{9}{160}\,\frac{A_j^2}{q_j^2}\,. \tag{4.3}$$

Both formulas come from substituting the three reference positions and the three corresponding slope or perturbation coefficients into the same three-by-three quadratic form defining the block score.

4.5. **Reflection favors matching block choices.** Reflection makes the interaction of the two halves depend on the difference of their signed block labels. Those labels distinguish all block words, so the maximum among uniform triples occurs exactly when the two halves make matching choices.

First set $\kappa = 0$. Write h_0 for the score and $\boldsymbol{B}_0, \boldsymbol{C}_0$ for the unperturbed endpoint moments of one half. Let $\mathcal{H}_\kappa$ denote the score of the full two-half word with perturbation parameter κ. Let s record the lens signs in the first half. For the second half, let t record the corresponding reflected signs but list them using the original first-half block order rather than their reversed geometric order. With this common indexing, the full unperturbed score is

$$\mathcal{H}_0(s,t) = h_0(s) + h_0(t) + \boldsymbol{B}_0(s)\cdot\boldsymbol{B}_0(t). \tag{4.4}$$

This is because $G(1-u,1-v) = G(u,v)$ and, when $u,v<1/2$, $G(u,1-v) = uv$.

Geometrically, the reflected half is encountered in the reverse block order. Our symbol τ nevertheless records its signs using the original first-half block indices, so that block j in the first half can be compared directly with its reflected copy. If one instead writes the signs in their actual left-to-right order along the path, the same abstract block word appears reversed. For example, the abstract word $(+,-)$ gives the first-half lens-sign list $(+,+,+,-,-,-)$, while its matching reflected copy appears along the second half as $(-,-,-,+,+,+)$.

For uniform block words $\sigma,\tau\in\{-1,1\}^m$, define

$$b_\sigma \;:=\; \sum_{j=0}^{m-1}\sigma_j b_j\,, \qquad d_\sigma \;:=\; \sum_{j=0}^{m-1}\sigma_j d_j\,, \qquad F_* \;:=\; \sum_{j=0}^{m-1} h_j + \frac{1}{2}\sum_{j=0}^{m-1} b_j^2\,.$$

Here b_σ is a scalar label of the block word, while F_* is independent of the word. The positive sequence $b_j = aq_j$ is superincreasing by construction, i.e. each term exceeds the sum of all preceding terms. This separation property prevents two distinct signed words from having the same label. Then

$$\boldsymbol{B}_0(\sigma) \;=\; (b_\sigma,0)\,, \quad \boldsymbol{C}_0(\sigma) \;=\; (-b_\sigma,d_\sigma)\,, \quad h_0(\sigma) \;=\; F_* - \frac{1}{2}b_\sigma^2\,.$$

The reflected cross term is $b_\sigma b_\tau$. The quantities d_σ, d_τ belong to the $\boldsymbol{C}_0$-moments and do not enter this cross term. Substituting the two half-scores into (4.4) therefore gives

$$\mathcal{H}_0(\sigma,\tau) \;=\; 2F_* - \frac{1}{2}(b_\sigma - b_\tau)^2\,. \tag{4.5}$$

Distinct sign words have $|b_\sigma - b_\tau| \ge 2aq$. To see this, divide their difference by $2aq$: the result is a nonzero integer $\sum_{j=0}^{m-1} e_j T^j$, $e_j\in\{-1,0,1\}$. Let $r\in\{0,\dots,m-1\}$ be the largest index for which $e_r\neq 0$. The term e_rT^r then has absolute value T^r, while the sum of the absolute values of all lower-order terms is at most $1+T+\cdots+T^{r-1} < T^r$. The leading base-T term is therefore too large to be cancelled by all lower-order terms, so this integer is nonzero. Returning to the original scaling gives $|b_\sigma - b_\tau| \ge 2aq$, and the square in (4.5) shows that unequal copies lose at least $2a^2q^2$. We call this positive loss the *matching gap*.

4.6. **Every triple is uniform at a maximum.** We now compare arbitrary lens-sign words with the uniform block words used above. Replacing one mixed triple by a uniform triple increases its internal score. The block scales ensure that this gain exceeds any possible decrease in its interactions with the rest of the chain. This is a comparison of sign words. It does not require changing a boundary path locally while keeping its other segments fixed.

At $\eta = 0$ the three reference vectors give internal scores $36/25$ for uniform triples, $33/25$ for four of the six mixed triples, and $24/25$ for the other two, so the minimum internal gap is $3/25$. The sum of the three reference slope-vector norms is less than 10, and the total change in those vectors caused by the parameter η has norm at most $10\eta/9$. Every coefficient $G(u, v)$ in the three-lens score has absolute value at most $1/4$. Expanding the difference between the perturbed and unperturbed quadratic forms and using these two norm bounds shows that the internal score changes by less than 3η whenever $0 < \eta \le 1$. Here $\eta_j \le (18/5)T^{-2} < 10^{-6}$, so replacing a mixed triple by a uniform triple raises its internal score by more than $1/10$.

For every sign triple, its reference moments have norms less than 5, hence its transformed moments satisfy

$$|\boldsymbol{B}_j| \;<\; 5q_j\,, \qquad |\boldsymbol{C}_j| \;<\; 5/q_j\,.$$

Replacing it by a uniform triple changes these moments by less than $10q_j$ and $10/q_j$. Its interaction with the remaining blocks of its own half therefore changes by less than

$$50 \sum_{0\le i<j} \frac{q_i}{q_j} + 50 \sum_{j<i\le m-1} \frac{q_j}{q_i} \;<\; \frac{100}{T-1}\,.$$

The total unperturbed $\boldsymbol{B}$-moment of either half has norm less than $5\sum_{i=0}^{m-1} q_i < 5/(T-1)$. By (4.4), the interaction with the other half changes by less than

$$\frac{50q_j}{T-1} \;\le\; \frac{50}{T(T-1)}\,.$$

The loss in interactions with all other blocks and with the reflected half is smaller than the internal gain. Quantitatively, the net increase in score from replacing any mixed triple by a suitable uniform triple is therefore greater than

$$\frac{1}{10} - \frac{100}{T-1} - \frac{50}{T(T-1)} \;>\; \frac{2}{25}\,. \tag{4.6}$$

4.7. **A uniform perturbation bound.** The arithmetic perturbation must preserve both comparisons: a maximum must still be uniform within each block and must still use matching words in the two halves. We bound its change to the score by one quantity that works for every lens-sign word.

Over all $6m$ lenses, the norms of the unperturbed slope vectors and the absolute values of the scalar perturbation coefficients satisfy

$$K_0 := \sum_{i=1}^{k} |\boldsymbol{L}_i^0| \;<\; \frac{40}{q}\,, \qquad K_p := \sum_{i=1}^{k} |p_i^0| \;<\; \frac{4S}{q^2}\,.$$

For the first inequality, each block contributes less than $10/q_j$, and $\sum_{j=0}^{m-1} 1/q_j < 2/q$. The mirror doubles this sum. For the second, each block contributes at most $5A_j/(4q_j^2)$, before doubling for the mirror. The different signs of the scalar perturbations in the two halves are immaterial to these absolute bounds.

For every lens-sign word, expanding the quadratic form gives

$$|\mathcal{H}_\kappa - \mathcal{H}_0| \;\le\; \frac{\kappa}{4}K_0K_p + \frac{\kappa^2}{8}K_p^2 \;<\; \frac{40\kappa S}{q^3} + \frac{2\kappa^2S^2}{q^4} \;=\; \frac{q^3}{25000} + \frac{q^8}{5\cdot 10^{11}} \;<\; \frac{q^2}{100}\,.$$

A comparison of two words incurs the perturbation error twice. Let $E = \sup_s |\mathcal{H}_\kappa(s) - \mathcal{H}_0(s)| < q^2/100$. If s' is obtained from s by replacing a mixed triple as above, then

$$\mathcal{H}_\kappa(s') - \mathcal{H}_\kappa(s) > \frac{2}{25} - 2E > \frac{2}{25} - \frac{q^2}{50} > 0.$$

Thus a maximizing word cannot contain a mixed triple. Second, when two uniform block words differ, the unperturbed matching loss in (4.5) is at least $2a^2q^2$; allowing the perturbation error for both the unequal pair and a competing equal pair still leaves a positive loss. The loss from any mismatch is therefore larger than the maximum possible gain supplied by the perturbation. Replacing a mismatched pair by the matching pair with the larger unperturbed score strictly increases the total score. Hence every global maximizer must use matching block words $\sigma = \tau$ in the two halves.

4.8. **Exact arithmetic penalty on the surviving words.** After the preceding two steps, every maximizing configuration has been reduced to a matching pair of uniform block words. The only remaining task is therefore an explicit calculation of the score for such a pair; this calculation is where the PARTITION residual appears. The opposite perturbations cancel the terms linear in κ. The remaining word-dependent term is the negative square of the PARTITION residual, so a positive arithmetic gap separates instances with a solution from instances without one.

For matching uniform words, the linear perturbations of the two internal half scores cancel exactly. The scalar perturbation coefficients have no interaction between different blocks because $C_{p,j} = 0$. Their internal contribution is therefore twice the sum in (4.3), multiplied by κ^2. The PARTITION *residual* $A_\sigma = \sum_{j=0}^{m-1} \sigma_j A_j$ is the difference of the sums assigned to the two signs. It vanishes precisely when the word specifies a solution of PARTITION. The two first-half-indexed moments used by the reflection interaction are

$$\boldsymbol{B}_+(\sigma) = (b_\sigma, \kappa A_\sigma), \qquad \boldsymbol{B}_-(\sigma) = (b_\sigma, -\kappa A_\sigma).$$

Their scalar product is $b_\sigma^2 - \kappa^2 A_\sigma^2$. Combining the two half-scores with this reflected cross term gives

$$\mathcal{H}_\kappa(\sigma, \sigma) \;=\; H_* \,-\, \kappa^2 A_\sigma^2\,, \tag{4.7}$$

where the rational constant is

$$H_* \;=\; 2\sum_{j=0}^{m-1} h_j + \sum_{j=0}^{m-1} b_j^2 \,+\, \frac{9\kappa^2}{80} \sum_{j=0}^{m-1} \frac{A_j^2}{q_j^2} \;=\; \frac{72m}{25} \,+\, \frac{1312}{125} \sum_{j=0}^{m-1} q_j^2 \,+\, \frac{144}{125} \sum_{j=0}^{m-1} q_j^4 \,+\, \frac{9\kappa^2}{80} \sum_{j=0}^{m-1} \frac{A_j^2}{q_j^2}\,.$$

When the $6m$ perturbed lenses are viewed as one ordered list, their full score is exactly the generic function $h(s)$ defined in §4.3. The notation $\mathcal{H}_\kappa$ is only a convenient way to write the same score after splitting that single sign word into its first and reflected halves. Combining all preceding bounds gives

$$H \;:=\; \max_{s\in\{-1,1\}^{6m}} h(s) \;=\; H_* \,-\, \kappa^2 \min_{\sigma\in\{-1,1\}^m} \Big(\sum_{j=0}^{m-1} \sigma_j A_j\Big)^2.$$

If the PARTITION instance is a yes instance, there is a sign word with $A_\sigma = 0$, and the formula above gives the maximum value $H = H_*$. If the instance is a no instance, every residual A_σ is a nonzero integer and therefore satisfies $|A_\sigma| \geq 1$. In that case every surviving score is at most $H_* - \kappa^2$, so $H \leq H_* - \kappa^2$. This is the *arithmetic* PARTITION *gap*: it compares the benchmark score H_* with the largest score possible when every residual is nonzero. It is not a best-word-versus-second-best-word gap; several words may have the same residual and hence the same score. In the geometric realization this arithmetic gap is multiplied by ε^4, so the length error must be smaller than the resulting scale $\varepsilon^4\kappa^2$.

The complete-path moments of surviving words are

$$\boldsymbol{B}_{\text{full}} = (0, d_\sigma + \kappa A_\sigma), \qquad \boldsymbol{C}_{\text{full}} = (0, d_\sigma - \kappa A_\sigma).$$

The reversed weights $d_{m-1}, \dots, d_0$ are superincreasing, so the endpoint labels d_σ distinguish the block words. In particular, distinct words with the same arithmetic sum A_σ are not identified.

5. Geometric realization lemma

5.1. **Setup.** We realize the score comparison as a boundary-distance comparison. Every shortest path stays on the lower graph and has small transverse displacement. A quadratic estimate then relates its length to the maximum score H, with a uniform error bound. Section 6 chooses a positive rational value of ε for which this error is smaller than the arithmetic Partition gap κ^2, after scaling by ε^4.

We use coordinates $(r_1, r_2, z, w) = (\boldsymbol{r}, z, w)$, with $\boldsymbol{r} \in \mathbb{R}^2$ transverse, z longitudinal, and w the actual height of the lower boundary. The pair $(\boldsymbol{r}, z)$ consists of the *base coordinates* of this graph. The constants have distinct roles. Let K be a uniform upper bound on the total slope size, and let consecutive lens positions be separated by at least Δ. Let the transverse box have half-width $W\varepsilon$, where $W := k + 1$. The small parameter $\varepsilon > 0$ controls the thinness of the polytope. The number H is the maximum finite sign-word score, while ℓ_ε is the length of the axis path, obtained by setting $\boldsymbol{r} = \boldsymbol{0}$ and letting z run from 0 to 1. Throughout this section, we have lens indices $i, h \in [k]$, and baseline indices j range over $0, \dots, k$, unless bounds are displayed explicitly. In particular, an unqualified sum over i runs over all k lenses.

Lemma 5.1. *Let $0 < t_1 < \ldots < t_k < 1$, and let $\boldsymbol{L}_1, \dots, \boldsymbol{L}_k \in \mathbb{R}^2$ with $\boldsymbol{L}_i \neq \boldsymbol{0}$. Let $t_0 := 0$, $t_{k+1} := 1$, $W := k + 1$, and choose*

$$K \geq \sum_{i=1}^{k} \max\{|\boldsymbol{L}_i|, 1\}, \qquad 0 < \Delta \leq \min\{t_{i+1} - t_i : 0 \leq i \leq k\}.$$

Suppose

$$0 < \varepsilon \leq \min\left\{\frac{1}{10K^2}, \frac{\Delta}{100WK}\right\}. \tag{5.1}$$

Define

$$P := \left\{ (\boldsymbol{r}, z, w) \in \mathbb{R}^4 \ : \ \begin{array}{l} \boldsymbol{r} = (r_1, r_2), \quad |r_1|, |r_2| \leq W\varepsilon, \quad -1 \leq z \leq 2, \\ \varepsilon \displaystyle\sum_{i=1}^{k} \max\{|\boldsymbol{L}_i \cdot \boldsymbol{r}|, |z - t_i|\} \leq w \leq 10 \end{array} \right\},$$

and set two points

$$v := \left(\boldsymbol{0}, 0, \varepsilon \sum_i t_i\right), \qquad x := \left(\boldsymbol{0}, 1, \varepsilon \sum_i (1 - t_i)\right).$$

Let

$$H := \max_{s \in \{-1,1\}^k} \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} G(t_i, t_j)\, s_i s_j\, (\boldsymbol{L}_i \cdot \boldsymbol{L}_j),$$

and

$$\ell_\varepsilon := \sum_{j=0}^{k} (t_{j+1} - t_j) \sqrt{1 + \varepsilon^2 (2j - k)^2}\,.$$

Then P is a 4-dimensional convex polytope with $3k + 8$ facets, such that

$$\left|\operatorname{dist}_{\partial P}(v, x) - (\ell_\varepsilon - \varepsilon^4 H)\right| \leq 10^6 K^{12} \varepsilon^6. \tag{5.2}$$

The rest of this section contains the proof of the lemma.

5.2. **An explicit list of supporting hyperplanes.** The lower boundary is the graph of the convex piecewise-affine function

$$f(\boldsymbol{r}, z) \,:=\, \varepsilon \sum_i \max\big\{|\boldsymbol{L}_i \cdot \boldsymbol{r}|\,,\, |z - t_i|\big\}.$$

The inequality $w \geq f(\boldsymbol{r}, z)$ is exactly the lower-boundary inequality in the definition of P. Therefore the untruncated lower region is the epigraph of f, and the four transverse walls, two longitudinal end caps, and top plane merely cut that epigraph down to a bounded polytope. We call one of the affine supports *active* at a point when it attains the maximum that defines the corresponding summand. Because the lens positions are well separated, at most one summand can use its transverse branch $|\boldsymbol{L}_i \cdot \boldsymbol{r}|$ at any point of the box. This observation gives both the short facet list and the local lens geometry.

Inside the transverse box,

$$|\boldsymbol{L}_i \cdot \boldsymbol{r}| \;\leq\; \sqrt{2}\, W\varepsilon |\boldsymbol{L}_i| \;<\; 2W\varepsilon K.$$

For the transverse branch $|\boldsymbol{L}_i \cdot \boldsymbol{r}|$ to dominate the longitudinal branch $|z - t_i|$, one must have $|z - t_i| \leq |\boldsymbol{L}_i \cdot \boldsymbol{r}|$. The preceding bound on the latter quantity is $2W\varepsilon K$. Hence the transverse branch of the ith lens can be active only inside the slab $|z - t_i| \leq 2W\varepsilon K$ between the two parallel hyperplanes $z = t_i \pm 2W\varepsilon K$. Condition (5.1) makes these slabs disjoint and contained in $(0, 1)$. Because the lens slabs are disjoint, at most one summand can use a transverse branch at any fixed point. All other summands are then on their longitudinal branches $|z - t_h|$. Expanding the resulting sum therefore leaves only the $k + 1$ longitudinal baseline supports and the two side supports for each of the k lenses. Hence the lower graph is the maximum of the following $3k + 1$ affine functions, multiplied by ε:

$$D_j(z) = (2j - k)z + \sum_{i \,>\, j} t_i - \sum_{i \,\leq\, j} t_i\,, \qquad\qquad 0 \;\leq\; j \;\leq\; k, \tag{5.3}$$

$$A_i^{\pm}(\boldsymbol{r}, z) = \pm \boldsymbol{L}_i \cdot \boldsymbol{r} + (2i - k - 1)z + \sum_{h \,>\, i} t_h - \sum_{h \,<\, i} t_h\,, \qquad\qquad 1 \;\leq\; i \;\leq\; k. \tag{5.4}$$

We use D_j and $A_i^{\pm}$ for these affine functions and also for their facets. The facet itself is the part of the supporting hyperplane $w = \varepsilon D_j$ or $w = \varepsilon A_i^{\pm}$ satisfying all the other inequalities.

For the local geometry near t_i, let $u = \boldsymbol{L}_i \cdot \boldsymbol{r}$, $\zeta = z - t_i$, and

$$c_i(z) = (2i - k - 1)z + \sum_{h \,>\, i} t_h - \sum_{h \,<\, i} t_h\,.$$

The four relevant affine functions are

$$D_{i-1} = c_i - \zeta, \qquad A_i^{+} = c_i + u, \qquad D_i = c_i + \zeta, \qquad A_i^{-} = c_i - u.$$

The incoming baseline D_{i-1} is active for $\zeta \leq -|u|$, the positive side A_i^{+} for $u \geq |\zeta|$, the outgoing baseline D_i for $\zeta \geq |u|$, and the negative side A_i^{-} for $-u \geq |\zeta|$. These four inequalities describe the four sectors in the (u, ζ)-plane, so moving once around the origin meets the four supporting facets in the cyclic order $D_{i-1}, A_i^{+}, D_i, A_i^{-}$. Each consecutive pair meets in a ridge. These face intersections and their dimensions are the *incidence relations* of the lens. All four meet at $u = \zeta = 0$, with $w = \varepsilon c_i(t_i)$. The transverse direction perpendicular to $\boldsymbol{L}_i$ remains free, so their common face is an edge. This four-facet neighborhood is the *lens*. A facet gallery from the incoming to the outgoing baseline that crosses only relative ridge interiors and repeats no facet is either (D_{i-1}, A_i^{+}, D_i) or (D_{i-1}, A_i^{-}, D_i). Passing directly between the two baseline facets would meet the common edge, see Figure 4.1(a).

Every one of these functions supports a facet. For a baseline, take $\boldsymbol{r} = \mathbf{0}$ and z strictly between its consecutive lens positions. For a side, take $z = t_i$ and a small $\boldsymbol{r}$ with the required nonzero sign of $\boldsymbol{L}_i \cdot \boldsymbol{r}$. Small neighborhoods preserve strict dominance.

The four transverse walls $r_a = \pm W\varepsilon$, the two end caps $z = -1, 2$, and the top $w = 10$ are also facets. Neighboring lenses share a baseline, which accounts for $k+1$ baselines and $2k$ side facets rather than four distinct facets per lens. The lower graph is below 1 throughout the box, by (5.1), so the polytope is nonempty and full dimensional. This proves the facet count.

At v, only the lower support D_0 is active, and at x only D_k is active. Both points are strictly inside the base box. All inequalities other than the displayed lower support are strict at these points. Therefore each point lies in the relative interior of the corresponding lower facet, rather than on one of its ridges or on a box wall. The preceding formulas also give an explicit halfspace description of the lower graph using only $3k+1$ supports.

5.3. **All shortest paths stay on the lower graph.** The extra walls, end caps, and top are needed only to make P bounded; we first show that no shortest path can use them. The comparison path is the *axis path*, the lift of $\boldsymbol{r} = \mathbf{0}$, $0 \le z \le 1$, to the lower graph. Its length is ℓ_ε, and

$$\ell_\varepsilon \;\le\; \sqrt{1 + k^2\varepsilon^2} \;<\; 2.$$

Any path that touches a transverse wall must travel from $r_a = 0$ to $r_a = \pm W\varepsilon$ and back while its z-coordinate changes from 0 to 1. Project the path to the (r_a, z)-plane and reflect the portion after its first wall contact across that wall. The reflected planar curve joins $(0,0)$ to $(2W\varepsilon, 1)$, so its length is at least $\sqrt{1 + 4W^2\varepsilon^2}$. Since $W = k+1$, this is strictly larger than $\sqrt{1+k^2\varepsilon^2}$, which is an upper bound for the axis-path length ℓ_ε.

Any boundary path that touches a transverse wall therefore has length strictly larger than the explicit axis path. Since a shortest path can never be longer than that competitor, no shortest path can touch a transverse wall. A path touching an end cap has total z-variation at least 3, while a path touching the top has total w-variation at least $20 - 2k\varepsilon > 19$. Each is longer than the axis path. We have now exhibited a boundary path shorter than any path that reaches a transverse wall, an end cap, or the top. A shortest path cannot use any of those truncating facets. It must therefore remain on the lower graph and stay in the interior of the base box. From this point on, every facet in its gallery is one of the lower baseline or side facets, so all subsequent unfolding may be carried out using only those facets.

By the regularity and unfolding results recalled in §4.1, a shortest path whose endpoints lie in facet interiors crosses each ridge through its relative interior and becomes a straight segment after unfolding two consecutive facets. It also cannot leave a facet and later return to it: the straight segment in that facet between the two return points is no longer, and is strictly shorter unless the intervening subpath was already that segment. Hence, if a shortest path left a facet and later returned to it, the intervening portion would have to lie in that same facet after all, contradicting the supposed departure. This rules out every genuine departure-and-return to the same facet. Hence a shortest path uses each facet at most once. Since P has finitely many facets, only finitely many facet galleries can occur; for fixed endpoints and a fixed gallery, the unfolded straight segment determines at most one candidate path.

5.4. **Ridge unfolding controls the transverse displacement.** Unfolding across a ridge rotates one facet hyperplane onto the next while fixing their common ridge. A locally shortest path is straight in this unfolded pair of facets. Its direction in the ambient space does change at the ridge, and the following formula controls the sum of those changes. Since all lower facets have small slopes, this control makes every shortest path longitudinally monotone and confines it to a much narrower region than the transverse box.

Orient a shortest path from v to x and parametrize each straight piece by arclength. Its *unit tangent* is its direction vector of length one. Let V_-, V_+ be these tangents on consecutive facets,

and let n_-, n_+ be the perpendicular unit vectors pointing into P. These inward normals have positive w component. Expressing the ridge-folding isometry in these coordinates gives

$$V_+ - V_- \;=\; -\frac{n_+ \cdot V_-}{1 + n_- \cdot n_+}\,(n_- + n_+), \qquad n_+ \cdot V_- \;<\; 0. \tag{5.5}$$

For example, this identity follows by making the correction parallel to $n_- + n_+$ so that the corrected vector is tangent to the second facet. This correction preserves norm and the component parallel to the ridge.

Each affine branch of the ith summand has slope norm at most $\max\{|\boldsymbol{L}_i|, 1\}$. Summing over all k lenses and using the definition of K, shows that the Euclidean norm of every lower-facet slope is at most $K\varepsilon$. We denote this uniform slope bound by $L := K\varepsilon$. In particular, we have $|(n_\pm)_{\text{base}}| \le L(n_\pm)_w$, and every unit tangent satisfies $|V_w| \le L$.

Formula (5.5) shows that every tangent jump has nonnegative w component. Its base-coordinate norm is at most L times its w component. Summing over all ridge crossings therefore telescopes in the w component and gives

$$\sum_{\text{ridges}} \big|(V_+ - V_-)_{\text{base}}\big| \;\le\; L\big((V_{\text{end}})_w - (V_{\text{start}})_w\big) \;\le\; 2L^2.$$

Thus the base component of the ambient unit tangent has total variation at most $2L^2$. Its arclength-weighted average transverse component is zero because the endpoints have the same transverse coordinate. Its averaged z component is $1/\operatorname{length}(\gamma)$, where γ is the path; since $\operatorname{length}(\gamma) \le \ell_\varepsilon$, this is at least $(1 + k^2\varepsilon^2)^{-1/2}$. Comparing every tangent with these averages gives

$$|V_{\text{transverse}}| \le 2K^2\varepsilon^2\,, \qquad V_z \;\ge\; (1 + k^2\varepsilon^2)^{-1/2} - 2K^2\varepsilon^2 \;>\; \frac{1}{2}\,.$$

The positive lower bound for the ambient tangent component V_z shows that the path is strictly increasing in z. Write it as

$$\boldsymbol{r}(z) = \varepsilon^2 \boldsymbol{R}(z), \qquad \boldsymbol{R}(0) = \boldsymbol{R}(1) = \mathbf{0}\,.$$

Here $\boldsymbol{R}$ is a finite polygonal chain, and its values and slopes obey

$$|\boldsymbol{R}(z)| \;\le\; Q_R, \qquad |\boldsymbol{R}'(z)| \;\le\; Q_R, \qquad Q_R = 4K^2. \tag{5.6}$$

The subscript in Q_R is only a reminder that this constant bounds both the transverse chain $\boldsymbol{R}$ and its slope; it is not an additional geometric parameter. Here $\boldsymbol{R}'(z)$ denotes the constant slope on the linear piece containing z; its value at a breakpoint is irrelevant.

5.5. **Quadratic comparison and completion of squares.** We now compare polygonal transverse chains by their Dirichlet energy. For a prescribed sign word, the Green-kernel tent functions give a chain whose slope jumps are exactly the chosen vectors. A finite integration by parts then turns the lens terms into a bilinear Dirichlet pairing, after which the required inequality is simply completion of squares.

For two polygonal chains $\boldsymbol{R}, \boldsymbol{U}$ on $[0,1]$, define their *Dirichlet pairing* by the standard integral

$$\mathcal{E}(\boldsymbol{R}, \boldsymbol{U}) := \int_0^1 \boldsymbol{R}'(z) \cdot \boldsymbol{U}'(z)\, dz, \qquad \mathcal{E}(\boldsymbol{R}) := \mathcal{E}(\boldsymbol{R}, \boldsymbol{R}).$$

The derivatives are constant on each linear piece, so this integral is just a finite sum. If $0 = z_0 < \ldots < z_N = 1$ is any common refinement of the breakpoints, then

$$\mathcal{E}(\boldsymbol{R}, \boldsymbol{U}) \;=\; \sum_{\nu=0}^{N-1} \frac{\big(\boldsymbol{R}(z_{\nu+1}) - \boldsymbol{R}(z_\nu)\big) \cdot \big(\boldsymbol{U}(z_{\nu+1}) - \boldsymbol{U}(z_\nu)\big)}{z_{\nu+1} - z_\nu}\,.$$

When $\boldsymbol{U} = \boldsymbol{R}$, the numerator on one linear piece is the squared norm of the increment of $\boldsymbol{R}$. Dividing by the length of the piece gives $|\boldsymbol{R}'|^2$ times that length. Summing over all pieces therefore gives $\int_0^1 |\boldsymbol{R}'(z)|^2\,dz$. Hence $\mathcal{E}(\boldsymbol{R})$ is the usual Dirichlet energy of the polygonal chain, and subdividing a linear piece further does not change its value.

For chains with $\boldsymbol{R}(0) = \boldsymbol{R}(1) = \mathbf{0}$, define the quadratic cost

$$\Phi(\boldsymbol{R}) \;:=\; \frac{1}{2}\,\mathcal{E}(\boldsymbol{R}) \;-\; \sum_i \big|\boldsymbol{L}_i \cdot \boldsymbol{R}(t_i)\big|.$$

The notation Φ is used only for this total cost. Its first term is the standard Dirichlet penalty for transverse motion, while the second is the leading length gain from entering the side facets. The comparison chain below shows that minimizing Φ is exactly the same finite optimization as maximizing the sign-word score h.

For a sign word s, define the *comparison chain*

$$\boldsymbol{R}_s(z) \;:=\; \sum_i G(z,t_i)\,s_i\,\boldsymbol{L}_i.$$

Because $G(0,t) = G(1,t) = 0$, this chain has endpoints $\mathbf{0}$. Moreover, the derivative of $G(z,t_i)$ has a jump of -1 at t_i, so the slope jump of $\boldsymbol{R}_s$ there is $-s_i\boldsymbol{L}_i$. The values and slopes of $\boldsymbol{R}_s$ have norm at most $\sum_i |\boldsymbol{L}_i| \le K$, hence (5.6) holds. Its lift also stays inside the transverse box because $\varepsilon Q_R \le 2/5 < W$.

To relate the comparison chain to the lens values, integrate by parts on each interval between consecutive lens positions and then add the results. The endpoint terms at 0 and 1 vanish because both chains have zero endpoints. At an interior lens position t_i, the only surviving boundary term is the jump $-s_i\boldsymbol{L}_i$ in the slope of $\boldsymbol{R}_s$. Equivalently, the same calculation can be viewed as a telescoping sum over the linear pieces. It gives

$$\mathcal{E}(\boldsymbol{R}_s, \boldsymbol{R}) = \sum_i s_i \boldsymbol{L}_i \cdot \boldsymbol{R}(t_i). \tag{5.7}$$

Now set $\boldsymbol{R} = \boldsymbol{R}_s$. The left side becomes the Dirichlet energy of the comparison chain, while the right side becomes $\sum_i s_i\boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i)$. Expanding $\boldsymbol{R}_s(t_i)$ from its definition produces twice the quadratic score $h(s)$. Therefore

$$\frac{1}{2}\mathcal{E}(\boldsymbol{R}_s) = h(s).$$

Now choose each s_i so that $s_i\boldsymbol{L}_i \cdot \boldsymbol{R}(t_i) = |\boldsymbol{L}_i \cdot \boldsymbol{R}(t_i)|$. By (5.7),

$$\Phi(\boldsymbol{R}) \;=\; \frac{1}{2}\mathcal{E}(\boldsymbol{R}) - \mathcal{E}(\boldsymbol{R}_s, \boldsymbol{R}) \;=\; \frac{1}{2}\mathcal{E}(\boldsymbol{R} - \boldsymbol{R}_s) - \frac{1}{2}\mathcal{E}(\boldsymbol{R}_s) \;=\; \frac{1}{2}\mathcal{E}(\boldsymbol{R} - \boldsymbol{R}_s) - h(s) \;\ge\; -H. \tag{5.8}$$

Conversely, let s maximize $h(s)$. Since $|x| \ge s_i x$ for every real x,

$$\Phi(\boldsymbol{R}_s) \;\le\; \frac{1}{2}\mathcal{E}(\boldsymbol{R}_s) - \sum_i s_i \boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i) \;=\; -H.$$

On the other hand, the general lower bound (5.8) is valid for every chain, so applying it to $\boldsymbol{R}_s$ gives $\Phi(\boldsymbol{R}_s) \ge -H$. Together with the preceding upper bound $\Phi(\boldsymbol{R}_s) \le -H$, this proves equality. Hence

$$\min_{\boldsymbol{R}(0)=\boldsymbol{R}(1)=\mathbf{0}} \Phi(\boldsymbol{R}) = -H, \qquad \Phi(\boldsymbol{R}_s) = -H$$

for every maximizing sign word s. This is the precise link between the finite score optimization and the geometry of transverse paths.

5.6. **A uniform Pythagorean length estimate.** We now pass from the quadratic cost $\Phi(\boldsymbol{R})$ to actual boundary length. The point is that every admissible chain has uniformly small transverse slope. On each affine piece, its Euclidean length is therefore obtained from a square root that can be expanded with a uniform error. The only places where the lifted chain differs from the axis path are the short intervals on which a side facet of a lens is active.

Let $\boldsymbol{R}$ be any polygonal chain satisfying (5.6), and let $\mathcal{L}_\varepsilon(\boldsymbol{R})$ be the length of its lift to the lower graph,

$$z \;\longmapsto\; \big(\varepsilon^2 \boldsymbol{R}(z), z, f(\varepsilon^2 \boldsymbol{R}(z), z)\big).$$

For the estimate below, use the same formula

$$\Phi(\boldsymbol{R}) = \frac{1}{2}\mathcal{E}(\boldsymbol{R}) - \sum_i |\boldsymbol{L}_i \cdot \boldsymbol{R}(t_i)|$$

even when the endpoints of $\boldsymbol{R}$ are not zero. We will prove

$$\big|\mathcal{L}_\varepsilon(\boldsymbol{R}) - \ell_\varepsilon - \varepsilon^4 \Phi(\boldsymbol{R})\big| \;\le\; (132K^7 + 14K^6)\varepsilon^6. \tag{5.9}$$

The estimate is uniform over every polygonal chain satisfying (5.6); it does not depend on the number or location of its breakpoints. The transverse chain associated with an actual shortest path satisfies exactly those bounds by the preceding subsection, so the estimate applies to that chain in particular.

For each lens, set

$$A_i(z) \;:=\; |\boldsymbol{L}_i \cdot \boldsymbol{R}(z)|, \qquad a_i \;:=\; A_i(t_i), \qquad M_i \;:=\; Q_R |\boldsymbol{L}_i|.$$

The bounds in (5.6) imply two useful estimates for A_i. First, by the Cauchy–Schwarz inequality, we have:

$$A_i(z) \;\le\; |\boldsymbol{L}_i| \cdot |\boldsymbol{R}(z)| \;\le\; Q_R\, |\boldsymbol{L}_i| \;=\; M_i\,.$$

Second, the piecewise-linear chain $\boldsymbol{R}$ is Q_R-Lipschitz: on each linear piece its slope has norm at most Q_R, and summing along the pieces gives

$$|\boldsymbol{R}(z) - \boldsymbol{R}(y)| \;\le\; Q_R\, |z - y|\,.$$

Therefore the reverse triangle inequality gives

$$\big|A_i(z) - A_i(y)\big| \;\le\; |\boldsymbol{L}_i| \cdot |\boldsymbol{R}(z) - \boldsymbol{R}(y)| \;\le\; M_i\, |z - y|.$$

Thus A_i is M_i-Lipschitz. In particular, wherever A_i is differentiable its derivative has absolute value at most M_i; at a zero of $\boldsymbol{L}_i \cdot \boldsymbol{R}$ it may have a corner, which is why we use the Lipschitz formulation below. The smallness condition on ε implies $\varepsilon^2 M_i \le 1/25 < 1$.

The i-th side facet is active exactly when

$$|z - t_i| \;\le\; \varepsilon^2 A_i(z).$$

Denote this active set by

$$I_i \;:=\; \big\{z \,:\, |z - t_i| \le \varepsilon^2 A_i(z)\big\}.$$

We now show that I_i is a single interval without differentiating A_i. If $t_i \le z_1 < z_2$, then the Lipschitz bound gives

$$\big[(z_2 - t_i) - \varepsilon^2 A_i(z_2)\big] - \big[(z_1 - t_i) - \varepsilon^2 A_i(z_1)\big] \;\ge\; (1 - \varepsilon^2 M_i)(z_2 - z_1) \;>\; 0.$$

Hence $|z - t_i| - \varepsilon^2 A_i(z)$ is strictly increasing to the right of t_i. The same argument, with the orientation reversed, shows that it is strictly increasing as one moves to the left away from t_i. Consequently it can cross zero at most once on each side, and I_i is a single interval containing t_i.

If $a_i = 0$, the Lipschitz bound forces $I_i = \{t_i\}$. Indeed, for $z \in I_i$ and $w = |z - t_i|$ we have $w \le \varepsilon^2 A_i(z) \le \varepsilon^2 M_i w$, and $\varepsilon^2 M_i < 1$ implies $w = 0$. Otherwise write w_i^- and w_i^+ for the left and right widths of I_i. At either endpoint, with $w = w_i^\pm$, equality gives

$$w = \varepsilon^2 A_i(t_i \pm w).$$

Since A_i is M_i-Lipschitz,

$$|A_i(t_i \pm w) - a_i| \;\le\; M_i w.$$

Thus

$$w \;\le\; \varepsilon^2(a_i + M_i w).$$

Therefore, because $\varepsilon^2 M_i \le 1/25$, we have:

$$w \;\le\; \frac{\varepsilon^2 a_i}{1 - \varepsilon^2 M_i} \;\le\; 2\varepsilon^2 a_i\,.$$

Using the endpoint identity once more,

$$|w - \varepsilon^2 a_i| \;=\; \varepsilon^2 |A_i(t_i \pm w) - a_i| \;\le\; \varepsilon^2 M_i w \;\le\; 2\varepsilon^4 M_i a_i \;\le\; 2\varepsilon^4 M_i^2\,.$$

Therefore

$$w_i^{\pm} \;\le\; 2\varepsilon^2 a_i \;\le\; 2\varepsilon^2 M_i\,, \qquad |w_i^{\pm} - \varepsilon^2 a_i| \;\le\; 2\varepsilon^4 M_i^2. \tag{5.10}$$

The intervals I_i lie in the pairwise disjoint lens slabs. We will also use

$$\sum_i M_i \;\le\; Q_R K, \qquad \sum_i M_i^2 \;\le\; Q_R^2 K^2.$$

Let $\rho(z)$ be the derivative of the lower-graph height divided by ε, computed on each affine piece of the lifted chain. We use ρ for this scalar slope to distinguish it from points or vectors denoted by v. Outside the intervals I_i, we have:

$$\rho(z) \;=\; \sum_i \operatorname{sgn}(z - t_i).$$

On I_i the ith summand is no longer $|z - t_i|$; it is the transverse quantity $|\boldsymbol{L}_i \cdot \varepsilon^2 \boldsymbol{R}(z)| = \varepsilon^2 A_i(z)$. The other $k-1$ summands remain on their longitudinal branches. Their derivatives add to the constant baseline value $b_i = 2i - k - 1$. Differentiating the transverse term then gives

$$\rho(z) = b_i + \varepsilon^2 A_i'(z), \qquad b_i = 2i - k - 1,$$

where $A_i'(z)$ denotes the constant derivative on the current linear piece. In all cases $|\rho(z)| \le k$.

On every affine piece, Pythagoras gives the length element

$$\sqrt{1 + \varepsilon^2 \rho(z)^2 + \varepsilon^4 |\boldsymbol{R}'(z)|^2}\, dz.$$

For $u, v \ge 0$,

$$\left|\sqrt{1 + u + v} - \sqrt{1 + u} - \frac{v}{2}\right| \;\le\; \frac{uv}{4} + \frac{v^2}{8}.$$

Applying this with $u = \varepsilon^2 \rho(z)^2$ and $v = \varepsilon^4 |\boldsymbol{R}'(z)|^2$, and summing over the affine pieces, gives

$$\left|\mathcal{L}_\varepsilon(\boldsymbol{R}) - \int_0^1 \sqrt{1 + \varepsilon^2 \rho(z)^2}\, dz - \frac{\varepsilon^4}{2}\mathcal{E}(\boldsymbol{R})\right| \;\le\; 6K^6\varepsilon^6. \tag{5.11}$$

To see the numerical bound in (5.11), integrate the two error terms in the preceding inequality separately. Since $|\rho(z)| \le k$ and $|\boldsymbol{R}'(z)| \le Q_R$ on an interval of total length one,

$$\frac{1}{4}\int_0^1 \varepsilon^6 \rho(z)^2 |\boldsymbol{R}'(z)|^2\, dz \;\le\; \frac{\varepsilon^6 k^2 Q_R^2}{4}, \qquad \frac{1}{8}\int_0^1 \varepsilon^8 |\boldsymbol{R}'(z)|^4\, dz \;\le\; \frac{\varepsilon^8 Q_R^4}{8}.$$

Thus the unsimplified error bound is the sum of these two quantities. Now $Q_R = 4K^2$ and $k \le K$. The smallness assumption $\varepsilon^2 Q_R^2 \le 16/100$ makes the second term lower order: explicitly, $\varepsilon^8 Q_R^4/8 \le (2/100)\varepsilon^6 Q_R^2$. Substituting $Q_R = 4K^2$ and using $K \ge 1$ then gives the convenient coarse bound $6K^6\varepsilon^6$ used in (5.11).

Equation (5.11) has separated the transverse Dirichlet energy from the part of the length that depends only on the scalar height slope $\rho(z)$. To finish the comparison with the reference axis

path, we now compare that remaining integral directly with the axis-path length ℓ_ε. Outside the I_i the two integrands agree exactly, so only the lens intervals contribute. On I_i, apply

$$|\sqrt{1+x} - 1 - x/2| \;\le\; x^2/8, \qquad x \ge 0,$$

to both the lifted path and the corresponding pieces of the axis path. The total Taylor error is at most

$$\frac{\varepsilon^4 k^4}{4} \sum_i |I_i| \;\le\; \varepsilon^6 k^4 Q_R K \;\le\; 4K^7\varepsilon^6.$$

After replacing the square roots by their quadratic terms, the contribution from I_i to the difference is

$$\frac{\varepsilon^2}{2}\left[(2b_i-1)w_i^- - (2b_i+1)w_i^+\right] + \varepsilon^4 b_i \int_{I_i} A_i'(z)\,dz + \frac{\varepsilon^6}{2}\int_{I_i}(A_i'(z))^2\,dz.$$

If both widths were exactly $\varepsilon^2 a_i$, the first line of the preceding display would simplify to $-\varepsilon^4 a_i$. This is the desired leading contribution of lens i, because $a_i = |\boldsymbol{L}_i \cdot \boldsymbol{R}(t_i)|$ is precisely the lens term appearing in $\Phi(\boldsymbol{R})$. The actual widths differ from $\varepsilon^2 a_i$ only by the small errors in (5.10). The remaining terms are controlled directly by (5.10), $|b_i| \le k$, $|A_i'| \le M_i$, and $|I_i| \le 4\varepsilon^2 M_i$. The three resulting errors, summed over i, are bounded respectively by

$$4k\varepsilon^6 \sum_i M_i^2 \;\le\; 64K^7\varepsilon^6, \qquad 4k\varepsilon^6 \sum_i M_i^2 \;\le\; 64K^7\varepsilon^6,$$

and

$$\frac{\varepsilon^6}{2}\sum_i M_i^2 \;\le\; 8K^6\varepsilon^6.$$

For the last estimate, the integral of $(A_i')^2$ contributes an extra factor $|I_i|$, and the width bound gives $|I_i| \le 4\varepsilon^2 M_i \le 1$. Adding the two width errors, the derivative error, the Taylor remainder on the lens intervals, and the Pythagorean remainder in (5.11) gives

$$|\mathcal{L}_\varepsilon(\boldsymbol{R}) - \ell_\varepsilon - \varepsilon^4\Phi(\boldsymbol{R})| \;\le\; (132K^7 + 14K^6)\varepsilon^6,$$

which is (5.9).

We can now finish the proof of Lemma 5.1. Let $\boldsymbol{R}$ be the transverse chain of an actual shortest path. The quadratic comparison proved $\Phi(\boldsymbol{R}) \ge -H$. Substituting this inequality into (5.9) shows that the shortest-path length cannot be much smaller than $\ell_\varepsilon - \varepsilon^4 H$; more precisely it gives the lower half of the two-sided estimate in (5.2). For the opposite inequality, choose a sign word s maximizing $h(s)$. Its comparison chain satisfies $\Phi(\boldsymbol{R}_s) = -H$. The lifted chain is an explicit boundary competitor, so applying (5.9) to it shows that the true geodesic distance is no larger than $\ell_\varepsilon - \varepsilon^4 H$ plus the same error term. This is the matching upper half of (5.2). Finally, $132K^7 + 14K^6 < 10^6 K^{12}$, which gives (5.2) and completes the proof of the lemma. □

6. The rational polytope and distance threshold

6.1. **Putting everything together.** We choose ε so that the geometric error is smaller than $\varepsilon^4\kappa^2$, the scaled arithmetic Partition gap. A rational approximation to the axis-path length then gives a rational threshold separating yes and no instances. We also bound the binary size of the polytope, endpoints, and threshold.

Apply Lemma 5.1 to the above $k = 6m$ lens positions and slope vectors. In this section, $i \in [k]$ indexes the full lens list, while $j = 0, \dots, m-1$ indexes the input blocks whenever it appears in A_j, q_j, η_j or $t_{j,\nu}$. When D_j denotes a baseline support, instead $0 \le j \le k$. With these conventions, set

$$W \;:=\; k+1, \qquad K \;:=\; \frac{100k}{q}, \qquad \Delta = \frac{q^2}{100}, \qquad \varepsilon \;:=\; \frac{\kappa\Delta}{10^{10} W K^{10}}.$$

Every slope vector is nonzero: its first component is a positive multiple of one of $-1-\eta_j$, $6/5$, and $3+\eta_j/9$. Also

$$\sum_i \max\{|\boldsymbol{L}_i|,1\} \;\le\; K_0+\kappa K_p+k \;<\; K.$$

We next verify the spacing hypothesis in Lemma 5.1. Fix a block j, $0\le j\le m-1$, and abbreviate $c:=q_j^2$. Applying the relocation map τ_c to the reference positions $1/10, 1/2, 9/10$ gives

$$t_{j,1}=\frac{c}{9+c}, \qquad t_{j,2}=\frac{c}{1+c}, \qquad t_{j,3}=\frac{9c}{1+9c}.$$

The gap from the left endpoint $z=0$ to the first lens is greater than $q^2/10$. Within a block, each of the two consecutive lens gaps is at least $2c/5$. For $0\le j<m-1$, the gap from the third lens of block j to the first lens of block $j+1$ is

$$\frac{c(T^2-81)}{(9+cT^2)(1+9c)} \;>\; c \;\ge\; q^2.$$

After the last first-half block, the gap to the first lens of the reflected half is greater than $1/2$. Every gap in the second half is the mirror image of a corresponding first-half gap. Therefore every consecutive lens gap, as well as the gaps from 0 to the first lens and from the last lens to 1, is at least $\Delta=q^2/100$, as required by Lemma 5.1.

In coordinates $(r_1,r_2,z,w)\in\mathbb{R}^4$, define

$$P=\left\{|r_1|,\,|r_2|\;\le\;W\varepsilon\,,\quad -1\;\le\;z\;\le\;2,\quad \varepsilon\sum_{i=1}^{k}\max\{|\boldsymbol{L}_i\cdot\boldsymbol{r}|,|z-t_i|\}\;\le\;w\;\le\;10\right\}.$$

This is a bounded, full-dimensional convex polytope with exactly $3k+8=18m+8$ facets. Its lower graph is the maximum of the $k+1$ baseline supports εD_j and the $2k$ side supports $\varepsilon A_i^{\pm}$, where

$$D_j \;:=\; (2j-k)z+\sum_{i>j}t_i-\sum_{i\le j}t_i\,, \qquad 0\le j\le k,$$

$$A_i^{\pm} \;:=\; \pm(\boldsymbol{L}_i\cdot\boldsymbol{r})+(2i-k-1)z+\sum_{h>i}t_h-\sum_{h<i}t_h\,, \qquad 1\le i\le k.$$

All these lower supports are active, since the lens slabs are disjoint and $\boldsymbol{L}_i\neq 0$. The other seven facets are the four transverse walls, two z caps, and the top. The two rational facet-interior endpoints are

$$v=(\mathbf{0},0,k\varepsilon/2), \qquad x=(\mathbf{0},1,k\varepsilon/2).$$

Each first-half lens at position t_i has one reflected partner at position $1-t_i$. The two positions in each pair therefore sum to one. There are $k/2$ such pairs, so the sum of all k lens positions is $k/2$. The endpoint heights in Lemma 5.1 are $\varepsilon\sum_i t_i$ and $\varepsilon\sum_i(1-t_i)$, and both are consequently equal to $k\varepsilon/2$.

For the rational approximation to the axis-path length, we collect the weighted second and fourth powers of the rescaled baseline slopes $2j-k$, obtained by dividing the actual slopes by ε. These are the coefficients that occur when each square root is replaced by its quadratic polynomial approximation. Let $t_0=0, t_{k+1}=1$. For $p=2,4$, define the weighted moments of the baseline slopes by

$$\beta_p \;:=\; \sum_{j=0}^{k}(t_{j+1}-t_j)(2j-k)^p.$$

Then set

$$L_0 \;:=\; 1+\frac{\varepsilon^2}{2}\beta_2-\frac{\varepsilon^4}{8}\beta_4\,.$$

For one longitudinal interval, the Taylor expansion is $\sqrt{1+x} = 1 + x/2 - x^2/8 + O(x^3)$ with $x = \varepsilon^2(2j-k)^2$. Multiplying the first three terms by the interval length $t_{j+1} - t_j$ and summing over j gives exactly the displayed formula for L_0. Thus L_0 is the rational fourth-order Taylor approximation to the complete axis-path length ℓ_ε; the symbols β_2 and β_4 merely collect the weighted quadratic and quartic coefficients that occur in this expansion. Write $d_{\partial P}(v,x) = \operatorname{dist}_{\partial P}(v,x)$. Lemma 5.1 and the elementary square-root remainder imply

$$\big|d_{\partial P}(v,x) - (L_0 - \varepsilon^4 H)\big| \;<\; 2\cdot 10^6 K^{12}\varepsilon^6 \;<\; \frac{\varepsilon^4\kappa^2}{16}\,.$$

Substituting the displayed value of ε shows that the geometric and Taylor errors together are less than one sixteenth of the scaled arithmetic gap $\varepsilon^4\kappa^2$. We can therefore place a rational decision threshold halfway between the leading-order value for a balanced partition and the leading-order value for the smallest possible nonzero residual:

$$\tau \;:=\; L_0 \;-\; \varepsilon^4 H_* \;+\; \frac{1}{2}\varepsilon^4\kappa^2.$$

Note that $\tau > 0$. Indeed, $\beta_4 \le k^4$ and $\varepsilon \le 1/(10K^2)$ give $L_0 > 1/2$, while the explicit formula for H_*, together with the chosen value of ε, gives $\varepsilon^4 H_* < 1/2$.

Here H_* is the word-independent benchmark obtained by setting the residual to zero in (4.7). If the integer residual is nonzero, its absolute value is at least 1, so the score falls by at least κ^2. Multiplication by ε^4 turns this into a leading-order distance gap of $\varepsilon^4\kappa^2$, and the final term in τ places the threshold at its midpoint. The error bound above then gives the following strict separation. If the Partition instance has a solution, then

$$d_{\partial P}(v,x) \;<\; \tau - 7\varepsilon^4\kappa^2/16.$$

If it has no solution, then

$$d_{\partial P}(v,x) \;>\; \tau + 7\varepsilon^4\kappa^2/16.$$

The yes-case upper bound lies strictly below the no-case lower bound, with the rational number τ between them. Therefore a query asking whether the intrinsic boundary distance is at most τ returns yes exactly when the original Partition instance has a solution. This completes the many-one reduction and proves NP-hardness for rational convex polytopes in $\mathbb{R}^4$.

All coordinates and halfspace coefficients are rational, computed by polynomially many arithmetic operations on integers with polynomially many bits. More explicitly, we have

$$\begin{aligned}
\varepsilon \;&=\; \frac{1}{10^{38}SWk^{10}T^{18m}} \;=\; \frac{1}{10^{72m+38}S(6m+1)(6m)^{10}}\,,\\
\varepsilon^4\kappa^2 \;&=\; \frac{1}{10^{164}S^6W^4k^{40}T^{84m}} \;=\; \frac{1}{10^{336m+164}S^6(6m+1)^4(6m)^{40}}\,,
\end{aligned}$$

where $S = A_0 + \ldots + A_{m-1}$.

In summary, each rational coefficient appearing in a single lens has bit length $O(m+\log S)$. The coefficients of the final supporting hyperplanes involve sums of $O(m)$ such rational numbers, and the threshold is obtained from polynomially many such arithmetic operations. Their bit lengths therefore remain polynomial in the input size.

6.2. **Why the rational approximation to the axis length suffices.** The axis-path length need not be rational, even though all its vertices are rational. To output a rational decision threshold, we approximate each square root in its explicit length formula. The remainder estimate below shows that the total error is smaller than the distance gap left by the integer partition residual.

The rational approximation L_0 uses the following elementary remainder bound for each square root in ℓ_ε:

$$1+\frac{u}{2}-\frac{u^2}{8} \;\le\; \sqrt{1+u} \;\le\; 1+\frac{u}{2}-\frac{u^2}{8}+\frac{u^3}{16}, \qquad 0 \;\le\; u \;\le\; 1.$$

Both sides follow by squaring nonnegative quantities. For the lower polynomial its square minus $1+u$ is $-u^3/8+u^4/64$. For the upper polynomial the difference is $5u^4/64-u^5/64+u^6/256$. Consequently,

$$0 \;\le\; \ell_\varepsilon - L_0 \;\le\; k^6\varepsilon^6/16.$$

Combined with Lemma 5.1, this bounds the distance error by $2\cdot 10^6 K^{12}\varepsilon^6$. Its ratio to $\varepsilon^4\kappa^2$ is at most

$$\frac{2\cdot 10^6 K^{12}\varepsilon^2}{\kappa^2} \;=\; \frac{2\cdot 10^6\Delta^2}{10^{20}W^2K^8} \;<\; \frac{1}{16}\,.$$

This proves the claimed strict threshold margins.

6.3. **Polynomial encoding and the exact reduction.** We bound the bit lengths of the output coefficients and the number of rational operations needed to compute them. The construction does not enumerate the candidate sign words.

Write N for the total binary length of the Partition input. We have $m \le N$ and $\log_2 S = O(N)$. Every $q_j, \eta_j, \kappa, t_i, \boldsymbol{L}_i, K, \Delta, \varepsilon$ is obtained from integers with $O(N+\log m)$ bits by a bounded number of arithmetic operations, except for the displayed sums over $O(m)$ terms. Adding these rational terms may increase denominators, but the resulting bit lengths are still polynomial in N. The same observation applies to $H_*, \beta_2, \beta_4, \tau$ and the explicit affine supports. The number of those supports is $18m+1$, with seven additional box facets.

The reduction algorithm computes these quantities using rational arithmetic and writes the inequalities (5.3) and (5.4), the four transverse walls, the two end caps, the top, the two points, and τ. It does not enumerate sign words. Every arithmetic operation is performed on rational numbers whose binary length remains polynomial in the input size, and only polynomially many such operations are used. Therefore both the running time of the construction and the binary length of the resulting polytope description are polynomial. The equivalence with Partition, including the two strict margins, proves Theorem 1.1. □

7. From unfolded segments to boundary geodesics

To prove Theorem 1.2, we need the following unfolding lemma. A path is called *locally minimizing*, if every sufficiently short subpath is shortest between its endpoints. The lemma shows that a strict sign condition for a prescribed gallery produces an actual regular locally minimizing path for all sufficiently small positive values of the thinness parameter.

In its statement and proof, the data $t_i, \boldsymbol{L}_i$ are fixed. We write P_ε for the thin polytope, $s=(s_1,\dots,s_k)$ for the prescribed lens choices, and $\boldsymbol{R}_s$ for the corresponding comparison chain. The number α_i below is the signed margin at lens i. Thus, positivity of all α_i is what keeps the formal unfolded segment inside the chosen side facets.

Lemma 7.1 (Strict galleries persist)**.** *Let $0<t_1<\ldots<t_k<1$, and let $\boldsymbol{L}_1,\dots,\boldsymbol{L}_k \in \mathbb{R}^2$ such that $\boldsymbol{L}_i \neq \boldsymbol{0}$. Throughout the statement and proof, $i\in[k]$ indexes these lenses, and every unqualified sum over i has this range. Use the thin graph polytope*

$$P_\varepsilon \;=\; \left\{ (\boldsymbol{r},z,w) \;:\; \begin{array}{c} |r_1|,\,|r_2| \;\le\; (k+1)\varepsilon, \quad -1 \;\le\; z \;\le\; 2, \\ \varepsilon\displaystyle\sum_i \max\left\{|\boldsymbol{L}_i\cdot\boldsymbol{r}|,\,|z-t_i|\right\} \;\le\; w \;\le\; 10 \end{array} \right\}$$

and endpoints

$$v_\varepsilon \,=\, \Big(\mathbf{0}, 0, \varepsilon \sum_i t_i\Big), \qquad x_\varepsilon \,=\, \Big(\mathbf{0}, 1, \varepsilon \sum_i (1-t_i)\Big).$$

For $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$, *let*

$$\boldsymbol{B}_s \,:=\, \sum_i t_i s_i \boldsymbol{L}_i\,, \qquad \boldsymbol{R}_s(z) \,:=\, \sum_i G(z, t_i) s_i \boldsymbol{L}_i\,.$$

Suppose

$$\alpha_i \,:=\, s_i \boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i) \,>\, 0 \quad \textit{for all} \quad 1 \,\le\, i \,\le\, k.$$

Then, for every sufficiently small positive ε, *the gallery*

$$\big(D_0, A_1^{s_1}, D_1, A_2^{s_2}, \ldots, A_k^{s_k}, D_k\big)$$

contains a regular locally minimizing path from v_ε *to* x_ε. *A finite collection of words satisfying the strict inequalities has one common interval of sufficiently small positive parameters on which all these conclusions hold.*

Proof. We begin by ignoring the finite boundaries of the facets and unfolding only their supporting hyperplanes. In the unfolded picture the prescribed gallery becomes a sequence of coplanar supports, so the candidate geodesic is simply the straight segment between the unfolded endpoints. At this stage the segment is only *formal*: after folding it back, we still have to verify that every segment actually lies inside the intended facet and that every crossing lies in the intended ridge. We then locate its ridge crossings and use the positive margins α_i to check all the facet inequalities.

Write a supporting plane as $w = \varepsilon(a \cdot X + c)$, with $X = (\boldsymbol{r}, z)$. Its upward unit normal and normalized offset are

$$n_a = \frac{(-\varepsilon a, 1)}{\sqrt{1 + \varepsilon^2 |a|^2}}\,, \qquad c_a = \frac{\varepsilon c}{\sqrt{1 + \varepsilon^2 |a|^2}}\,.$$

The normalized equation is $n_a \cdot (\boldsymbol{r}, z, w) = c_a$. Normalization means that the normal has length one, and does not require its coordinates to remain rational. For a point p on the old supporting plane, the ridge-folding isometry, written in the present coordinates, maps it to the new supporting plane with normal n_b and offset c_b by

$$\Phi_{a,b}(p) = p - \frac{n_b \cdot p - c_b}{1 + n_a \cdot n_b}(n_a + n_b). \tag{7.1}$$

It fixes the intersection of the two planes and is an isometry between them. Both facts follow directly by substitution and the identity $n_a \cdot (p - p') = 0$ for two points on the old plane.

Let $\delta = \varepsilon^2$. On bounded sets, the base coordinates of (7.1) satisfy

$$(\Phi_{a,b}(p))_{\text{base}} = X + \frac{\delta}{2}\big((a-b) \cdot X + c - d\big)(a+b) + O(\delta^2),$$

where d is the constant term of the new affine support. This estimate follows by rationalizing the square roots in the exact formula.

Unfold the source through the prescribed sequence of supporting planes, initially disregarding facet boundaries, and call its final copy ν_s. This formal unfolded image lies in the affine span of the final facet, but need not lie in the facet itself. Its base coordinates remain $O(\delta)$. In the affine functions after division by ε, the two differences of successive constant terms in lens i are both t_i, and the transverse gradient is $s_i \boldsymbol{L}_i$ on the intermediate plane and zero on the two baselines. The preceding formula therefore gives

$$(\nu_s)_{\boldsymbol{r}} = \delta \sum_i t_i s_i \boldsymbol{L}_i + O(\delta^2) = \delta \boldsymbol{B}_s + O(\delta^2), \qquad (\nu_s)_z = O(\delta).$$

The straight segment in the final supporting plane from ν_s to x_ε has terminal transverse slope, relative to z,

$$-\delta \boldsymbol{B}_s + O(\delta^2).$$

Fold this segment back through the prescribed planes. All formal ridge intersections are well defined for small δ: after cancelling ε in the two plane equations, the intersection denominators have nonzero constant term, since the successive z-gradients differ by one. Their z coordinates are $t_i + O(\delta)$.

The tangent identity

$$V_+ - V_- = -\frac{n_+ \cdot V_-}{1 + n_- \cdot n_+}(n_- + n_+)$$

shows that either of the two hinges in lens i changes the transverse slope by

$$-\frac{\delta}{2} s_i \boldsymbol{L}_i + O(\delta^2).$$

Starting from the terminal slope and proceeding backwards gives, on the baseline intervals,

$$\boldsymbol{r}(z) = \delta \boldsymbol{R}_s(z) + O(\delta^2), \qquad \frac{d\boldsymbol{r}}{dz} = \delta \frac{d\boldsymbol{R}_s}{dz} + O(\delta^2).$$

The same positional estimate holds at every ridge crossing. The ridge equations at the entrance and exit of lens i are respectively

$$z = t_i - s_i \boldsymbol{L}_i \cdot \boldsymbol{r}, \qquad z = t_i + s_i \boldsymbol{L}_i \cdot \boldsymbol{r}.$$

Substituting the expansion $\boldsymbol{r}(z) = \delta \boldsymbol{R}_s(z) + O(\delta^2)$ into the two ridge equations and using $\alpha_i = s_i \boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i)$ gives the corresponding crossing coordinates

$$z_i^- = t_i - \delta\alpha_i + O(\delta^2), \qquad z_i^+ = t_i + \delta\alpha_i + O(\delta^2).$$

The portion of the path in one facet is a *flight*. Since $\alpha_i > 0$, the side-facet flight has positive length and its entry precedes its exit. All baseline flights have positive length as well, since the fixed gaps between consecutive t_i are positive. The chosen-side inequalities are strict between the two ridge crossings. At each crossing the two prescribed supports are equal, and all other supports are strictly lower: for the other two supports of the same lens this follows from $\alpha_i > 0$, and for other lenses from the fixed gaps between the t_i. Hence at every crossing exactly the two supports defining the intended ridge are active, while all other supports are strict. This is precisely the condition that the crossing point lie in the relative interior of the intended ridge. The segment joining consecutive ridge points lies in its facet by convexity. Its interior lies in the facet interior, since every other support is strict for at least one endpoint. The transverse coordinates are $O(\delta)$, so the chain stays strictly inside the transverse box. It also avoids the end caps and the top.

These checks make the gallery *valid*: the formal back-folded segment is an actual path in the stated gallery, with positive flights and crossings in relative interiors of ridges. Its unfoldings are straight, so it is locally length minimizing. Its z coordinate is strictly increasing. Only finitely many formulas and strict inequalities are involved, so finitely many words admit one common sufficiently small parameter interval. □

8. Exponentially many globally shortest paths

We now prove Theorem 1.2 by making the selected geodesic lengths exactly equal. The main geometric step is to place the three-lens blocks so that all their selected unfolded copies lie on one line in the affine span of the last baseline facet. Reflection of the half-chain produces a parallel line of target images, and matching images are joined by perpendicular segments of one common length. Rational rotations and translations make this construction exact. We then use the earlier

strict facet inequalities to realize the selected galleries and the finite score comparison to exclude shorter paths.

The notation in this section separates three levels of motion. The single-lens unfolding is written $C(\xi)$ using Cayley parameters. The product through one three-lens block is a rotation R_{blk}. After a rational alignment, the two uniform block routes become inverse screw motions $S_j^{\pm 1}$. Finally, Φ_i denotes a fixed template chart, Ψ_j a moving chart in the assembled half-chain, and $\mathcal{A}_j$ the ambient rigid motion used to place block j.

Fix $m \geq 1$ and let $k = 3m$. Use the k unperturbed first-half lenses of Section 4, with the parameters $t_i, \boldsymbol{L}_i$, the correction η_j, and $\kappa = 0$. Here k counts the lenses in one half, so the full reflected chain will have $2k$ lenses. Throughout this section, $i \in [k]$ indexes first-half lenses and $j = 0, \ldots, m-1$ indexes blocks, unless the full mirrored lens list is explicitly invoked; on that list the lens index runs from 1 to $2k$. For the positive uniform triple in block j, the moments are

$$\boldsymbol{B}_j = b_j \boldsymbol{e}_1, \qquad \boldsymbol{C}_j = -b_j \boldsymbol{e}_1 + d_j \boldsymbol{e}_2, \qquad b_j = aq_j, \quad d_j = a/q_j. \tag{8.1}$$

Let $\delta = \varepsilon^2$. Throughout this section m is fixed, and $O(\varepsilon^a)$ denotes a bound by $C_m \varepsilon^a$, uniform over the finite set of lenses and sign words. The final ε is an ordinary sufficiently small positive rational number. No bound on its binary encoding length is needed for this counting theorem.

8.1. **Rational lenses with prescribed unfolding rotations.** We prescribe the unfolding rotation of a lens using rational unit normals. All calculations use orthonormal coordinates in the three-dimensional facet hyperplanes. Normals are oriented with positive w component. Rational supporting equations need not have rational unit normals, so the construction specifies the latter directly.

For a rational Cayley parameter $\xi = (u, v, 0)$, define

$$C(\xi) := (I + [\xi]_\times)(I - [\xi]_\times)^{-1}, \qquad [\xi]_\times x = \xi \times x.$$

Here $[\xi]_\times$ is the matrix of the cross product with ξ. It is skew-symmetric, meaning that its transpose is its negative. The matrix $C(\xi)$ is the *Cayley rotation*; see [HLW06, §IV.8.3]. It has rational entries when u, v are rational. For $\xi \neq 0$ it rotates about the line in direction ξ through angle $2 \arctan |\xi|$. Rationality of the matrix does not mean rationality of that angle. The Cayley parameter ξ is local to this unfolding calculation and is unrelated to the scalar block scales q_j.

Consider two baseline hyperplanes in $\mathbb{R}^4$. In orthonormal coordinates (r_1, r_2, z, w) based at their intersection, suppose their unit normals are

$$n_- = (\mathbf{0}, -s, c), \qquad n_+ = (\mathbf{0}, s, c), \qquad c^2 + s^2 = 1, \quad c > 0, \quad s \neq 0.$$

An *orthonormal tangent frame* is a basis of three perpendicular unit vectors in the hyperplane. Together with an origin, it gives a distance-preserving affine coordinate chart. Use the tangent vectors $(\mathbf{0}, c, s)$ and $(\mathbf{0}, c, -s)$ for the third vectors of the two frames, with the standard first two transverse vectors. If a side hyperplane through the same base point has unit normal $n = (\boldsymbol{v}_r, z, w)$, then unfolding through this side plane has the matrix $C(\xi)$, where

$$\xi = \frac{s}{c + w} (\boldsymbol{v}_{r,2}, -\boldsymbol{v}_{r,1}, 0). \tag{8.2}$$

Changing the two transverse components of the normal to their negatives changes this matrix to $C(-\xi) = C(\xi)^{-1}$. Formula (8.2) follows by applying the rotation

$$\mathcal{R}(n, n') = I + Z + \frac{Z^2}{1 + n \cdot n'}, \qquad Z = n' n^{\mathsf{T}} - n (n')^{\mathsf{T}},$$

at the two hinges and multiplying in the displayed tangent frames. This ambient rotation takes n to n' and fixes the directions parallel to the common ridge. With a point of that ridge as origin, it carries the old facet hyperplane to the new one by the ridge-folding isometry; its inverse unfolds

the new facet into the old one. Its action on the normal two-plane is a rotation, not a reflection. All denominators are nonzero for the nearby normals used below.

Conversely, prescribed rational $\xi = (u, v, 0)$ can be realized with a rational unit side normal. Let

$$B := \frac{(1+c)(u^2+v^2)}{2s^2}, \quad \boldsymbol{v}_r := \frac{1+c}{s(1+B)}(-v, u), \quad z := \frac{sB}{1+B}, \quad w := \frac{1-cB}{1+B}. \tag{8.3}$$

Direct substitution gives $|\boldsymbol{v}_r|^2 + z^2 + w^2 = 1$ and (8.2). Write $\boldsymbol{L} = (L_1, L_2)$ and suppose

$$s = -\varepsilon + O(\varepsilon^3), \qquad \xi = \frac{\delta}{2}(L_2, -L_1, 0) + O(\delta^2).$$

Then we have:

$$\boldsymbol{v}_r = -\varepsilon \boldsymbol{L} + O(\varepsilon^3), \qquad z = O(\varepsilon^3), \qquad w = 1 + O(\varepsilon^2).$$

After division of the graph equations by ε, the leading terms of the rational-lens supports agree with the leading terms of the graph-lens supports, and the remaining coefficients differ by $O(\delta)$. Hence the rational lens is an $O(\delta)$ perturbation of the corresponding graph lens in the normalized support coordinates used in the comparison.

Here are explicit rational baseline normals and intercepts. For $i = 0, \dots, k$, let $\bar{p}_i = 2i - 2k$ and set

$$U_i = \frac{1}{1+\alpha_i^2}\begin{pmatrix} 1-\alpha_i^2 & -2\alpha_i \\ 2\alpha_i & 1-\alpha_i^2 \end{pmatrix}, \qquad \alpha_i = \frac{\varepsilon \bar{p}_i}{4}.$$

Write e_z, e_w for the coordinate unit vectors in the (z, w) plane. In these coordinates take the unit normal of baseline i to be $U_i^2 e_w$ and its unit tangent to be $U_i^2 e_z$. The baseline slope is

$$P_i = -\frac{(U_i^2 e_w)_z}{(U_i^2 e_w)_w} = \varepsilon \bar{p}_i + O(\varepsilon^3).$$

Set $b_0^{\mathrm{bas}} = \varepsilon k$ and

$$b_i^{\mathrm{bas}} = b_{i-1}^{\mathrm{bas}} + (P_{i-1} - P_i) t_i \,.$$

Then the consecutive baseline planes $w = P_i z + b_i^{\mathrm{bas}}$ meet at $z = t_i$. Their unit normal bisector, the direction halfway between their two unit normals, is $U_{i-1} U_i e_w$, so its orthonormal coordinates, and the numbers s, c in (8.2), are rational. With the baselines ordered by increasing longitudinal position, $s = -\varepsilon + O(\varepsilon^3)$ and $c = 1 + O(\varepsilon^2)$. The squared planar rotations U_i^2 were chosen precisely to make these bisecting frames rational. The side planes from (8.3) are placed through $(\mathbf{0}, t_i, P_{i-1} t_i + b_{i-1}^{\mathrm{bas}})$. Their two signs preserve exactly the four-facet incidence of a lens.

8.2. **One rational correction in each block.** The three unfolding rotations of a block must satisfy one additional alignment condition. We impose it by changing only the third lens, by an amount smaller than the leading lens rotation. This leaves the earlier strict inequalities available while placing the image of the first coordinate direction exactly in the transverse plane. The required correction reduces to selecting a rational point on an explicitly parametrized circle.

Write $\boldsymbol{L}_i = (L_{i,1}, L_{i,2})$. For the first two lenses of each block use

$$\xi_i^0 = \frac{\delta}{2}(L_{i,2}, -L_{i,1}, 0).$$

For the third lens we make an $O(\delta^2)$ rational correction so that the block rotation $R_{\mathrm{blk}} = C(\xi_3) C(\xi_2) C(\xi_1)$ satisfies

$$(R_{\mathrm{blk}})_{31} = 0. \tag{8.4}$$

The symbol R_{blk} is reserved for this three-lens rotation; the tent kernel continues to be denoted by $G(u, v)$. Since $(R_{\mathrm{blk}})_{31}$ is the third coordinate of $R_{\mathrm{blk}} \boldsymbol{e}_1$, this equation puts $R_{\mathrm{blk}} \boldsymbol{e}_1$ in the transverse plane. It permits the common-axis description of the two block routes in the next subsection.

Write $y = C(\xi_2)C(\xi_1)\boldsymbol{e}_1 = (y_1, y_2, y_3)$ and $\xi_3^0 = (u_0, v_0, 0)$. The vector y is rational and has norm one. Moreover $y_1 = 1 + O(\delta^2)$, $y_2 = O(\delta^2)$ and $y_3 = O(\delta)$. Its first nonzero term is

$$y_3 = \delta(L_{1,1} + L_{2,1}) + O(\delta^2).$$

For the first two reference vectors in each block, $L_{1,1} + L_{2,1} = -3/(10q_j) - (137/50)q_j - (9/25)q_j^3 < 0$. The leading term of y_3 is nonzero and the error is of strictly higher order in ε. Hence, after decreasing the fixed small parameter if necessary, the error cannot cancel the leading term and $y_3 \neq 0$. The desired uncorrected third vector already satisfies

$$(C(\xi_3^0)y)_3 = O(\delta^2),$$

because (8.1) gives $\sum_{i \text{ in block}} L_{i,1} = 0$.

For $\xi_3 = (u, v, 0)$, equation (8.4) is precisely

$$-2vy_1 + 2uy_2 + (1 - u^2 - v^2)y_3 = 0. \tag{8.5}$$

This rational circle has center $(y_2/y_3, -y_1/y_3)$ and radius $1/|y_3|$. Both are rational. The familiar parametrization of the unit circle by $(2t/(1+t^2), (1-t^2)/(1+t^2))$ therefore gives rational points on this circle. The following choice is sufficiently close to ξ_3^0 :

$$t = \frac{y_3u_0 - y_2}{2}, \qquad u = \frac{y_2 + 2t/(1+t^2)}{y_3}, \qquad v = \frac{-y_1 + (1-t^2)/(1+t^2)}{y_3}. \tag{8.6}$$

The equation follows by substituting these expressions and using $|y| = 1$. Since $t = O(\delta^2)$, we have $u - u_0 = O(\delta^5)$. Subtracting (8.5) at (u, v) and (u_0, v_0) then gives $v - v_0 = O(\delta^2)$: the coefficient of $v - v_0$ tends to -2. This proves the required bound directly from the explicit rational formula. Use (8.3) with this corrected third vector.

8.3. **A rational affine alignment lemma.** We next express the two uniform routes through a block in coordinates with a common invariant line. To keep the maps distinct, R denotes the linear rotation of the positive block route, K is a common outgoing alignment, T is a translation used only to change coordinates, and S is the resulting screw motion. Each route will rotate about this line and translate along it. Such an isometry is a *screw motion*. The two signs give inverse motions after a common change of the outgoing coordinates. Their translations along the line will then add when blocks are placed consecutively.

Let $J = \operatorname{diag}(-I_2, 1)$ on $\mathbb{R}^3$. This map reverses the two transverse coordinates and fixes the longitudinal coordinate. It is a half-turn about the third axis, so $J^2 = I$. It is not reflection in a single hyperplane. Use rational orthonormal tangent charts Φ_i of the baselines, with origin $(\mathbf{0}, 0, b_i^{\text{bas}})$, the standard first two vectors, and the third vector $U_i^2 e_z$. A chart sends a coordinate vector to its origin plus the corresponding linear combination of these three tangent vectors. Let $g(x) = Rx + t$ be the positive uniform three-lens unfolding of one block in these charts. This is an affine isometry: $R^{\mathsf{T}}R = I$ and t is a translation vector. The negative uniform route is obtained by reversing the two transverse coordinates before and after the positive route. In the chosen charts its unfolding map is therefore JgJ. Because J, R, and t are rational, both uniform unfolding maps have rational affine coefficients.

By (8.4), the first column of R is $(R_{11}, R_{21}, 0)$, a rational unit vector in the transverse plane. Thus the planar rotation

$$V = \begin{pmatrix} R_{11} & -R_{21} \\ R_{21} & R_{11} \end{pmatrix} \tag{8.7}$$

has rational entries, satisfies $V^{\mathsf{T}}V = I$, and sends $\boldsymbol{e}_1$ to $(R_{11}, R_{21})^{\mathsf{T}}$. Multiplying R on the left by $\operatorname{diag}(V^{\mathsf{T}}, 1)$ therefore rotates the transverse part of $R\boldsymbol{e}_1$ back to the first coordinate direction.

Hence $\operatorname{diag}(V^{\mathsf{T}},1)R$ fixes $\boldsymbol{e}_1$, and it has the form

$$\operatorname{diag}(V^{\mathsf{T}},1)R=\begin{pmatrix}1&0&0\\0&d&\beta\\0&-\beta&d\end{pmatrix},\qquad d=R_{33},\quad \beta=-R_{32},\quad d^2+\beta^2=1.$$

The 2×2 lower-right block is a planar rotation: d is its cosine parameter and β is its signed sine parameter. For the small lens rotations used here, d is close to 1 and $\beta\neq 0$. The important point is that both numbers are already rational matrix entries. We have therefore aligned the rotation axis without normalizing a vector whose length might introduce an unwanted square root.

Let $v=V^{\mathsf{T}}t_U$, where t_U consists of the first two coordinates of t, and define the rational affine isometries

$$K(x)=\operatorname{diag}(V,1)x+\left(t_3+\frac{\beta v_2}{1+d}\right)e_3,\qquad T(x)=x-\frac{v_2}{\beta}e_3. \tag{8.8}$$

Both commute with J. A direct calculation shows that

$$S=T^{-1}K^{-1}gT,\qquad S(x_1,x_2,x_3)=(x_1+p,\,dx_2+\beta x_3,\,-\beta x_2+dx_3),\qquad p=v_1. \tag{8.9}$$

The expression $T^{-1}K^{-1}gT$ first removes the common outgoing motion K and then changes coordinates by the translation T. The map S therefore expresses the same block isometry g in coordinates where its distinguished axis and base point have been normalized. In particular S is a screw motion about the first coordinate axis. Its rotation acts in the (x_2,x_3) plane, and p is its *axial displacement*, the translation along the axis in one application. This axis is transverse to the longitudinal direction of the lens chain. Screw motions with a common axis are called *coaxial*. The same calculation gives

$$g=KTST^{-1},\qquad JgJ=KTS^{-1}T^{-1}.$$

We also need the size of the motions in (8.8). The finite hinge calculation used in Lemma 7.1 gives, for one block,

$$R\;=\;I+\delta\left[\left(\sum_{i\text{ in block}}L_{i,2}\,,-\sum_{i\text{ in block}}L_{i,1}\,,0\right)\right]_{\times}+O(\delta^2),\qquad t_U\;=\;\delta\boldsymbol{B}_j+O(\delta^2),\quad t_3\;=\;O(\delta).$$

The subscript $\times$ denotes the skew-symmetric matrix defined above. By (8.1),

$$\beta\;=\;-\delta d_j+O(\delta^2),\quad v_2\;=\;O(\delta^2),\quad p\;=\;\delta b_j+O(\delta^2),\quad K\;=\;I+O(\delta),\quad T\;=\;I+O(\delta). \tag{8.10}$$

All error estimates are finite matrix estimates. The condition $\boldsymbol{B}_j\cdot(\boldsymbol{B}_j+\boldsymbol{C}_j)=0$ is what makes the translation in T small.

8.4. **Rigid assembly of the half-chain.** We now place the blocks so that their two uniform unfolding maps act as the coaxial inverse screws just obtained. A moving frame records the orthonormal coordinates of each outgoing baseline and determines the placement of the next block. Each placement is an ambient rigid motion, so it preserves the metric and the incidences within the block. We shall define the final convex polytope by its placed supporting halfspaces after completing the assembly.

There are two kinds of coordinate charts in this assembly. The maps Φ_i are the fixed tangent charts of the unplaced template lenses; the maps Ψ_j are the moving charts after the first j blocks have been placed. The ambient rigid motion $\mathcal{A}_j$ places block j so that its incoming template chart agrees with the current moving chart. All these charts map $\mathbb{R}^3$ isometrically onto baseline hyperplanes; they are coordinate devices, not additional dimensions of the polytope.

Start with $\Psi_0 = \Phi_0$, its upward unit normal μ_0, and source $v = \Psi_0(0)$. For block j, $0 \le j \le m-1$, let K_j, T_j, S_j be the maps just constructed. Place the block by the unique ambient rigid motion $\mathcal{A}_j$ such that

$$\mathcal{A}_j \Phi_{3j} T_j = \Psi_j \tag{8.11}$$

and its linear part carries the incoming normal to μ_j. Then set

$$\Psi_{j+1} = \mathcal{A}_j \Phi_{3j+3} K_j T_j, \qquad \mu_{j+1} = (\mathcal{A}_j)_{\mathrm{lin}} n_{3j+3}.$$

An ambient rigid motion is uniquely determined by the image of one point and of an orthonormal basis at that point. Equation (8.11) specifies the image of the incoming tangent frame, while the condition on the normal specifies the fourth orthogonal direction. These requirements therefore determine a unique ambient rigid motion $\mathcal{A}_j$. Every frame and every map in these formulas is rational, so $\mathcal{A}_j$ is rational. The incoming plane of the next block is exactly the outgoing plane of the preceding piece. The two prescribed routes, expressed in the placed frames, act by S_j and S_j^{-1}.

The assembly commutes with the ambient transverse sign reversal $(\boldsymbol{r}, z, w) \mapsto (-\boldsymbol{r}, z, w)$, so its baseline planes remain independent of $\boldsymbol{r}$. We verify the size of the placements. Here the *template* is the original collection of rational lenses before placement. The linear parts of K_j and T_j are $\mathrm{diag}(V_j, 1)$ and I, and their translations are longitudinal and $O(\delta)$. The first placement extends T_0^{-1} from the initial baseline. At each subsequent step, the additional change in the shared baseline chart is $K_{j-1} T_{j-1} T_j^{-1}$. Its ambient extension rotates only the transverse plane and translates along the baseline tangent. That tangent has vertical component $O(\varepsilon)$.

Induction over the fixed number of blocks therefore gives a transverse linear part $I + O(\delta)$, identity linear part on the (z, w)-plane, a longitudinal translation $O(\delta)$, and a vertical translation $O(\varepsilon^3)$ for each $\mathcal{A}_j$. In particular, $\mathcal{A}_j = I + O(\delta)$. Baseline normals are unchanged by these placements. A side normal has transverse component $O(\varepsilon)$, so it changes by $O(\varepsilon\delta) = O(\varepsilon^3)$. The offset also changes by $O(\varepsilon^3)$: the normal's longitudinal component is $O(\varepsilon)$ and the translation has the bounds just established. Thus every moved graph equation, after division by ε, differs from its template by $O(\delta)$ in all coefficients.

For a word $\sigma \in \{-1, 1\}^m$, the formal unfolded image in the final baseline is consequently

$$\nu_\sigma = \Psi_m(p_\sigma \boldsymbol{e}_1), \qquad p_\sigma = \sum_{j=0}^{m-1} \sigma_j p_j. \tag{8.12}$$

Here p_j is the axial displacement of block j in (8.9), so $p_j = \delta b_j + O(\delta^2)$ by (8.10). Equation (8.12) is exact, not asymptotic. In the normalized coordinates, each block acts on the common first coordinate axis by translation through $+p_j$ or $-p_j$ according to its sign. The composition therefore sends the origin to the point with first coordinate $p_\sigma = \sum_{j=0}^{m-1} \sigma_j p_j$ and with the other two coordinates zero. After applying the final chart Ψ_m, all unfolded images lie on the same affine line. The numbers p_σ are distinct for sufficiently small ε, because

$$p_\sigma - p_\tau = \delta a \sum_{j=0}^{m-1} (\sigma_j - \tau_j) q_j + O(\delta^2)$$

and different signed base-T sums are distinct.

The outgoing baseline has slope $O(\varepsilon^3)$ and a rational unit normal. A rational rotation in the (z, w) plane, of size $O(\varepsilon^3)$, makes this baseline horizontal. A rational longitudinal translation puts the source at $z = 0$. Apply these common motions to the assembled half-chain and keep the same notation. They preserve all the preceding identities and coefficient estimates.

8.5. **Reflection gives exact equality.** The collinear unfolded copies now give an equal-length family by a Euclidean reflection. The reflected images lie on a parallel line, and corresponding points are joined by perpendicular segments of one fixed length. Both original endpoints remain fixed throughout: the different points on the two lines are their unfolded images for different facet galleries.

Reflect the entire half-chain in the hyperplane $z = 1/2$, and let $\mathcal{R}(\boldsymbol{r}, z, w) = (\boldsymbol{r}, 1 - z, w)$. The horizontal outgoing baseline is fixed by this reflection. The second half therefore shares that baseline exactly. Let $x = \mathcal{R}v$ be the target. The reflected copies, unfolded backwards from x into the middle baseline, are exactly $\mathcal{R}\nu_\tau$.

The line in (8.12) has transverse direction $\Psi_{m,\mathrm{lin}}\boldsymbol{e}_1$ and longitudinal coordinate $a_\varepsilon = O(\delta)$. Its reflected line has the same direction and coordinate $1 - a_\varepsilon$. Their longitudinal coordinates differ by $D_\varepsilon = 1 - 2a_\varepsilon > 0$, so the two lines are distinct parallel lines and the reflected half-chain lies strictly on the other side of the midpoint. The transverse coordinate along either line is measured by the scalar parameter p_σ. Pythagoras in the plane spanned by the common line direction and the longitudinal direction therefore gives

$$|\nu_\sigma - \mathcal{R}\nu_\tau|^2 = D_\varepsilon^2 + (p_\sigma - p_\tau)^2. \tag{8.13}$$

When $\sigma = \tau$, the second term in (8.13) vanishes, so every matching pair has the same unfolded length D_ε. Distinct words have distinct values p_σ, and their perpendicular middle segments therefore meet the midplane $z = 1/2$ at distinct transverse points. Once the corresponding galleries are verified to be valid and globally shortest, these distinct midpoint crossings show that the construction contains 2^m genuinely different shortest paths.

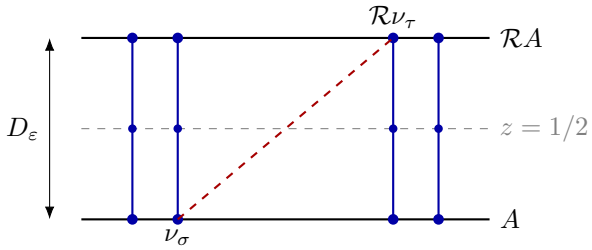


FIGURE 8.1. The two lines of unfolded images in the middle facet. The solid segments join matching labels and have common length D_ε. The dashed segment illustrates the additional transverse displacement $p_\sigma - p_\tau$ for a mismatch in (8.13).

8.6. **The convex polytope and the global comparison.** The algebraic placement has produced the desired candidate galleries and made their lengths equal, but two geometric points still require proof. First, the placed supporting planes must actually occur as facets of one convex polytope. Second, no other boundary path may be shorter than the candidate galleries. We verify these two assertions next. We first check which supporting functions are active. We then show that every shortest path stays on the lower graph and follows one of the lens sign words. Finally, the earlier strict inequalities validate all matching uniform words, while the mixed-block repair gap excludes mixed blocks. The exact identity (8.13) compares the remaining uniform words.

Take the maximum of all the assembled first-half and reflected second-half affine supporting functions as the lower graph. Intersect its epigraph with

$$|r_1|, |r_2| \le M\varepsilon, \qquad -1 \le z \le 2, \qquad w \le 10, \qquad M = 10(k+1).$$

Call the resulting polytope P_ε. It is a bounded rational convex 4-polytope and is full-dimensional for sufficiently small ε. The two points v, x are rational and lie in the interiors of the first and last baseline facets.

We now make the facet comparison under assembly quantitative. Write the graph equations, divided by ε, as $\widetilde{D}_i(z)$ for baselines and $\widetilde{A}_i^{\pm}(\boldsymbol{r}, z)$ for sides. The tildes distinguish the assembled coefficients from the original template data. Rigidly moving a whole lens preserves its incidence. Since the motions commute with $\boldsymbol{r} \mapsto -\boldsymbol{r}$, its equations have the exact form

$$\begin{aligned} \widetilde{A}_i^{\pm} &= (1-\lambda_i)\widetilde{D}_{i-1} + \lambda_i \widetilde{D}_i \pm \widetilde{\boldsymbol{L}}_i \cdot \boldsymbol{r}, \\ \widetilde{D}_i - \widetilde{D}_{i-1} &= 2\rho_i (z - \widetilde{t}_i). \end{aligned} \tag{8.14}$$

where

$$\lambda_i = \tfrac{1}{2} + O(\delta), \quad \rho_i = 1 + O(\delta), \quad \widetilde{\boldsymbol{L}}_i = \boldsymbol{L}_i + O(\delta), \quad \widetilde{t}_i = t_i + O(\delta)$$

for the full mirrored list. The numbers $\widetilde{t}_i$ remain strictly ordered. A side plane can exceed its two adjacent baselines only in $|z - t_i| \le C_m \varepsilon$. These slabs are disjoint. In each slab all nonneighboring supports have a fixed strict gap. Each listed baseline and each side support is active on an open patch of the lower graph. No two of these open patches represent the same supporting hyperplane. Hence there are exactly $6k+1$ lower facets, four transverse walls, two end caps and the top, giving

$$f_3(P_\varepsilon) = 6k + 8 = 18m + 8. \tag{8.15}$$

Here f_3 denotes the number of three-dimensional faces, namely the facets. There are $2k$ lenses in the full reflected chain.

Every shortest path stays on this lower graph. The axis path has height slope at most $(2k+1)\varepsilon$ for sufficiently small ε, so its length is at most $\sqrt{1 + (2k+1)^2\varepsilon^2}$. A path reaching a transverse wall has length at least $\sqrt{1 + 4M^2\varepsilon^2}$, by reflection in that wall after projection onto the corresponding (r_a, z) plane. Our choice of M makes this strictly longer. A path reaching an end cap has length at least 3, and the top is excluded by a still larger fixed gap.

A shortest path between these facet-interior endpoints avoids the faces of codimension at least two in the boundary, by MP's regularity property recalled in §4.1. It cannot revisit a facet: the straight chord in that convex facet would shorten a detour. There are therefore at most $6k$ lower-facet crossings. All the normals are $e_w + O(\varepsilon)$, and the exact tangent-jump formula (5.5) changes a base tangent by $O(\delta)$ at each crossing. Summing these finitely many changes and using the endpoint displacement gives $dz/ds > 1/2$ and $dr/dz = O(\delta)$ for small ε, where s in dz/ds is the arclength parameter. The transverse coordinates of every segment and ridge crossing are bounded by a constant times δ, while the longitudinal tangent component stays positive. Hence the whole path lies in $\boldsymbol{r} = O(\delta)$ and is strictly increasing in z. In this tube the lens neighborhoods are disjoint and retain the displayed four-facet incidence. A direct transition between the two baselines, or between the two side facets, would meet their common edge, which is a face of codimension two in the boundary. The strict inequalities exclude all other supports at that point. Consequently every shortest path chooses exactly one side in each of the $2k$ lenses, in their prescribed order.

We next compare all these words, including words whose formal unfoldings do not give valid paths. For a full lens sign word s, let d_s be the distance from its formally unfolded source to the target in the last baseline plane. We claim that, for any two words s, u,

$$d_s^2 - d_u^2 = -2\delta^2 (h_s - h_u) + O(\delta^3), \tag{8.16}$$

where h_s, h_u are the scores of the original unperturbed mirrored template. This estimate is uniform over the finite set of words. The estimate will separate words with unequal scores. The equal lengths of matching words continue to follow from the exact identity (8.13), rather than from equality of their leading score terms.

Here is a finite matrix proof that also explains why the small motions do not alter the leading comparison. Use canonical orthonormal charts on the moved baseline planes: their first two vectors are $\boldsymbol{e}_1, \boldsymbol{e}_2$, their third vector has positive z component, and their origins have base coordinates $(\boldsymbol{r}, z) = (\boldsymbol{0}, 0)$. To multiply affine maps as matrices, append the coordinate 1: a map $y \mapsto Ay + b$ acts on $(y, 1)$ by the block matrix $\left(\begin{smallmatrix} A & b \\ 0 & 1 \end{smallmatrix}\right)$. These are *homogeneous coordinates*. The extra entry

records translation and does not change the dimension of the polytope. In coordinates $(\boldsymbol{r}, z, 1)$, write the exact affine lens maps as

$$T_i^{s_i} = E_i + s_i O_i, \qquad E_i = \tfrac{1}{2}(T_i^+ + T_i^-), \qquad O_i = \tfrac{1}{2}(T_i^+ - T_i^-).$$

The even part E_i is unchanged when the lens sign is reversed. The odd part O_i changes sign. They are summands of the two isometry matrices, and need not themselves be isometries. Extend J by fixing the homogeneous coordinate. Since $T_i^- = JT_i^+J$, E_i preserves the transverse space and its complement spanned by the longitudinal and homogeneous coordinates, whereas O_i interchanges them. The two-hinge formula gives

$$E_i = I + O(\delta), \qquad O_i = \delta\Omega_i + O(\delta^2), \qquad \Omega_i(\boldsymbol{r}, z, \chi) = ((t_i\chi - z)\boldsymbol{L}_i,\ \boldsymbol{L}_i \cdot \boldsymbol{r},\ 0). \tag{8.17}$$

The symbol χ is the auxiliary homogeneous coordinate, equal to 1 on affine points. The map Ω_i is the leading coefficient of the odd part. Indeed, its rotation part has first-order angular vector $(L_{i,2}, -L_{i,1}, 0)$ and its transverse translation is $t_i\boldsymbol{L}_i$. These are precisely the finite fold expansions in the proof of Lemma 7.1. The coefficient changes under the symmetry-preserving assembly affect only the $O(\delta^2)$ remainder in the odd part; the leading positions and slope vectors are unchanged.

Apply the finite product $T_{2k}^{s_{2k}} \cdots T_1^{s_1}$ to the source, whose initial coordinates are $(\mathbf{0}, 0, 1)$. Terms with an odd number of odd factors are transverse. Terms with an even number have zero transverse component. A single odd factor takes the source off the longitudinal axis in a transverse direction. Two odd factors can contribute a longitudinal displacement. This accounts for the pairwise sign products below. For the final source coordinates this gives

$$\boldsymbol{r}_s = \delta\boldsymbol{B}_s + O(\delta^2), \qquad z_s = c_\delta + \delta^2 \sum_{1 \le j < i \le 2k} t_j s_i s_j \boldsymbol{L}_i \cdot \boldsymbol{L}_j + O(\delta^3),$$

where $c_\delta = O(\delta)$ is independent of s. The two-odd-factor term follows directly from $[\Omega_i\Omega_j(\mathbf{0}, 0, 1)]_z = t_j\boldsymbol{L}_i \cdot \boldsymbol{L}_j$. The target has zero transverse coordinates and longitudinal coordinate $1 + O(\delta)$ in its final chart. Because this coefficient tends to one as $\delta \to 0$, it is nonzero for all sufficiently small δ. Thus the coefficient of $s_i s_j \boldsymbol{L}_i \cdot \boldsymbol{L}_j$, for $j < i$, in its squared distance from the unfolded copy is

$$2\delta^2(t_i t_j - t_j) = -2\delta^2 G(t_i, t_j).$$

All diagonal terms and all terms with no odd factors are independent of s. Subtracting two squared distances proves (8.16). This proof uses only finite affine maps and their products.

We next justify the use of strict galleries after assembly. The exact incidence relations in (8.14), not just closeness of individual supporting coefficients, are needed here.

Lemma 8.1 (Strict galleries after assembly)**.** *Fix a finite collection of sign words s of the unperturbed mirrored template, each satisfying*

$$\alpha_i := s_i \boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i) > 0 \qquad (1 \le i \le 2k).$$

For all sufficiently small positive ε, each formal straight development folds back to a regular locally minimizing path on P_ε. Its transverse path and ridge crossings satisfy, uniformly over the collection,

$$\boldsymbol{r}(z) = \delta\boldsymbol{R}_s(z) + O(\delta^2), \qquad z_i^\pm = \widetilde{t}_i \pm \delta\alpha_i + O(\delta^2), \qquad \delta = \varepsilon^2.$$

Proof. The source expansion above gives terminal transverse slope $-\delta\boldsymbol{B}_s + O(\delta^2)$ for the formal segment. Each of the two hinges at lens i changes this slope by $-\delta s_i \boldsymbol{L}_i/2 + O(\delta^2)$, by (5.5) and (8.14). Indeed, the successive divided longitudinal gradients differ by $1 + O(\delta)$, and the side gradient is $s_i\boldsymbol{L}_i + O(\delta)$ in the transverse directions. The formal ridge intersections are well defined for small δ, since their longitudinal denominators tend to nonzero values. Folding backwards, as in the proof of Lemma 7.1, gives $\boldsymbol{r}(z) = \delta\boldsymbol{R}_s(z) + O(\delta^2)$ on baseline flights and at ridge crossings. The positional estimate also holds on each formal side segment: its longitudinal width is $O(\delta)$ and its transverse slope is $O(\delta)$.

For the transverse crossing coordinates $\boldsymbol{r}_i^-, \boldsymbol{r}_i^+$, the exact entrance and exit equations from (8.14) are

$$z_i^- = \widetilde{t}_i - \frac{s_i \widetilde{\boldsymbol{L}}_i \cdot \boldsymbol{r}_i^-}{2\lambda_i \rho_i}, \qquad z_i^+ = \widetilde{t}_i + \frac{s_i \widetilde{\boldsymbol{L}}_i \cdot \boldsymbol{r}_i^+}{2(1-\lambda_i)\rho_i}.$$

Both denominators are $1+O(\delta)$. Substituting the positional estimate, together with $\widetilde{t}_i = t_i + O(\delta)$ and $\widetilde{\boldsymbol{L}}_i = \boldsymbol{L}_i + O(\delta)$, proves the two stated crossing expansions. In particular, $z_i^+ - z_i^- = 2\delta\alpha_i + O(\delta^2) > 0$. The positive limiting gaps between consecutive lenses give positive baseline flights.

At either crossing, $s_i \widetilde{\boldsymbol{L}}_i \cdot \boldsymbol{r}_i^{\pm} = \delta\alpha_i + O(\delta^2) > 0$. The exact incidence equations then put the opposite side support and the unused adjacent baseline strictly below the two crossing supports. All nonneighboring supports retain the fixed gaps established above. Convexity places the segments joining consecutive crossings in their intended facets. Every other supporting inequality is strict for at least one endpoint, so the segment interiors lie in facet interiors. The transverse coordinates are $O(\delta)$, so the path avoids the truncating facets. It therefore realizes the prescribed gallery and is locally minimizing, since its development is straight. For finitely many words, take the minimum of their positive margins and the maximum of the finitely many remainder constants. One sufficiently small positive ε then works for all of them. □

For the mirrored template with $\kappa = 0$, equations (4.4) and (4.6) give maximum score $H_0 = 2F_*$, attained by every matching uniform word; any word with a mixed triple has score less than $H_0 - 2/25$. To check the strict margins, let $s^{(i)}$ be obtained from a maximizing word s by reversing its ith sign. Expanding the quadratic score gives

$$h(s) - h(s^{(i)}) = 2s_i \boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i) - 2G(t_i,t_i)|\boldsymbol{L}_i|^2.$$

The left side is nonnegative, so

$$\alpha_i = s_i \boldsymbol{L}_i \cdot \boldsymbol{R}_s(t_i) \geq t_i(1-t_i)|\boldsymbol{L}_i|^2 > 0.$$

Lemma 8.1 therefore applies simultaneously to all 2^m matching galleries. They are actual regular locally minimizing paths, and their lengths are exactly D_ε by (8.13). If u is one such matching word and s has a mixed triple, then (8.16) gives

$$d_s^2 - D_\varepsilon^2 = 2\delta^2(H_0 - h_s) + O(\delta^3) \;>\; \frac{4}{25}\delta^2 - O(\delta^3) \;>\; 0$$

for sufficiently small ε. Such a word cannot belong to a global shortest path. A shortest path must therefore have two uniform block words σ, τ. Its unfolding has length (8.13), which is at least D_ε. This proves global shortestness of every matching gallery. Different block words choose different side facets in at least one lens, so these galleries are pairwise distinct. Each matching path meets the relative interior of the middle baseline facet, and its prefix to that point is globally shortest between its own endpoints. Replacing that prefix by a shorter route would shorten the whole path, which is impossible. Hence the distinct points ν_σ in (8.12) are genuine source images for that facet in MP's terminology, introduced in §1.3.

Finally, choose a sufficiently small positive rational ε satisfying all the finitely many strict inequalities above. Every operation in the construction is rational: the baseline frames, side normals, correction (8.6), affine factors, rigid placements, reflection, box and endpoints. Every supporting hyperplane in the construction has rational coefficients, and the chosen endpoints are obtained by solving rational linear equations in those supports. Hence the polytope P_ε and the two endpoints v, x are all described by rational coordinates. The common length $D_\varepsilon = 1 - 2a_\varepsilon$ is rational because the moving-frame origin has rational coordinates. Distinct axial displacements give distinct paths, and monotonicity makes these paths simple. Set $R_m = P_\varepsilon$ and $(v_m, x_m) = (v, x)$. Together with (8.15), this proves Theorem 1.2. □

Remark 8.2 (*The complete source-unfolding algorithm*)**.** Algorithm 6.1 in [MP08] inserts one source image at each iteration and continues until the source unfolding is complete. The family in Theorem 1.2 has at least 2^m source images in one facet and $n = 18m + 8$ facets. The algorithm therefore makes at least $2^{(n-8)/18}$ insertions on this family, even with unit-cost exact real arithmetic. This counts operations of that complete unfolding algorithm, not operations required by every algorithm for a single distance.

9. Arithmetic structure and exact complexity

We prove the arithmetic theorems and exact upper bounds from Section 3. Throughout this section, the facets of the input polytope are indexed by $j \in [m]$, with inequality $a_j \cdot x \le b_j$ for facet F_j. When a gallery of length $r \le m$ is fixed, its facets are relabeled in traversal order as $F_1, \dots, F_r$; within that local discussion, $i \in [r]$ indexes gallery pieces and crossings, and the maps T_i develop its facets into the first facet. The parameter λ parametrizes the resulting straight segment. This convention separates combinatorial geometry from arithmetic-circuit bookkeeping: geometric indices always refer to facets or crossings, while the algebraic field notation is introduced separately when needed.

The proofs use a nonrepeating facet gallery and its unfolding. The resulting formulas give the sign verifier and prove the multiquadraticity and total positivity asserted in Theorem 3.1. All computational bounds use explicit binary rational input.

9.1. Facet pieces and the existential formula.

Fix $d \ge 2$ and write

$$P = \{x \in \mathbb{R}^d : a_j \cdot x \le b_j, \quad 1 \le j \le m\}, \qquad F_j = P \cap \{x : a_j \cdot x = b_j\},$$

with one inequality for each facet, after removing redundancies. Let $s, t \in \partial P$ and $k \ge 0$ be rational. A yes certificate needs only a boundary path of length at most k. It need not certify global shortestness. The following chord argument bounds the number of pieces required, see also [MP08, proof of Lemma 2.4].

Lemma 9.1. *A shortest path on ∂P is a union of at most m straight segments, each contained in a facet. More precisely, its intersection with any facet is either empty or a connected straight subpath.*

Proof. Let $\gamma : [0, L] \to \partial P$ be a shortest path, parametrized by arclength. If $\gamma(u), \gamma(v) \in F_j$ with $u < v$, convexity gives

$$[\gamma(u), \gamma(v)] \subset F_j \subset \partial P.$$

The restriction of a shortest path to a subinterval is shortest. Comparison with this chord therefore gives

$$v - u = \|\gamma(v) - \gamma(u)\|.$$

Equality in the length–chord inequality forces the intervening subpath to be the straight segment. Therefore the set of parameter values for which the path lies in one facet is connected. Since the facet is closed, $\gamma^{-1}(F_j)$ is a closed interval or is empty. These at most m intervals cover $[0, L]$. Starting at 0, choose at each stage an interval containing the current parameter and extending furthest to the right. This partitions $[0, L]$ into at most m subintervals, each assigned to one facet. The assertion is also valid when $s = t$, using one zero-length segment. □

Proof of Proposition 3.3. Introduce real points $x_0 = s, x_1, \dots, x_{m-1}, x_m = t$ and real numbers $\ell_1, \dots, \ell_m$. Require $x_i \in P$ for every i, and impose

$$\ell_i \ge 0, \qquad \sum_{\nu=1}^{d} (x_{i,\nu} - x_{i-1,\nu})^2 \le \ell_i^2, \qquad \sum_{i=1}^{m} \ell_i \le k. \tag{9.1}$$

For every segment index $i \in [m]$, impose also

$$\bigvee_{j=1}^{m} \big(a_j \cdot x_{i-1} = b_j \ \text{ and } \ a_j \cdot x_i = b_j\big). \tag{9.2}$$

The last condition puts both endpoints of each segment in one facet. Convexity places the entire segment in that facet and hence in ∂P. Consequently, every satisfying assignment describes a boundary path of length at most $\sum_{i=1}^{m} \ell_i \le k$. Conversely, Lemma 9.1 provides such an assignment whenever the distance is at most k. Zero-length segments at t supply any unused pieces.

The formula has polynomial size and polynomial-bit rational coefficients. Clearing denominators gives an instance of the existential theory of the reals. If a conjunction is preferred, replace each disjunction by variables z_{ij}, where $1 \le i, j \le m$, and the quadratic constraints

$$z_{ij}(z_{ij} - 1) = 0, \qquad \sum_{j=1}^{m} z_{ij} = 1,$$

$$z_{ij}(a_j \cdot x_{i-1} - b_j) = 0, \qquad z_{ij}(a_j \cdot x_i - b_j) = 0.$$

This finishes the proof of the first inclusion. The second inclusion $\exists\mathbb{R} \subseteq \mathsf{PSPACE}$ is Canny's theorem [Can88]. □

The formula includes equality at the threshold and endpoints on lower-dimensional faces. For a vertex input in fixed dimension, enumerating affinely independent d-tuples, testing their hyperplanes for support, and removing duplicates gives a polynomial-size facet description, so the same conclusion holds.

9.2. **Sign tests for radical circuits.** The arithmetic ingredient is a form of Tiwari's separation and approximation method [Tiw92]. More general algebraic-expression bounds appear in [B+09]. A circuit retains intermediate values without expanding their integer numerators and denominators. This distinction allows polynomial-size descriptions even when exponentially many accuracy bits would be needed in expanded form. An *algebraic integer* is a root of a monic polynomial with integer coefficients.

Lemma 9.2. *Suppose an arithmetic circuit has rational constants, additional inputs $\sqrt{r_1}, \dots, \sqrt{r_h}$ for positive binary-encoded rational numbers r_i, and operations $+, -, \times, \div$. If its divisions are defined, positivity of its output reduces in polynomial time to* PosSLP. *Its sign, including equality to zero, can therefore be determined in* $\mathsf{P}^{\text{PosSLP}}$.

Proof. For positive integers p, q, replace $\sqrt{p/q}$ by $\sqrt{pq}/q$. Represent every gate by a numerator–denominator pair. This eliminates divisions with only a constant-factor increase in circuit size. Since $\operatorname{sign}(A/B) = \operatorname{sign}(AB)$ when $B \neq 0$, it suffices to test positivity of

$$\alpha = C(\sqrt{a_1}, \dots, \sqrt{a_h}),$$

where the a_i are positive integers and C is division-free with integer constants. Let $N \ge 2$ bound the total description length of this circuit, including its constants and radicands.

Set

$$T = 2^{2^{2N}}, \qquad \gamma = 2^{-2^{3N}}, \qquad \delta = 2^{-2^{5N}}.$$

Every input has absolute value at most 2^N. At an addition or subtraction gate, replacing a bound M by $2M$ is sufficient, so the base-two logarithm of the magnitude bound increases by at most one. At a multiplication gate the bound is squared, so that logarithm at most doubles. Inducting over the at most N circuit gates gives the displayed doubly exponential value T as a valid uniform bound for every intermediate quantity. The same gate-by-gate estimates hold after

applying any field embedding, because embeddings preserve addition and multiplication. Thus every intermediate value has absolute value at most T under every embedding of the field

$$K = \mathbb{Q}(\sqrt{a_1}, \ldots, \sqrt{a_h}).$$

Here a field *embedding* is an injective field map fixing $\mathbb{Q}$.

The number α is an algebraic integer and $[K : \mathbb{Q}] \le 2^h \le 2^N$. The *field norm* $\mathrm{Norm}_{K/\mathbb{Q}}(\alpha)$ is the product of the images of α under all embeddings of K into $\mathbb{C}$. If $\alpha \neq 0$, the product of all its Galois conjugates is a nonzero integer, so its absolute value is at least one. Combining this lower bound with the upper bounds for the remaining conjugates gives

$$1 \;\le\; \left|\mathrm{Norm}_{K/\mathbb{Q}}(\alpha)\right| \;\le\; |\alpha| T^{[K:\mathbb{Q}]-1}, \qquad \text{so} \qquad |\alpha| \;\ge\; \gamma.$$

Each $\sqrt{a_i}$ has an approximation to absolute error at most δ represented by a polynomial-size rational arithmetic circuit. Indeed, start Newton iteration at $y_0 = 2^N$ and set

$$y_{j+1} = \frac{1}{2}\left(y_j + \frac{a_i}{y_j}\right).$$

After N iterations, $1 \le y_j/\sqrt{a_i} \le 2$. If $e = y_j/\sqrt{a_i} - 1$, the next relative error is $e^2/(2(1+e)) \le e^2/2$. A further $5N + 1$ iterations therefore give the required absolute error. All divisions in this iteration are by positive numbers.

Evaluate C on these rational approximations, retaining the answer as a rational circuit $\widetilde{\alpha}$. The same magnitude bound T holds for the perturbed computation. At each arithmetic gate, the largest input error increases by a factor of at most $2T$, and hence

$$|\widetilde{\alpha} - \alpha| \;\le\; (2T)^N \delta \;<\; \gamma/4.$$

The separation bound now gives

$$\alpha \;>\; 0 \quad\Longleftrightarrow\quad \widetilde{\alpha} \;>\; \gamma/2.$$

This equivalence includes the case $\alpha = 0$. The number $\gamma/2$ has a circuit of size $O(N)$, using repeated squaring. Clearing divisions in $\widetilde{\alpha} - \gamma/2$ reduces its positivity to one integer arithmetic-circuit sign test. Apply the same procedure to $-\alpha$ when the full sign is needed. □

The lemma does not claim polynomial time without an oracle. It reduces an exact sign, including zero, to the specified arithmetic comparison problem.

9.3. **Verification of unfolded galleries.** The guessed gallery determines fixed affine unfolding maps. We use the preceding sign tests to check the developed ridge intersections and the length, allowing degenerate crossings to handle arbitrary boundary endpoints.

Proof of Theorem 3.4. We first describe the unfolding maps and their arithmetic. Let $H_i = \{x : a_i \cdot x = b_i\}$ and $H_j = \{x : a_j \cdot x = b_j\}$ be the hyperplanes of adjacent facets. Choose a rational point $w \in H_i \cap H_j$, and let

$$u_i = \frac{a_i}{\sqrt{a_i \cdot a_i}}, \qquad u_j = \frac{a_j}{\sqrt{a_j \cdot a_j}}, \qquad v = u_i + u_j.$$

The affine reflection

$$R_{ij}(x) = w + \left(I - \frac{2vv^{\mathsf{T}}}{v^{\mathsf{T}}v}\right)(x - w) \tag{9.3}$$

fixes their common ridge pointwise. It maps H_j onto H_i and puts the image of F_j on the opposite side of the ridge from F_i. This is exactly the geometric operation needed to flatten the two adjacent facets into one common Euclidean hyperplane. Thus its restriction to F_j is the required unfolding. The vector v is nonzero because adjacent facet hyperplanes are not parallel.

Guess a sequence of distinct facets $F_1, \dots, F_r$, with $r \le m$, such that consecutive facets are adjacent, $s \in F_1$, and $t \in F_r$. These incidences can be checked in polynomial time in fixed dimension. Define the cumulative development maps

$$T_1 = I, \qquad T_{i+1} = T_i \circ R_{i,i+1}.$$

By composing the ridge reflections along the prefix of the gallery, T_i carries facet F_i isometrically into the affine span of the first facet. The developed facets $T_i(F_i)$ lie in H_1. Every entry of every T_i has a polynomial-size arithmetic circuit involving rational numbers and the square roots of the positive rationals $a_j \cdot a_j$. Their inverses have circuits of the same kind, since their linear parts are orthogonal. No square root of an intermediate algebraic expression is needed.

Consider the straight segment between the developed endpoints, parametrized by

$$z(\lambda) := s + \lambda(T_r(t) - s), \qquad 0 \;\le\; \lambda \;\le\; 1.$$

The parameter λ records position along this single developed segment; the verifier only has to determine where it crosses the developed ridges. For $1 \le i < r$, let

$$I_i = \{\lambda \in [0,1] : z(\lambda) \in T_i(F_i \cap F_{i+1})\}.$$

Writing this condition in the original coordinates using T_i^{-1} gives linear equations and inequalities in λ, with coefficients represented by the circuits above. Their solution I_i is a closed interval, possibly empty or a singleton. Its endpoints and its emptiness can be determined with polynomially many sign tests from Lemma 9.2. A coefficient is tested for zero before division by it.

Check whether there are parameters

$$0 \;\le\; \lambda_1 \;\le\; \cdots \;\le\; \lambda_{r-1} \;\le\; 1, \qquad \lambda_i \in I_i.$$

This can be done greedily: choose each parameter as the larger of the previous parameter and the current interval's lower endpoint, and reject if it exceeds the upper endpoint. Let $\lambda_0 = 0$ and $\lambda_r = 1$. The two endpoints of the piece $z([\lambda_{i-1}, \lambda_i])$ lie in $T_i(F_i)$, so convexity places the entire piece there. Ridge membership makes consecutive pieces agree when folded back. Their total length is $\|T_r(t) - s\|$. A final sign test checks

$$\|T_r(t) - s\|^2 \;\le\; k^2.$$

Every accepted sequence therefore gives an actual boundary path of length at most k.

For completeness, first suppose that s and t lie in facet interiors. A shortest path then crosses facet boundaries through ridge interiors and unfolds to a straight segment. These are the standard regularity and unfolding facts in [MP08, Corollary 1.4 and Lemma 1.7]. Its facet sequence has no repetitions by Lemma 9.1, so it is among the certificates just described.

For arbitrary endpoints, fix incident facets $F_s \ni s$ and $F_t \ni t$, and take $s_n \in \operatorname{relint} F_s$ and $t_n \in \operatorname{relint} F_t$ converging to s and t. Paths within the endpoint facets give

$$|\operatorname{dist}_{\partial P}(s_n, t_n) - \operatorname{dist}_{\partial P}(s, t)| \;\le\; \|s_n - s\| + \|t_n - t\| \longrightarrow 0.$$

There are finitely many nonrepeating facet sequences. Pass to a subsequence for which shortest paths between s_n and t_n have one fixed sequence. Its unfolding maps are fixed. Pass further to a subsequence along which their crossing parameters converge in the compact simplex $0 \le \lambda_1 \le \cdots \le \lambda_{r-1} \le 1$. The developed segments converge, and membership in each developed ridge is closed. Their limit is a valid certificate of length $\operatorname{dist}_{\partial P}(s,t)$. This is why the verifier allows zero-length pieces and equality between successive crossing parameters.

The nondeterministic guess has polynomial length, and the verifier runs in $\mathsf{P}^{\textsc{PosSLP}}$. This proves the first containment. If $\textsc{PosSLP} \in \mathsf{C}_h\mathsf{P}$ for a fixed h, then

$$\mathsf{NP}^{\textsc{PosSLP}} \subseteq \mathsf{NP}^{\mathsf{C}_h\mathsf{P}} \subseteq \mathsf{PP}^{\mathsf{C}_h\mathsf{P}} = \mathsf{C}_{h+1}\mathsf{P} \subseteq \mathsf{CH}.$$

The remaining containment $\mathsf{CH} \subseteq \mathsf{PSPACE}$ is standard.

For the recent refinement in (3.1), use the claimed containment PosSLP $\in \mathsf{BPP}^{\mathsf{PP}}$ from [Mei26]. Fixing the nondeterministic witness, amplify every oracle simulation so that the total error over all polynomially many adaptive queries is exponentially small. Merlin supplies the witness and Arthur performs these amplified simulations with a PP oracle. The nondeterministic machine can therefore guess the gallery and all combinatorial choices, while delegating only the exact radical sign comparisons to the PosSLP oracle. Hence

$$\mathsf{NP}^{\mathrm{PosSLP}} \subseteq \mathsf{MA}^{\mathsf{PP}} \subseteq \mathsf{PP}^{\mathsf{PP}} = \mathsf{C}_2\mathsf{P},$$

where the middle inclusion is the standard relativization of $\mathsf{MA} \subseteq \mathsf{PP}$. □

9.4. **Multiquadraticity and total positivity of squared distances.** The gallery verifier gives more than a sign test. A shortest gallery expresses the squared distance as a sum of squares in a field generated by square roots of rational normal norms. We use this expression to prove the arithmetic assertions of Theorem 3.1, including for endpoints on lower-dimensional faces.

Proof of Theorem 3.1. Let

$$q_i := a_i \cdot a_i\,, \qquad K_P := \mathbb{Q}(\sqrt{q_1}, \ldots, \sqrt{q_m}).$$

Because $q_i = a_i \cdot a_i$, the number q_i is the squared Euclidean length of the ith rational facet normal, and $\sqrt{q_i}$ is the normalizing factor used by the unfolding reflections. Consequently K_P is exactly the multiquadratic field generated by those facet-normal lengths, as defined in Theorem 3.1. Every affine reflection in (9.3) has coefficients in K_P. The proof of Theorem 3.4 provides a nonrepeating shortest gallery and a corresponding affine map T with

$$z = T(t) - s \in K_P^d\,, \qquad D^2 = z_1^2 + \ldots + z_d^2\,. \tag{9.4}$$

For general boundary endpoints this is the limiting gallery obtained there from facet-interior endpoints. Its maps are fixed along the chosen subsequence, so (9.4) remains an exact identity. It is not a limit in an increasing sequence of number fields.

The field K_P is generated by square roots of positive rationals, so every embedding $\sigma : K_P \to \mathbb{C}$ sends each generator to one of its two real square roots and therefore has real image. Since D^2 is the sum in (9.4), we have $D^2 \in K_P$. Moreover, $s \neq t$ implies $D > 0$, so at least one z_j is nonzero. Injectivity of σ gives

$$\sigma(D^2) = \sum_{j=1}^{d} \sigma(z_j)^2 > 0.$$

Every embedding of $\mathbb{Q}(D^2)$ into $\mathbb{C}$ extends to an embedding of K_P, since these are finite extensions in characteristic zero. Hence every conjugate of D^2 is positive, proving total positivity.

A nonrepeating gallery uses at most $m-1$ affine reflections. Their coefficients have polynomial-size circuits in the rational data and the roots $\sqrt{q_i}$. Composing the fixed-dimensional affine maps, forming the squared norm, and taking its principal square root therefore gives a polynomial-size circuit for D. These compositions use only polynomially many arithmetic operations, so the circuit is constructed in polynomial time. The construction uses the supplied gallery; it does not find that gallery. □

9.5. **Rational path lengths and irrational crossings.** A prescribed rational polygonal path gives a square-root-sum instance directly. A shortest path on a rational polytope need not have rational crossing points, so this observation does not replace the unfolding verifier. We give the reduction and an explicit example.

Rational polygonal paths. Suppose a path is supplied by rational vertices $q_0, \ldots, q_r \in \mathbb{Q}^d$, together with a rational threshold $k \geq 0$. Let D be the product of the denominators in the coordinates and in k, and let

$$A_i = \sum_{\nu=1}^{d} \big(D(q_{i,\nu} - q_{i-1,\nu})\big)^2, \qquad B = Dk.$$

These are nonnegative integers of polynomial binary length, and

$$\sum_{i=1}^{r} \|q_i - q_{i-1}\| \;\leq\; k \quad \Longleftrightarrow \quad \sum_{i=1}^{r} \sqrt{A_i} \;\leq\; B.$$

The left side is the Euclidean length of the prescribed rational polygonal path, while the right side is a sum of square roots of explicit nonnegative integers. Therefore comparing the path length with k is exactly one instance of SqrtSum, and the instance can be constructed in polynomial time. This is also the arithmetic issue in verifying a prescribed tour in the Euclidean traveling salesman problem. A polynomial-length list of visited vertices leaves an exact radical comparison to perform [ABKM09, §1.4].

For a convex polygon, the reduction already gives $\textsc{GeoDist}_2 \in \mathsf{P}^{\textsc{SqrtSum}}$: there are two simple boundary arcs between the endpoints, their vertices are rational, and two threshold queries suffice. On a higher-dimensional boundary the bend points of a shortest path are not part of the input and need not be rational.

Remark 9.3 (*An irrational crossing in dimension four*)**.** Consider the rational convex 4-polytope

$$P_0 \;=\; \big\{(x,y,z,w) : 0 \;\leq\; y \;\leq\; 10,\; -y \;\leq\; x \;\leq\; 10,\; |z| \;\leq\; 10,\; |w| \;\leq\; 10\big\}$$

and the rational points $s = (1,0,0,0)$ and $t = (-1,1,1,0)$. They lie in the relative interiors of the facets supported by $y = 0$ and $x + y = 0$. Unfold these facets along the ridge $x = y = 0$. In coordinates consisting of signed distance from that ridge and the coordinates z, w, the endpoints become $(1,0,0)$ and $(-\sqrt{2},1,0)$. The developed straight segment crosses the ridge at the point

$$q = (0, 0, \sqrt{2} - 1, 0)$$

of the original polytope, and its length is $\sqrt{4 + 2\sqrt{2}} < 3$. Every other facet is at Euclidean distance at least 9 from s, so a competing shortest path cannot reach one. The union of the two facets unfolds to a rectangular box, where the straight segment is uniquely shortest in the developed union of the two facets. Folding back is an isometry on each facet and along their common ridge, so uniqueness is preserved. Hence the unique shortest boundary path has an irrational ridge-crossing point, despite all input data being rational.

10. Exact encoding of square-root sums

We prove Lemma 3.6 and Theorem 3.5. A strip whose middle facet normal has squared norm a contributes a translation proportional to $\sqrt{a}$ after unfolding. Reflected pairs cancel the transverse translations. The rational moving-frame construction of Section 8 then makes the longitudinal translations add. The distance identity is exact, so no separation bound for square-root sums is needed.

We use $i \in [n]$ for the input radicands. After setting $m = 2n$ in §10.3, we use $j \in [m]$ for the individual strips. Each radicand A_i is represented by an integer vector of squared norm A_i. A strip is described by rational data (u, b, η): the vector u is perpendicular to the ridges within the incoming baseline, so their direction space is $u^{\perp} \times \{0\}$. The scalar b completes the middle normal, and η scales the perpendicular width of the middle strip, which is $\eta\sqrt{q_u a}$ in the notation below. After unfolding, these data produce a translation of length proportional to $\sqrt{A_i}$. The global parameters λ, σ, R_0 introduced later control, respectively, the scale of that translation, the small common tilt, and the size of the truncating box.

Fix positive radicands $A_1, \dots, A_n$ together with supplied vectors $v_i \in \mathbb{Z}^4$, $v_i \cdot v_i = A_i$, as in Lemma 3.6. Unfolding terminology follows [MP08, Defs 1.5–1.6 and Lemma 1.7].

10.1. **A two-hinge strip with an exact radical translation.** We first construct a *three-facet strip*, meaning three consecutive facets bounded by two successive ridges. Its two ridges are parallel, so the unfolding calculation reduces to a *two-dimensional normal section*, the plane perpendicular to their common ridge direction. In rational tangent coordinates on the incoming and outgoing facets, the linear part is the identity and the translation contains exactly one prescribed square root.

Work in coordinates $(x, w) \in \mathbb{R}^3 \times \mathbb{R}$. Let $u \in \mathbb{Q}^3$ be nonzero and $b \in \mathbb{Q}_{>0}$. Write

$$q_u := u \cdot u, \qquad a := b^2 + q_u.$$

Because $q_u = u \cdot u$, it is the squared Euclidean length of the ridge-normal component u. The other scalar b is the component of the middle-facet normal in the fourth coordinate direction. Therefore the full middle normal is $(-u, b)$ and its squared norm is $|u|^2 + b^2 = q_u + b^2 = a$. Define the rational orthogonal matrix

$$R(u,b) = \begin{pmatrix} I_3 - 2uu^{\mathsf{T}}/a & -2bu/a \\ 2bu^{\mathsf{T}}/a & (b^2 - q_u)/a \end{pmatrix}. \tag{10.1}$$

It fixes all $(y, 0)$ with $u \cdot y = 0$. On the plane spanned by $(u, 0)$ and e_w, it is a rotation through $2\arctan(\sqrt{q_u}/b)$. On this two-plane the displayed matrix is an ordinary planar rotation, and on its orthogonal complement it is the identity. Therefore it preserves the Euclidean inner product on all of $\mathbb{R}^4$, i.e. $R^{\mathsf{T}}R = I_4$.

Choose a rational $\eta > 0$ and set

$$c := \eta(bu, q_u). \tag{10.2}$$

Consider the three supporting hyperplanes

$$H_0 = \{w = 0\}, \qquad H_1 = \{-u \cdot x + bw = 0\}, \qquad H_2 = \{(Re_w) \cdot ((x, w) - c) = 0\}.$$

Their inward normals are, respectively,

$$e_w\,, \quad \frac{(-u, b)}{\sqrt{a}}\,, \quad Re_w\,.$$

The middle one bisects the other two. The first ridge contains 0. The second contains c. The ridges have common direction $u^\perp \times \{0\}$. The regions between them form the strip. Rational orthonormal charts for its baselines are

$$\Phi_0(x) = (x, 0), \qquad \Phi_2(x) = c + R(x, 0).$$

Lemma 10.1 (Exact strip development)**.** *Unfolding the strip into H_0, with the incoming chart fixed, sends*

$$\Phi_2(x) \longmapsto x + \eta\sqrt{a}\,u\,. \tag{10.3}$$

The developed middle facet lies between the parallel planes

$$u \cdot x = 0, \qquad u \cdot x = \eta\sqrt{a}\,q_u\,.$$

Proof. In the normal plane use the orthonormal basis $\{u/\sqrt{q_u}\,, e_w\}$. The middle facet makes angle $\theta = \arctan(\sqrt{q_u}/b)$ with the first baseline, and the last baseline makes angle 2θ. The first hinge rotates by θ and the second by the complementary amount needed to return the final baseline to the original orientation. Their product therefore cancels the matrix R. In the displayed incoming and outgoing charts, the complete two-hinge strip has identity linear part; only a translation remains.

The vector joining the two ridge base points is $c = \eta(bu, q_u)$ and has norm $\eta\sqrt{q_u a}$. It points along the middle facet and is perpendicular to the ridges. After unfolding, it is therefore the vector

$\eta\sqrt{a}\,u$ in H_0. Its two endpoints are the developed positions of the entering and exiting ridges, so this single vector calculation gives both the translation of the strip and the two developed ridge equations. The argument fixes every direction parallel to the ridges, so it proves the whole affine identity, not only its restriction to the normal plane. □

A reflected pair. Write $x = (r_1, r_2, z)$ and $u = (\boldsymbol{r}, \gamma)$ with $\gamma > 0$. Replace u in a second strip by $u^* = (-\boldsymbol{r}, \gamma)$, keeping b and η unchanged. Its squared normal norm is still a. The two translation vectors add to

$$\eta\sqrt{a}\,(u + u^*) = 2\eta\gamma\sqrt{a}\,e_z. \tag{10.4}$$

Choosing $\eta = \lambda/(2\gamma)$ makes the sum in (10.4) equal to $\lambda\sqrt{a}\,e_z$. The parameter η itself is rational because λ and γ are rational. Hence every supporting plane, coordinate chart, and rigid placement used to build the strip is rational. The irrational quantity $\sqrt{a}$ appears only after the two hinges are unfolded and the resulting rational data are interpreted as the Euclidean length of the developed translation.

We separate the two strips by a positive baseline interval to ensure the developed ridges occur in the prescribed order. We place a unit baseline interval between them. §10.5 verifies this order explicitly.

10.2. **Rational norm data and normalization.** We first explain how distance queries supply the integer vectors used in the construction. We then apply rational orthogonal transformations to make every fold turn slightly in the same longitudinal direction. These transformations preserve each squared norm exactly.

Obtaining the norm data. For each positive integer a, Lagrange's theorem guarantees that

$$\big\{v \in \mathbb{N}^4 \,:\, v \cdot v = a\big\}$$

is nonempty, see e.g. [HW08, §20.5] and [IR82, §17.7]. Each coordinate of such a vector has at most $\lceil \log_2(a+1) \rceil$ bits. The question whether a prescribed binary prefix extends to a vector in this set is in NP: guess the remaining bits and check the integer identity. By Theorem 1.1, one GeoDist$_4$ query decides this question. Starting with the empty prefix, test a zero next bit and retain it when an extension exists. Otherwise retain a one. At every step the current prefix has an extension, so this procedure finds a suitable vector after $O(\log(a+1))$ queries. All queries have polynomial binary length. Applying it to the input radicands supplies the norm data for Lemma 3.6.

Normalizing the direction. Square roots are approximated only to choose a rational orthogonal transformation. Orthogonality preserves the squared norm exactly, so the final distance has no approximation error.

Lemma 10.2 (Small rational tilt)**.** *Given $v \in \mathbb{Z}^4$ with $|v|^2 = a \geq 1$ and a rational $0 < \sigma \leq 10^{-3}$, one can compute in deterministic polynomial time a rational vector (u, b) of squared norm a. Writing $u = (\boldsymbol{r}, u_z)$ with $\boldsymbol{r} \in \mathbb{R}^2$, we have*

$$b \;>\; 0, \qquad \sigma \;\leq\; \frac{u_z}{b} \;\leq\; 2\sigma, \qquad \frac{|\boldsymbol{r}|}{b} \;\leq\; \sigma^2. \tag{10.5}$$

The running time and output size are polynomial in the input length and $\log(1/\sigma)$.

Proof. Permute coordinates and, if necessary, negate v, so that its last coordinate is positive and at least $\sqrt{a}/2$. Let $y = v/\sqrt{a}$. Let $h = 3\sigma/4$ and let

$$y_* = \left(\mathbf{0}, \frac{2h}{1+h^2}, \frac{1-h^2}{1+h^2}\right).$$

This is an exactly rational unit vector. Its ratio of the third to fourth coordinate is between 1.49σ and 1.51σ.

Compute a rational $\widetilde{y}$ with $|\widetilde{y}-y| \le \sigma^2/10^4$. This uses a square-root approximation of polynomial precision. For example, choose a power of two $D \ge 2\cdot 10^4/\sigma^2$, let $r = \lfloor D\sqrt{a}\rfloor/D$, and take $\widetilde{y} = v/r$. Increasing D by a fixed factor supplies the stated bound. The integer square root is computed by bisection.

Let $z = \widetilde{y} + y_*$ and define

$$Q \;:=\; -I_4 \;+\; \frac{2zz^{\mathsf{T}}}{z\cdot z}\,.$$

The matrix Q is a Householder-type reflection, so it is rational and exactly orthogonal. If we could use the exact vector $z_0 = y + y_*$, the corresponding reflection would send y exactly to y_*. We instead use the rational approximation $z = \widetilde{y} + y_*$. The hypotheses $y_w \ge 1/2$ and $y_*\cdot y > 0.49$ imply $|z_0| > 1$, so perturbing the normal vector from z_0 to z changes its normalized direction by a controlled amount. The elementary normalization estimate

$$\left|\frac{z}{|z|} - \frac{z_0}{|z_0|}\right| \;\le\; \frac{2|z-z_0|}{|z_0|}$$

shows that $|Qy - y_*| \le 8|\widetilde{y}-y| < \sigma^2/1000$.

Set $(u,b) = Qv$. Orthogonality gives $|u|^2 + b^2 = a$ exactly. The last coordinate of Qy exceeds 0.99, and the preceding approximation gives (10.5). All rational numbers used have polynomial bit length. Neither an exact value of $\sqrt{a}$ nor a square-root-sum comparison is computed. □

10.3. **Rational assembly and the developed endpoints.** We assemble two strips for each radicand, inserting a unit baseline gap between consecutive strips. The incoming and outgoing charts are rational. Since each strip has identity developed linear part, the affine composition adds its translation vector in the same developed coordinates. We now give the rational supporting data and the endpoints.

We use two strips for each radicand, so set

$$m := 2n, \qquad \lambda := \frac{1}{nA}, \qquad \sigma := \frac{\lambda}{10^{12}(m+2)^4}, \qquad R_0 := 100(m+2). \tag{10.6}$$

Here m is the total number of strips, λ is the common scale of the radical translations, σ is a much smaller tilt parameter used to order all folds, and R_0 is the half-width of the large truncating box. These roles, rather than the particular constants, are what matter in the subsequent estimates. For each A_i, apply Lemma 10.2 and write the resulting vector as (u_i, b_i), where $u_i = (\boldsymbol{r}_i, \gamma_i)$. Use the following two *strip records*, where a strip record simply means the data triple (u, b, η) specifying one strip:

$$(u_i, b_i, \eta_i), \quad (u_i^*, b_i, \eta_i), \qquad \eta_i = \frac{\lambda}{2\gamma_i}, \quad u_i^* = (-\boldsymbol{r}_i, \gamma_i).$$

List all m records in pair order as (u_j, b_j, η_j). In the formulas below, $a_j = b_j^2 + |u_j|^2$ is the corresponding input radicand, repeated twice. Set $q_j = |u_j|^2$, $R_j = R(u_j, b_j)$, and $c_j = \eta_j(b_j u_j, q_j)$.

An ambient rational orthonormal frame M together with an origin o gives the incoming baseline chart $x \mapsto o + M(x, 0)$. Start with $M_0 = I_4$, $o_0 = 0$, and the supporting halfspace $w \ge 0$.

For strip $j \in [m]$, add the middle supporting halfspace

$$M_{j-1}(-u_j, b_j)\cdot(p - o_{j-1}) \;\ge\; 0. \tag{10.7}$$

Then let

$$M_j = M_{j-1}R_j, \qquad o_j^* = o_{j-1} + M_{j-1}c_j \tag{10.8}$$

and add the outgoing baseline halfspace

$$(M_j e_w)\cdot(p - o_j^*) \;\ge\; 0. \tag{10.9}$$

If $j < m$, place the next strip after a unit gap by setting

$$o_j = o_j^* + M_j e_z.$$

The incoming baseline of strip $j+1$ is the same supporting hyperplane as the outgoing baseline of strip j. Only its chart origin changes. Every number in these constructions is rational.

There are $2m+1$ lower supporting planes. Intersect their inward halfspaces with

$$|r_1|, |r_2|, |z| \ \le \ R_0, \qquad w \ \le \ R_0, \tag{10.10}$$

and call the resulting set P. We call the facets arising from these box inequalities the *truncating facets*, and the box $|r_1|, |r_2|, |z| < R_0$ in the base coordinates the *base box*. Define

$$s = -e_z, \qquad t = o_m^* + M_m e_z. \tag{10.11}$$

We will prove that P is a full-dimensional convex polytope and that these points lie in its first and last lower facets.

Exact endpoint development. Lemma 10.1 gives translation $\eta_j \sqrt{a_j} u_j$ for strip j in its incoming coordinates, and each gap contributes e_z. The final chart origin develops to

$$D \ = \ (m-1)e_z \ + \ \sum_{j=1}^{m} \eta_j \sqrt{a_j} u_j \ = \ \Big(m - 1 + \lambda \sum_{i=1}^{n} \sqrt{A_i}\Big) e_z \, .$$

The initial unit baseline segment places the developed source one unit before the first strip, at $-e_z$. The strip translations then sum to D, and the final unit baseline segment places the developed target one unit beyond that total translation. Hence the developed target is $D + e_z$. Their developed distance is exactly

$$L \ = \ m + 1 \ + \ \lambda \sum_{i=1}^{n} \sqrt{A_i} \, , \qquad m+1 \ < \ L \ \le \ m+2. \tag{10.12}$$

The strict lower bound uses positive radicands. This computation does not yet prove global shortestness: the planes must define the intended boundary strip, and the developed straight segment must stay in its facets. The next two subsections prove these facts.

10.4. **Convexity and the order of all facets.** All lower normals remain close to vertical. We show that consecutive graph slopes increase in the longitudinal direction and that their ridges stay ordered throughout the transverse box. Each proposed support therefore defines a facet of the halfspace intersection, in the required order.

All matrix norms below are operator norms. For one record, write $v = u/b$. From (10.5),

$$|v| \ \le \ 3\sigma, \qquad \sigma \ \le \ v_z \ \le \ 2\sigma, \qquad |v_r| \ \le \ \sigma^2.$$

Equation (10.1) implies $|R_j - I| \le 6\sigma$. Telescoping products of orthogonal matrices therefore gives

$$|M_j - I| \ \le \ 6m\sigma. \tag{10.13}$$

Every inward unit normal of a lower support lies within $9m\sigma$ of e_w, and in particular has positive w component at least $3/4$. Write its graph plane as

$$w = f_\ell(r, z) = g_\ell \cdot (r, z) + b_\ell, \qquad 0 \ \le \ \ell \ \le \ 2m.$$

We have $|g_\ell| \le 12m\sigma$.

Lemma 10.3 (Ordered ridges)**.** *Throughout the box $|r_1|, |r_2| \le R_0$, the planes of equality between consecutive supports have equations $z = \beta_\ell(r)$, $0 \le \ell < 2m$, with*

$$-\tfrac{1}{2} \ < \ \beta_0(r) \ < \ \cdots \ < \ \beta_{2m-1}(r) \ < \ m + \tfrac{1}{2}, \qquad \beta_{\ell+1}(r) - \beta_\ell(r) \ \ge \ \lambda/5. \tag{10.14}$$

Every lower support is active on an open set. On the interval between its two consecutive ridges, it is the maximum of all lower supports.

Proof. We verify the explicit choice in (10.6) by quantitative estimates.

In local coordinates the three consecutive unit normals are

$$e_w, \qquad \frac{(-v,1)}{\sqrt{1+|v|^2}}, \qquad \frac{(-2v,1-|v|^2)}{1+|v|^2}.$$

Each of the two successive normal differences is $(-v,0)+E$, with $|E| \le 25\sigma^2$. For the first difference the error norm is $\sqrt{1+|v|^2}-1 \le |v|^2/2$. For the second, subtract the first difference from the total difference, whose error relative to $(-2v,0)$ has norm at most $2|v|^2$. After multiplication by M_{j-1}, the error relative to $(-v,0)$ is at most $50m\sigma^2$, by (10.13).

For unit normals μ, μ' with graph slopes $g = -\mu_{\text{base}}/\mu_w$, one has

$$g' - g \;=\; -\frac{\mu'_{\text{base}} - \mu_{\text{base}}}{\mu'_w} + \frac{\mu_{\text{base}}(\mu'_w - \mu_w)}{\mu_w \mu'_w}.$$

We now substitute the estimates for the two normal vectors into the exact formula for $g'-g$. The base components of μ and μ' are $O(m\sigma)$, their fourth coordinates are bounded below by $3/4$, and the difference $\mu'-\mu$ is $(-v,0)$ up to an $O(m\sigma^2)$ error. Using these bounds term by term gives

$$|(g'-g)-v| \;\le\; 200m\sigma^2.$$

More explicitly, after subtracting the target vector v, the three contributions in the preceding identity are bounded by $36m\sigma^2$, $(200/3)m\sigma^2$, and $800m^2\sigma^3$, respectively. The chosen value of σ makes the cubic term smaller than the spare room in the constant $200m\sigma^2$. Therefore the complete error is at most $200m\sigma^2$. Since $v_z \ge \sigma$ and $|v_r| \le \sigma^2$, every successive slope difference satisfies

$$(g'-g)_z \;\ge\; \sigma/2, \qquad |(g'-g)_r| \;\le\; 250m\sigma^2, \qquad \frac{|(g'-g)_r|}{(g'-g)_z} \;\le\; 500m\sigma. \tag{10.15}$$

To locate the ridges, set $h_j = \lambda\sqrt{a_j}/2$. Then

$$\lambda/2 \;\le\; h_j \;\le\; \tfrac{1}{2}, \qquad \sum_{j=1}^{m} h_j \;\le\; 1.$$

(The stronger bound $h_j \le 1/(2n)$ also holds.) From $u_z/b \ge \sigma$, $|u_r|/u_z \le \sigma$, and $|u|/b \le 3\sigma$, equation (10.2) gives

$$|c_j - h_j e_z| \;\le\; 12h_j\sigma, \qquad |c_j| \;\le\; 2h_j.$$

Combining with (10.13),

$$|M_{j-1}c_j - h_j e_z| \;\le\; 24m\sigma h_j, \qquad |M_j e_z - e_z| \;\le\; 6m\sigma. \tag{10.16}$$

The translation estimate controls the displacement accumulated inside one strip, while the frame estimate controls the direction of each inserted unit gap. Summing these errors over at most m preceding strips and gaps shows that every designated ridge point p, whether an entering origin o_{j-1} or an exiting origin o_j^*, differs by at most

$$E_0 = 30m^2\sigma$$

from its ideal position Ze_z, where Z is the number of preceding unit gaps plus the sum of preceding h_j's. All these ideal positions belong to $[0,m]$. The actual successive z coordinates are separated by at least $\lambda/4$: a middle-facet step advances by at least $h_j/2$, and a gap advances by at least $1/2$, using (10.16).

The equality of the two support planes meeting at p is

$$z = p_z - \frac{(g' - g)_r}{(g' - g)_z} \cdot (r - p_r).$$

For $|r_1|, |r_2| \le R_0$, its height differs from p_z by at most

$$J_0 = 1500 m\sigma R_0 \ < \ \lambda/1000.$$

The bound on the ridge tilt in (10.15) shows that moving across the whole transverse box can change a ridge height by at most J_0. Here we used $|r| \le 2R_0$ and $E_0 \le R_0$. In addition, $E_0 < \lambda/1000$ and $J_0 < \lambda/1000$ by the parameter choice. Two consecutive actual designated ridge heights are separated by at least $\lambda/4$, so the two perturbations can reduce their separation by at most $2J_0$. The actual separation is therefore at least $\lambda/4 - 2J_0 > \lambda/5$. The same estimates at the two ends give the extreme bounds in (10.14).

Finally, for fixed r the graph slopes in z increase strictly. We have

$$f_{\ell+1}(r, z) - f_\ell(r, z) = (g_{\ell+1} - g_\ell)_z \big(z - \beta_\ell(r)\big).$$

The ordering of the β_ℓ implies, by summing these differences, that f_ℓ is larger than every earlier and later support between its adjacent ridge heights. All intervening intervals are nonempty, so every plane has an open active patch. □

Lemma 10.3 shows that, for each fixed transverse point, the active lower supports occur in the intended order and every support is active on a nonempty interval. Hence the lower boundary of P is exactly the graph of $f = \max_\ell f_\ell$. We next bound its height uniformly. On the whole base box, $|f| \le E_0 + 36 m\sigma R_0 < R_0/2$. To obtain this estimate, fix one lower support. It contains a designated ridge point whose height differs from its ideal height by at most E_0 and whose base norm is at most $m + E_0$. Moving from that point to any point of the base box changes the support value by at most the slope norm $12m\sigma$ times the base displacement. The chosen parameter bounds make the sum of these two contributions smaller than $R_0/2$.

The preceding estimate gives a uniform upper bound on the height of every point of the lower graph inside the base box. The chosen top-plane height is larger than this bound, so the top plane does not meet the lower graph and lies strictly above it throughout the box. The polytope is bounded and full dimensional. It has precisely $2m + 1$ lower facets, four transverse walls, two longitudinal caps, and the top: altogether $2m + 8 = 4n + 8$ facets.

The endpoints (10.11) are inside the truncation box. The source is placed one unit of developed longitudinal distance before the first strip, so it precedes the first ridge. In the final baseline chart the target is the point e_z measured from the last outgoing origin. The final translation has positive longitudinal component $u_{m,z} > 0$, so this target lies strictly beyond the last ridge. Lemma 10.3 makes all other lower inequalities strict there. Both endpoints are therefore in facet interiors.

10.5. **The valid straight development is globally shortest.** The ordered supports define a convex boundary whose nonconsecutive lower facets do not meet. We check that the developed segment crosses the ridges in order. Its development, together with the distance to the truncating facets, then proves global shortestness.

Crossing order within a pair. At the beginning of an input pair, its incoming origin develops to ce_z for some real c. Write

$$u = \gamma(v, 1), \qquad |v| \ \le \ \sigma, \qquad h = \lambda\sqrt{A_i}/2.$$

The first strip contributes $h(v, 1)$, the gap contributes e_z, and the reflected strip contributes $h(-v, 1)$. By Lemma 10.1, their four ridge intersections with the developed axis $x = \theta e_z$ occur at

$$c, \quad c + h(1 + |v|^2), \quad c + h + 1 - h|v|^2, \quad c + 2h + 1. \tag{10.17}$$

Subtracting consecutive values in (10.17), the first and third intervals both have length $h(1+|v|^2) > 0$. The middle interval has length $1 - 2h|v|^2$, which is positive because $h \le 1/2$ and $|v| \le \sigma < 1$. After the fourth crossing, the construction inserts another unit baseline gap before the next pair. The initial and final baseline intervals also have length one. Every consecutive difference is therefore positive, so the developed straight segment from $-e_z$ to $D + e_z$ crosses all prescribed ridges in their stated order.

Fold the developed straight segment back through the supporting hyperplanes. Consecutive folded pieces share the prescribed ridge point, so they join continuously, and unfolding preserves their total length L. Every point of the resulting polygonal path lies within Euclidean distance L of the source s; hence every coordinate has absolute value at most $L + 1 \le m + 3 < R_0$, so the entire path stays inside the transverse truncation box. Consider one segment between two consecutive ridge crossings. By construction it lies in the supporting hyperplane of the intended facet and satisfies the two adjacent ridge inequalities. Lemma 10.3 says that, between those two ridge heights, these adjacent inequalities force all other lower supports to lie below the intended support. Therefore the segment lies in the actual facet of the polytope, not merely in its supporting hyperplane. At every internal crossing the other supports and all truncation constraints are strict. This is a regular boundary geodesic, of length exactly L.

All competing paths. Every truncating facet in (10.10) is at Euclidean distance at least $R_0 - 1$ from the source, and the parameter choice gives $R_0 - 1 > L$. Reaching any truncating facet would therefore already require more length than the entire constructed path. Hence every boundary path of length at most L avoids all truncating facets and remains on the lower graph over the open base box.

The lower facets occur in a chain, and nonconsecutive ones do not meet in this region. The successive unfolding maps agree on each common ridge and therefore define a continuous development into $\mathbb{R}^3$. This map is a local intrinsic isometry on facet interiors and across ordinary two-facet hinges. The compact image of any rectifiable curve is covered by finitely many such isometric neighborhoods. A finite subdivision of its parameter interval places each subcurve in one neighborhood, where its length is preserved. Adding these lengths proves length preservation for the whole curve, even when it crosses ridges infinitely often. No global injectivity of the development is needed.

Under the continuous development just described, the source and target become $-e_z$ and $D + e_z$. Any rectifiable developed curve joining these two points has length at least their Euclidean distance, which is L. The constructed developed curve is the straight segment between them and has exactly that length. Since development preserves lengths for all competing curves that avoid the truncating facets, the constructed boundary path is globally shortest. Therefore

$$\operatorname{dist}_{\partial P}(s, t) = L.$$

This proves the exact identity in Lemma 3.6, including equality cases in (3.2).

10.6. **Binary size and the rational threshold.** All radicals above occur in the proof of the metric identity. The actual output consists only of rational inequalities and endpoints. We verify polynomial bit size and construct the rational threshold used in the reduction.

The parameters λ, σ, R_0 have bit length $O(\log n + \log A)$. Lemma 10.2 has polynomial bit complexity. Its rational orthogonal matrices preserve A_i exactly. The matrices R_j, vectors c_j, and records η_j therefore have polynomial bit length. Products of m matrices of fixed size and sums of their translates also have polynomial bit length: one can use the product of their denominators as a common denominator, and numerator bit lengths grow by the sum of the individual bit lengths plus $O(m)$.

There are $4n+8$ inequalities. Formulas (10.7)–(10.11) compute them and the endpoints by polynomially many rational operations. For the distance query use

$$k = 2n + 1 + \lambda B. \tag{10.18}$$

No exact radical comparison occurs in constructing this rational threshold.

10.7. **The deterministic oracle reduction.** The exact distance identity now gives the Turing reduction in Theorem 3.5.

Proof of Theorem 3.5. Delete zero radicands and decide an empty sum directly. The oracle preprocessing in §10.2 supplies the auxiliary vectors required by Lemma 3.6 using polynomially many GeoDist$_4$ queries. Apply that theorem to obtain P, s, t, and set

$$k = 2n + 1 + \lambda B, \qquad \lambda = \frac{1}{n \max_{1\le i\le n} A_i}.$$

The input data and k have polynomial binary encoding length. By the exact distance formula,

$$\sum_{i=1}^{n} \sqrt{A_i} \;\le\; B \quad \Longleftrightarrow \quad \operatorname{dist}_{\partial P}(s,t) \;\le\; k.$$

One final distance query therefore decides the original instance, including equality at the threshold. Both the number of queries and the total binary length of the constructed instances are polynomial. This completes the proof. □

Remark 10.4. In the proof above, the preprocessing uses Theorem 1.1, while the final query uses the exact geometric encoding. The oracle answers are therefore tied to explicitly constructed rational polytopes, not to an assumed algebraic property of SqrtSum. Thus the argument does not infer SqrtSum-hardness from NP-hardness alone, and it does not assume SqrtSum ∈ NP.

11. Final remarks

11.1. **The role of convexity.** There are other notable examples in which convexity imposes a strong geometric restriction yet fails to prevent computational hardness as the dimension grows. Notable examples include $\exists\mathbb{R}$-completeness of realizations of convex 4-polyhedra [Ric96] and NP-hardness of finding small triangulations of 3-polytopes [BDR04]. This also a theme of our series of papers with Nguyen [NP18, NP21, NP22]. There, we show that finding and counting integer points in convex polyhedra (after their projections, intersections and complements) can be NP- and #P-hard, respectively, already in a small fixed dimension.

11.2. **Distances in permutation groups.** A *configuration graph* has one vertex for each configuration and one edge for each allowed local move. Such a graph can have exponentially many vertices even when configurations and local moves have short descriptions. Its distance problems separate the length of a path certificate from the size of the graph, providing a discrete comparison with our facet galleries. One nice family of such examples is the class of Cayley graphs of permutation groups.

Let S_n be the group of permutations of n elements, and let $\langle A\rangle$ be the subgroup generated by $A \subseteq S_n$. Its *directed Cayley graph* has vertices $g \in \langle A\rangle$ and edges $g \to g\sigma$ for $\sigma \in A$. For $\pi \in \langle A\rangle$, minimum word length is the distance from the identity to π in this graph. Inverse moves are available only when the corresponding inverses belong to A. A threshold k is *unary* when its encoding uses k symbols, rather than the $O(\log(k+1))$ bits of a binary encoding.

Even and Goldreich proved NP-hardness [EG81]. With a unary threshold the problem is NP-complete, while Jerrum proved PSPACE-completeness for binary thresholds even with two generators [Jer85]. This latter theorem concerns positive words. For connected undirected Cayley graphs, Lohrey and Rosowski prove NP-hardness of distinguishing distances k and $k+1$, with unary k, even for commuting involutions [LR23, Thm 4.1].

Finally, in *token swapping*, each vertex of a graph carries one token, and a move exchanges the tokens on adjacent vertices. Finding a shortest sequence reaching a specified placement is known to be NP-complete [M+16]. For a connected underlying graph, these configuration graphs are Cayley graphs of S_n generated by the transpositions corresponding to its edges.

11.3. **Distances in geometric graphs.** A *flip* replaces a diagonal of a convex quadrilateral formed by two adjacent triangles with the other diagonal. The *flip distance* between triangulations is the minimum number of flips joining them. See [STT88] for the remarkable bound on the diameter in the convex polygon case, and [Pak09, §14] for further background and higher-dimensional analogues. The flip distance is NP-complete for planar point sets and polygons with holes [LP15] and NP-complete for simple polygons [AMP15] and, more recently, convex polygons [Dor26].

Most recently, Black and Steiner obtain NP-hard edge distance and monotone-path problems for simple polytopes described by rational inequalities [BS26]. A d-polytope is *simple* if each vertex belongs to exactly d facets. A path is *monotone* for a linear objective when that objective increases along each edge. These polytope results use variable dimension and count edges. In fixed dimension the explicitly described vertex-edge graph has polynomial size, whereas our Euclidean boundary paths cross facet interiors.

11.4. **Complexity of radical circuits.** The problem GeoSign asks whether the output of a valid radical circuit is nonnegative. The circuit conventions are those of §3.2. Comparison with a rational threshold is included by subtracting that threshold. Meiburg claims that GeoSign and PosSLP each admit a deterministic polynomial-time many-one reduction to the other [Mei26, App. O.5, Prop. 14]. The corresponding gallery-verification bound in Theorem 3.4 can be written as

$$\textsc{GeoDist}_d \in \mathsf{NP}^{\textsc{GeoSign}}.$$

A supplied gallery permits exact arithmetic verification. Finding a shortest gallery is a separate task. The following natural inclusions remain difficult challenges:

$$\textsc{PosSLP} \in^{?} \mathsf{P}^{\textsc{GeoDist}_4}, \qquad \textsc{GeoSign} \in^{?} \mathsf{P}^{\textsc{GeoDist}_4} \tag{11.1}$$

The second inclusion would strengthen Theorem 3.5. Note that the arithmetic restrictions in Theorem 3.1 obstruct some constructions that preserve numerical values. They do not rule out a reduction that preserves only the answer to a comparison.

11.5. $\exists\mathbb{R}$ **conjectures.** Note that Proposition 3.3 proves only membership, not hardness for $\exists\mathbb{R}$. *Cardinal's Conjecture* states that

$$\exists\mathbb{R} = \mathsf{NP}^{\textsc{PosSLP}}, \tag{11.2}$$

see [SCM26]. Recall that $\mathsf{NP}^{\textsc{PosSLP}} \subseteq \exists\mathbb{R}$, see [ABKM09, SCM26]. Since the paper already places $\textsc{GeoDist}_4$ in $\mathsf{NP}^{\textsc{PosSLP}}$, proving $\exists\mathbb{R}$-hardness of $\textsc{GeoDist}_4$ would force equality throughout and would therefore imply Cardinal's conjecture (11.2).

Assuming in addition Meiburg's claimed bound $\textsc{PosSLP} \in \mathsf{BPP}^{\mathsf{PP}}$, that equality would place $\exists\mathbb{R}$ in the second level of the counting hierarchy, rather than the claimed generic fourth level. Our SqrtSum-hardness theorem is weaker: neither it nor the gallery verifier supplies a reduction from arbitrary $\exists\mathbb{R}$ instances. This is all to say that we are deeply sceptical that $\textsc{GeoDist}_4$ is $\exists\mathbb{R}$-hard. The arithmetic constraints in Theorem 3.1 should be the first warning. Worse, even for much broader classes of distance problems $\exists\mathbb{R}$-hardness remains unreachable.

Most relevant to this paper, Kisfaludi-Bak discusses shortest paths among open, convex, pairwise disjoint obstacles in $\mathbb{R}^3$ and geodesic diameter of polygonal surfaces among the natural candidates for $\exists\mathbb{R}$-completeness [Kis24, slide 18], see also [Kis26, §5.3]. As we mentioned above, Theorem 3.1 fails for the obstacle problem. It does not directly apply to geodesic diameter, because a diametral pair need not consist of rational points.[1] For obstacle avoidance, Canny–Reif prove NP-hardness [CR87], and Canny's existential-real formulation gives membership in $\exists\mathbb{R}$ [Can88, Cor. 3.4]. The stronger completeness question remains open. Curiously, our Main Theorem 1.1 does not apply to the geodesic diameter, although we believe that some variation of our construction should work in this case.

[1] It is easy to see that this fails already for convex quadrilaterals in the plane.

11.6. **Convex approximation.** For convex polyhedral boundaries in $\mathbb{R}^3$, Hershberger and Suri give a linear-time factor-two approximation [HS98]. Agarwal, Har-Peled, Sharir, and Varadarajan approximate distances using a small enclosing polytope and convert exterior paths to boundary paths without increasing length [AHSV97]. Har-Peled gives fast approximate queries for pairs of points [HP99a]. Agarwal, Har-Peled, and Karia give a graph-based $(1+\eta)$-approximation that returns a boundary path [AHK02]. Chazelle, Liu, and Magen obtain sublinear expected time for fixed accuracy, without preprocessing, using smaller enclosing approximations [CLM05, §6]. Their guarantee concerns distance and an exterior path. A boundary path can require linearly many segments. These results exploit global convexity in intrinsic dimension two.

Asano, Kirkpatrick, and Yap give obstacle-avoidance approximations in every fixed dimension [AKY04]. The resulting fully polynomial-time approximation scheme for a convex boundary is compatible with exact NP-hardness: its running time is polynomial in $1/\eta$, rather than in the number of accuracy bits $\log(1/\eta)$.

We include a short, self-contained proof of the fixed-dimensional approximation scheme in Appendix A. The proof specializes the convex-cover discretization of Bandyapadhyay, Lochet, Saurabh, and Xue [BLSX26] to the facets of a rational polytope. In this setting, feasibility follows from facet convexity and the approximation estimate follows from the triangle inequality. We give the rational-arithmetic details needed for a bit-complexity bound. No improvement in the approximation guarantee or running time is claimed.

AI declaration. This paper gave the author a sense of closure on a problem he had long hoped to see resolved. The original strategy behind Theorem 1.2 was conceived several years ago, but working out the technical details proved exceedingly difficult, and would have been unlikely to succeed without AI assistance. Main Theorem 1.1 is based on closely related ideas and was obtained only after Theorem 1.2 had been proved and properly deconstructed.

Much of the paper resulted from an extensive collaboration with various models of ChatGPT, which assisted with research, calculations, writing, proofreading, formatting, and references. Some parts of the paper were written entirely by AI and left essentially untouched by the author, notably both appendices and the translation of [Cay46]. Other parts were written collaboratively. At some point, it became difficult to tell who did what and who wrote what, so we will leave it at that.

Acknowledgements. The author is grateful to Robert Andrews, Joe O'Rourke, Colleen Robichaux, Maurice Rojas, Daniel Soskin and Sylvester Zhang for interesting discussions and helpful remarks.

References

[AAOS97] Pankaj K. Agarwal, Boris Aronov, Joseph O'Rourke and Catherine A. Schevon, Star unfolding of a polytope with applications, *SIAM J. Comput.* **26** (1997), 1689–1713; extended abstract in *Proc. 2nd SWAT* (1990), 251–263.

[AHK02] Pankaj K. Agarwal, Sariel Har-Peled and Meetesh Karia, Computing approximate shortest paths on convex polytopes, *Algorithmica* **33** (2002), 227–242; extended abstract in *Proc. 16th SCG* (2000), 270–279.

[AHSV97] Pankaj K. Agarwal, Sariel Har-Peled, Micha Sharir and Kasturi R. Varadarajan, Approximating shortest paths on a convex polytope in three dimensions, *J. ACM* **44** (1997), 567–584; extended abstract in *Proc. 12th SCG* (1996) 329–338.

[AMP15] Oswin Aichholzer, Wolfgang Mulzer and Alexander Pilz, Flip distance between triangulations of a simple polygon is NP-complete, *Discrete Comput. Geom.* **54** (2015), 368–389; extended abstract in *Proc. 21st ESA*, Springer, Berlin, 2013, 13–24.

[ABKM09] Eric Allender, Peter Bürgisser, Johan Kjeldgaard-Pedersen and Peter Bro Miltersen, On the complexity of numerical analysis, *SIAM J. Comput.* **38** (2009), 1987–2006; extended abstract in *Proc. 21st CCC*, IEEE, 2006, 331–339.

[AO92] Boris Aronov and Joseph O'Rourke, Nonoverlap of the star unfolding, *Discrete Comput. Geom.* **8** (1992), 219–250.

[AB09] Sanjeev Arora and Boaz Barak, *Computational complexity. A modern approach*, Cambridge Univ. Press, Cambridge, 2009, 579 pp.

[AKY04] Tetsuo Asano, David Kirkpatrick and Chee Yap, Pseudo approximation algorithms with applications to optimal motion planning, *Discrete Comput. Geom.* **31** (2004), 139–171; extended abstract in *Proc. 18th SCG* (2002), 170–178.

[Baj85] Chanderjit Bajaj, The algebraic complexity of shortest paths in polyhedral spaces, in *Proc. 23rd Allerton Conf. Comm., Control and Comput.* (1985), 510–517; available at docs.lib.purdue.edu/cstech/442/.

[Bar82] Francisco Barahona, On the computational complexity of Ising spin glass models, *J. Phys. A: Math. Gen.* **15** (1982), 3241–3253.

[BLSX26] Sayan Bandyapadhyay, William Lochet, Saket Saurabh and Jie Xue, FPTAS for polygonal multi-robot motion planning, preprint (2026), 17 pp.; arXiv:2610.20244; to appear in *Proc. 38th SODA* (2027).

[BDR04] Alexander Below, and Jesús A. De Loera and Jürgen Richter-Gebert, The complexity of finding small triangulations of convex 3-polytopes, J. Algorithms **50** (2004), no. 2, 134–167; extended abstract in *Proc. 11th SODA* (2000), 65–66.

[BCJS26] C. S. Bhargav, Radu Curticapean, Gorav Jindal and Himanshu Shukla, Multivariate resultants: Complexity and applications to real feasibility, preprint (2026), 49 pp.; submitted to *59th STOC* (2027); available at sigact-stoc.github.io/stoc2027/.

[BS26] Alexander E. Black and Raphael Steiner, Finding short paths on simple polytopes, preprint (2026), 30 pp.; arXiv:2603.05482; to appear in *Proc. 38th SODA* (2027).

[BCSS98] Lenore Blum, Felipe Cucker, Michael Shub and Steve Smale, *Complexity and real computation*, Springer, New York, 1998, 453 pp.

[BMSW11] Prosenjit Bose, Anil Maheshwari, Chang Shu and Stefanie Wuhrer, A survey of geodesic paths on 3D surfaces, *Comput. Geom.* **44** (2011), 486–498.

[B+09] Christoph Burnikel, Stefan Funke, Kurt Mehlhorn, Stefan Schirra and Susanne Schmitt, A separation bound for real algebraic expressions, *Algorithmica* **55** (2009), 14–28; extended abstract in *Proc. 9th ESA* (2001), 254–265.

[Can88] John Canny, Some algebraic and geometric computations in PSPACE, in *Proc. 20th STOC*, ACM, 1988, 460–467.

[CR87] John Canny and John H. Reif, Lower bounds for shortest path and related problems, preprint (1987), 36 pp., available at tinyurl.com/4j9n925y; extended abstract in *Proc. 28th FOCS* (1987), 49–60.

[Cay46] Arthur Cayley, Sur quelques propriétés des déterminants gauches (in French), *J. Reine Angew. Math.* **32** (1846), 119–123; modern English translation available at tinyurl.com/yw2n66at.

[CLM05] Bernard Chazelle, Ding Liu and Avner Magen, Sublinear geometric algorithms, *SIAM J. Comput.* **35** (2005), 627–646; extended abstract in *Proc. 35th STOC* (2003), 531–540.

[CH96] Jindong Chen and Yijie Han, Shortest paths on a polyhedron, Part I: Computing shortest paths, *Int. J. Comput. Geom. Appl.* **6** (1996), 127–144; extended abstract in *Proc. 6th SCG* (1990), 360–369.

[CSY97] Joonsoo Choi, Jürgen Sellen and Chee-Keng Yap, Approximate Euclidean shortest paths in 3-space, *Int. J. Comput. Geom. Appl.* **7** (1997), 271–295; extended abstract in *Proc. 10th SCG* (1994), 41–48.

[Cla87] Kenneth L. Clarkson, Approximation algorithms for shortest path motion planning, in *Proc. 19th STOC* (1987), 56–65.

[DHK20] Erik D. Demaine, Adam Hesterberg and Jason S. Ku, Finding closed quasigeodesics on convex polyhedra, in *Proc. 36th SoCG* (2020), Paper No. 33, 13 pp.

[DO07] Erik D. Demaine and Joseph O'Rourke, *Geometric folding algorithms: Linkages, origami, polyhedra*, Cambridge Univ. Press, Cambridge, 2007, 472 pp.

[DO25] Satyan L. Devadoss and Joseph O'Rourke, *Discrete and computational geometry* (second ed.), Princeton Univ. Press, Princeton, NJ, 2025, 280 pp.

[Dij59] Edsger W. Dijkstra, A note on two problems in connexion with graphs, *Numer. Math.* **1** (1959), 269–271.

[Dor26] Joseph Dorfer, Flip distance between triangulations of convex polygons is NP-complete, to appear in *Proc. 67th FOCS* (2026); preprint, 32 pp.; arXiv:2602.22874.

[EHS24] Friedrich Eisenbrand, Matthieu Haeberle and Neta Singer, An improved bound on sums of square roots via the subspace theorem, in *Proc. 40th SoCG* (2024), Paper No. 54, 8 pp.; arXiv:2312.02057.

[EG81] Shimon Even and Oded Goldreich, The minimum-length generator sequence problem is NP-hard, *J. Algorithms* **2** (1981), 311–313.

[GJ79] Michael R. Garey and David S. Johnson, *Computers and Intractability: A Guide to the Theory of* NP*-completeness*, Freeman, San Francisco, CA, 1979, 338 pp.

[Gol08] Oded Goldreich, *Computational complexity. A conceptual perspective*, Cambridge Univ. Press, Cambridge, UK, 2008, 606 pp.

[HLW06] Ernst Hairer, Christian Lubich and Gerhard Wanner, *Geometric numerical integration* (second ed.), Springer, Berlin, 2006, 644 pp.

[HP99a] Sariel Har-Peled, Approximate shortest paths and geodesic diameter on a convex polytope in three dimensions, *Discrete Comput. Geom.* **21** (1999), 217–231.

[HP99b] Sariel Har-Peled, Constructing approximate shortest path maps in three dimensions, *SIAM J. Comput.* **28** (1999), 1182–1197; extended abstract in *Proc. 14th SCG* (1998), 383–391.

[HW08] Godfrey H. Hardy and Edward M. Wright, *An introduction to the theory of numbers* (sixth ed., revised), Oxford Univ. Press, Oxford, UK, 2008, 621 pp.

[HS98] John Hershberger and Subhash Suri, Practical methods for approximating shortest paths on a convex polytope in $\mathbb{R}^3$, *Comput. Geom.* **10** (1998), 31–46; extended abstract in *Proc. 6th SODA* (1995), 447–456.

[HS99] John Hershberger and Subhash Suri, An optimal algorithm for Euclidean shortest paths in the plane, *SIAM J. Comput.* **28** (1999), 2215–2256; extended abstract in *Proc. 34th FOCS* (1993), 508–517.

[IR82] Kenneth F. Ireland and Michael I. Rosen, *A classical introduction to modern number theory*, Springer, New York, 1982, 341 pp.

[JLY02] Christian U. Jensen, Arne Ledet and Noriko Yui, *Generic polynomials: Constructive aspects of the inverse Galois problem*, Cambridge Univ. Press, Cambridge, UK, 2002, 258 pp.

[Jer85] Mark R. Jerrum, The complexity of finding minimum-length generator sequences, *Theoret. Comput. Sci.* **36** (1985), 265–289; extended abstract in *Proc. 11th ICALP* (1984), 270–280.

[Kar72] Richard M. Karp, Reducibility among combinatorial problems, in *Complexity of Computer Computations*, Plenum Press, New York, 1972, 85–103.

[Kis24] Sándor Kisfaludi-Bak, $\exists\mathbb{R}$-completeness in geometric algorithms, lecture slides (Feb. 16, 2024), 78 pp.; available at tinyurl.com/4mxuhcka.

[Kis26] Sándor Kisfaludi-Bak, Precision of continuous distance problems, in *Precision in Geometric Algorithms*, *Dagstuhl Rep.* **15** (2026), no. 9, §5.3, p. 34; available at tinyurl.com/yf5wb5rr

[LR23] Markus Lohrey and Andreas Rosowski, On the complexity of diameter and related problems in permutation groups, in *Proc. 50th ICALP* (2023), Paper No. 134, 18 pp.

[Luc14] Andrew Lucas, Ising formulations of many NP problems, *Front. Physics* **2** (2014), Art. 5, 15 pp.

[LP15] Anna Lubiw and Vinayak Pathak, Flip distance between two triangulations of a point set is NP-complete, *Comput. Geom.* **49** (2015), 17–23; extended abstract in *Proc. 24th CCCG* (2012), 127–132.

[Mei26] Alex Meiburg, $\exists\mathbb{R} \subseteq$ CH, preprint (2026), 114 pp.; arXiv:2610.10514.

[MP08] Ezra Miller and Igor Pak, Metric combinatorics of convex polyhedra: cut loci and nonoverlapping unfoldings, *Discrete Comput. Geom.* **39** (2008), 339–388.

[M+16] Tillmann Miltzow, Lothar Narins, Yoshio Okamoto, Günter Rote, Antonis Thomas and Takeaki Uno, Approximation and hardness of token swapping, in *Proc. 24th ESA* (2016), Paper No. 66, 15 pp.

[Mit96] Joseph S. B. Mitchell, Shortest paths among obstacles in the plane, *Int. J. Comput. Geom. Appl.* **6** (1996), 309–332.

[Mit09] Joseph S. B. Mitchell (scribe), Open problem session, in *Computational Geometry*, Dagstuhl Seminar Proc. **09111** (2009), 1–3.

[MMP87] Joseph S. B. Mitchell, David M. Mount and Christos H. Papadimitriou, The discrete geodesic problem, *SIAM J. Comput.* **16** (1987), 647–668.

[Mou87] David M. Mount, Storing the subdivision of a polyhedral surface, *Discrete Comput. Geom.* **2** (1987), 153–174.

[Mou90] David M. Mount, The number of shortest paths on the surface of a polyhedron, *SIAM J. Comput.* **19** (1990), 593–611.

[NP18] Danny Nguyen and Igor Pak, Enumerating projections of integer points in unbounded polyhedra, *SIAM J. Discrete Math.* **32** (2018), 986–1002; extended abstract in *Proc. 19th IPCO* (2017), 417–429.

[NP21] Danny Nguyen and Igor Pak, On the number of integer points in translated and expanded polyhedra, *Discrete Comput. Geom.* **65** (2021), 405–424.

[NP22] Danny Nguyen and Igor Pak, Short Presburger arithmetic is hard, *SIAM J. Comput.* **51** (2022), no. 2, 30 pp.; extended abstracts in *Proc. 49th STOC* (2017), 812–820, and *Proc. 58th FOCS* (2017), 37–48.

[OAI26] OpenAI, Existential–universal real sentences in the counting hierarchy, preprint (2026), 35 pp.

[O'Rou81] Joseph O'Rourke, Advanced Problem 6369, *Amer. Math. Monthly* **88** (1981), 768–769.

[Pak09] Igor Pak, *Lectures on discrete and polyhedral geometry*, monograph draft (2009), 440 pp.; available at math.ucla.edu/~pak/book.htm

[Pak20] Igor Pak, *What if they are all wrong?*, WordPress blog post (Dec. 10, 2020); wp.me/p211iQ-uT

[Pap85] Christos H. Papadimitriou, An algorithm for shortest-path motion in three dimensions, *Inform. Process. Lett.* **20** (1985), 259–263.

[Ric96] Jürgen Richter-Gebert, Realization spaces of polytopes, *Lecture Notes in Math.* **1643**, Springer, Berlin, 1996, 187 pp.

[SCM26] Marcus Schaefer, Jean Cardinal and Tillmann Miltzow, The existential theory of the reals as a complexity class: a compendium, in *Courses in Discrete and Computational Geometry*, J. Pach and G. Tóth, eds., Springer, Cham, 2026, 167–313; arXiv:2407.18006.

[Sch10] Yevgeny Schreiber, An optimal-time algorithm for shortest paths on realistic polyhedra, *Discrete Comput. Geom.* **43** (2010), 21–53; extended abstract in *Proc. 23rd SCG* (2007), 74–83.

[SS08] Yevgeny Schreiber and Micha Sharir, An optimal-time algorithm for shortest paths on a convex polytope in three dimensions, *Discrete Comput. Geom.* **39** (2008), 500–579; extended abstract in *Proc. 22nd SCG* (2006), 30–39.

[SCY00] Jürgen Sellen, Joonsoo Choi and Chee-Keng Yap, Precision-sensitive Euclidean shortest path in 3-space, *SIAM J. Comput.* **29** (2000), 1577–1595; extended abstract in *Proc. 11th SCG* (1995), 350–359.

[Sha87] Micha Sharir, On shortest paths amidst convex polyhedra, *SIAM J. Comput.* **16** (1987), 561–572; extended abstract in *Proc. 2nd SCG* (1986), 193–206.

[SS86] Micha Sharir and Amir Schorr, On shortest paths in polyhedral spaces, *SIAM J. Comput.* **15** (1986), 193–215; extended abstract in *Proc. 16th STOC* (1984), 144–153.

[STT88] Daniel D. Sleator, Robert E. Tarjan and William P. Thurston, Rotation distance, triangulations, and hyperbolic geometry, *Jour. AMS* **1** (1988), 647–681.

[SH11] Ivar Stakgold and Michael Holst, *Green's functions and boundary value problems* (third ed.), John Wiley, Hoboken, NJ, 2011, 855 pp.

[Tiw92] Prasoon Tiwari, A problem that is easier to solve on the unit-cost algebraic RAM, *J. Complexity* **8** (1992), 393–397.

[Wang23] Haitao Wang, A new algorithm for Euclidean shortest paths in the plane, *J. ACM* **70** (2023), Paper No. 11, 62 pp.; extended abstract in *Proc. 53rd STOC* (2021), 975–988.

[Wang25] Haitao Wang, Shortest paths on convex polyhedral surfaces, preprint (2025), 38 pp.; arXiv:2512.11299; extended abstract in *Proc. 66th FOCS* (2025), 2306–2327.

[Yam25] Yoshikazu Yamagishi, The farthest point map on the 4-cube, *Eur. J. Math.* **11** (2025), Paper No. 48, 23 pp.

Appendix A. Geodesic distance approximation

We give a self-contained binary-arithmetic proof of a known fixed-dimensional approximation result. The proof produces a rational boundary path and a certified rational upper bound on its length. Theorem A.1 is included to explain how exact NP-hardness coexists with a fully polynomial-time approximation scheme. The final subsection compares the required accuracy with the gap in the hardness reduction.

A.1. **Relative approximation and exact precision.** The hardness statements concern exact comparisons. In this appendix, n denotes the number of facet inequalities, B the total binary input length, and η the requested relative error. The output path is γ and W is a rational certificate satisfying $\text{length}(\gamma) \le W$.

For every fixed dimension there is a fully polynomial-time approximation scheme (FPTAS), meaning polynomial dependence on the input length and on $1/\eta$, rather than on the number of accuracy bits $\log(1/\eta)$. The following statement makes the rational input and output explicit.

Theorem A.1 (Fixed-dimensional approximation)**.** *Fix $d \ge 2$. Let $P \subset \mathbb{R}^d$ be a bounded full-dimensional convex polytope given by n rational linear inequalities, let $s, t \in \partial P$ be rational points, and let $0 < \eta \le 1$ be rational. One can compute a polygonal boundary path γ with rational vertices and a rational number W such that*

$$\text{dist}_{\partial P}(s,t) \;\le\; \text{length}(\gamma) \;\le\; W \;\le\; (1+\eta)\,\text{dist}_{\partial P}(s,t). \tag{A.1}$$

The number of bit operations is

$$(n/\eta)^{O(d)}\,\text{poly}(B),$$

where B is the total binary encoding length of the inequalities, the endpoints, and η. For fixed dimension d, the exponent hidden in $O(d)$ is a constant and the running time is polynomial both in the input size B and in $1/\eta$. This is exactly the definition of a fully polynomial-time approximation scheme, so the result is an FPTAS in every fixed dimension.

Relation to earlier approximation results. Asano, Kirkpatrick, and Yap extend obstacle-avoidance approximation to every fixed dimension, without requiring convex obstacles, and count algebraic operations [AKY04, §4, p. 152]. For one convex obstacle, nonexpanding nearest-point projection identifies exterior and boundary distances, see [HP99a, Thm 2.1] and Proposition A.2 below. Our proof specializes the convex-cover sampling method of Bandyapadhyay, Lochet, Saurabh, and Xue [BLSX26] to the facets of P.

The purpose of the proof is to supply the binary-arithmetic details in this setting: representatives in rational face intersections, polynomial coordinate bounds, rational upper approximations to edge lengths, and a search for the distance scale. Neither the approximation guarantee nor this rational implementation is claimed as new, and no improvement in running time is asserted.

This FPTAS excludes NP-hardness with a fixed or inverse-polynomial relative gap in fixed dimension unless P = NP. The Partition reduction instead uses an exponentially small gap with polynomially many encoding bits. The square-root-sum identity is exact and needs no separation estimate at all. The exact reduction therefore remains valid even when the relevant algebraic quantities are extraordinarily close. By contrast, an approximation algorithm with running time polynomial in $1/\eta$ may become exponentially slow when η is chosen as small as the exact decision gap. For that reason, both hardness statements are fully compatible with polynomial dependence on $1/\eta$. The quantitative comparison in §A.4 distinguishes polynomial dependence on $1/\eta$ from polynomial dependence on its encoding length $\log(1/\eta)$.

A.2. **Exterior distance and boundary distance.** For a full-dimensional convex polytope $P \subset \mathbb{R}^d$, write $E = \mathbb{R}^d \setminus \operatorname{int} P$ for its closed exterior. The metric dist_E is the infimum of lengths of paths in E. Nearest-point projection gives the following comparison in every dimension. The three-dimensional statement appears in [HP99a, Thm 2.1].

Proposition A.2. *For $s, t \in \partial P$,*

$$\operatorname{dist}_E(s,t) = \operatorname{dist}_{\partial P}(s,t).$$

Proof. Let $\pi_P(x)$ be the point of P nearest to x in Euclidean distance. It exists by compactness and is unique by strict convexity of squared Euclidean distance. If $x \in E$, then $\pi_P(x) \in \partial P$, and π_P fixes every point of ∂P.

We check directly that projection cannot increase distances. Put $u = \pi_P(x)$ and $v = \pi_P(y)$. Minimality of u along the segment $[u, v] \subseteq P$, and minimality of v along the same segment, give

$$\langle x - u, v - u \rangle \;\le\; 0, \qquad \langle y - v, u - v \rangle \;\le\; 0.$$

For example, the first inequality follows by expanding $\|x-u-h(v-u)\|^2 \ge \|x-u\|^2$ for $0 < h \le 1$ and letting h decrease to zero. Adding the two inequalities gives

$$\|u - v\|^2 \;\le\; \langle x - y, u - v \rangle \;\le\; \|x - y\|\,\|u - v\|.$$

If $u \neq v$, division by $\|u - v\|$ gives $\|\pi_P(x) - \pi_P(y)\| \le \|x - y\|$; for $u = v$ the same inequality is immediate. Thus π_P is 1-Lipschitz. Applying this inequality to every polygonal sum in the definition of curve length shows that projection does not increase the length of any rectifiable path, meaning a path of finite length. Projecting an exterior path therefore produces a boundary path with the same endpoints and no greater length. The reverse inequality follows because every boundary path is also an exterior path. □

A.3. **Approximation by a finite rational graph.** We now implement the convex-cover discretization on the facets, including face feasibility, rational edge-weight bounds, and the search for an appropriate distance scale. The number of cells and arithmetic operations is polynomial for fixed ambient dimension.

Proof of Theorem A.1. If $s = t$, return the constant path and $W = 0$. Henceforth assume $s \neq t$. Throughout the proof $\|\cdot\|$ denotes the Euclidean norm. We write the inequalities as $a_i \cdot x \le b_i$. Discard any inequality whose normal a_i is zero. Such an inequality is a tautology because the input is a nonempty polytope. For each remaining inequality form

$$F_i = P \cap \{x : a_i \cdot x = b_i\},$$

and omit empty sets. These are convex faces contained in ∂P, and their union is ∂P. They include every facet, although a redundant inequality can also cut out a lower-dimensional face rather than a facet. We keep all nonempty sets F_i instead of first removing redundant inequalities. This convention allows arbitrary rational halfspace descriptions, including redundant ones, without changing any subsequent covering or sampling argument. There are at most n of them. Here and below, i, j range over the retained face labels unless a different range is displayed explicitly.

Rational points in face intersections. All feasibility questions below concern bounded intersections of rational halfspaces and hyperplanes in the fixed space $\mathbb{R}^d$. A nonempty such intersection has a vertex. At a vertex its active constraint normals span $\mathbb{R}^d$: otherwise a small displacement in a nonzero direction orthogonal to all active normals would remain feasible in both directions, contradicting the definition of a vertex. Thus we can find a rational point by enumerating sets of d linearly independent constraint hyperplanes, solving their linear systems, and retaining a solution satisfying every constraint. If no solution is retained, the bounded intersection is empty. This works also for intersections of lower dimension. With h constraints it requires $h^{O(d)}$ rational

operations, and Cramer's rule bounds the output bit lengths by a polynomial in the input bit length. We use this elementary procedure for all the selections below.

A finite initial upper bound. Make a graph with the nonempty faces F_i as vertices, joining two when their intersection is nonempty. This graph is connected: a disconnection would partition the connected boundary ∂P into two disjoint nonempty compact unions of faces. Choose a simple chain from a face containing s to one containing t, and choose a rational point in each consecutive intersection. The segments between successive selected points, with s and t appended, lie in the corresponding faces and form a boundary path. The sum U of their lengths in the norm $\|x\|_1 = \sum_{a=1}^{d} |x_a|$ is a rational upper bound for their Euclidean length and hence for $L = \operatorname{dist}_{\partial P}(s,t)$. The number of segments and the bit length of U are polynomial in B.

A shortest path uses at most n segments. The preceding path shows that L is finite. Take paths with lengths decreasing to L, parametrized at constant speed on $[0,1]$. Their Lipschitz constants are uniformly bounded, and their images lie in the compact set ∂P, so a uniformly convergent subsequence exists. Its limit is a boundary path from s to t. For every finite partition of $[0,1]$, the sum of the limit path's chord lengths is the limit of the corresponding sums and is at most L. Taking the supremum over partitions shows that the limit path has length at most L, hence exactly L. Thus a shortest boundary path exists.

Parametrize it at constant speed on $[0,1]$. If two of its points belong to the same face F_i, their joining segment lies in F_i. The intervening subpath is shortest between those points, so its length equals their Euclidean distance. Equality in the Euclidean triangle inequality forces the intervening subpath to be the straight segment joining the two points. That segment lies in the convex face F_i. Therefore, once the shortest path enters a given face and later returns to it, the entire intervening portion also lies in that face. The set of parameter values belonging to any one F_i is consequently an interval, possibly empty or a single point. These at most n closed intervals cover the parameter interval. Because the intervals cover $[0,1]$, we can start with an interval containing 0 and repeatedly choose, among intervals meeting the current right endpoint, one whose right endpoint is furthest to the right. Each step introduces a new face, so at most n steps are needed before reaching 1. Taking the path points at the successive overlap parameters gives

$$s = q_0, q_1, \ldots, q_r = t, \qquad r \;\le\; n, \tag{A.2}$$

where each segment $[q_{i-1}, q_i]$ lies in one face and each internal q_i lies in the two consecutive faces. No regularity assumption on the endpoints or crossings is needed.

Sampling at one distance scale. Suppose for the moment that a rational number D satisfies $L \le D < 2L$. Every point of a shortest path lies in the box $\mathcal{B}_D = s + [-D, D]^d$. Let $J := \lceil 16dn/\eta \rceil$, and divide this box into J^d closed cubes of side length $2D/J$. For every pair of faces F_i, F_j and every cube C, choose one rational representative in $F_i \cap F_j \cap C$ when this intersection is nonempty. Include all representatives, together with s, t, as vertices of a new graph G_D. Join any two graph vertices belonging to a common face. Each such edge represents their straight segment in that face, so every graph path is a boundary path. The graph has at most $2 + n^2 J^d$ vertices.

For an internal crossing q_i in (A.2), choose the representative p_i for its two consecutive faces in one cube containing q_i. Let $p_0 = s$ and $p_r = t$. Both points of every consecutive pair p_{i-1}, p_i belong to the corresponding face, and

$$\|p_i - q_i\| \;\le\; \frac{2\sqrt{d}\,D}{J}.$$

After deleting consecutive repetitions, the selected points form a graph walk whose length satisfies

$$\sum_{i=1}^{r} \|p_i - p_{i-1}\| \;\le\; L + 2\sum_{i=1}^{r-1} \|p_i - q_i\| \;\le\; L + \frac{4n\sqrt{d}\,D}{J} \;\le\; L + \frac{\eta D}{4} \;\le\; (1+\eta/2)L. \tag{A.3}$$

Deleting cycles, if any, gives a graph path no longer than this walk. Very thin faces cause no difficulty. A representative is selected inside the exact face intersection with a cube. It need not be a vertex of the grid. The approximation argument uses only the ambient mesh size and the fact that every exact face intersection can be represented by a nearby rational sample point. It never assumes that an ambient Cartesian grid itself hits the relative interior of every face. Consequently very thin faces and very small dihedral angles cause no separate difficulty, and no lower bound on face widths or angles is used.

Rational edge weights. An edge has Euclidean length $w = \sqrt{a}$, where $a \geq 0$ is rational. For $w > 0$, compute a rational upper approximation $\widehat{w}$ with

$$w \;\leq\; \widehat{w} \;\leq\; (1+\eta/4)w.$$

Such an approximation follows by rational bisection: comparisons between a candidate $u \geq 0$ and $\sqrt{a}$ reduce to comparing u^2 with a. The polynomial bit bound on a bounds both its magnitude and, when positive, its distance from zero by powers of two with polynomial exponents. It therefore takes only polynomially many bits and bisection steps, including $O(\log(1/\eta))$ accuracy bits, to obtain the displayed multiplicative bounds. Give zero-length edges weight zero.

Run Dijkstra's algorithm with the nonnegative rational weights $\widehat{w}$. If W_D is the total weight of the path returned, its actual length is at most W_D. By (A.3),

$$W_D \;\leq\; (1+\eta/4)(1+\eta/2)L \;\leq\; (1+\eta)L,$$

where the last inequality uses $0 < \eta \leq 1$. Every graph path is feasible, so $W_D \geq L$. All graph comparisons use rational numbers. No exact comparison between sums of square roots is required.

Finding a suitable scale. The number $\ell = \|s-t\|_\infty = \max_{1\leq a\leq d} |s_a - t_a|$ is a positive rational lower bound for L. Apply the graph construction at the scales

$$D_j = 2^j \ell, \qquad 0 \;\leq\; j \;\leq\; \lceil \log_2(U/\ell) \rceil .$$

The upper bound U and lower bound ℓ have polynomial bit length, so this list has polynomial length. Its first scale at least L lies in $[L, 2L)$. Discard any scale whose graph has no s–t path, and choose the path with smallest rational total weight among the remaining scales. Every path considered is a boundary path. At least one of the tried dyadic scales lies within the factor-two window required by the preceding analysis. For that scale the graph-distance guarantee applies, and taking the minimum over all tried scales can only improve the answer. This gives (A.1).

At each scale there are at most $n^2 J^d$ feasibility questions, each with $n + O(d)$ constraints. The number of graph edges is at most the square of its vertex count. All coefficients have polynomial bit length in B, and the rational path-weight sums have polynomial bit length in B and the graph size. The feasibility enumeration, edge construction, rational shortest-path search, and polynomially many scales together take $(n/\eta)^{O(d)}\,\mathrm{poly}(B)$ bit operations, as claimed. □

Remark A.3. The common Euclidean embedding is essential to the facet shortcut: a chord is no longer than any competing subpath in that ambient space. Abstract complexes with arbitrary face identifications need not have this property. If the dimension is part of the input, the displayed bound no longer gives a polynomial-time scheme.

A.4. **Accuracy and the hardness gap.** Theorem A.1 prevents an extension of the construction to hardness for a fixed relative approximation factor in any fixed dimension, unless $\mathsf{P} = \mathsf{NP}$. More generally, a polynomial reduction promising $L \leq A$ in yes instances and $L \geq (1+\delta)A$ in no instances, with rational $A > 0$ and $1/\delta$ polynomially bounded in the input length, can be decided by taking $\eta < \delta$. This also excludes hardness at inverse polynomial relative separation under the same assumption.

The gap in Section 6 has a different size. The reduction produces a rational threshold τ and a positive rational gap parameter g, both of polynomial binary encoding length, such that the yes

and no distances satisfy, respectively, $L < \tau - 7g/16$ and $L > \tau + 7g/16$, with $1 \le L < 2$. The explicit value $g = \varepsilon^4\kappa^2$ and these bounds are proved in that section. Suppose an approximation algorithm returned an upper bound within relative error $1+\eta$ with $\eta = g/8$. Because $1 \le L < 2$, its additive overestimate would be less than $g/4$, which is smaller than the $7g/16$ margin on either side of τ. Comparing the approximation with τ would therefore distinguish the yes and no cases. The number of accuracy bits $\log_2(1/g)$ is polynomial in the PARTITION input length, whereas $1/g$ can be exponentially large. Polynomial dependence on $1/\eta$ is therefore compatible with the exact-comparison hardness of Theorem 1.1. Polynomial dependence on the number of accuracy bits would not be, unless $\mathsf{P} = \mathsf{NP}$.

Finally, exponentially many shortest paths do not obstruct this approximation algorithm. A finite graph can itself have exponentially many shortest paths, while Dijkstra's algorithm computes one shortest distance without listing them. The graph above approximates the minimum length directly. It does not enumerate unfolded source images or facet galleries.

Appendix B. A two-obstacle example

We give an explicit two-obstacle example to show why the arithmetic conclusion of Theorem 3.1 does not extend to obstacle distances. The example concerns the optimal length itself, not only its contact coordinates: its square has nonreal conjugates and therefore is not a real multiquadratic number. Here a real algebraic number is *constructible* if it can be obtained from rational numbers by arithmetic operations and successive real square roots. It is *totally real* if all its conjugates are real.

Proposition B.1 (Two-obstacle arithmetic contrast)**.** *There are two bounded open rational convex polyhedral obstacles in $\mathbb{R}^3$ with disjoint closures and rational endpoints such that the shortest obstacle-avoiding length L is neither constructible nor totally real. Its square L^2 is not a real multiquadratic number. In the construction below, the number $Q = L^2/4$ has minimal polynomial*

$$T^4 - 6T^3 + 10T^2 - 18T + 2. \tag{B.1}$$

Proof. We break the proof into two parts.

The obstacles and the global minimum. In coordinates (X, Y, Z), define the open obstacles

$$O_1 = \{\, |Y| < \tfrac{1}{4},\quad 4|Y| < X < 100,\quad |Z| < 100 \,\}, \tag{B.2}$$

$$O_2 = \{\, |Y-2| < \tfrac{1}{4},\quad 4|Y-2| < Z < 100,\quad |X| < 100 \,\}. \tag{B.3}$$

They are bounded open convex polyhedra given by rational linear inequalities: for example, $X > 4|Y|$ means $X > 4Y$ and $X > -4Y$. Their closures are elongated pentagonal prisms in parallel, disjoint slabs, since their Y-ranges are separated. Their wedge edges lie on $X = Y = 0$ and $Y = 2,\ Z = 0$, respectively, and run in the Z- and X-directions. The cutoff 100 keeps the corresponding truncating faces far from the short paths considered below.

Let

$$s = (0, -1, 1), \qquad t = (1, 3, 0). \tag{B.4}$$

Note that obstacle boundaries are allowed on a path. Only their interiors have to be avoided. For $0 < r < 1$, the polygonal path

$$s \longrightarrow (0, 0, r) \longrightarrow (r, 2, 0) \longrightarrow t \tag{B.5}$$

is feasible. On the middle segment,

$$X = \frac{r}{2}Y < 4Y \quad (0 < Y < \tfrac{1}{4}), \qquad Z = \frac{r}{2}(2-Y) < 4(2-Y) \quad (\tfrac{7}{4} < Y < 2).$$

Its initial segment has $X = 0$ and its final segment has $Z = 0$, so these segments stay on the safe sides of the corresponding obstacles. The length is

$$\phi(r) = 2\sqrt{(r-1)^2+1} + \sqrt{2r^2+4}. \tag{B.6}$$

At $r = 1/2$ it is $\sqrt{5} + 3/\sqrt{2} < 5$.

Consider any competing path of length less than 5. Every point of this path is at Euclidean distance less than 5 from s, so $|X| < 5$ and $|Z| < 6$. In particular, the path cannot reach the distant truncating faces at coordinates ± 100. Choose its first point on $Y = 2$ and an earlier point on $Y = 0$. Such points exist by continuity, since the path starts at $Y = -1$ and ends at $Y = 3$; no monotonicity in Y is required. At $Y = 0$, obstacle avoidance forces $X \le 0$, and at $Y = 2$ it forces $Z \le 0$. Thus the two points have the form

$$p = (u, 0, z), \quad u \le 0, \qquad q = (x, 2, v), \quad v \le 0.$$

The triangle inequality on the three intervening portions bounds the competitor's length from below by

$$F(u,z,x,v) = \sqrt{u^2+1+(z-1)^2} + \sqrt{(x-u)^2+4+(v-z)^2} + \sqrt{(x-1)^2+1+v^2}. \tag{B.7}$$

This lower bound does not require the three straight segments used in the comparison to avoid the obstacles. The function F is differentiable and convex. On $u = v = 0$, it is symmetric in x, z and strictly convex, so its unique minimizer has $x = z = r$. The resulting function ϕ is strictly convex, with $\phi'(0) < 0$ and $\phi'(1) > 0$. Its unique minimizer lies in $(0, 1)$. At $(u, z, x, v) = (0, r, r, 0)$,

$$F_z = F_x = 0, \qquad F_u = F_v = -\frac{r}{\sqrt{2r^2+4}} < 0.$$

The supporting-hyperplane inequality for the convex function F gives, for $u, v \le 0$,

$$F(u,z,x,v) \ge F(0,r,r,0) - \frac{r}{\sqrt{2r^2+4}}(u+v) \ge F(0,r,r,0).$$

Path (B.5) attains this lower bound. Its length is less than 5, whereas all longer competitors are irrelevant. Thus the global obstacle distance is $L = \phi(r)$ for this unique minimizing r.

The squared length and its conjugates. The stationary equation is

$$\frac{1-r}{\sqrt{(r-1)^2+1}} = \frac{r}{\sqrt{2r^2+4}}. \tag{B.8}$$

Squaring and simplifying gives

$$r^4 - 2r^3 + 4r^2 - 8r + 4 = 0. \tag{B.9}$$

The unsquared equation also gives

$$L = \sqrt{2r^2+4}\,\frac{2-r}{r}.$$

Let $Q = L^2/4$ and set

$$f(R) = R^4 - 2R^3 + 4R^2 - 8R + 4, \qquad g(T) = T^4 - 6T^3 + 10T^2 - 18T + 2.$$

For monic f, its *resultant* with a polynomial h is $\mathrm{Res}_R(f,h) = \prod_{f(\xi)=0} h(\xi)$, with roots counted with multiplicity. Eliminating r gives

$$\mathrm{Res}_R\big(f(R), 4TR^2 - (2R^2+4)(2-R)^2\big) = 4096\,g(T). \tag{B.10}$$

Consequently $g(Q) = 0$.

Modulo 3, the polynomial is $T^4 + T^2 + 2$. It has no root, and none of the three monic irreducible quadratics

$$T^2 + 1, \qquad T^2 + T + 2, \qquad T^2 + 2T + 2$$

over $\mathbb{F}_3$ divides it. Thus it is irreducible modulo 3, and g is the minimal polynomial of Q over $\mathbb{Q}$. For a monic polynomial with roots $\xi_1, \ldots, \xi_n$, its *discriminant* is $\prod_{1 \le i < j \le n} (\xi_i - \xi_j)^2$. Here

$$\operatorname{disc}(g) = -16 \cdot 43^3 \ < \ 0. \tag{B.11}$$

A real squarefree quartic with negative discriminant has exactly one pair of nonreal roots. Hence Q is not totally real. Since $L^2 = 4Q$, the number L^2 also has nonreal conjugates. Every real multiquadratic field is totally real, so L^2 cannot belong to such a field. This is the claimed failure of the convex-boundary arithmetic conclusion. If L were totally real, then $L^2/4$ would be totally real as well, so L also fails this property.

It remains to prove the additional assertion that L is not constructible. For this we use the factorization of g modulo 11:

$$g(T) \equiv (T-1)(T^3 - 5T^2 + 5T - 2) \pmod{11}. \tag{B.12}$$

The cubic has no root modulo 11: its values at $0, 1, \ldots, 10$ are

$$9, 10, 7, 6, 2, 1, 9, 10, 10, 4, 9.$$

It is therefore irreducible. Since 11 does not divide the discriminant, Dedekind's factorization theorem [JLY02, Thm 3.3.3] implies that the factorization of type $(1, 3)$ gives a 3-cycle in the Galois group of g.

Here the *splitting field* is the field generated by all roots of g, and its *Galois group* consists of the field automorphisms fixing $\mathbb{Q}$. The splitting field of the minimal polynomial of a constructible number has degree a power of two, so its Galois group cannot contain an element of order three. Thus Q is not constructible, and neither is L. □

Numerically,

$$r \approx 0.6786642452, \qquad L \approx 4.3190916070.$$

These values only identify the relevant root. Neither global optimality nor the algebraic conclusions depend on them. The quartic g is solvable by radicals, and the final square root preserves that property. The example proves nonconstructibility, not non-solvability by arbitrary radicals.

Remark B.2. For rational polygonal obstacles in the plane and rational endpoints, a shortest path can bend only at obstacle vertices. All vertices of the path are therefore rational. Thus, each segment length is the square root of a nonnegative rational, and the total length belongs to a field generated by square roots of positive rationals. Therefore, such a field is multiquadratic, totally real, and obtained by quadratic extensions, as in Theorem 3.1, see also the planar shortest-path setting in [Mit96, HS99].

(Igor Pak) Department of Mathematics, UCLA, Los Angeles, CA 90095.
Email address: pak@math.ucla.edu