3. Geometric properties of systoles of simple triangle surfaces

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

Band (Jahr): **47 (2001)**

Heft 1-2: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **18.04.2024**

3. GEOMETRIC PROPERTIES OF SYSTOLES OF SIMPLE TRIANGLE SURFACES

This section is devoted to a description of some geometric properties of the systoles on a simple triangle surface $S = S(p;k)$ and its associated ideal surface S_∞. We continue to use the notations from Section 2.

The canonical triangulation of the surface S is invariant under the group Γ of isometries of S, and its vertices $0, A, B$ are fixed points for the action of $\widetilde{\Gamma}$. The quotient S/Γ is a topological 2-sphere with a singular hyperbolic metric which is isometric to two equilateral hyperbolic triangles with angles π/p glued at their boundaries. Every closed geodesic on S which does not pass through any of the vertices $A, B, 0$ projects to a closed geodesic on S/Γ. We first observe that this is the case for the projection to S/Γ of a systole on S.

LEMMA 3.1. *A systole of S does not pass through a vertex of the canonical triangulation.*

Proof. Let γ be a geodesic in S which passes through one of the vertices of the canonical triangulation, say through the vertex 0. Assume that we obtain S from side pairing transformations of a fundamental $2p$-gon Ω in such a way that the center of Ω projects to the point 0.

The lift of γ to the polygon Ω has to intersect the boundary $\partial\Omega$ of Ω and hence its length is not smaller than twice the distance between the center of Ω and $\partial\Omega$. In particular, if α is any geodesic arc in Ω of minimal length which connects the edge 1 to an edge $r \neq p+1$, then α is necessarily shorter than γ.

Let $k < p$ be such that the side pairings for Ω which define S identify the edge 1 with the edge $2k$. If $2k \neq p+1$ then the above shows that the closed geodesic on S which is the projection of the arc of minimal length in Ω connecting the edges 1 and $2k$ is shorter than γ.

On the other hand, if $2k = p+1$, then we obtain from Lemma 2.2 that the side pairings which define Ω with center at the point A identify the edge 1 with an edge $2m$ for some $m \neq (p+1)/2$. Again we conclude that the arc γ is longer than a systole on S. $\square$

Let Ω be a fundamental $2p$-gon and let γ be the geodesic arc through the center 0 of Ω which connects the vertex $2p$ to the vertex p. Let Ψ be the reflection in $\mathbf{H}^2$ along γ. Then Ψ leaves Ω invariant and maps a pair of edges of the form $\{2i+1, 2i+2k\}$ to the pair $\{2p-2i, 2p-2i-2k+1\}$ of the same form. In other words, Ψ descends to an orientation reversing isometry of S. The group $\widetilde{\Gamma}$ of isometries of S generated by Ψ and the basic group Γ has order $p+1$ and contains the group Γ as a normal subgroup of index 2. The orientation reversing isometry Ψ of S descends to an orientation reversing isometry $\widehat{\Psi}$ of order 2 of S/Γ which exchanges the two triangles.

Let $\triangle$ be an equilateral hyperbolic triangle with angle π/p. The triangle $\triangle$ will be viewed as a billiard table. A billiard orbit consists of geodesic arcs inside $\triangle$ which are joined at points of the boundary $\partial\triangle$ according to the rule that the angle of incidence equals the angle of reflection. We view a billiard orbit as unparametrized and unoriented.

A closed geodesic on S/Γ not passing through one of the singular points $\widehat{0}, \widehat{A}, \widehat{B}$ corresponds to a periodic billiard orbit in $\triangle$ of one of the following three types:

a) A periodic billiard orbit with an odd number of collisions with the boundary of $\triangle$, none of them perpendicular.

In the sequel we call such a billiard orbit an *A-orbit*. An A-orbit $\widetilde{\gamma}$ admits a lift to a closed geodesic $\widehat{\gamma}$ on S/Γ, unique up to reparametrization, which is freely homotopic as a curve on the thrice punctured sphere $S/\Gamma \setminus \{\widehat{0}, \widehat{A}, \widehat{B}\}$ to its image under the isometry $\widehat{\Psi}$. Its trace is invariant under $\widehat{\Psi}$. The lift of every collision point of the billiard orbit with $\partial\triangle$ is a transverse intersection of $\widehat{\gamma}$ with the common boundary of the two triangles forming S/Γ. The length of $\widehat{\gamma}$ is twice the length of $\widetilde{\gamma}$.

b) A periodic billiard orbit whose trace consists of one piecewise geodesic arc which meets the boundary $\partial\triangle$ orthogonally at its endpoints.

We call such an orbit a *B-orbit* in the sequel. A B-orbit $\widetilde{\gamma}$ admits a lift to S/Γ, unique up to reparametrization, which is freely homotopic to the image $\widehat{\Psi}(\widehat{\gamma}^{-1})$ under $\widehat{\Psi}$ of its inverse $\widehat{\gamma}^{-1}$. Its trace is invariant under $\widehat{\Psi}$ and its length is twice the length of $\widetilde{\gamma}$.

c) A periodic billiard orbit with an even number of collisions with the boundary of $\triangle$, none of them perpendicular.

We call such an orbit a *C-orbit*. A C-orbit $\widetilde{\gamma}$ admits two different lifts $\widehat{\gamma}_1, \widehat{\gamma}_2$ to closed geodesics on S/Γ whose traces intersect transversely and whose lengths coincide with the length of the billiard orbit. The geodesic $\widehat{\gamma}_2$ is the image of $\widehat{\gamma}_1$ under the isometry $\widehat{\Psi}$ of S/Γ. Neither the geodesic $\widehat{\gamma}_i$ nor its inverse $\widehat{\gamma}_i^{-1}$ is freely homotopic to $\widehat{\Psi}(\widehat{\gamma}_i)$.

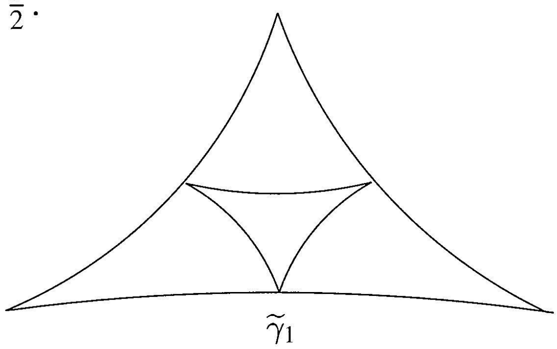
Call a periodic billiard orbit $\widetilde{\gamma}$ on $\triangle$ as above *liftable to* S if there is a closed geodesic γ on S whose projection to S/Γ is a lift $\widehat{\gamma}$ of $\widetilde{\gamma}$ to S/Γ. We then call γ a *lift of* $\widetilde{\gamma}$ *to* S.

The group $\widetilde{\Gamma}$ also acts as a group of isometries on the ideal surface S_∞ associated to S. The quotient of S_∞ unter the basic group Γ is the thrice punctured sphere S_∞/Γ with the complete hyperbolic metric of finite volume. The orientation reversing involution $\widehat{\Psi}$ acts on S_∞/Γ as the natural reflection which leaves each of the punctures fixed. Every closed geodesic on S_∞ projects to a closed geodesic on S_∞/Γ.

Let $\triangle_\infty$ be an ideal triangle. Once again we can view $\triangle_\infty$ as a billiard table. The above definition for billiard orbits in $\triangle$ can also be made for billiard orbits in $\triangle_\infty$. We call a billiard orbit $\widetilde{\gamma}$ in $\triangle_\infty$ *liftable* to the ideal surface S_∞ if there is a closed geodesic γ on S_∞ which projects to $\widetilde{\gamma}$. In the remainder of this section the ideal triangle, its billiard orbits and their lifts to the ideal surface S_∞ are always included in our considerations without further comments. More precisely, even though for simplicity we formulate all our statements only for billiard orbits in $\triangle$ and the surface S it is immediately clear from the proofs that they are equally valid for $\triangle_\infty$ and the ideal surface S_∞.

A first example of a liftable billiard orbit is given in the next lemma.

LEMMA 3.2. *There is a unique A-orbit* $\widetilde{\gamma}_1$ *in* $\triangle$ *with* 3 *collisions with the boundary, and this orbit is liftable. The length of a lift of* $\widetilde{\gamma}_1$ *to* S *is not bigger than* $6 \operatorname{arccosh} \frac{3}{2}$.



Proof. Let $S = S(p;k)$ and let Ω be a fundamental $2p$-gon. Connect the midpoint of the edge 1 in Ω with the midpoint of the edge 3 by a simple arc, and connect the midpoint of the edge $2k$ with the midpoint of the edge $2k+2$ by a simple arc. These two arcs together project to a simple closed curve on S which is freely homotopic to a closed geodesic γ on S. The geodesic γ is necessarily a lift of an A-orbit $\widetilde{\gamma}_1$ in $\triangle$ of period 3. Notice that there are exactly p lifts of $\widetilde{\gamma}_1$, and every such lift intersects exactly 6 other lifts, with each of these intersections consisting of a single point. The length ℓ_1 of a lift of $\widetilde{\gamma}_1$ to S is twice the length of $\widetilde{\gamma}_1$.

To give a sharp upper bound for ℓ_1 notice that $\ell_1/2$ is just the smallest circumference of a hyperbolic triangle with vertices on the sides of $\triangle$ and hence $\ell_1/2$ is not larger than the smallest circumference of a hyperbolic triangle T_∞ with vertices on the boundary of an ideal triangle. This circumference is the limit as $k \to \infty$ of the circumferences of hyperbolic triangles T_k whose vertices are the midpoints of the sides of an equilateral triangle $\triangle_k$ with angle π/k.

To give a formula for the circumference of T_k let λ_k be the length of the sides of $\triangle_k$, and let ℓ_k be the length of the sides of T_k.

Hyperbolic trigonometry (see [I]) gives $\cosh \frac{\lambda_k}{2} = \frac{\cos \pi/2k}{\sin \pi/k}$ and

$$\cosh \ell_k = (\cosh \frac{\lambda_k}{2})^2 - (\sinh \frac{\lambda_k}{2})^2 \cos \frac{\pi}{k} = \frac{(1-\cos \pi/k)(\cos \pi/2k)^2}{(\sin \pi/k)^2} + \cos \frac{\pi}{k}.$$

This shows that as $k \to \infty$ we have $\cosh \ell_k \to \frac{3}{2}$ and $6\ell_k \to 6 \operatorname{arccosh} \frac{3}{2} \sim 5.775$. This completes the proof of our lemma. $\square$

As an immediate consequence of Lemma 3.2, the length of the systole of a simple triangle surface and its associated ideal surface does not exceed $6 \operatorname{arccosh} \frac{3}{2} < 5.8$. In particular, for large genus such triangle surfaces are never globally maximal [BS].

LEMMA 3.3. *A lift to S of an A-orbit $\widetilde{\gamma}$ which is different from $\widetilde{\gamma}_1$ is not a systole.*

Proof. By Lemma 3.2 it suffices to show that the length of every A-orbit $\widetilde{\gamma}$ in $\triangle$ is not smaller than the length of the A-orbit $\widetilde{\gamma}_1$ from Lemma 3.1, with equality if and only if $\widetilde{\gamma} = \widetilde{\gamma}_1$.

For this recall from the definition that an A-orbit $\widetilde{\gamma}$ is a closed curve in $\triangle$ with an odd number of collisions with the boundary, none of them perpendicular. This implies that for every pair of sides of the boundary of $\triangle$ there is a geodesic arc of $\widetilde{\gamma}$ with endpoints on these sides.

Thus we can find three points E_1, E_2, E_3 which lie on the three different sides of the boundary of $\triangle$ and are contained in $\widetilde{\gamma}$ in this order with respect to the choice of some fixed orientation and some fixed initial point. Since $\widetilde{\gamma}$ is closed, its length is not smaller than the circumference of the triangle T inscribed in $\triangle$ with vertices E_1, E_2, E_3 with equality if and only if $\widetilde{\gamma}$ coincides with the boundary of T. However the length of the orbit $\widetilde{\gamma}_1$ from Lemma 3.2 is the smallest circumference of any triangle with vertices on the three different sides of $\triangle$. From this the lemma is immediate. $\square$

B-orbits and C-orbits in $\triangle$ are more difficult to control. For their investigation let S_* be a thrice punctured sphere. We equip S_* with the (noncomplete) hyperbolic metric which we obtain by glueing two equilateral hyperbolic triangles T_1, T_2 with angle π/p along their boundaries. Thus S_* with this metric is just the space $S/\Gamma - \{\widehat{0}, \widehat{A}, \widehat{B}\}$. The sides of T_1, T_2 are geodesics a, b, c in S_* which connect a pair of punctures of S_*. We call a, b, c the *edges* of S_*. Define a curve α in S_* to be *admissible* if α is a closed curve with the additional property that every connected component of an intersection of α with one of the triangles T_i consists of a single geodesic arc in T_i. We call these components the *segments* of α. Thus α is composed of a finite number of geodesic arcs with endpoints on the edges of S_*, and no two consecutive such segments are contained in the same triangle T_i. In the sequel we identify two such curves if they coincide up to an orientation preserving reparametrization.

An *admissible homotopy* of an admissible curve α is a free homotopy of α through admissible curves. We call the admissible curve α on S_* *essential* if α can not be homotoped into one of the punctures. An *admissible subcurve* of α is a connected subarc β of α such that there exists an admissible homotopy of α which deforms β into a closed admissible curve. For every admissible subcurve β of α we can write $\alpha = \beta\gamma$ for an admissible subcurve γ. We say that α is *irreducible* if for every essential admissible subcurve β of α the curve $\gamma = \alpha - \beta$ is not essential. A curve which is not irreducible is called *reducible*. An irreducible essential curve α is called *minimal* if α does not contain any nontrivial essential closed subcurve.

There are two obvious types of minimal closed curves which can be described as follows. The first type consists of curves which are freely homotopic to a lift of the A-orbit $\widetilde{\gamma}_1$ from Lemma 3.2. We call such a curve a *minimal curve of type A*. The second type consists of curves which are freely homotopic to a curve of the form $\alpha\beta$ where α and β are simple closed curves in S_* which generate the fundamental group of S_*. Up to orientation there are three different free homotopy classes of such minimal curves which correspond to a choice of two of the three punctures.

LEMMA 3.4. *Every minimal admissible closed curve is either a minimal curve of type A or a minimal curve of type B.*

Proof. Let α be a minimal admissible closed curve. If α contains two consecutive geodesic segments with endpoints on the same pair of edges of S_* then α contains a nontrivial non-essential admissible subcurve β and necessarily $\alpha = \beta\gamma$ where γ is non-essential. Since α is essential, β and γ are homotopic to different punctures. The same argument can be applied to any subarc of γ which consists of two consecutive geodesic segments and shows that γ has exactly two segments. This means that α is of type B.

On the other hand, if there are no two consecutive segments of α hitting the same edges of S_* then α is necessarily homotopic to a multiple of the lift of the A-orbit $\widetilde{\gamma}_1$ from Lemma 3.2. By minimality, α is of type A. This shows the lemma. $\square$

Let now α be any irreducible closed curve. A *simplification* of α is an admissible essential subcurve β of α such that α can be written in the form $\alpha = \beta\gamma$ where γ is non-essential. A *minimal model* is a minimal closed curve which can be obtained from α by finitely many simplifications. Clearly every irreducible closed curve has a minimal model which is not necessarily unique.

Recall that S_* admits a natural orientation reversing isometry $\widehat{\Psi}$ which fixes pointwise the edges of S_*. This isometry acts on the space of admissible curves. We have

LEMMA 3.5. *Let α be an irreducible admissible curve which admits a minimal model of type B. Then α is freely homotopic to $\widehat{\Psi}(\alpha^{-1})$.*

Proof. Let α be an irreducible admissible closed curve. Assume that α admits a minimal model β of type B. We have to show that $\widehat{\Psi}(\alpha^{-1})$ is freely homotopic to α.

By definition of a minimal model, with respect to a suitable numbering of the edges of S_* the curve β can be written in the form $\beta = \beta_1\beta_2\beta_3\beta_4$ where β_1 connects the edge a to the edge b, β_2 connects the edge b to the edge a, β_3 connects a to c and β_4 connects c to a. Notice that β has exactly 4 intersection points with the edges of S_*.

Since β is a minimal model for α, the curve α can be represented in the form $\alpha = \beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\alpha_4$ where α_i is an admissible closed curve. By assumption α is irreducible and therefore the curves α_i are non-essential.

We distinguish three cases.

1) *The curve $\beta_1\alpha_1\beta_2$ is essential.*

Then α_1 consists of an even number of geodesic arcs which connect the edges b and c. Moreover the subcurve $\alpha_2\beta_3\alpha_3\beta_4\alpha_4$ has to be non-essential and therefore $\alpha = \beta_1\alpha_1\beta_2(\beta_3\beta_4)^m$ for some $m \geq 1$. In particular, α is freely homotopic to $\widehat{\Psi}(\alpha^{-1})$.

2) *The curve* $\beta_3\alpha_3\beta_4$ *is essential.*

As above we conclude that then $\alpha = (\beta_1\beta_2)^m\beta_3\alpha_3\beta_4$ and α is freely homotopic to $\widehat{\Psi}(\alpha^{-1})$.

3) $\beta_1\alpha_1\beta_2 = (\beta_1\beta_2)^{m_1}$ *and* $\beta_3\alpha_3\beta_4 = (\beta_3\beta_4)^{m_2}$ *for some* $m_1, m_2 \geq 1$.

Since the curves α_2 and α_4 are non-essential and have their endpoints on the side a this implies that α can be represented in the form $\alpha = (\beta_1\beta_2)^{\ell_1}(\beta_3\beta_4)^{\ell_2}$ for some $\ell_1, \ell_2 \geq 1$. Once again we conclude that α is homotopic to $\widehat{\Psi}(\alpha^{-1})$. $\square$

REMARK. The proof of Lemma 3.5 also shows the following: Let α be an irreducible admissible essential closed curve on S_* which admits a minimal model of type B. Then with respect to a suitable labeling of the edges of S_*, α is freely homotopic to a curve of the form $(\beta_1\beta_2)^k\beta_3\zeta^m\beta_4$ where $k \geq 1$, $m \geq 0$ and β_1 is an arc joining the edge a to the edge b, β_2 connects b to a, β_3 joins b to c, ζ is nonessential and β_4 connects c to a.

LEMMA 3.6. *The projection to* $S/\Gamma - \{\widehat{0}, \widehat{A}, \widehat{B}\}$ *of a systole on a simple triangle surface* $S = S(p;k)$ *is irreducible.*

Proof. By Lemma 3.2 it suffices to show that the length of every admissible reducible closed curve α in S_* is bigger than twice the length of the A-orbit $\widetilde{\gamma}_1$. For this let α be reducible and write $\alpha = \alpha_1\alpha_2$ where the curves α_1, α_2 are essential.

Let β be an irreducible admissible essential subcurve of α_1. If β has a minimal model of type A, then we can cut from β finitely many non-essential closed curves to obtain a shorter curve which is homotopic to two copies of the A-orbit $\widetilde{\gamma}_1$ from Lemma 3.2. Since the lift $\widehat{\gamma}_1$ of $\widetilde{\gamma}_1$ to S/Γ has minimal length in its free homotopy class and since α is homotopic to $\beta\gamma$ for some closed curve γ, the length of α is bigger than the length of the lift $\widehat{\gamma}_1$ of $\widetilde{\gamma}_1$ to S_*. Thus by Lemma 3.2 α can not lift to a systole on S.

We are left with the case that all minimal models of irreducible subcurves α_1, α_2 of α are of type B. Then we can cut away finitely many closed curves from α which shortens the length of α to end up with a closed curve β of the form $\beta = \beta_1\gamma\beta_2\delta$ where β_1, β_2 are minimal curves of type B and γ, δ are possibly trivial arcs connecting the edges containing the endpoints of β_1, β_2. If γ, δ are not trivial then we can replace $\gamma\beta_2\delta$ by a minimal curve $\gamma\widetilde{\beta}_2\delta$ of type B where $\widetilde{\beta}_2$ is an admissible subcurve of β_2. In other words, we may as well assume that $\beta = \beta_1\beta_2$.

Now we distinguish two cases.

1) *The curves* β_1, β_2 *are homotopic.*

Then there are simple closed generators η, ζ of the fundamental group of S_* such that β is freely homotopic to $\eta\zeta\eta\zeta$. In particular there is a closed geodesic ρ on S_* which is freely homotopic to β, whose length is not bigger than the length of β and which is not a prime geodesic. This geodesic is the double of a minimal curve γ of type B. The length of ρ equals twice the length of γ. However, since the length ℓ_1 of the A-orbit $\widetilde{\gamma}_1$ from Lemma 3.2 is the minimal length of any closed curve in the triangle $\triangle$ which intersects the three sides of $\triangle$, the length of $\widetilde{\gamma}_1$ is strictly smaller than the length of γ. Thus ρ is longer than a lift of $\widetilde{\gamma}_1$ and α can not lift to a systole on S.

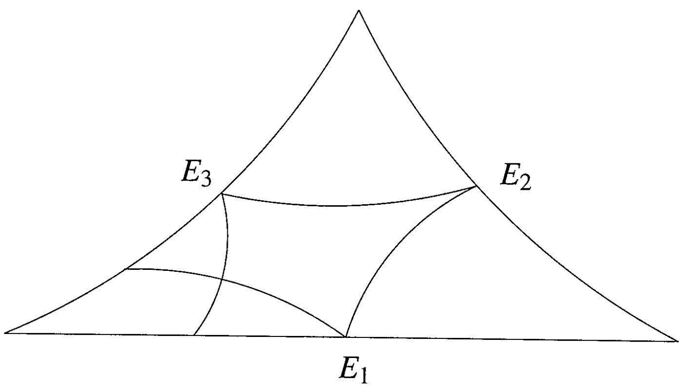
2) *The curves* β_1, β_2 *are not homotopic.*

Let $\widetilde{\zeta}$ be the B-orbit in $\triangle$ whose lift to $S/\Gamma - \{\widehat{A}, \widehat{B}, \widehat{0}\} = S_*$ is freely homotopic to $\beta_1\beta_2$. The length of $\widetilde{\zeta}$ is not bigger than half the length of $\beta_1\beta_2$ and $\widetilde{\zeta}$ consists of four arcs $\widetilde{\zeta}_1, \widetilde{\zeta}_2, \widetilde{\zeta}_3, \widetilde{\zeta}_4$. The arc $\widetilde{\zeta}_1$ meets one of the sides, say the side a, perpendicularly, and $\widetilde{\zeta}_4$ meets a different side, say the side b, perpendicularly.

We denote by E_1, E_2, E_3 the endpoints of $\widetilde{\zeta}_2$ and $\widetilde{\zeta}_3$; they lie on the three different sides of $\triangle$.

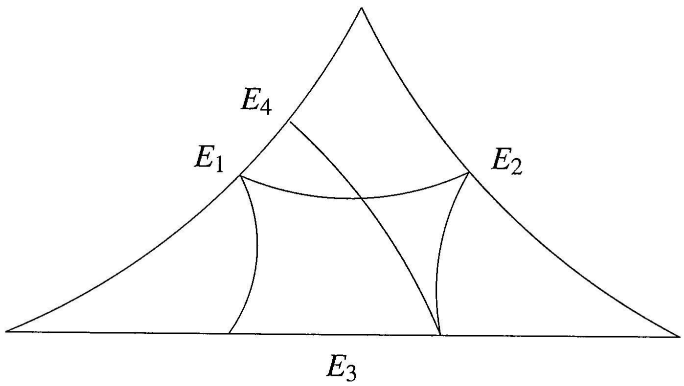
Once again we distinguish two cases:

a) *The arcs* $\widetilde{\zeta}_1$ *and* $\widetilde{\zeta}_4$ *intersect.*



Then the length of $\widetilde{\zeta}$ is bigger than the length of the triangle inscribed in $\triangle$ with vertices E_1, E_2, E_3. In particular, the length of $\widetilde{\zeta}$ is bigger than the length of the A-orbit $\widetilde{\gamma}_1$ from Lemma 3.2.

b) *The arcs* $\widetilde{\zeta}_1$ *and* $\widetilde{\zeta}_4$ *do not intersect.*



In this case either the arc $\widetilde{\zeta}_1$ intersects the arc $\widetilde{\zeta}_3$ or the arc $\widetilde{\zeta}_4$ intersects the arc $\widetilde{\zeta}_2$. Assume that the second case holds.

Let again E_1, E_2 be the endpoints of $\widetilde{\zeta}_2$ where E_1 lies on the edge b and let E_4 be the endpoint of the arc $\widetilde{\zeta}_4$ on the edge b. Since $\widetilde{\zeta}_4$ meets b orthogonally at E_4 and has its second endpoint E_3 on the side a, the angle at E_4 of the triangle with vertices E_1, E_4, E_2 is strictly bigger than $\frac{\pi}{2}$. This means that the distance between E_2 and E_4 is smaller than the length of the arc $\widetilde{\zeta}_2$ and therefore the length of $\widetilde{\zeta}$ is bigger than the circumference of the triangle with vertices E_2, E_3, E_4. In particular, this length is bigger than the length of the A-orbit $\widetilde{\gamma}_1$.

This completes the proof of our lemma. □

As an immediate corollary of Lemma 3.6 and Lemma 3.5 we obtain

COROLLARY 3.7. *A C-orbit in $\triangle$ does not lift to a systole on S.*

4. LENGTH ESTIMATES FOR SYSTOLES

In this section we complete the geometric description of the systoles of a simple triangle surface and its associated ideal surface. As a consequence we obtain that a simple triangle surface which is different from one of the three surfaces listed in the introduction is not maximal.

We resume the assumptions and notations from Section 3. Our goal is to describe all B-orbits in the equilateral triangle $\triangle$ with angle π/p or in an ideal triangle $\triangle_\infty$ which lift to a systole on a simple triangle surface S or its associated ideal surface S_∞. For this it is convenient to consider any piecewise geodesic α in $\triangle$ with the following properties: