The Spheres of Sol

Matei P. Coiculescu and Richard Evan Schwartz *

April 25, 2021

Abstract

Let Sol be the 3-dimensional solvable Lie group whose underlying space is $\boldsymbol{R}^3$ and whose left-invariant Riemannian metric is given by

$$e^{-2z}dx^2 + e^{2z}dy^2 + dz^2.$$

Let $E : \boldsymbol{R}^3 \to$ Sol be the Riemannian exponential map. Given $V = (x, y, z) \in \boldsymbol{R}^3$, let $\gamma_V = \{E(tV) | t \in [0, 1]\}$ be the corresponding geodesic segment. Let AGM stand for the arithmetic-geometric mean. We prove that γ_V is a distance minimizing segment in Sol if and only if

$$\mathrm{AGM}\left(\sqrt{|xy|}, \frac{1}{2}\sqrt{(|x| + |y|)^2 + z^2}\right) \leq \pi.$$

We use this inequality to precisely characterize the cut locus in Sol, prove that the metric spheres in Sol are topological spheres, and almost exactly characterize their singular sets.

1 Introduction

1.1 Background

Sol is one of the 8 Thurston geometries [**Th**], the one which uniformizes torus bundles which fiber over the circle with Anosov monodromy. Sol

*Supported by N.S.F. Grant DMS-1807320

has sometimes been the topic of studies in coarse geometry and geometric group theory. The deep and difficult work of A. Eskin, D. Fisher, and K. Whyte [**EFW**], a landmark of geometric group theory, shows that any quasi-isometry of Sol is boundedly close to an isometry. As another example, N. Brady [**B**] proves that lattices in Sol are not asynchronously automatic.

The metric geometry of Sol is intriguing and mysterious. Sol has two totally geodesic foliations by hyperbolic planes, meeting at right angles, but somehow the two foliations are "turned upside down" with respect to each other. This engenders a kind of topsy-turvy feel. Another complicating feature is that Sol has sectional curvatures of both signs, causing an interplay of focus and dispersion. A number of authors have studied the differential geometry of Sol, with an emphasis on mean curvature surfaces. See the work by R. López and M. I. Munteanu [**LM**] and the references therein.

In [**T**], M. Troyanov integrates the geodesic equations for Sol and gets explicit formulas for the geodesics in terms of elliptic integrals. He uses these expressions to determine what he calls the *horizon* of Sol: the topological space of equivalence classes of geodesics, where two geodesics are equivalent iff they have finite Hausdorff distance. The horizon gives information about the large-scale organization of the Sol geodesics. This theme is further pursued by S. Kim in [**K**]. In [**BS**], A. Bölcskei and B. Szilágyi take a related approach to the geodesics in Sol, with the view towards drawing pictures of the spheres in Sol. Their paper has pictures of the spheres of radius 1 and 2.

Matt Grayson's 1983 Princeton PhD thesis [**G**] takes a different approach to studying the geodesics. Working in a special frame of reference, Grayson converts the geodesic flow on Sol to a particular Hamiltonian flow on the 2-sphere and then gives a detailed, penetrating analysis of the geodesics in Sol. We think that Grayson had many of the ingredients needed to establish the results in our paper, but he doesn't quite go in that direction. In any case, [**G**] was a tremendous inspiration for us.

The Hamiltonian flow approach, which we also take, goes back at least to V. I. Arnold's work [**A**] on hydrodynamics. See also the book by V. I. Arnold and B. Khesin [**AK**]. In a related direction, A. V. Bolsinov and I. A. Taimanov [**BT**] use the same formalism to study the geodesic flow on a 3-dimensional solv-manifold and construct an integrable geodesic flow with positive topological entropy.

In a different direction, R. Coulon, E. A. Matsumoto, H. Segerman, and S. Trettel [**CMST**] recently made a virtual reality ray-tracing program for Sol. We can say, from firsthand experience, that this thing is amazing.

1.2 Main Results

The AGM, or *arithmetic-geometric mean*, is defined for $0 \leq \alpha_0 \leq \beta_0$, as follows. We iteratively define

$$\alpha_{n+1} = \sqrt{\alpha_n \beta_n}, \qquad \beta_{n+1} = \frac{\alpha_n + \beta_n}{2}. \tag{1}$$

Then

$$\mathrm{AGM}(\alpha_0, \beta_0) = \lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \beta_n. \tag{2}$$

This definition gives a rapidly converging sequence. See [**BB**] for details.

Given $V = (x, y, z) \in \boldsymbol{R}^3$ we define

$$\mu(V) = \mathrm{AGM}\left(\sqrt{|xy|}, \frac{1}{2}\sqrt{(|x| + |y|)^2 + z^2}\right). \tag{3}$$

We note several properties of μ.

- $\mu(V) = 0$ iff $xy = 0$.
- $\mu(rV) = |r|\mu(V)$.
- $\mu(V) = \mathrm{AGM}(x, y)$ when $x, y \geq 0$ and $z = 0$.

We equip Sol with the left invariant metric

$$e^{-2z}dx^2 + e^{2z}dy^2 + dz^2. \tag{4}$$

This is a canonical choice because it agrees with the usual dot product at the identity of Sol. Given $V \in \boldsymbol{R}^3$ as above, we let $\gamma_V = \{E(tV) | t \in [0, 1]\}$ be the corresponding geodesic segment. Here E denotes the Riemannian exponential map.

We call V and γ_V *small*, *perfect*, or *large* whenever we have $\mu(V) < \pi$, $\mu(V) = \pi$, or $\mu(V) > \pi$, respectively. A typical geodesic in Sol looks like a corkscrew and in this case the three conditions above respectively say that the geodesic segment makes less than one, exactly one, or more than one twist. We will discuss geometric interpretations of our inequalities more formally and in more detail in §2.2.

Theorem 1.1 (Main) *A geodesic segment in Sol is a distance minimizer if and only if it is small or perfect. That is, γ_V is a distance minimizing geodesic segment if and only if $\mu(V) \leq \pi$.*

The Main Theorem is a concise way of writing a more extensive result, which we call the Cut Locus Theorem. We now describe this result. We will identify the Lie algebra of Sol with $\boldsymbol{R}^3$ in a canonical way. See §2.1. Let $\Pi' \subset \boldsymbol{R}^3$ be the plane $\{z = 0\}$ in the Lie algebra of Sol. Let Π denote the plane $\{z = 0\}$ in Sol. Really Π' and Π are the same set of points, but we make the distinction to avoid confusion. We define sets

$$\partial_0 N' \subset \partial N' \subset \boldsymbol{R}^3 \text{ and } N' \subset \boldsymbol{R}^3, \qquad \partial_0 N \subset \partial N \subset \text{Sol} \text{ and } N \subset \text{Sol}$$

as follows.

- Let $N' \subset \boldsymbol{R}^3$ be the set of small vectors.
- Let $\partial N' \subset \boldsymbol{R}^3$ be the set of perfect vectors.
- Let $\partial_0 N' = \partial N' \cap \Pi'$.
- Let $\partial_0 N = E(\partial_0 N') \subset \Pi$.
- Let $\partial N \subset \Pi$ be the closure of the set of points which $\partial_0 N$ separates from the origin.
- Let $N = \text{Sol} - \partial N$.

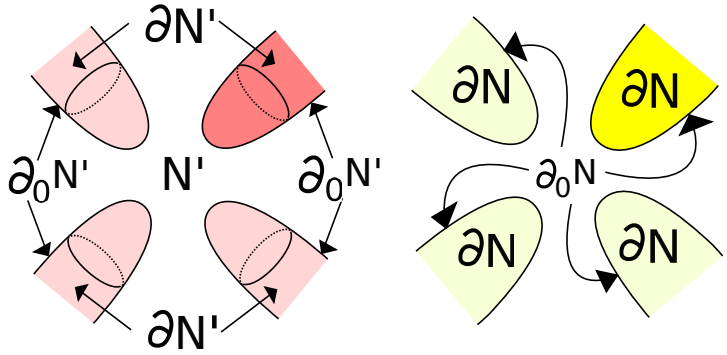

Figure 1: A schematic picture of the important sets.

Figure 1 shows a schematic picture of these sets. The picture on the left is in the Lie algebra. Our viewpoint is that we are looking down on the plane Π'. This is supposed to be a 3-dimensional picture. The set $\partial N'$ is a union of 4 topological planes, Each component of $\partial N'$ intersects each sphere of radius $L > \pi\sqrt{2}$ in a topological circle. The circles shrink to the points $(\pm\pi, \pm\pi, 0)$

as $L \to \pi\sqrt{2}$. Each component of $\partial N'$ bounds a solid pink region consisting entirely of large vectors. The region N' is the component of $\boldsymbol{R}^3 - \partial N'$ which is not pink. From the picture it may be hard to tell that the arrows for $\partial N'$ point to the surfaces of the pink sets and not the solid interior. The set $\partial_0 N'$ is a union of 4 curves, each dividing a component of $\partial N'$ in half.

The picture on the right is in Π, the plane $\{z = 0\}$ in Sol. This is a planar picture. The set N is not shown; it is the complement of the 4 yellow planar regions. Note the sets N and ∂N are defined entirely from the 1-dimensional set $\partial_0 N$. It turns out that $\partial_0 N$ is the disjoint union of 4 properly embedded curves, each diffeomophic to a line and the graph of a function in polar coordinates. See Lemma 3.5.

Our coloring in Figure 1 highlights the components on each side which lie in what we call the positive sector. The *positive sector* is the set of points (x, y, z), either in the Lie algebra or in Sol, with $x, y > 0$ While accurate topologically, our schemetic pictures are somewhat misleading geometrically. Figure 3 below shows accurate plots of $\partial N'$ and ∂N.

Theorem 1.2 (Cut Locus) *The following is true:*

1. *E induces a diffeomorphism from N' to N.*
2. *E induces a 2-to-1 local diffeomorphism from $\partial N' - \partial_0 N'$ to $\partial N - \partial_0 N$.*
3. *E induces a diffeomorphism from $\partial_0 N'$ to $\partial_0 N$.*

The Cut Locus Theorem gives ∂N as the cut locus of the identity in Sol.

Our main motivation for understanding the cut locus is to understand something about the spheres in Sol. We think that opinion had been divided as to whether or not the metric spheres in Sol are topological spheres. In §4.4 we deduce the following easy corollary of the Main Theorem.

Theorem 1.3 (Sphere) *Metric spheres in Sol are topological spheres. For the sphere S_L of radius L centered at the identity in Sol the following holds.*

1. *When $L < \pi\sqrt{2}$, the sphere S_L is smooth.*
2. *When $L = \pi\sqrt{2}$, the sphere S_L is smooth except (perhaps) at the 4 points $(x, y, 0)$ where $|x| = |y| = \pi$.*
3. *When $L > \pi\sqrt{2}$, the sphere S_L is smooth away from 4 disjoint arcs, all sayisfying $z = 0$ and $|xy| = H_L^2$ for some $H_L > \pi$.*

Remarks:
(1) We do not know whether the sphere $S_{\pi\sqrt{2}}$ is smooth at the 4 points $(x, y, 0)$ where $|x| = |y| = \pi$.

(2) The function $L \to H_L$ is defined by the following property.

$$L = \sqrt{8 + 8m}\mathcal{K}(m) \quad \Longrightarrow \quad H_L = \frac{4\mathcal{E}(m)}{\sqrt{1-m}} - \sqrt{4 - 4m}\mathcal{K}(m). \tag{5}$$

The range of m is $[0, 1)$. Here $\mathcal{K}$ and $\mathcal{E}$ are the complete elliptic integrals of the first and second kind:

$$\mathcal{K}(m) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - m\sin^2(\theta)}}, \qquad \mathcal{E}(m) = \int_0^{\pi/2} \sqrt{1 - m\sin^2(\theta)}\; d\theta. \tag{6}$$

See [**AS**, Eq. 16.1.1] and [**AS**, Eq. 17.3.7] respectively. Because we did many numerical experiments with Mathematica [**W**], we note that the integrals above agree with the functions called `EllipticK` and `EllipticE` in Mathematica [**W**, p 774]. One can derive Equation 5 from the formulas in [**G**] or [**T**], but we will not do so. We do not need this formula for our proofs.

(3) We have $H_L \exp(-L/4) \to 1$ as $L \to \infty$. This asymptotic formula is also mentioned on [**G**, p 75], though with a typo: a spurious factor of 2 in the formula.

1.3 Some Computer Plots

We include some computer plots of the sets that figure in our results. The Java program one of us wrote [**S**] generates these pictures and shows animations.

Figure 2 shows two different projections of small portions of the Sol sphere S_5 of radius 5 centered at the origin. The black arc is one of the singular arcs mentioned in the Sol Sphere Theorem. The grey curves are images of lines of longitude under the exponential map. The projections are designed to highlight the geometry of the singular arc.

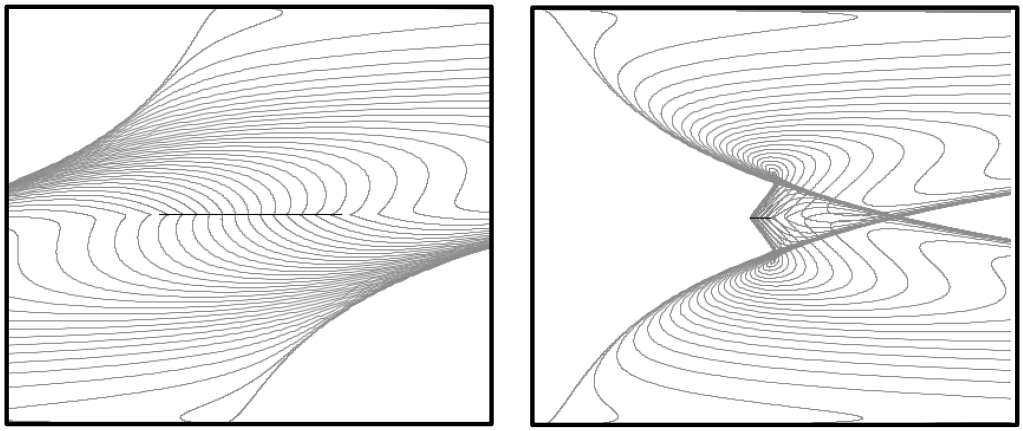

Figure 2: Two projections of the Sol metric sphere S_5.

Both projections reveal the planar nature of the singular set. The second projection also reveals a kind of concavity to the ball B_5 whose boundary is S_5. Were we to plot B_5, it would appear to the left of the plot in the second projection, in the white portion.

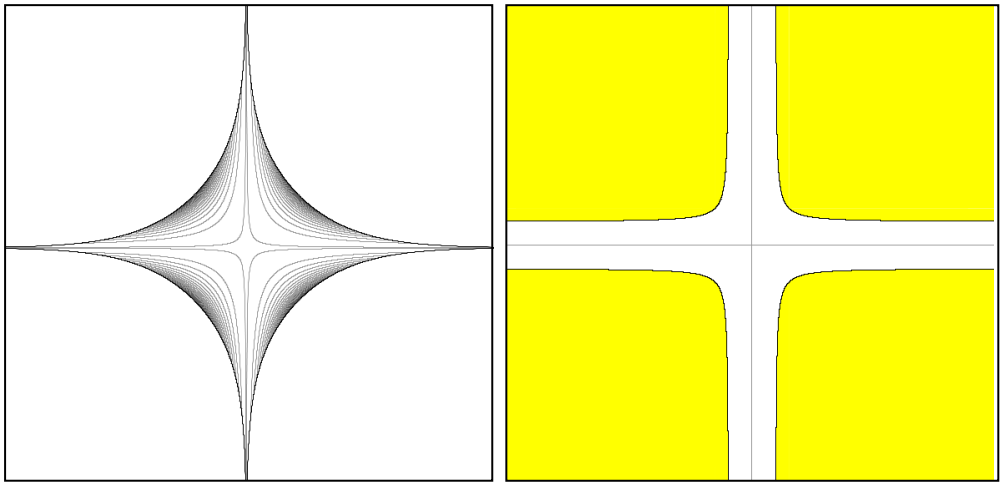

Figure 3: Plots of $\partial N'$ and ∂N.

The left side of Figure 3 shows a plot of the orthogonal projection of $\partial N'$ into the plane Π'. Each hyperbola-arc in the picture is the projection of $\partial N' \cap S'_r$ for some $r > \pi\sqrt{2}$. The larger the value of r, the closer the hyperbola-arc comes to the origin. Here S'_r is the sphere of radius r in the Lie algebra, and we let B'_r is the ball it bounds. The surface $\partial N'$ really hugs

the union $\Psi' = \{x = 0\} \cup \{y = 0\}$. For any $\epsilon > 0$ there is some r such that $\partial N' - B'_r$ is contained in the ϵ-tubular neighborhood of $\Psi' - B'_r$.

The right side of Figure 3 shows a plot of ∂N. Though we do not prove it in this paper the curves of $\partial_0 N$ turn out to be asymptotic to the lines $x = \pm 2$ and $y = \pm 2$. See [**S2**] for a proof.

1.4 Proof Outline

We first recall several standard definitions from Riemannian geometry. See e.g. [**KN**, §8] for details. A geodesic segment is a *minimizer* if it is the shortest geodesic segment connecting its endpoints. It is a *unique minimizer* if it is the only such geodesic of minimal length connecting its endpoints. Two points on a geodesic segment γ_0 are said to be *conjugate* if there is some nontrivial Jacobi field that vanishes at the two distinct points. A basic fact from Riemannian geometry is that if a geodesic segment is a minimizer, then every proper sub-segment is a unique minimizer without conjugate points. Call this the *restriction principle*. Now we can give the sketch.

Step 1: We call $V_+ = (x, y, z)$ and $V_- = (x, y, -z)$ *partners*. Note that V_+ is perfect if and only if V_- is perfect. Moreover, if $V_\pm$ is perfect, we prove that $E(V_+) = E(V_-)$. This is a surprising[1] result because the map $(x, y, z) \to (x, y, -z)$ is not an isometry of Sol. By the restriction principle (and a bit of fussing with the case $z = 0$), no large geodesic segment is a minimizer. We carry out this step in §2.

Step 2: This is the crucial step. We show that $E(N') \subset N$. This is equivalent to the statement that $E(N') \cap \partial N = \emptyset$. By symmetry, it suffices to prove this for vectors $V \in N'$ having non-negative coordinates. The condition that $E(V) \in \Pi$ places constraints on V. It turns out that there is a 2-parameter family of such vectors. We study the E-images of these vectors *via* a certain non-linear O.D.E. Our main result about this O.D.E., the Bounding Triangle Theorem, establishes that E does not map any of these vectors into ∂N. We carry out this step in §3.

Step 3: We show that $E(\partial N') \subset \partial N$. Combining this with step 2, we

[1]We are not the first to notice this kind of phenomenon. [**K**, Lemma 4.1] is the less precise result that geodesics tangent to partner vectors meet "at some point". Sungwoon Kim proves this by analytic methods that differ from our more geometric approach.

see that $E(\partial N') \cap E(N') = \emptyset$. The key point in showing that $E(\partial N') \subset \partial N$ is showing that E is injective on the closure of each component $\partial N' - \partial_0 N'$. This follows from our Corollary 2.10. We carry out this step in §4.1, though we prove Corollary 2.10 at the end of §2.

Step 4: Step 3 tells us that $E(N') \subset N$. Steps 1 and 3 tell us that if a perfect geodesic segment γ is not a minimizer, then the actual minimizer γ^* with the same endpoints must also be perfect. The injectivity result in Step 3 then implies that γ and γ^* are the geodesic segments associated to partner perfect vectors, and hence have the same length, a contradiction. Hence, perfect geodesic segments are minimizers. We also carry out this step in §4.1.

Step 5: By the restriction principle, small geodesic segments are unique minimizers without conjugate points. Now we can say that the cut locus is ∂N. The rest of the proof is quite easy. We finish the proof of the Cut Locus Theorem in §4.3. At the end of §4 we deduce the Sphere Theorem from the Cut Locus Theorem, and then the Main Theorem from the Cut Locus Theorem and Equation 21.

Finally, we mention that we defer some of the technical calculations until §5. The material in §5.1 and §5.2 just reproves results in [**G**], and we include it for the convenience of the reader. The material in §5.3 is new.

1.5 Acknowledgements

We thank ICERM for their fabulous Fall 2019 program, Illustrating Mathematics, during which this work was done. We thank Rémi Coulon, David Fisher, Bill Goldman, Alexander Holroyd, Boris Khesin Jason Manning, Greg McShane, Saul Schleimer, Henry Segerman, Sergei Tabachnikov, and Steve Trettel for many interesting discussions about Sol. We thank Matt Grayson for his great work on Sol. Finally, we thank the anonymous referees for their helpful suggestions.

2 Basic Structure

2.1 The Metric and its Symmetries

The underlying space for Sol is $\boldsymbol{R}^3$ and the group law is

$$(x, y, z) * (a, b, c) = (e^z a + x, e^{-z} b + y, c + z). \tag{7}$$

This is compatible with the left invariant metric on Sol given in Equation 4. For the sake of calculation, we mention two additional formulas:

$$(x, y, z)^{-1} = (-e^{-z} x, -e^z y, -z), \tag{8}$$

$$(x, y, z)^{-1} * (a, b, 0) * (x, y, z) = (e^{-z} a, e^z b, 0). \tag{9}$$

We identify $\boldsymbol{R}^3$ with the Lie algebra of Sol in such a way that the standard basis elements $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ respectively generate the 1-parameter subgroups $t \to (tx, 0, 0)$, $t \to (0, ty, 0)$ and $t \to (0, 0, tz)$. See §5.1 for a discussion of the left invariant vector fields extending the standard basis elements.

Sol has 3 interesting foliations.

- The xy foliation is by (non-geodesically-embedded) Euclidean planes.
- The xz foliation is by geodesically embedded hyperbolic planes.
- The yz foliation is by geodesically embedded hyperbolic planes.

The complement of the union of the two planes $x = 0$ and $y = 0$ is a union of 4 *sectors*. One of the sectors, the *positive sector*, consists of vectors of the form (x, y, z) with $x, y > 0$. The sectors are permuted by the Klein-4 group generated by isometric reflections in the planes $x = 0$ and $y = 0$. The Sol isometry $(x, y, z) \to (y, x, -z)$ permutes the sectors and preserves the positive sector. Because the coordinate planes $x = 0$ and $y = 0$ are geodesically embedded, the Riemannian exponential map E carries each open sector of the Lie algebra into the same open sector of Sol. We will abbreviate this by saying that E is *sector preserving*.

There are 3 kinds of geodesics in Sol:

1. Certain straight lines contained in xy planes.
2. Hyperbolic geodesics contained in the xz and yz planes.
3. The rest. We call these *typical*.

We discuss the nature of typical geodesics in Sol in the next section.

2.2 The Geodesic Flow

Let S' denote the sphere of unit vectors in the Lie algebra of Sol. Given a unit speed geodesic γ, the tangent vector $\gamma'(t)$ determines a left invariant vector field on Sol, and we let $\gamma^*(t) \in S'$ be the restriction of this vector field to $(0,0,0)$. Given an element $\sigma \in \text{Sol}$ we define LEFT_σ to be the left multiplication map by σ on Sol. In terms of left multiplication, we get the formula

$$\gamma^*(t) = d\text{LEFT}_{\gamma(t)^{-1}}(\gamma'(t)). \tag{10}$$

Here dLEFT is the differential of LEFT.

In §5.1 we verify that γ^* satisfies the following differential equation.

$$\frac{d\gamma^*(t)}{dt} = \Sigma(\gamma^*(t)), \qquad \Sigma(x,y,z) = (+xz, -yz, -x^2+y^2). \tag{11}$$

This is the point of view taken in [**A**] and [**G**]. This system in Equation 11 is really just the geodesic flow on the unit tangent bundle of Sol, viewed in a left-invariant reference frame. Our formula agrees with the one in [**G**] up to sign, and the difference of sign comes from the fact that our group law differs by a sign change from the one there.

This vector field Σ has Klein-4 symmetry and vanishes at the 6 points: $(0,0,\pm 1)$ and $(\pm 1/\sqrt{2}, \pm 1/\sqrt{2}, 0)$. The first two points are saddle singularities and the rest are elliptic. The geodesics corresponding to the elliptic singularities are straight (diagonal) lines in the plane $z = 0$. The geodesics corresponding to the saddle singularities are vertical geodesics in the xz and yz planes.

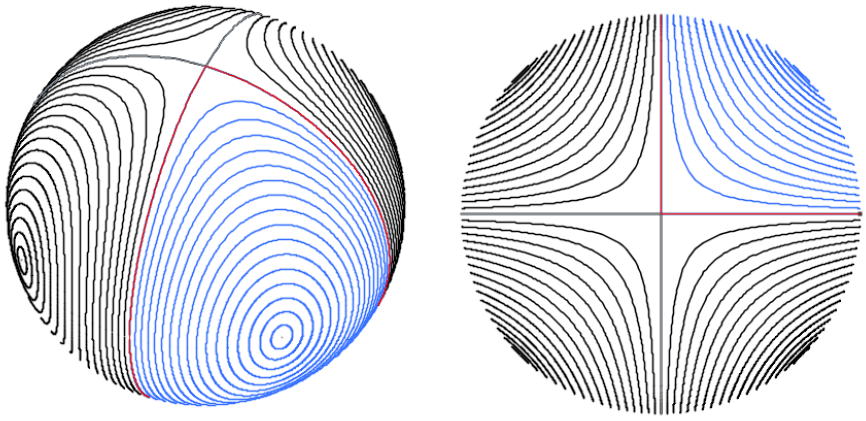

Figure 4: Trajectories of the vector field Σ.

The geodesics corresponding to the flowlines connecting the saddle singularities lie in the xz and yz planes; these are all geodesics of the second kind. The rest of the geodesics are typical. The flowlines corresponding to the typical geodesics lie on closed loops. Figure 4 shows 2 viewpoints of these level sets on S'. We have highlighted one of the sectors in blue.

Now we will restrict our attention to the typical geodesic segment. Let $F(x, y, z) = xy$. The restriction of F to S^2 gives a function on the sphere. The *symplectic gradient* X_F is defined by taking the gradient of this function (on the sphere) and rotating it 90 degrees counterclockwise. Up to sign $X_F = \Sigma$. By construction, the flow lines of Σ lie in the level sets of F.

Each loop level set Θ has an associated *period* $L = L_\Theta$, which is the time it takes a flowline – i.e., an integral curve – in Θ to flow exactly once around. Equation 21 below gives a formula. We can compare L to the length T of a geodesic segment γ associated to a flowline that starts at some point of Θ and flows for time T. In view of Equation 21, the geodesic segment γ is small, perfect, or large according as $T < L$, or $T = L$, or $T > L$. In other words, γ is small, perfect, or large according as the corresponding flowline travels less than once, exactly once, or more than once around its loop level set.

Geometric Interpretation: Here is a more direct geometric interpretation of what small, perfect, and large mean for the typical geodesic segment. Let $\widehat{\gamma}$ be a typical geodesic segment. We will see in §5.2 that $\widehat{\gamma}$ lies on the surface of a certain cylinder $C = C_{\widehat{\gamma}}$, which we call a *Grayson cylinder*.

The isometry group of C contains with finite index a copy of $\boldsymbol{R}$. That is, C has "translation symmetry". The quotient $C/\boldsymbol{R}$ is a topological circle. Let $\xi : C \to C/\boldsymbol{R}$ be the quotient map. The map ξ is locally injective on $\widehat{\gamma}$, which means essentially that $\widehat{\gamma}$ is winding around C in a nontrivial way, like a corkscrew. Let $\gamma \subset \widehat{\gamma}$ be a geodesic segment.

- The segment γ winds less than once around C if ξ is injective on γ.
- The segment γ winds exactly once around C if ξ is injective on the interior of γ but identifies the endpoints.
- Otherwise γ winds more than once around C.

The geodesic segment γ turns out to be small, perfect, or large according as γ winds less than once, exactly once, or more than once around its Grayson cylinder.

2.3 Concatenation

Let g be the flowline given by

$$g(t) = (x(t), y(t), z(t)), \qquad t \in [0, T]. \tag{12}$$

The corresponding geodesic segment is $\gamma_{Tg(0)}$. This geodesic has length T. We call g *small*, *perfect*, or *large* according as the corresponding initial tangent vector $Tg(0)$ is small, perfect, or large. We define

$$\Lambda_g = E(Tg(0)) \tag{13}$$

Here Λ_g is the far endpoint of the geodesic segment corresponding to g when this segment starts at the origin.

We use the notation $g = u|v$ to indicate that we are splitting the flowline g into sub-flowlines u and v. Here u is some initial part of g and v is the final part. It follows from the left invariant nature of the geodesics that

$$\Lambda_g = \Lambda_u * \Lambda_v \tag{14}$$

This is also a consequence of Equation 17 below.

While the elements Λ_u and Λ_v do not necessarily commute, their vertical displacements commute. This gives us

$$\pi_z \circ \Lambda_g = \pi_z \circ \Lambda_u + \pi_z \circ \Lambda_v. \tag{15}$$

Here π_z is projection onto the z-coordinate. Equation 15 has a nice integral form:

$$\pi_z(\Lambda_g) = \int_0^T z(t)\ dt. \tag{16}$$

Remark: The Arnold-Grayson point of view suggests a method for numerically simulating geodesics in Sol. We choose equally spaced times

$$0 = t_0 < t_1 < ... < t_n = T,$$

and consider the corresponding points $g_0, ..., g_n$ along the flowline g. We then have

$$\Lambda_g = \lim_{n\to\infty} (\epsilon_n g_0) * ... * (\epsilon_n g_n), \qquad \epsilon_n = T/(n+1). \tag{17}$$

In practice, we first pick some large n and then use Euler's method to find approximations to $g_0, ..., g_n$. We then take the product in Equation 17. We

used this method to reproduce the numerics in [**G**]. It is possible that there are more efficient numerical schemes for simulating the geodesics in Sol, but this method works well for our purposes and it makes the concatenation rule transparent.

Let us deduce some consequences from the equations above. We call g a *symmetric flowline* if the endpoints of g have the form $(x, y, +z)$ and $(x, y, -z)$. We may otherwise say that the endpoints of g are *partners*. Let Π be the plane $z = 0$.

Lemma 2.1 *A small flowline g is symmetric if and only if $\Lambda_g \in \Pi$.*

Proof: If g is symmetric, then the integral in Equation 16 vanishes, by symmetry. Hence $\pi_z(\Lambda_g) = 0$. If g is a small flowline having both endpoints on the same side of Π then $\pi_z(\Lambda_g) \neq 0$ because the integrand in Equation 16 either is an entirely negative function or an entirely positive function. In general, if g is not symmetric then we can write $g = u|w|v$ where u, v are either symmetric or empty, and w lies entirely above or entirely below Π. But then $\pi_z(\Lambda_g) = \pi_z(\Lambda_w) \neq 0$ by Equation 15. ♠

Lemma 2.2 *If g is a perfect flowline then $\Lambda_g \in \Pi$. If g_1 and g_2 are perfect flowlines in the same loop level set, and $\Lambda_{g_j} = (a_j, b_j, 0)$, then $a_1 b_1 = a_2 b_2$.*

Proof: We can write $g = u|v$ where u and v are both small symmetric flowlines. But then by Equation 15 and Lemma 2.1,

$$\pi_z(\Lambda_g) = \pi_z(\Lambda_u) + \pi_z(\Lambda_v) = 0 + 0 = 0.$$

Hence $\Lambda_g \in \Pi$ and we can write $\Lambda_g = (a, b, 0)$. We can write $g_1 = u|v$ and $g_2 = v|u$ for suitable choices of small flowlines u and v. Then $\Lambda_{g_1} = \Lambda_u * \Lambda_v$ and Then $\Lambda_{g_2} = \Lambda_v * \Lambda_u$. Hence Λ_{g_1} and Λ_{g_2} are conjugate in Sol. The second statement now follows from Equation 9. ♠

Our Theorem 2.3 below strengthens [**K**, Lemma 4.1], but the method of proof is completely different. Let E be the Riemannian exponential map.

Theorem 2.3 *If V_+ and V_- are perfect partners, then $E(V_+) = E(V_-)$.*

Proof: Let $g_\pm \subset S'$ be the flowline corresponding to $V_\pm$. We can write $g_+ = u|v$ and $g_- = v|u$ where u and v are small flowlines. Since V_+ and V_- are partners, we can take u and v both to be symmetric. But then the elements Λ_u and Λ_v both lie in the plane $z = 0$ and hence commute. Hence, by Equation 14, we have $E(V_+) = \Lambda_{g_+} = \Lambda_u * \Lambda_v = \Lambda_v * \Lambda_u = \Lambda_{g_-} = E(V_-)$. ♠

2.4 Large Geodesic Segments are not Minimizers

Now we complete Step 1 of our proof outline. Our result is essentially a corollary of Theorem 2.3, but we have to bring in some other results to handle special cases.

Lemma 2.4 *If $x, z > 0$ and (x, x, z) is perfect, then $E(x, x, z) = (h, h, 0)$ for some h.*

Proof: This is a result of [**G**]. Here is another proof. We can write our given perfect vector as $(x, x, z) = (Tm, Tm, Tn)$, where $(m, m, n) \in S'$ is a unit vector. Let g be the flowline corresponding to (m, m, n) and suppose our geodesic now has unit speed. We can write $g = u|v$ where u is the flowline starting at (m, m, n) and ending at $(m, m, -n)$ and v is the flowline starting at $(m, m, -n)$ and ending at (m, m, n). Both u and v are small symmetric arcs.

The map $\iota(x, y, z) = (y, x, -z)$ is an isometry both of $\boldsymbol{R}^3$ (the Lie algebra) and of Sol, and the exponential map commutes with this map. Acting on the Lie Algebra, ι swaps u and v. Hence $\Lambda_u = (\alpha, \beta, 0)$ and $\Lambda_v = (\beta, \alpha, 0)$ for some α, β. But then

$$E(x, x, z) = \Lambda_g = (\alpha, \beta, 0) * (\beta, \alpha, 0) = (h, h, 0),$$

with $h = \alpha + \beta$. ♠

Corollary 2.5 *A large geodesic segment is not a length minimizer.*

Proof: If this is false then, by the restriction principle, we can find a perfect geodesic segment γ, corresponding to a perfect vector $V = (x, y, z)$, which is a unique geodesic minimizer without conjugate points. If $z \neq 0$ we immediately

contradict Theorem 2.3. If $z = 0$ and $|x| \neq |y|$ we consider the variation, $\epsilon \to \gamma(\epsilon)$, through same-length perfect geodesic segments $\gamma(\epsilon)$ corresponding to the vector $V_\epsilon = (x_\epsilon, y_\epsilon, \epsilon)$. Here x_ϵ and y_ϵ are chosen to keep us on the same loop level set. The vectors V_ϵ and $V_{-\epsilon}$ are partners, so $\gamma(\epsilon)$ and $\gamma(-\epsilon)$ have the same endpoint. Hence, this variation corresponds to a conjugate point on γ, a contradiction.

It remains only to consider the segments connecting $(0,0,0)$ to $(t, \pm t, 0)$ with $|t| > \pi$. By symmetry it suffices to show that the segment connecting $(0,0,0)$ to $(t,t,0)$ is not a distance minimizer when $t > \pi$. This is proved in [**G**]. For the sake of completeness, we give another proof. It follows from Equation 21 below that there are values $h \in (\pi, t)$ such that $(h,h,0) = E(V)$ for some perfect vector V of the form (x,x,z) with $z \neq 0$. Hence, by the restriction principle, the segment connecting $(0,0,0)$ to $(t,t,0)$ is not a distance minimizer. ♠

2.5 The Reciprocity Lemma

In this section we prove a technical result which is a crucial ingredient for Step 2 of our outline. We discovered this result experimentally. It does not appear in [**G**]. The result strengthens Lemma 2.4.

Lemma 2.6 (Reciprocity) *Let $V = (x,y,z)$ be any perfect vector. Then there some $\lambda \neq 0$ such that $E(V) = \lambda(y,x,0)$.*

Proof: By symmetry it suffices to work in the positive quadrant. We write $\zeta = \zeta(t)$ for any quantity ζ which depends on t. Let $p = (x,y,z)$ be a flowline for the structure field Σ with initial conditions $x(0) = y(0)$ and (say) $z(0) > 0$. Let $(a,b,0) = E(x,y,z)$. We want to show that $a/b = y/x$ for all t. We do this by showing that the two functions satisfy the same O.D.E. and have the same initial conditions. We get the same initial conditions by Lemma 2.4: we have $a(0)/b(0) = 1 = y(0)/x(0)$.

We get the O.D.E. for y/x using the definition of Σ and the product rule:

$$\frac{d}{dt}\frac{y}{x} = \frac{y'x - x'y}{x^2} = \frac{-yzx - xzy}{x^2} = -2z \times \frac{y}{x}. \tag{18}$$

Now we work out the O.D.E. satisfied by a/b. By definition,

$$\frac{d}{dt}\frac{a}{b} = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\left(\frac{a(t+\epsilon)}{b(t+\epsilon)} - \frac{a(t)}{b(t)}\right).$$

Let $p(t,\epsilon)$ denote the minimal flowline connecting $p(t)$ to $p(t+\epsilon)$. Referring to the definition in §2.3, we have

$$\Lambda_{p(t,\epsilon)} \approx \epsilon(x,y,z). \tag{19}$$

Here the approximation means that we have equality up to order ϵ^2. Hence

$$(a(t+\epsilon), b(t+\epsilon), 0) = \Lambda^{-1}_{p(t,\epsilon)} * (a,b,0) * \Lambda_{p(t,\epsilon)} \approx$$

$$(\epsilon x, \epsilon y, \epsilon z)^{-1} * (a,b,0) * (\epsilon x, \epsilon y, \epsilon z) = (ae^{-\epsilon z}, be^{+\epsilon z}, 0).$$

The first equality is Equation 14. The approximation (to order ϵ^2) comes from Equation 19. The last equality is Equation 9. But then

$$\frac{d}{dt}\frac{a}{b} = \lim_{\epsilon \to 0} \frac{e^{-2\epsilon z}-1}{\epsilon} \times \frac{a}{b} = -2z \times \frac{a}{b}. \tag{20}$$

Therefore a/b satisfies the same O.D.E. as does y/x. ♠

2.6 The Period Function

Let L_a be the period of the loop level set containing $U_a = (a, a, \sqrt{1-2a^2})$.

Lemma 2.7 *$dL_a/da < 0$.*

Proof: This is part of [**G**, Lemma 3.2.1], and it also follows from the formula

$$L_a = \frac{\pi}{\mathrm{AGM}(a, \frac{1}{2}\sqrt{1+2a^2})}. \tag{21}$$

We derive this formula in §5.2. ♠

Remarks:
(1) The denominator on the right side of Equation 21 is just $\mu(U_a)$. Here μ is the formula that makes its appearance in our main results.
(2) What we see from Equation 21 is that the period of a loop level set increases monotonically from $\pi\sqrt{2}$ to ∞ as the level sets move outward from the equilibrium points of Σ, namely $(\pm\sqrt{2}/2, \pm\sqrt{2}/2, 0)$. Thus, the range of periods is $(\pi\sqrt{2}, \infty)$.

Lemma 2.8 *Let $V_0 = (x, y, z)$ be a perfect vector with $x, y, z > 0$. Then E is a local diffeomorphism in a neighborhood of V_0.*

Proof: By the Inverse Function Theorem, this is the same as saying that the differential dE has full rank at V_0. Let S' be the sphere in the Lie algebra centered at the origin and containing V_0. Let T_0 be the tangent plane to S' at V_0. Let N_0 be the orthogonal complement of T_0. As is well known, the images $dE|_{V_0}(T_0)$ and $dE|_{V_0}(N_0)$ are orthogonal, and the latter space is 1-dimensional. So, we just need to show that $dE|_{V_0}(T_0)$ contains 2 linearly independent vectors. Below, the symbols $O(t)$ and $O(1)$ denote quantities which respectively are bounded below by positive constants times t and 1. Both our variations below consist of vectors all having the same length. Before we proceed, we define the *unit normalization* of a nonzero vector V to be $V/\|V\|$.

Let $V_t \in S'$ be a unit speed curve of perfect vectors which moves at unit speed away from V_0 and which have the property that the unit normalizations stay in the a single loop level set. The projection of V_t into Π' moves at speed $O(1)$ because $V_t \notin \partial_0 M$. This point moves monotonically along a hyperbola. By the Reciprocity Lemma, $E(V_t)$ moves with speed $O(1)$ in Π. Hence $dE_{V_0}(T_0)$ contains a nonzero vector of the form $(x_1, y_1, 0)$.

As above, let Π' be the plane $\{z = 0\}$ in the Lie algebra. Now let $V_t \in S'$ be the curve of constant-length vectors moving at unit speed whose unit normalizations move orthogonally to the loop level sets. We choose the direction so that V_t is a small vector for $t > 0$. Let g_t be the flowline corresponding to $V_t/\|V_t\|$. Let Θ_t be the loop level set containing g_t. Let h_t be the complementary flowline, so that $g_t|h_t$ is a perfect flowline in Θ_t. By Equation 16, we have $\pi_z \circ E(V_t) = -\pi_z(\Lambda_{h_t})$. Let $L(t)$ be the period of Θ_t. By Lemma 2.7 we have $dL/dt > 0$. Hence h_t travels for time $O(t)$. The distance from Π' to h_t is $O(1)$. Therefore, by Equation 15, we have $|\pi_z(\Lambda_{h_t})| = O(t)$. Hence $dE|_{V_0}(T_0)$ contains a vector (x_2, y_2, z_2) with $z_2 \neq 0$. ♠

2.7 The Holonomy Function

If V is a perfect vector, and $(a, b, 0) = E(V)$, then we let $H(V) = \sqrt{|ab|}$. We call $H(V)$ the *holonomy invariant* of V. By Lemma 2.2, this only depends on the loop level set. Thus H is a function of L, the level set period. By

Equation 21 we have $L \geq \pi\sqrt{2}$. Since $(\pi, \pi, 0)$ is a perfect vector, we have and $H(\pi\sqrt{2}) = \pi$.

The next result is stated on [**G**, p 78]. We give an independent proof.

Lemma 2.9 *$dH/dL \geq 0$, with strict inequality when when $L > \pi\sqrt{2}$. Also, H is a proper monotone increasing function of L.*

Proof: Let us first show that H is an unbounded function. Pick an arbitrary $R > 0$ and let V be the shortest vector such that $E(V) = (R, R, 0)$. Corollary 2.5 says that V is either small or perfect. In either case, there is some $\lambda \geq 1$ such that λE is perfect. Geodesic segments in the positive sector cannot be tangent to the coordinate planes $x = x_0$ or $y = y_0$. Hence $E(\lambda V) = (a, b, 0)$ with $a, b \geq R$. Hence $H(\|\lambda V\|) \geq R$.

Now we know that H is unbounded. Suppose there is $L > \pi\sqrt{2}$ where $H'(L) = 0$. Consider the perfect vectors $U_t = (x_t, x_t, z_t)$, with positive coordinates, such that $\|U_t\| = L + t$. By Lemma 2.4, we have $E(U_t) = (a_t, a_t, 0)$. Since $H'(L) = 0$ we have $da/dt(0) = 0$. This shows that dE is singular at U_0. But this contradicts Lemma 2.8. Hence H' has just one sign on $(\pi\sqrt{2}, \infty)$. Since H is unbounded, the sign must be positive. Since H is monotone and unbounded, H is proper. ♠

Our final result is not in [**G**].

Corollary 2.10 *The map E is injective on the set of perfect vectors having all non-negative coordinates.*

Proof: Since V_1 and V_2 have the same holonomy invariant, Lemma 2.9 implies $\|V_1\| = \|V_2\|$. But then Lemma 2.7 implies that $U_1 = V_1/\|V_1\|$ and $U_2 = V_2/\|V_2\|$ lie in the same loop level set. Hence $U_{11}U_{12} = U_{21}U_{22}$. But then $V_{11}V_{12} = V_{21}V_{22}$. Here U_{ij} and V_{ij} respectively are the jth coordinates of U_i and V_i. The Reciprocity Lemma says that $V_{12}/V_{11} = V_{22}/V_{21}$. Hence $V_{11} = V_{21}$ and $V_{21} = V_{22}$. Since $\|V_1\| = \|V_2\|$ we have $V_{13} = V_{23}$ as well. ♠

3 Controlling Small Geodesic Segments

3.1 The Main Argument

The goal of this chapter is to prove that $E(N') \subset N$, where N' and N are as in the Cut Locus Theorem. This is Step 2 of our outline. Given any set S, either in the Lie algebra or in Sol, let S_+ denote the intersection of S with the positive sector, i.e. the set of points $(x, y, z) \in S$ with $x, y > 0$.

Our next result refers to the interaction between the pink triangle and the yellow region in Figure 5 below. Given a point $p = (a, b, 0) \in \partial_0 N_+$ with $a > b > 0$ let $\Delta(p)$ be the right triangle with vertices $(0, 0, 0)$, $(a, 0, 0)$ and $(a, b, 0)$. We remark on the condition $a > b$ in §3.2, where we prove:

Theorem 3.1 (Triangle Avoidance) $\partial N_+ \cap \Delta(p) = \{p\}$.

Let $(N')^{\rm symm}_+ \subset N'$ denote those vectors having all positive coordinates which correspond to small symmetric flowlines in the sense of Lemma 2.1. We write

$$(N')^{\rm symm}_+ = \bigcup_{L > \pi\sqrt{2}} (N')^{\rm symm}_{L,+} \tag{22}$$

where $(N')^{\rm symm}_{L,+}$ is the subset of those vectors V associated to the loop level set of period L. That is, the unit normalizations lie in the loop level set of period L. Define

$$\Lambda_L = E((N')^{\rm symm}_{L,+}). \tag{23}$$

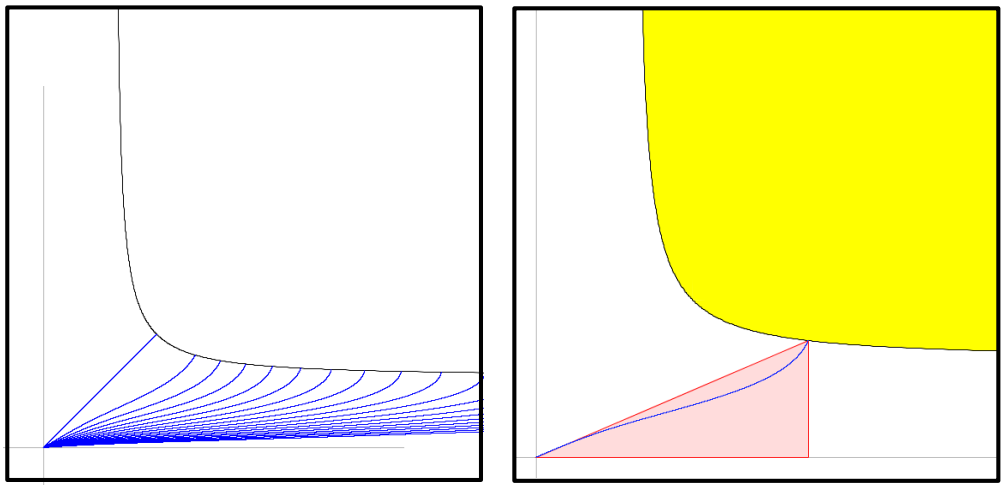

Figure 5: $\partial_0 N_+$ (black), ∂N_+ (yellow), Λ_L (blue), and Δ_L (pink).

The blue curves in Figure 5 are various curves Λ_L. (Technically, the blue segment lying in the diagonal is the limit of Λ_L as $L \to \pi\sqrt{2}$.) The left side shows many of these curves and the right side shows Λ_5. We define

$$\Delta_L = \Delta(\Lambda_L(\ell)). \tag{24}$$

The pink triangle in Figure 5 is Δ_5. In the §3.3 we prove:

Theorem 3.2 (Bounding Triangle) $\Lambda_L \subset \text{interior}(\Delta_L)$ *for all* $L > \pi\sqrt{2}$.

Corollary 3.3 $\Lambda_L \cap \partial N_+ = \emptyset$ *for all* $L > \pi\sqrt{2}$.

Proof: Once we know that Δ_L satisfies the hypotheses of the Triangle Avoidance Theorem, this result is an immediate consequence of the Bounding Triangle Theorem and the Triangle Avoidance Theorem. Let us check the needed fact about Δ_L. By the Reciprocity Lemma and symmetry we have

$$\frac{a(\ell)}{b(\ell)} = \frac{y(\ell)}{x(\ell)} = \frac{x(0)}{y(0)} > 1.$$

This means that $\Lambda_L(\ell) = (a(\ell), b(\ell), 0)$ satisfies $a(\ell) > b(\ell) > 0$. ♠

Corollary 3.4 $E(N') \subset N$.

Proof: We suppose that $E(N') \not\subset N$ and derive a contradiction. This is equivalent to the statement that there is some $V = (x, y, z) \in N'$ such that $E(V) \in \partial N$. By symmetry, it suffices to assume that $x, y, z \geq 0$. Since E is sector-preserving, we must have $E(V) \in \partial N_+$ and moreover $x, y > 0$. Either V is associated to a small flowline in a loop level set or else $V/\|V\|$ is a critical point of the vector field Σ. In this latter case, $V = (x, x, 0)$ and $E(V) = (x, x, 0)$ and $x < \pi$. But this point does not lie in ∂N. Hence V is associated to some small flowline in a loop level set.

Recall that Π is the plane $\{z = 0\}$ in Sol. Since $\partial N_+ \subset \Pi$, we must have $E(V) \in \Pi$. But then, V is associated to a small *symmetric* flowline, by Lemma 2.1. In this case, we must have $z > 0$ because the endpoints of small symmetric flowlines are partner points in the sense of §2.3. So, we have produced $V \in (N')^{\text{symm}}_+$ such that $E(V) \in \partial N_+$. That is, we have produced some $L > \pi\sqrt{2}$ such that Λ_L intersects ∂N_+. This contradicts Corollary 3.3. ♠

3.2 Proof of the Triangle Avoidance Theorem

If we knew that $\partial_0 N_+$ was a graph in Cartesian coordinates, we could give a simpler proof of the Triangle Avoidance Theorem, and we would not even need the hypothesis that $a > b$. Since we don't know this, we have to work harder. The rays through the origin and the hyperbolas of the form $xy = C$ make a kind of coordinate grid in Π_+. When $a > b > 0$ the two coordinate curves through $(a, b, 0)$ turn out to separate $\partial_0 N_+$ from $\Delta(p) - \{p\}$. Our first result (which does not need the hypothesis $a > b$) establishes the desired separation for the rays and the second result (which crucially uses $a > b$) establishes the desired separation for the hyperbolas.

Lemma 3.5 *The set $\partial_0 N_+$ is the graph of a function in polar coordinates, diffeomorphic to $\boldsymbol{R}$, and properly embedded in Π.*

Proof: First of all, the set $\partial_0 N'_+$ is the graph of a smooth function in polar coordinates. By Equation 21, the function is

$$f(\theta) = \frac{\pi}{\text{AGM}(\sin(\theta), \cos(\theta))}. \tag{25}$$

Now we turn to $\partial_0 N_+$. The polar defining function g for $\partial_0 N_+$ is

$$g(\theta) = \sqrt{2/\sin(2\theta)} \times H(f(\theta)) \geq \frac{\pi\sqrt{2}}{\sqrt{\sin(2\theta)}}. \tag{26}$$

Here H is the holonomy function from Lemma 2.9. The statements in the lemma follow from this formula.

Here we derive Equation 26. Let $V = (x, y, 0) \in \partial_0 N'_+$ be the vector which makes angle θ with the X-axis. By the Reciprocity Lemma we have $E(V) = \lambda(y, x, 0)$. This gives us

$$g(\theta) = \|E(V)\| = \lambda\sqrt{x^2 + y^2}, \qquad H(f(\theta)) = H(\|V\|) = \lambda\sqrt{xy}.$$

We also have the trig identity:

$$\frac{\sqrt{x^2 + y^2}}{\sqrt{xy}} = \sqrt{2/\sin(2\theta)},$$

Equation 26 comes from these relations and a bit of algebra. The inequality in Equation 26 comes from Lemma 2.9 and the fact $H(\pi\sqrt{2}) = \pi$. ♠

Lemma 3.6 *Let $(\theta, g(\theta))$ be the polar parametrization of $\partial_0 N_+$. Let $(x_\theta, y_\theta, 0)$ be the corresponding point in cartesian coordinates. Then the product $x_\theta y_\theta$ is strictly monotonically increasing as θ decreases from $\pi/4$ to 0.*

Proof: Let f and g be the polar functions considered in previous lemma. Let $\theta^* = \pi/2 - \theta$. The Reciprocity Lemma says that $E : \partial_0 N'_+ \to \partial_0 N_+$, when expressed in these polar parametrizations, maps $(\theta^*, f(\theta^*))$ to $(\theta, g(\theta))$. As θ^* increases from $\pi/4$ to $\pi/2$ the period $L(\theta^*)$ corresponding to the point $(\theta^*, f(\theta^*))$ increases strictly monotonically from $\pi\sqrt{2}$ to ∞. Therefore, by Lemma 2.9, the holomomy $H(\theta) = \sqrt{x_\theta y_\theta}$ increases strictly monotonically from π to ∞ as θ decreases from $\pi/4$ to 0. ♠

Proof of the Triangle Avoidance Lemma: Let θ_0 be the angle that the hypotenuse of $\Delta(p)$ makes with the x-axis. Since $a > b$ we have $\theta_0 < \pi/4$. The point $(\theta_0, g(\theta_0))$ in Cartesian coordinates is $(a, b, 0)$. When $\theta > \theta_0$ the point $(\theta, g(\theta))$ lies above the ray through the origin extending the hypotenuse of $\Delta(p)$ and hence is not contained in $\Delta(p)$. When $\theta > \theta_0$ we have $x_\theta y_\theta > ab$ by Lemma 3.6. Hence the hyperbola $xy = ab$ separates such points from $\Delta(p)$. We have shown that $\partial_0 N_+$ intersects $\Delta(p)$ only in the vertex $(a, b, 0)$. But $\partial_0 N_+$ is the boundary of ∂N_+. Hence ∂N_+ has this property as well. ♠

3.3 Proof of the Bounding Triangle Theorem

We first outline the proof.

1. We parametrize the curve Λ_L.
2. We examine the parametrization and work out the differential equation satified by Λ_L and the associated quantities.
3. We reduce the Bounding Triangle Theorem to an inequality which essentially says that Λ_L lies beneath the hypotenuse of Δ_L.
4. We take a preliminary step towards the inequality, showing that Λ_L is tangent to the hypotenuse of Δ_L at the origin. Compare Figure 5.
5. We use the differential equation to establish the inequality.

1. Parametrizing the Curve: We first explain how to parametrize Λ_L. Let Θ_L be the loop level set of period L in the positive sector. Each vector in $(N')^{\text{symm}}_{+,L}$ corresponds to a small symmetric flowline of Θ_L whose initial point has positive z-coordinate. Figure 6 shows these flowlines schematically.

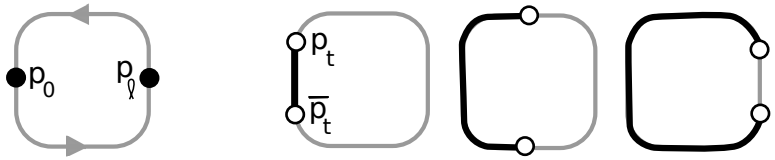

Figure 6: Increasingly long symmetric flowlines.

We let $\ell = L/2$. We let p_0 be the point on Θ_L having coordinates $(x, y, 0)$ with $x > y$. Let

$$p_t = (x(t), y(t), z(t)). \tag{27}$$

be the point such that the flowline of length t starting at p_t ends at p_0. In other words, we get to p_t by flowing *backwards* along the vector field Σ by t units. The small symmetric flowline g_t associated to p_t is the one of length $2t$. It starts at p_t and ends at the partner point $\overline{p}_t$.

Let $V_t \in (N')^{\text{symm}}_{+,L}$ be the vector corresponding to g_t. Let

$$\Lambda_L(t) := \Lambda_{g_t} = E(V_t) = (a(t), b(t), 0) \qquad t \in (0, \ell]. \tag{28}$$

As $t \in (0, \ell)$ the image $\Lambda_L(t)$ sweeps out Λ_L.

2. The Differential Equations: We work out the differential equations governing our parametrization. Remembering that we flow backwards to get from p_0 to p_t, we see that p satisfies the following O.D.E.

$$\frac{dp}{dt} = (x', y', z') = -\Sigma(x, y, z) = (-xz, +yz, x^2 - y^2).$$

To work out the formulas for a' and b' we proceed as we did during the proof of the Reciprocity Lemma. We write $g_{t+\epsilon} = u|g_t|v$, where u is the flowline connecting $p_{t+\epsilon}$ to p_t and v is the flowline connecting $\overline{p}_t$ to $\overline{p}_{t+\epsilon}$. We have

$$(a', b', 0) = \Lambda_L'(t) = \lim_{\epsilon \to 0} \frac{\Lambda_L(t+\epsilon) - \Lambda(t)}{\epsilon},$$

$$\Lambda_L(t+\epsilon) \approx (\epsilon x, \epsilon y, \epsilon z) * (a, b, 0) * (\epsilon x, \epsilon y, -\epsilon z).$$

The approximation is true up to order ϵ^2 and $(*)$ denotes multiplication in Sol. A calculation gives $a' = 2x + az$ and $b' = 2y - bz$. In summary:

$$a' = 2x + za, \quad b' = 2y - zb, \quad x' = -xz, \quad y' = yz, \quad z' = x^2 - y^2. \tag{29}$$

Using Equation 29 we compute

$$a'' = (+x^2 - y^2 + z^2)a, \qquad b'' = (-x^2 + y^2 + z^2)b. \tag{30}$$

3. Reduction to an Inequality: Here we reduce the Bounding Triangle Theorem to an inequality. Referring to Equations 27 and 28, we have $a, b, x, y, z > 0$ on $(0, \ell)$. Since $b > 0$, the curve Λ_L avoids the bottom edge of Δ_L. From Equation 29 we have $a' > 0$. This implies that Λ_L is the graph of a function and hence avoids the vertical edge of Δ_L.

We define

$$f(t) = \phi(t) - \phi(\ell), \qquad \phi(t) = \frac{b(t)}{a(t)}. \tag{31}$$

We will prove that

$$f(t) < 0, \qquad t \in (0, \ell). \tag{32}$$

This implies that $\Lambda_L(t)$ lies beneath the hypotenuse of Δ_L.

Given the truth of Equation 32, we see that Λ_L avoids all three sides of Δ_L and also lies beneath the hypotenuse of Δ_L. These facts imply that $\Lambda_L \subset \text{interior}(\Delta_L)$. To finish the proof of the Bounding Triangle Theorem, we just need to establish Equation 32.

4. A Preliminary Step: We claim that f extends continuously to 0 and $f(0) = 0$. What this means geometrically is that the hypotenuse of Δ_L is tangent to Λ_L at the origin. To verify our claim, we note that $a(0) = b(0) = 0$ and then we use L'Hopital's rule:

$$\phi(0) = \lim_{t \to 0} \frac{b'(t)}{a'(t)} = \lim_{t \to 0} \frac{2y(t) - b(t)z(t)}{2x(t) + a(t)z(t)} = \frac{y(0)}{x(0)} = \frac{x(\ell)}{y(\ell)} = \frac{b(\ell)}{a(\ell)} = \phi(\ell). \tag{33}$$

The penultimate equality above is due to the Reciprocity Lemma, applied to the perfect vector V_ℓ. We also claim that the function

$$\psi = ab' - ba' \tag{34}$$

extends continuously to 0: To see this claim, we note that $\psi(0) = 0$ because $a(0) = b(0) = 0$ and a', b' do not blow up at 0.

5. Proof of the Inequality: Now we establish Equation 32. Consider the endpoints first: We have $f(0) = f(\ell) = 0$. If $f \geq 0$ somewhere on $(0, \ell)$ then f has a local maximum at some $t_0 \in (0, \ell)$. We have $f'(t_0) = 0$ and $f''(t_0) \leq 0$. Recalling that $\psi = ab' - ba'$, and that $\phi = b/a$, and that ϕ and f differ by a constant, we have

$$\psi = ab' - ba' = a^2\phi' = a^2 f', \qquad \psi' = 2aa' f' + a^2 f''. \tag{35}$$

Hence $\psi(t_0) = 0$ and $\psi'(t_0) \leq 0$. We are going to look at the differential equations and show that in fact $\psi(t_0) < 0$. This gives us the contradiction that establishes Equation 32.

Combining Equation 30 with the fact that $\psi' = ab'' - ba''$, we get

$$\psi' = 2ab(y^2 - x^2). \tag{36}$$

We note that $a, b > 0$ and that $y(t)^2 - x(t)^2$ is monotone increasing and vanishes at $t = \ell/2$. Therefore $\psi' < 0$ on $(0, \ell/2)$ and $\psi' > 0$ on $(\ell/2, \ell)$. Since $\psi'(t_0) \leq 0$ we have $t_0 \leq \ell/2$. But then, using the fact that $\psi(0) = 0$ and $\psi' < 0$ on $(0, \ell/2)$ we have

$$\psi(t_0) = \int_0^{t_0} \psi'(t)dt < 0.$$

This is the contradiction we wanted. Our proof is done.

4 The Main Results

4.1 Separating Small and Perfect Vectors

Here we carry out Step 3 of the outline, where we show that the Riemannian exponential map separates the small and perfect vectors. That is, we show that $E(\partial N') \cap E(N') = \emptyset$.

We begin by fixing some notation. Recall that Π is the plane $\{z = 0\}$ in Sol. Let $\partial N'_+$ be set of perfect vectors of the form (x, y, z) with $x, y > 0$ and $z \geq 0$. Let $\partial_0 N'_+$ be the set of perfect vectors of the form $(x, y, 0)$ with $x, y > 0$ Referring to Figure 1, the set $\partial N'_+$ is "half" the boundary of the highlighted pink region.

Lemma 4.1 $E(\partial N'_+ - \partial_0 N'_+) \subset \partial N_+ - \partial_0 N_+$.

Proof: For ease of notation write $M' = \partial N'_+ - \partial_0 N'_+$ and $M = \partial N_+ - \partial_0 N_+$. By Corollary 2.10, the map E is injective on $\partial N'_+$. Also, $E(\partial_0 N'_+) = \partial_0 N_+$. Hence

$$E(M') \subset \Pi - \partial_0 N_+. \tag{37}$$

The set on the right hand side of Equation 37 has M as one of its components. Since M' is connected, $E(M')$ either lies in M or is disjoint from M. By Lemma 2.4 and Lemma 2.9 we have $E(V) \in M$ when $V \in M'$ has the form (x, x, z) and $\|V\|$ is large. Since $E(M')$ intersects M, we have $E(M') \subset M$. ♠

Corollary 4.2 $E(\partial N') \cap E(N') = \emptyset$.

Proof: By the previous result, and symmetry, we have $E(\partial N' - \partial_0 N') \subset \partial N$. By definition, $E(\partial_0 N') = \partial_0 N \subset \partial N$. Therefore

$$E(\partial N') \subset \partial N. \tag{38}$$

By Step 2 of our outline, $E(N') \subset N$. This corollary now follows from the fact that N and ∂N are disjoint. ♠

4.2 Proof of the Main Theorem

We already know from Step 1 of the outline that a geodesic segment is a distance minimizer only if it is small or perfect. We finish the proof of the Main Theorem by establishing the converse. By the Restriction Principle, it suffices to show that perfect geodesic segments are length minimizing.

Suppose $V_1 \in \partial N'$ and $E(V_1) = E(V_2)$ for some V_2 with $\|V_2\| < \|V_1\|$. We take V_2 to be the shortest vector with this property. By symmetry we can assume both V_1 and V_2 have all coordinates non-negative. By Corollary 2.5, we have $V_2 \in N' \cup \partial N'$. By Corollary 4.2 we have $V_2 \in \partial N'$. But then $V_1 = V_2$ by Corollary 2.10. This is a contradiction. This completes the proof.

4.3 Proof of the Cut Locus Theorem

For convenience we repeat the statement of the Cut Locus Theorem.

Theorem 4.3 (Cut Locus) *The following is true:*

1. *E induces a diffeomorphism from N' to N.*
2. *E induces a 2-to-1 local diffeomorphism from $\partial N' - \partial_0 N'$ to $\partial N - \partial_0 N$.*
3. *E induces a diffeomorphism from $\partial_0 N'$ to $\partial_0 N$.*

Proof of Statement 1: By Step 2 of the outline, $E(N') \subset N$. By the Restriction Principle, members of N', corresponding to small geodesic segments, are unique length minimizers without conjugate points. Hence $E : N' \to N$ is an injective local diffeomorphism. We just need to show E is proper. Suppose $\{V_n\}$ is a sequence in N' which exits every compact subset of N'. If $\|V_n\| \to \infty$ then, since vectors in N' correspond to distance minimizing geodesics, $\{E(V_n)\}$ diverges to ∞. If $\|V_n\|$ remains bounded then $V_n \to \partial N'$ and, by continuity, $E(V_n) \to \partial N$. Hence, in both cases, $\{E(V_n)\}$ exits every compact subset of N. ♠

Proof of Statement 2: By symmetry and Theorem 2.3 it suffices to prove that $E : M \to M'$ is a diffeomorphism, where M and M' are as in Lemma 4.1. By Corollary 2.10, the map $E : M' \to M$ is injective. By Lemma 2.8, the map $E : M' \to M$ is a local diffeomorphism. A properness argument just like the one above finishes the proof. ♠

Proof of Statement 3: By Lemma 3.5 and symmetry, both $\partial_0 N'$ and $\partial_0 N$ are unions of 4 curves, each one the graph of a smooth function in polar coordinates. The map $E : \partial_0 N' \to \partial_0 N$ is a local diffeomorphism by symmetry and the Reciprocity Lemma, and proper for the same reasons as in the previous cases. Hence this map is a diffeomorphism. ♠

4.4 Proof of The Sphere Theorem

Let S'_L and S_L respectively denote the spheres of radius L about the origin in the Lie algebra and in Sol. For convenience we repeat the statement of the Sphere Theorem.

Theorem 4.4 (Sphere) *Metric spheres in Sol are topological spheres. For the sphere S_L of radius L centered at the identity in Sol the following holds.*

1. *When $L < \pi\sqrt{2}$, the sphere S_L is smooth.*
2. *When $L = \pi\sqrt{2}$, the sphere S_L is smooth except (perhaps) at the 4 points $(x, y, 0)$ where $|x| = |y| = \pi$.*
3. *When $L > \pi\sqrt{2}$, the sphere S_L is smooth away from 4 disjoint arcs, all sayisfying $z = 0$ and $|xy| = H_L^2$ for some $H_L > \pi$.*

Proof of Statement 1: When $L < \pi\sqrt{2}$, we have $S'_L \subset N'$. Hence $S_L = E(S'_L)$, and the restriction of E to S'_L is a diffeomorphism by Statement 1 of the Cut Locus Theorem. ♠

Proof of Statement 2: Let $T' = \{(x, y, 0) |\ |x| = |y| = \pi\}$. Let $T = E(T')$. Really T and T' are the same set of 4 points. When $L = \pi\sqrt{2}$, we have $S'_L \subset N' \cup T'$. By Statements 1 and 3 of the Cut Locus Theorem, the map

$$E : N' \cup T' \to N \cup T$$

is a homeomorphism. Hence $S_L = E(S'_L)$ is a topological sphere. Again by Statement 1 of the Cut Locus Theorem, $S_L - T = E(S'_L - T')$ is smooth. ♠

Proof of Statement 3: Suppose $L > \pi\sqrt{2}$. Define

$$S''_L = S'_L \cap (N' \cup \partial N'). \tag{39}$$

The space S''_L is a 4-holed sphere. The boundary $\partial S''$ consists of 4 loops, each contained in $\partial N'$, each homothetic to the loop level set of period L, each having holonomy invariant H_L. It follows from the Cut Locus Theorem that $S_L = E(S''_L)$ and that E is a diffeomorphism when restricted to $S''_L - \partial S''_L$. On $\partial S''_L = S''_L \cap \partial N'$, the map E is a 2-to-1 folding map which identifies partner points within each component. Thus, we see that S_L is obtained from a 4-holed sphere by gluing together each boundary component (to itself) in a 2-to-1 fashion. This reveals S_L to be a topological sphere. By Statement 1 of the Cut Locus Theorem, S_L is smooth away from $E(\partial S''_L)$. Finally, from our description of $\partial S''_L$ and by definition of the quantity H_L, we see that $E(\partial S''_L)$ lies in the union of 4 planar arcs satisfying $z = 0$ and $|xy| = H_L^2$. ♠

5 Technical Calculations

5.1 The Structure Field

In this section we derive Equation 11. The derivation is a bit different from the one on [**G**, pp 62-65]. Let $\{e_1, e_2, e_3\}$ denote the standard Euclidean orthonormal basis. Let E_j be the left invariant vector field which agrees with e_j at $(0,0,0)$. The triple $\{E_1, E_2, E_3\}$ is a left-invariant orthonormal framing of Sol. If we express the derivative γ' of a unit speed geodesic γ in terms of our left-invariant framing, namely

$$\gamma'(t) = \sum u_i(t) E_i,$$

then Equation 11 describes the evolution of the coefficients. For convenience, we have set $x(t) = u_1(t)$ and $y(t) = u_2(t)$ and $z(t) = u_3(t)$.

Let ∇ denote the covariant derivative for Sol. The fact that γ is a geodesic means that the covariant derivative of γ' along γ vanishes. That is,

$$0 = \nabla_{\gamma'}(\gamma') = \sum_i \frac{du_i}{dt} E_i + \sum_{ij} u_i u_j \nabla_{E_j} E_i. \tag{40}$$

Parallel translation along any curve contained in a totally geodesic plane Π preserves the unit normals to Π along that curve, and thus the covariant derivative of that unit normal along the curve vanishes. Hence $\nabla_{E_j} E_i = 0$ for $(j,i) = (1,2), (2,1), (3,1), (3,2)$. Also, $\nabla_{E_3} E_3 = 0$ because the curves integral to E_3 are geodesics. Below we will show that

$$\nabla_{E_1} E_1 = +E_3, \quad \nabla_{E_1} E_3 = -E_1, \quad \nabla_{E_2} E_2 = -E_3, \quad \nabla_{E_2} E_3 = +E_2. \tag{41}$$

Plugging all this information into Equation 40, we get

$$0 = \left(\frac{du_1}{dt} - u_1 u_3\right) E_1 + \left(\frac{du_2}{dt} + u_2 u_3\right) E_2 + \left(\frac{du_3}{dt} + u_1^2 - u_2^2\right) E_3. \tag{42}$$

This is equivalant to Equation 11.

It only remains to establish Equation 41. Since E_1, E_3 are parallel to the totally geodesic plane $x_2 = 0$ and form an orthonormal framing of this plane,

and since parallel translation along the curves integral to E_1 is an isometry, there is some constant λ such that $\nabla_{E_1} E_1 = \lambda E_3$ and $\nabla_{E_1} E_3 = -\lambda E_1$. By left invariance, we have $\lambda = \Gamma^3_{11}(0,0,0)$, the Christoffel symbol with respect to $\{e_1, e_2, e_3\}$, evaluated at $(0,0,0)$. Let g^{ij} be the (ij)th entry of g^{-1}. Using the facts that, at $(0,0,0)$,

$$g^{31} = 0, \quad g^{32} = 0, \quad g^{33} = 1, \quad \frac{dg_{1i}}{dx_1} = 0, \quad \frac{dg_{11}}{dx_3} = -2,$$

we have

$$\Gamma^3_{11}(0,0,0) = \frac{1}{2}\sum_{i=1}^{3} g^{3i}\left(\frac{dg_{1i}}{dx_1} + \frac{dg_{1i}}{dx_1} - \frac{dg_{11}}{dx_i}\right) = 1.$$

This deals with the first two equalities in Equation 41. The last two have similar treatments, and indeed follow from the first two and the existence of the isometry $(x_1, x_2, x_3) \to (x_2, x_1, -x_3)$.

5.2 Grayson's Cylinders and Period Formula

Let $U_a = (a, a, \sqrt{1-2a^2})$ and let L_a be the period of the loop level set containing U_a. The following result bundles together some of the results on [**G**, pp 67-75].

Proposition 5.1 *When $a \in (0, \sqrt{2}/2)$ and $r \in \boldsymbol{R}$, we have $E(rU_a) \in C_a$, where*

$$C_a = \{(x,y,z) | w^2 + \cosh 2z = \frac{1}{2a^2}\}, \qquad w = \frac{x-y}{\sqrt{2}}.$$

The geodesic segment corresponding to the perfect vector $L_a U_a$ winds once around C_a. Moreover,

$$L_a = \int_a^{t_a} \frac{4dt}{\sqrt{1 - 2a^2 \cosh 2t}}, \qquad t_a = \frac{1}{2}\cosh^{-1}\left(\frac{1}{2a^2}\right). \tag{43}$$

One can deduce from symmetry and from Proposition 5.1 that every typical geodesic lies on some cylinder isometric to C_a, and that a typical geodesic segment is small, perfect, or large according as it winds less than once, exactly once, or more than once around the cylinder that contains it.

Our one remaining goal is to prove Equation 21. For this we don't need Proposition 5.1 but we do need Equation 43. For the sake of completeness, we essentially repeat the proof given on [**G**, p 68]. In our derivation, the symbol $\cdot$ denotes a quantity we don't need to compute.

Lemma 5.2 *Equation 43 is true.*

Proof: Let u denote the flow line for the structure field Σ corresponding to the vector $\frac{1}{4}L_aU_a$. Referring to Equation 11 the flowline u starts at U_a and ends the first time it reaches Π, the plane $Z = 0$. The loop level sets are level sets of the function $F(x, y, z) = xy$, and they lie on the unit sphere. Hence

$$u = S([0, t_a]), \qquad S(t) = (ae^t, ae^{-t}, \sqrt{1 - 2a^2 \cosh 2t}). \tag{44}$$

Referring to Equation 11, the two quantities $S'(t)$ and $\Sigma(S(t))$ are scalar multiples. Setting $S(t) = (x_t, \cdot, z_t)$, and noting that $dx_t/dt = x_t$, we have

$$S'(t) = (x_t, \cdot, \cdot) = (1/z_t) \times (x_t z_t, \cdot, \cdot) = (1/z_t) \times \Sigma(S(t)). \tag{45}$$

Let γ be the geodesic corresponding to u. Let $\gamma(t)$ be the point of γ corresponding to $S(t)$. By definition, the unit tangent field $\boldsymbol{T}(t)$ along $\gamma(t)$ lies in the same left invariant vector field as $S(t)$. By the Chain Rule and Equation 45,

$$\frac{d\gamma}{dt}(t) = \frac{1}{z_t}\boldsymbol{T}(t). \tag{46}$$

By symmetry and by definition, L_a is 4 times the length of the geodesic segment γ just considered. Noting that $\|\boldsymbol{T}(t)\| = 1$, and integrating Equation 46, we have

$$L_a = 4 \text{ Length}(\gamma) = 4 \int_a^{t_a} \left\| \frac{d\gamma}{dt} \right\| dt = 4 \int_a^{t_a} \frac{dt}{z_t} = \int_a^{t_a} \frac{4dt}{\sqrt{1 - 2a^2 \cosh(t)}}.$$

This completes the proof ♠

5.3 The AGM Period Formula

Now we manipulate Equation 43 until it is equivalent to Equation 21. Using the relations

$$\cosh(2t) = 2\sinh^2(t) + 1, \qquad m = \frac{1 - 2a^2}{1 + 2a^2}, \qquad \mu = \sqrt{\frac{m}{1 - m}} = \frac{\sqrt{1 - 2a^2}}{2a},$$

we see that Equation 43 is equivalent to the following:

$$L_a = \frac{4}{\sqrt{1 + 2a^2}} \times I_a, \qquad I_a = \frac{1}{\sqrt{m}} \int_a^{\sinh^{-1}(\mu)} \frac{dt}{\sqrt{1 - (\sinh(t)/\mu)^2}}. \tag{47}$$

To get further we relate this expression to something more classical. Let

$$\mathcal{K}(m) = \mathcal{F}(\pi/2, m), \qquad \mathcal{F}(\phi, m) := \int_a^\phi \frac{d\theta}{\sqrt{1 - m\sin^2\theta}}. \tag{48}$$

These quantities respectively are called the complete and incomplete elliptic integrals of the first kind.

Lemma 5.3 $I_a = \mathcal{K}(m)$.

Proof: This is related to Equation 19.7.7 in the Electronic Handbook of Mathematical Functions. The substitution

$$u = \tan^{-1}\sinh(t), \qquad du = dt/\cosh(t) = dt\cos(u)$$

gives

$$I_a = \frac{1}{\sqrt{m}} \times \mathcal{F}(\tan^{-1}(\mu), \frac{1}{m}).$$

The substitution $t = \sin(\theta)$ gives

$$I_a = \frac{1}{\sqrt{m}} \int_a^{\sqrt{m}} \frac{dt}{\sqrt{(1-t^2)(1-t^2/m)}}, \qquad \mathcal{K}(m) = \int_a^1 \frac{dt}{\sqrt{(1-t^2)(1-mt^2)}}.$$

The substitution $u = t/\sqrt{m}$ converts I_a into $\mathcal{K}(m)$. ♠

See e.g. [**BB**] for a proof of the following classic identity:

$$\mathcal{K}(m) = \frac{\pi/2}{\mathrm{AGM}(\sqrt{1-m}, 1)}, \qquad m \in (0,1). \tag{49}$$

Combining Lemma 5.3 and Equation 49, we get Equation 21:

$$L_a = \frac{4}{\sqrt{1+2a^2}} \times \frac{\pi/2}{\mathrm{AGM}(1, \sqrt{1-m})} = \frac{\pi}{\mathrm{AGM}(a, \frac{1}{2}\sqrt{1+2a^2})}.$$

6 References

[**A**] V. I. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits.* Ann. Inst. Fourier Grenoble, (1966).

[**AK**] V. I. Arnold and B. Khesin, *Topological Methods in Hydrodynamics*, Applied Mathematical Sciences, Volume 125, Springer (1998)

[**AS**], M. Abramovitz and I. A. Stegun (editors), *Hendbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series **55** (1964)

[**B**] N. Brady, *Sol Geometry Groups are not Asynchronously Automatic*, Proceedings of the L.M.S., 2016 vol 83, issue 1 pp 93-119

[**BB**] J. M. Borwein and P. B. Borwein, *Pi and the AGM*, Monographies et Études de la Société Mathématique du Canada, John Wiley and Sons, Toronto (1987)

[**BS**] A. Bölcskei and B. Szilágyi, *Frenet Formulas and Geodesics in Sol Geometry*, Beiträge Algebra Geom. 48, no. 2, 411-421, (2007).

[**BT**], A. V. Bolsinov and I. A. Taimanov, *Integrable geodesic flow with positive topological entropy*, Invent. Math. **140**, 639-650 (2000)

[**CMST**] R. Coulon, E. A. Matsumoto, H. Segerman, S. Trettel, *Noneuclidean virtual reality IV: Sol*, math arXiv 2002.00513 (2020)

[**EFW**] D. Fisher, A. Eskin, K. Whyte, *Coarse differentiation of quasi-isometries II: rigidity for Sol and Lamplighter groups*, Annals of Mathematics 176, no. 1 (2012) pp 221-260

[**G**], M. Grayson, *Geometry and Growth in Three Dimensions*, Ph.D. Thesis, Princeton University (1983).

[**K**] S. Kim, *The ideal boundary of the Sol group*, J. Math Kyoto Univ 45-2 (2005) pp 257-263

[**KN**] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Volume 2*, Wiley Classics Library, 1969.

[**LM**] R. López and M. I. Muntaenu, *Surfaces with constant curvature in Sol geometry*, Differential Geometry and its applications (2011)

[**S**] R. E. Schwartz, *Java Program for Sol*, download (in 2019) from
http://www.math.brown.edu/∼res/Java/SOL.tar

[**S2**] R. E. Schwartz, *Area Growth in Sol*, arXiv 2004.10622 (2021)

[**T**] M. Troyanov, *L'horizon de SOL*, Exposition. Math. 16, no. 5, 441-479, (1998).

[**Th**] W. P. Thurston, *The Geometry and Topology of Three Manifolds*, Princeton University Notes (1978). (See
http://library.msri.org/books/gt3m/PDF/Thurston-gt3m.pdf
for an updated online version.)

[**W**] S. Wolfram, *The Mathematica Book, 4th Edition*, Wolfram Media and Cambridge University Press (1999).