

POINCARÉ'S CLOSED GEODESIC ON A CONVEX SURFACE

WILHELM P. A. KLINGENBERG

ABSTRACT. We present a new proof for the existence of a simple closed geodesic on a convex surface M . This result is due originally to Poincaré. The proof uses the $2k$ -dimensional Riemannian manifold ${}_k\Lambda M =$ (briefly) Λ of piecewise geodesic closed curves on M with a fixed number k of corners, k chosen sufficiently large. In Λ we consider a submanifold $\tilde{\Lambda}_0$ formed by those elements of Λ which are simple regular and divide M into two parts of equal total curvature 2π . The main burden of the proof is to show that the energy integral E , restricted to $\tilde{\Lambda}_0$, assumes its infimum. At the end we give some indications of how our methods yield a new proof also for the existence of three simple closed geodesics on M .

1. THE THEOREM

1.1. Let M be a convex surface. We may assume that M lies in \mathbb{R}^3 . $\langle \cdot, \cdot \rangle$ denotes the Riemannian metric (scalar product) on M and $d(\cdot, \cdot)$ the distance. We choose on M the usual orientation with the normal field N pointing into the interior of M . Put $\mathbb{R}/\mathbb{Z} = S$. A regular smooth simple closed curve $c : S \rightarrow M$ on M bounds to its left a domain $D(c)$ which is the image of an orientation preserving injective homeomorphism $\varphi : D \rightarrow M$ of the plane disc D with $\partial D = S$. φ is differentiable of rank 2 in the interior of D .

With K the Gaussian curvature on M we define for such a c the function

$$(1.1) \quad \Omega(c) = \int \int_{D(c)} K dM - 2\pi = - \int_S \kappa(t) |\dot{c}(t)| dt \in] - 2\pi, +2\pi[.$$

Here, the second equation is the Gauss-Bonnet formula with $\kappa(t)$ the geodesic curvature of $c(t)$.

$\Omega(c) = 0$ means that the image curve $N \circ c$ of c under the normal mapping $N : M \rightarrow S^2$ divides S^2 into two parts of equal area. We ask whether there exists a c with minimal energy among the simple closed curves with this property. In a famous paper, Poincaré [1905] formulated the following theorem.

Theorem. *On a convex surface M there exists a simple closed curve of minimal energy among the set of simple closed curves which divide M into two parts, each having total curvature 2π . Such a curve is a closed geodesic, which we also call a Poincaré closed geodesic.*

Received by the editors April 16, 2003 and, in revised form, June 3, 2003.

2000 *Mathematics Subject Classification.* Primary 53A05, 53C22; Secondary 34C25, 58G30, 58E10.

Remark. For every simple closed geodesic c on M we have $\Omega(c) = 0$. A Poincaré closed geodesic has, in addition, minimal energy and minimal length among the simple closed geodesics on M . There may exist on M geodesic loops which are shorter than a Poincaré closed geodesic.

The problem of finding a shortest curve among the simple closed curves on M which divide M into two parts of equal total curvature constitutes a modified isoparametric problem where area is replaced by total curvature.

1.2. A proof of the Theorem could run along the following lines (cf. also Blaschke [1930]): We consider the variation $D\Omega(c) \cdot X$ of $\Omega(c)$ by moving the boundary c of the domain $D(c)$ in the direction of the vector field $X(t)$ along $c(t)$. One finds

$$(1.2) \quad D\Omega(c) \cdot X = \int_S -\dot{c}(t)K(t) \wedge X(t)dt = - \int_S |\dot{c}(t)|K(t)\langle X(t), n(t) \rangle dt.$$

Here, $K(t)$ is the Gaussian curvature in $c(t)$ and $n(t)$ is the normal field along $c(t)$, pointing into the interior of c . Putting

$$(1.3) \quad \text{grad } \Omega(c)(t) = -|\dot{c}(t)|K(t)n(t),$$

we may write (1.2) also in the form

$$(1.4) \quad D\Omega(c) \cdot X = \int_S \langle \text{grad } \Omega(c)(t), X(t) \rangle dt.$$

Consider a c with $\Omega(c) = \rho \in]-2\pi, +2\pi[$. The maximal decrease of $E(c)$ goes along the vector field

$$(1.5) \quad -\text{grad } E(c)(t) = \nabla \dot{c}(t) = \alpha(t)\dot{c}(t) + \kappa(t)|\dot{c}(t)|^2 n(t).$$

Here, $\kappa(t)$ is the geodesic curvature of c in $c(t)$.

In general, the decrease of $E(c)$ in the direction of $-\text{grad } E(c)$ will not preserve the property $\Omega(c) = \rho$. Therefore, we consider the modification

$$(1.6) \quad -\overline{\text{grad}} E(c) = -\text{grad } E(c) - \mu(c)\text{grad } \Omega(c)$$

of the field $-\text{grad } E(c)$ such that $-\overline{\text{grad}} E(c)$ becomes orthogonal to $\text{grad } \Omega(c)$. This means

$$(1.7) \quad \begin{aligned} \mu(c) &= \langle -\text{grad } E(c), \text{grad } \Omega(c) \rangle / |\text{grad } \Omega(c)|^2 \\ &= - \int_S \kappa(t)K(t)|\dot{c}(t)|^3 dt / \int_S |\dot{c}(t)|^2 K(t)^2 dt. \end{aligned}$$

If c has minimal E -value with $\Omega(c) = \rho$, this implies

$$(1.8) \quad -\overline{\text{grad}} E(c) = 0, \text{ i.e., } \nabla \dot{c}(t) = -\mu(c)|\dot{c}(t)|K(t)n(t).$$

Forming the scalar product with $\dot{c}(t)$ we get $\langle \nabla \dot{c}(t), \dot{c}(t) \rangle = 0$, i.e., $|\dot{c}(t)| = \text{const} = L$ = length of c . The scalar product of (1.8) with $n(t)$ yields

$$(1.9) \quad \kappa(t)|\dot{c}(t)| = -\mu(c)K(t).$$

If, in particular, $\Omega(c) = \rho = 0$, we get with (1.1) from (1.9) $\mu(c) = 0$, i.e., $\kappa(t) = 0$, c is a closed geodesic.

In case $\Omega(c) = \rho \neq 0$, it still can be said that c is parametrized proportionally to arc length and $\kappa(t)$ is proportional to $K(t)$.

1.3. For a rigorous proof, many details must be filled in. In a carefully written paper, Croke [1982] did this. He worked not with the family of smooth closed curves on the surface M but with a family of piecewise geodesic simple closed curves.

In the proof given in Section 2, we too employ piecewise geodesic closed curves on M , with an a priori fixed number k of corners, k chosen sufficiently large. In Klingenberg [2002] one finds an exposition of the $2k$ -dimensional Riemannian manifold, obtained in this way. Strangely enough, this very natural object, associated to M and, more generally, to any compact Riemannian manifold, does not seem to have been considered before in the literature. A manifold considered by Bott [1983] is different from our manifold.

The manifold of regular piecewise geodesic closed curves has two components; cf. Smale [1958]. We consider the component $\tilde{\Lambda}$ which contains the submanifold $\tilde{\Lambda}_0$ of simple regular piecewise geodesic closed curves whose image under the normal mapping divides the sphere into two parts of equal area. The Theorem states that $E|_{\tilde{\Lambda}_0}$ assumes its infimum. A $c \in \tilde{\Lambda}_0$ with $E(c) = \inf E|_{\tilde{\Lambda}_0}$ is a simple closed geodesic on M having shortest length among all simple closed geodesics.

I wish to thank Chris Croke for the helpful comments which he made through various versions of this manuscript.

2. THE PROOF

2.1. The symbol $\langle \cdot, \cdot \rangle$ for the scalar product on the convex surface M will be used also for the scalar product (Riemannian metric) induced on the k -fold product M^k of M . k will be a positive number divisible by 4. Below we will prescribe a lower bound for k . By ι we denote the injectivity radius of M .

A **piecewise geodesic** (briefly: p.g.) **closed curve with k corners** is a continuous mapping $c : S \rightarrow M$ satisfying the following conditions: Put $c(j/k) = p_j$. p_j is called the **j -th corner** of c with $p_k = p_0$. $d(p_j, p_{j+1}) < \iota/2$ for $j = 0, \dots, k-1$. $c[j/k, (j+1)/k] =$ (briefly) c_j is the unique shortest geodesic from p_j to p_{j+1} , called **j -th edge** of c , $j = 0, \dots, k-1$. A p.g. closed curve c is determined by its k corners $\{p_0, \dots, p_{k-1}, p_k = p_0\}$. Hence, the set ${}_k\Lambda M$ of p.g. closed curves on M with k corners is an open, nonempty subset of M^k and in this way becomes a $2k$ -dimensional Riemannian manifold with scalar product $\langle \cdot, \cdot \rangle$. Having fixed M and k , we also write Λ instead ${}_k\Lambda M$. On M we have the **energy (integral)** E and the **length** L :

$$(2.1) \quad \begin{aligned} E(c) &= \frac{1}{2} \int_S |\dot{c}(t)|^2 dt = \frac{k}{2} \sum_j d(p_j, p_{j+1})^2, \\ L(c) &= \int_S |\dot{c}(t)| dt = \sum_j d(p_j, p_{j+1}). \end{aligned}$$

$L(c)^2 \leq 2E(c)$ with $=$ if and only if all edges of c have the same length $|\dot{c}|/k = L(c)/k$. Put $\dot{c}(j/k+) = D_j^+$, $-\dot{c}(j/k-) = D_j^-$. Then one gets for the differential $DE(c) \cdot Y$ on the vector $Y \in T_c\Lambda$ with components $Y_j \in T_{p_j}M$ the expression

$$(2.2) \quad DE(c) \cdot Y = - \sum_j \langle D_j^- + D_j^+, Y_j \rangle;$$

cf. Klingenberg [2002].

Hence, the vector $-\text{grad } E(c) \in T_c\Lambda$ has in $T_{p_j}M$ the component $D_j^- + D_j^+$. This vector points towards the midpoint of the shortest geodesic from p_{j-1} to p_{j+1} . This is the direction in which one has to go if one wants to decrease with maximal effect the energy of the broken geodesic from p_{j-1} over p_k to p_{p+1} . We have

$$(2.3) \quad \text{grad } E(c) = 0 \iff c \text{ is a closed geodesic.}$$

Indeed, $D_j^- + D_j^+ = 0$, all j , means $\dot{c}(j/k-) = \dot{c}(j/k+)$, all j . Thus, c is smooth and $\dot{c}(t) = \text{const} = L(c) \geq 0$. The converse is obvious.

Denote by $\phi_s c$, $s \geq 0$, the positive integral curve of the field $-\text{grad } E$ on Λ with $\phi_0 c = c$. The following statements are easy to prove; cf. Klingenberg [2002]:

- (i) $E(\phi_s c) \leq E(c)$.
- (ii) $\phi_s c$ is defined for all $s \geq 0$.
- (iii) For a fixed s is $\phi_s : \Lambda \rightarrow \Lambda$ continuous.

2.2. An element $c \in \Lambda$ is called **regular**, if all edges of c are nonconstant and if at each corner p_j of c the oriented angle β_j from the incoming direction $\dot{c}_{j-1}(j/k)$ to the outgoing direction $\dot{c}_j(j/k)$ belongs to the interval $] -\pi, +\pi[$.

The regular p.g. closed curves on M form an open and dense subset of Λ , hence, a $2k$ -dimensional Riemannian manifold. It is well known (cf. Smale [1958]) that this manifold possesses two connected components. The component which contains the simple regular p.g. closed curves will be denoted by $\tilde{\Lambda}$. $\tilde{\Lambda}$ contains also all odd coverings of simple closed p.g. curves and among the closed geodesics it contains those which have an even number of selfintersections. Thus, $\tilde{\Lambda}$ contains, in particular, the simple closed geodesics—if there are any. The other component contains the even coverings of simple elements in Λ and those closed geodesics with an odd number of selfintersections.

If $c \in \tilde{\Lambda}$ is simple, it bounds to its left a domain $D(c)$ on M . $D(c)$ is the image $\varphi(D)$ of an injective orientation preserving homeomorphism $\varphi : D \rightarrow M$ of the plane disc D with $\partial D = S$, which in its interior is differentiable of rank 2. We use this to define the function

$$\Omega(c) = \int \int_{D(c)} K dM - 2\pi = - \sum_j \beta_j(c) \in] -2\pi, +2\pi[.$$

Here, the second equation is the Gauss-Bonnet formula.

If $E(c)$ is small, $\Omega(c)$ will have a value near -2π or near $+2\pi$. To make sure that all values in $] -2\pi, +2\pi[$ occur, we must choose k sufficiently large. Recall that the length of an element in ${}_k\Lambda M$ is bounded from above by $k\iota/2$. Let $N : M \rightarrow S^2$ be the normal mapping. The length of the counter image on M of a great circle on S^2 is bounded from above by some $a > 0$. Such a closed curve on M is simple. Assume it to be parametrized by S , proportional to arc length. If we choose $k > 2a/\iota$ sufficiently large, then the points on such a curve with parameter values $t = j/k$, $j = 0, \dots, k-1$, will be the corners of a simple c in $\tilde{\Lambda}$ with $\Omega(c)$ near 0. Small modifications of c yield simple curves in $\tilde{\Lambda}$ with Ω -value 0.

We want to extend Ω to all of $\tilde{\Lambda}$. For that purpose observe that an arbitrary $c = c_1 \in \tilde{\Lambda}$ can be joined in $\tilde{\Lambda}$ by a regular homotopy $\{c_s, 0 \leq s \leq 1\}$ to a simple $c_0 \in \tilde{\Lambda}$. Starting from the injective homeomorphism $\varphi_0 : D \rightarrow M$ which has as its image the domain $D(c_0)$ bounded by c_0 , the curve $\{c_s, 0 \leq s \leq 1\}$, determines a family $\{\varphi_s : D \rightarrow M, 0 \leq s \leq 1\}$ of mappings depending continuously on s . Putting

$\varphi_s(D) = D(c_s)$ we define

$$\Omega^*(c_s) = \int \int_{D(c_s)} K dM + \sum_j \beta_j(c_s).$$

Ω^* can assume as values only integer multiples of 2π , i.e., $\Omega^*(c_s) = 2\pi$, for all s . Writing $D(c)$ instead $D(c_1)$ we define

$$(2.4) \quad \Omega : \tilde{\Lambda} \rightarrow \mathbb{R}; c \mapsto \Omega(c) = \int \int_{D(c)} K dM - 2\pi = - \sum_j \beta_j(c).$$

One could define Ω also on the second component of regular p.g. closed curves by taking the negative sum of the exterior angles. Among the closed curves in this component taking the value 0, there will be arbitrarily short ones, formed like the number eight. This is quite different for a $c \in \tilde{\Lambda}$ with $\Omega(c) = 0$: In this case, the integral over $D(c)$ of $K dM$ has the value 2π . Hence, $L(c)$ and $E(c)$ are bounded away from 0.

2.3. As a differentiable function, $\Omega : \tilde{\Lambda} \rightarrow \mathbb{R}$ possesses on the Riemannian manifold $\tilde{\Lambda}$ a gradient field $\text{grad } \Omega$. This means

$$(2.5) \quad D\Omega(c) \cdot X = \langle \text{grad } \Omega(c), X \rangle, \quad \text{for all } X \in T_c \tilde{\Lambda}.$$

We want to calculate $\text{grad } \Omega$. For that purpose consider the normal unit field n_j along the edge c_j , where $\{\dot{c}_j/|\dot{c}_j|, n_j\}$ forms a positive orthonormal frame along c_j . Along $c_j(t)$ define the Jacobi fields $n_j^+(t)$ and $n_j^-(t)$ by the boundary conditions

$$(2.6) \quad \begin{aligned} n_j^+((j+1)/k) &= n_j, & n_j^+(j/k) &= 0, \\ n_j^-((j+1)/k) &= 0, & n_j^-(j/k) &= n_j. \end{aligned}$$

With this we define

$$(2.7) \quad \begin{aligned} \sigma_j^+(c) &= (\text{briefly}) \sigma_j^+ = |\dot{c}_j| \int_{I_j} \langle K_j(t) n_j^+(t), n_j \rangle dt, \\ \sigma_j^-(c) &= (\text{briefly}) \sigma_j^- = |\dot{c}_j| \int_{I_j} \langle K_j(t) n_j^-(t), n_j \rangle dt. \end{aligned}$$

Here, $I_j = [j/k, (j+1)/k]$, $K_j(t)$ the Gauss curvature in $c_j(t)$.

Proposition 1.

$$(2.8) \quad \text{grad } \Omega(c)_j = -\sigma_j^-(c) n_j - \sigma_{j-1}^+(c) n_{j-1} \in T_{p_j} M,$$

where $\sigma_j^-(c)$ and $\sigma_{j-1}^+(c)$ are both positive. Hence, $\text{grad } \Omega(c)_j$ points into the exterior of c .

Proof. We first consider the case where c is simple. Then

$$(2.9) \quad D\Omega(c) \cdot X = \sum_j \int_{I_j} -\dot{c}_j(t) \wedge K_j(t) X_j(t) dt.$$

Here, $X_j(t)$, $t \in I_j$, is the Jacobi field along $c_j(t)$ determined by the boundary conditions $X_j(j/k) = X_j$, $X_j((j+1)/k) = X_{j+1}$, where X_j is the component of X in $T_{p_j} M$.

The integrand in (2.9) can be written as

$$(2.10) \quad -|\dot{c}_j| \langle n_j(t), K_j(t) X_j(t) \rangle = (\text{briefly}) -|\dot{c}_j| K_j(t) x_j(t)$$

with

$$(2.11) \quad \ddot{x}_j(t) + |\dot{c}_j|^2 K_j(t) x_j(t) = 0,$$

$$x_j(j/k) = \langle X_j, n_j \rangle, \quad x_j((j+1)/k) = \langle X_{j+1}, n_j \rangle.$$

Hence,

$$x_j(t) = \langle X_{j+1}, n_j \rangle \langle n_j^+(t), n_j \rangle + \langle X_j, n_j \rangle \langle n_j^-(t), n_j \rangle,$$

which yields (2.8). \square

Complement to Proposition 1. Using the Jacobi functions $x_j(t) = \langle X_j(t), n_j \rangle$ along $c_j(t)$ we also can write

$$(2.12) \quad D\Omega(c) \cdot X = \sum_j (x'_{j-1}(j/k) - x'_j(j/k)).$$

Here, the dash denotes derivation with an arc length parameter.

Proof. From (2.10) and (2.11) we have

$$D\Omega(c) \cdot X = \sum_j \int_{I_j} \ddot{x}_j(t) dt / |\dot{c}_j| = \sum_j (\dot{x}_j((j+1)/k) - \dot{x}_j(j/k)) / |\dot{c}_j|.$$

\square

Remark. The field $\text{grad } \Omega$ has a continuous extension from $\tilde{\Lambda}$ to the closure of $\tilde{\Lambda}$ in Λ : Whenever $\dot{c}_j = 0$, put $\sigma_j^+(c) = \sigma_j^-(c) = 0$.

2.4. Since $\text{grad } \Omega \neq 0$ on $\tilde{\Lambda}$, Ω has regular values only. Hence, for $\rho \in]-2\pi, +2\pi[$, $\Omega^{-1}(\rho) =$ (briefly) $\tilde{\Lambda}_\rho$ is a submanifold of codimension 1 in $\tilde{\Lambda}$. $\tilde{\Lambda}_\rho$ is connected. $\tilde{\Lambda}_0$ contains the closed geodesics of $\tilde{\Lambda}$. We consider the orthogonal projection $-\overline{\text{grad}} E(c)$ of $-\text{grad } E(c)$ in $T_c \tilde{\Lambda}_\rho$, $c \in \tilde{\Lambda}_\rho$. This projection can be written as

$$(2.13) \quad -\overline{\text{grad}} E(c) = -\text{grad } E(c) - \mu(c) \text{grad } \Omega(c)$$

with

$$(2.14) \quad \mu(c) = \langle -\text{grad } E(c), \text{grad } \Omega(c) \rangle / \langle \text{grad } \Omega(c), \text{grad } \Omega(c) \rangle.$$

Proposition 2. Let c be a critical point of $\overline{\text{grad}} E(c)$ on $\tilde{\Lambda}_\rho$. This means: $\Omega(c) = \rho$ and

$$(2.15) \quad -\dot{c}_{j-1} + \dot{c}_j = -\mu(c)(\sigma_j^-(c)n_j + \sigma_{j-1}^+(c)n_{j-1}), \quad \text{for all } j.$$

If $\rho = 0$, then $\mu(c) = 0$ and c is a closed geodesic in $\tilde{\Lambda}$. If $\rho \neq 0$, then $\mu(c) \neq 0$ and $\mu(c)$ and ρ have the same sign. Moreover,

$$(2.16) \quad \text{tg } \beta_j(c)/2 = -\mu(c)(\sigma_j^-(c) + \sigma_{j-1}^+(c)) / (|\dot{c}_{j-1}| + |\dot{c}_j|).$$

Hence, all exterior angles $\beta_j(c)$ have the same sign equal to (-1) sign ρ .

Proof. Forming the scalar product of (2.15) with n_j and n_{j-1} , respectively, we get

$$(2.17) \quad \begin{aligned} |\dot{c}_{j-1}| \sin \beta_j(c) &= -\mu(c)(\sigma_j^-(c) + \sigma_{j-1}^+(c) \cos \beta_j(c)), \\ |\dot{c}_j| \sin \beta_j(c) &= -\mu(c)(\sigma_j^-(c) \cos \beta_j(c) + \sigma_{j-1}^+(c)). \end{aligned}$$

Taking the sum we get (2.16). Thus, if $\mu(c) \neq 0$, all $\beta_j(c)$ are $\neq 0$ with the sign (-1) sign $\mu(c)$. If $\rho = 0$, $\sum_j \beta_j(c) = 0$, hence, $\mu(c) = 0$, $\text{grad } E(c) = 0$, c is a closed geodesic. \square

2.5. Denote by $\overline{\phi}_s c$, $s \geq 0$, the positive flow line on $\tilde{\Lambda}_\rho$ of the field $-\overline{\text{grad}} E$, starting from $c \in \tilde{\Lambda}_\rho$.

Proposition 3. (i) $E(\overline{\phi}_s c) \leq E(c)$.

(ii) $\overline{\phi}_s c$ is defined for all $s \geq 0$.

(iii) For a fixed $s \geq 0$, $\overline{\phi}_s : \tilde{\Lambda}_\rho \rightarrow \tilde{\Lambda}_\rho$ is continuous.

Proof. (i) follows from

$$\frac{dE}{ds}(\overline{\phi}_s c)|_{s=0} = -|\text{grad } E(c)|^2 + \langle \text{grad } E(c), \text{grad } \Omega(c) \rangle^2 / |\text{grad } \Omega(c)|^2 \leq 0.$$

To prove (ii) we derive a contradiction from the assumption that $\overline{\phi}_s c$ is defined on $[0, s^*[$ only, with some finite $s^* > 0$. Let $\{s_m\}$ be a monotonously increasing sequence in $[0, s^*[$ with limit s^* . One easily shows that $\{\overline{\phi}_{s_m} c\} = (\text{briefly}) \{c_m\}$ is a Cauchy sequence, cf. Klingenberg [2002]. Let c^* be its limit. We must prove that c^* is regular.

We first derive a contradiction from the assumption that c^* has a constant edge. Since not all edges of c^* are constant, we may assume that $\dot{c}_{j-1}^* = 0$ and $\dot{c}_j^* \neq 0$. Hence, $-\text{grad } E(c^*)_j = \dot{c}_j^*$. This vector tends to push apart the coinciding corners $p_j^* = p_{j-1}^*$ of c^* . Hence, the same will hold for the vector $-\text{grad } E(c_m)_j$, for large m . However, it is not the vector $-\text{grad } E(c^*)_j$ but the vector $-\overline{\text{grad}} E(c^*)_j = -\text{grad } E(c^*)_j - \mu(c^*) \text{grad } \Omega(c^*)_j$ which is operating on $p_j^* = p_{j-1}^*$. Observe now that $\text{grad } \Omega(c^*)_j = -\sigma_j^-(c^*) n_j^*$. This vector is orthogonal to $-\text{grad } E(c^*)_j$. Hence, it still remains true that $-\overline{\text{grad}} E(c^*)_j$ is pushing apart $p_j^* = p_{j-1}^*$ and the same is true for $-\overline{\text{grad}} E(c_m)_j$, m large.

Next we show that $\beta_j(c^*)$, which is the limit of the sequence $\{\beta_j(c_m)\}$, in absolute value is different from π . Otherwise, $-\text{grad } E(c^*)_j = -\dot{c}_{j-1}^* + \dot{c}_j^*$, where the two summands point into the same direction. Hence, for large m , $\beta_j(c_m)$ in absolute value is pushed away from π . Again, it is not $-\text{grad } E(c^*)_j$ but $-\overline{\text{grad}} E(c^*)_j$ which is operating on $\beta_j(c^*)$. However, since $\text{grad } \Omega(c^*)_j$ is a multiple of $-n_j^* = n_{j-1}^*$, the effect of $-\overline{\text{grad}} E(c^*)_j$ on $\beta_j(c^*)$ is the same as the effect of $-\text{grad } E(c^*)_j$ and again we get the desired contradiction. This shows that $c^* \in \tilde{\Lambda}_\rho$.

With standard arguments we now prove that the flow line $\overline{\phi}_s c$ can be extended to values $s \geq s^*$; cf. Klingenberg [2002].

The proof of (iii) is obvious. \square

2.6. Next we prove the existence of critical points.

Proposition 4. Fix $\rho \in]-2\pi, +2\pi[$. Then $\inf E|_{\tilde{\Lambda}_\rho} = \lambda_\rho > 0$. Let $\{c_m\}$ be a sequence in $\tilde{\Lambda}_\rho$ with $E(c_m) \geq E(c_{m+1})$ and $\lim E(c_m) = \lambda_\rho$. Then $\{c_m\}$ has a subsequence with limit a regular c^* with $E(c^*) = \lambda_\rho$. Possibly, c^* has some edges with length $= \iota/2$ instead of $< \iota/2$. In any case, c^* is a critical point of E in the canonical extension of $\tilde{\Lambda}_\rho$, formed by the p.g. regular closed curves where the edges have length $< (1 + \varepsilon)\iota/2$ instead of $< \iota/2$, some small fixed $\varepsilon > 0$. At least if $\rho = 0$ or if $|2\pi - \rho|$ is sufficiently small, c^* belongs to the original $\tilde{\Lambda}_\rho$. According to Proposition 3 this means that, for $\rho = 0$, c^* is a closed geodesic. For $\rho \neq 0$ see Proposition 3 for the properties of c^* .

Proof. $\lambda_\rho > 0$, since we consider curves with Ω -value $\rho \in]-2\pi, +2\pi[$. Each c_m is determined by its k corners. M is compact. Therefore, we have a subsequence of

$\{c_m\}$ which again we denote by $\{c_m\}$ which converges to a c^* in the closure of $\tilde{\Lambda}_\rho$. The sequence $\{\bar{\phi}_1 c_m\}$ then will converge to $\bar{\phi}_1 c^*$. $E(\bar{\phi}_1 c^*) = E(c^*) = \lambda_\rho$ implies

$$\overline{\text{grad}} E(c^*) = \text{grad } E(c^*) + \mu(c^*) \text{grad } \Omega(c^*) = 0.$$

Here we use the fact that the field $\text{grad } \Omega$ can be extended to the closure of $\tilde{\Lambda}_\rho$; cf. the remark at the end of 2.3.

We claim that c^* is regular. Indeed, if we had $\dot{c}_{j-1}^* = 0$, $\dot{c}_j^* \neq 0$, $\text{grad } E(c^*)_j = -\dot{c}_j^*$, $\text{grad } \Omega(c^*)_j = \sigma_j^-(c^*)n_j^*$. Hence, $\mu(c^*) = 0$ and $\text{grad } E(c^*) = 0$. That is, c^* is a closed geodesic and therefore regular. Thus, all edges of c^* are nonconstant. If the exterior angle at the corner p_j of c^* has absolute value π , then $\text{grad } E(c^*)_j = -\dot{c}_j^* + \dot{c}_{j-1}^* = \alpha \dot{c}_j^* \neq 0$ cannot be a multiple of $\text{grad } \Omega(c^*)_j = \beta n_j^*$. Thus, c^* is regular, with some edges possibly having length $= \iota/2$ instead of $< \iota/2$.

This possibility is excluded, if $\rho = 0$, since in this case a critical point c^* in $\tilde{\Lambda}_0$ is a closed geodesic. All edges of c^* then have the same length equal to $(2E(c^*))^{1/2}/k \leq (2E(c_m))^{1/2}/k < \iota/2$. Also, if $|2\pi - \rho|$ is sufficiently small, c^* will lie in a disc on M where the curvature is almost constant, say K_0 . Now, on the sphere of constant curvature K_0 , the critical points of $E|_{\tilde{\Lambda}_\rho}$ are regular polygons. The length of the edge of such a polygon will be small since $|2\pi - \rho|$ is small. Hence, an edge of $c^* \in \tilde{\Lambda}_\rho$ will have length $< \iota/2$. \square

2.7. In Proposition 4 we proved the existence of closed geodesics in $\tilde{\Lambda}_0$ with E -value $= \inf E|_{\tilde{\Lambda}_0}$. $\tilde{\Lambda}_0$ contains also curves with selfintersections. To prove the Theorem, it remains to show that E assumes its infimum also on the open submanifold $\tilde{\tilde{\Lambda}}_0$ of $\tilde{\Lambda}_0$ which is formed by the simple p.g. closed curves with Ω -value 0.

We start by assuming that the compact set C_0 where $E|_{\tilde{\tilde{\Lambda}}_0}$ assumes its infimum is not empty. C_0 is formed by simple closed geodesics; cf. the proof of Proposition 4.

Since k is fixed, $E|\Lambda = E|_k \Lambda M$ is bounded from above. We will show that a simple closed geodesic c not only does not touch itself but actually, if t_0 and t_1 are different points on S , $\dot{c}(t_0)$ and $\pm \dot{c}(t_1)$ are uniformly bounded away from each other, for all simple closed geodesics in Λ . In particular, this holds for the $c \in C_0$.

In other words, if I_0 and I_1 are disjoint, sufficiently small intervals on S , having t_0 and t_1 as midpoints, respectively, then $c|I_0$ and $c|I_1$ are bounded away from each other uniformly, for all $c \in C_0$. This means, in particular, that C_0 is bounded away from the boundary $\partial \tilde{\tilde{\Lambda}}_0$ of $\tilde{\tilde{\Lambda}}_0$ in $\tilde{\Lambda}_0$. Elements in $\partial \tilde{\tilde{\Lambda}}_0$ will touch each other without having a selfintersection.

This is a qualitative statement. We therefore may consider not just M , but a family $\{M_s, 0 \leq s \leq \varepsilon\}$ of convex surfaces M_s with $M_0 = M$. Here, the scalar product $\langle \cdot, \cdot \rangle_s$ on M_s shall depend smoothly on s . $d_s(\cdot, \cdot)$ and ι_s denote the distance and the injectivity radius for M_s . On ${}_k \Lambda M_s = (\text{briefly}) {}_s \Lambda$ we have the functions E_s and Ω_s defined in the same way as the functions $E_0 = E$ and $\Omega_0 = \Omega$ on $M_0 = M$. ${}_s \tilde{\Lambda}_0$ denotes the manifold of regular p.g. closed curves on M_s having Ω_s -value 0. We also consider the submanifold ${}_s \tilde{\tilde{\Lambda}}_0 \subset {}_s \tilde{\Lambda}_0$ of simple regular p.g. closed curves on M_s as well as the boundary $\partial {}_s \tilde{\tilde{\Lambda}}_0 \subset {}_s \tilde{\Lambda}_0$.

Proposition 5. *The compact set ${}_s C_0 \subset {}_s \tilde{\tilde{\Lambda}}_0$ of simple closed geodesics having minimal E_s -value is uniformly bounded away from $\partial {}_s \tilde{\tilde{\Lambda}}_0$, $0 \leq s \leq \varepsilon$.*

Proof. We may assume that ${}_sC_0$ is not empty. The norm of the initial vector of an element $c \in {}_sC_0$ belongs to a compact interval on \mathbb{R} not containing 0, for all s . We follow c until it closes up. Hence, if the initial vectors of two closed geodesics c and \bar{c} of the set ${}_sC_0$ are sufficiently close to each other, c and \bar{c} will stay close to each other. Choose a $c : S \rightarrow M_s$ in ${}_sC_0$. Let t_0 and t_1 be different points on S . We assume $t_0 = 0$ and $0 < t_1 < 1$. If $\dot{c}(t_1)$ is getting very near to $\dot{c}(0)$, then $c|_{[t_1, 1]}$ must squeeze through the arcs of c near $t = 0$ and $t = t_1$ since c is simple closed. This means that there exists t'_1 , $t_1 < t'_1 < 1$, such that $-\dot{c}(t'_1)$ is even closer to $\dot{c}(0)$ than $\dot{c}(t_1)$. Write again t_1 instead t'_1 and consider the geodesics $c|_{[0, t_1]}$ and $\bar{c}|_{[0, t_1]}$ with $\bar{c}(t) = c(t_1 - t)$. The initial vectors $\dot{c}(0)$ and $\dot{\bar{c}}(0) = -\dot{c}(t_1)$ are close to each other. Therefore, $c|_{[0, t_1]}$ and $\bar{c}|_{[0, t_1]}$ will stay close to each other. In case they have an intersection point, the tangent vectors at this point will form a small angle. Observe now that the two geodesics c and \bar{c} intersect at $t = t_1/2$ with opposite tangent vectors. This shows, in reality, the arcs of c near $t = t_0$ and $t = t_1$ cannot get very close. This means that ${}_sC_0$ is bounded away from $\partial_s \tilde{\Lambda}_0$ uniformly for all $s \in [0, \varepsilon]$. \square

2.8. We now have everything prepared for the proof of the Theorem. We start with a 1-parameter family $\{M_s, 0 \leq s \leq 1\}$ of convex surfaces where M_0 is a sphere of constant curvature K_0 and M_1 is the given convex surface M . Such a family can be obtained by viewing M as a surface in \mathbb{R}^3 and choosing for M_0 the smallest sphere containing M . Every ray starting from the midpoint o of M_0 will meet M_0 and $M = M_1$ in exactly two points, say p_0 and p_1 . Consider the homotopy $p_s = (1-s)p_0 + sp_1$, $0 \leq s \leq 1$. Define M_s to be the set of points p_s , for all rays starting from o .

In 2.2 we chose the integer k sufficiently large such that the function Ω assumes all values in $] - 2\pi, +2\pi[$. By choosing, if necessary, k even larger, we can make sure that $\text{im } \Omega_s =] - 2\pi, +2\pi[$, all $0 \leq s \leq 1$. Denote by s^* the supremum of the $s \in [0, 1]$ with the property that the set ${}_sC_0 \subset {}_s\tilde{\Lambda}_0$ of simple closed geodesics with E_s -value equal to $\inf E_s|_{{}_s\tilde{\Lambda}_0}$ is not empty. ${}_0C_0$ is formed by the parametrized great circles on the sphere M_0 . Thus, s^* is well defined and ≥ 0 . In Klingenberg [2002] there is derived the estimate $E(\tilde{c}_s) \geq 2\pi^2 / \max K_s$, K_s the curvature on M_s and \tilde{c}_s a geodesic biangle. Using this estimate, one can show $s^* > 0$. But we don't need this.

We derive a contradiction from the assumption that ${}_{s^*}C_0 = \emptyset$. Choose a sequence $\{s_m\}$ in $[0, s^*[$ with limit s^* and ${}_{s_m}C_0 \neq \emptyset$, all m . For simplicity we write M_m instead M_{s_m} , ${}_m\tilde{\Lambda}_0$ instead ${}_{s_m}\tilde{\Lambda}_0$, ${}_mC_0$ instead ${}_{s_m}C_0$, E_m instead E_{s_m} , ι_m instead ι_{s_m} and $d_m(\cdot, \cdot)$ for the distance on M_m . Similarly, we write M_* instead M_{s^*} , ${}_*\tilde{\Lambda}_0$ instead ${}_{s^*}\tilde{\Lambda}_0$, ${}_*C_0$ instead ${}_{s^*}C_0$, E_* instead E_{s^*} , ι_* instead ι_{s^*} and $d_*(\cdot, \cdot)$ for the distance on M_* .

For each m choose $c_m \in {}_mC_0$. The k corners of c_m can be viewed as points on M_* . A subsequence of $\{c_m\}$ which we denote again by $\{c_m\}$ will on M_* converge to a set $\{p_0^*, \dots, p_{k-1}^*, p_k^* = p_0^*\}$ of k points. Since the c_m are simple closed geodesics, these points can be viewed as the corners $c^*(j/k) = p_k^*$ of a simple closed geodesic c^* on M_* . By hypothesis, ${}_*C_0 = \emptyset$. Hence, $d_*(p_j^*, p_{j+1}^*) = \iota_*/2$. This means, c^* is not an element of ${}_k\Lambda M_*$ where the edges of an element must have distance

$< \iota_*/2$. But as we did already in 2.6, we consider the canonical extension of ${}_k\Lambda M_*$ by considering the p.g. closed curves where the edges are allowed to have length $< (1 + \varepsilon)\iota_*/2$ instead $< \iota_*/2$, for some small $\varepsilon > 0$. We employ for this extension the same notation as for the case where we restricted the length of the edges to $< \iota_*/2$.

Besides a sequence $\{c_m\}$, $c_m \in {}_mC_0$, we also consider a sequence $\{\tilde{c}_m\}$, $\tilde{c}_m \in {}_m\tilde{\Lambda}_0$. Again we may assume the corners of the \tilde{c}_m converge in M_* to, say, $\{\tilde{p}_0^*, \dots, \tilde{p}_{k-1}^*, \tilde{p}_k^* = \tilde{p}_0^*\}$. Here, $d_*(\tilde{c}_j^*, \tilde{p}_{j+1}^*) \leq \iota_*/2$. Thus, we get on M_* a p.g. closed curve \tilde{c}^* with $\tilde{c}^*(j/k) = \tilde{p}_j^*$. Since $E_m(\tilde{c}_m) \geq E_m(c_m)$ we have $E_*(\tilde{c}^*) = (k\iota_*/2)^2/2$.

Therefore, $d_*(\tilde{p}_j^*, \tilde{p}_{j+1}^*) = \iota_*/2$ and $E_*(\tilde{c}^*) = E_*(c_*)$, i.e., $E_*(\tilde{c}^*) = \inf E_*|_{\tilde{\Lambda}_0}$. According to Proposition 4, \tilde{c}^* is a simple closed geodesic on M_* of minimal E_* -value. From the construction of \tilde{c}^* as the limit of a sequence $\{\tilde{c}_m\}$, $\tilde{c}_m \in {}_m\tilde{\Lambda}_0$, subject only to the condition that the corners in M_* converge, we get: The set ${}_sC_0$ of simple closed geodesics on M_* with minimal E_* -value is not bounded away from $\partial_*\tilde{\Lambda}_0$, a contradiction to Proposition 5.

To complete the proof of the Theorem we derive a contradiction from the assumption $1 - s^* = \varepsilon > 0$. With s instead of $s - s^*$ consider the family $\{M_s, 0 \leq s \leq \varepsilon\}$ of convex surfaces with ${}_0C_0 \neq \emptyset$. From Proposition 5 we know that for a $\tilde{c} \in {}_0\tilde{\Lambda}_0 \setminus {}_0C_0$, but near ${}_0C_0$, $E_0(\tilde{c}) > E_0|_{{}_0C_0}$. This is a stable situation. That means, if we consider, for sufficiently small $s > 0$, the function E_s on ${}_s\tilde{\Lambda}_0$, then $E_s|_{{}_s\tilde{\Lambda}_0}$ still assumes its infimum on ${}_s\tilde{\Lambda}_0$, i.e., ${}_sC_0 \neq \emptyset$. If we write s again instead of $s + s^*$, we have that there exists $s > s^*$ with ${}_sC_0 \neq \emptyset$. \square

3. CONCLUDING REMARKS

In his paper Poincaré [1905] indicated that actually there exist three simple closed geodesics on a convex surface. The example of an ellipsoid with three different axes, all having approximately the same length, shows that three is the optimal number of simple closed geodesics. In a short note, Lusternik and Schnirelmann [1929] sketched a proof of the fact that actually three simple closed geodesics exist on any orientable closed surface of genus 0, i.e., on a surface given by S^2 , endowed with an arbitrary Riemannian metric. A complete proof of the theorem of Lusternik-Schnirelmann is notoriously complicated. Over the years, many papers have appeared on this subject and only a few are satisfactory; cf. Taimanov [1992]. It seems therefore worthwhile to indicate how the arguments used in Section 2 yield yet another proof of the Theorem of the Three Closed Geodesics.

We start with an $(\mathbb{Z}_2 \times \mathbb{Z}_2)$ -action on the manifold $\Lambda = {}_k\Lambda M$ with generators

$$(3.1) \quad \begin{aligned} (-1): \quad \Lambda &\longrightarrow \Lambda; \quad c(t) \longmapsto c(1-t), \\ (1/2): \quad \Lambda &\longrightarrow \Lambda; \quad c(t) \longmapsto c(t+1/2). \end{aligned}$$

(-1) reverses the orientation of c while $(1/2)$ moves the initial point of c from $c(0)$ to $c(1/2)$. One easily verifies that the $(\mathbb{Z}_2 \times \mathbb{Z}_2)$ -action consists of isometries which leave E invariant; cf. Klingenberg [2002]. Moreover, if $c \in \tilde{\Lambda}$, we have $\Omega((1/2) \cdot c) = \Omega(c)$ and $\Omega((-1) \cdot c) = -\Omega(c)$. Thus, the $(\mathbb{Z}_2 \times \mathbb{Z}_2)$ -action can be restricted to $\tilde{\Lambda}_0$ and here, the action is free and commutes with the $\bar{\phi}_s$ -flow.

On $\tilde{\Lambda}_0/(\mathbb{Z}_2 \times \mathbb{Z}_2)$ we define a 1-dimensional \mathbb{Z}_2 -homology class w_1 by the cycle which for the standard sphere S^2 is given as follows: Take all parametrized great circles starting from $(1, 0, 0) \in S^2$ with initial direction tangent to the closed upper half sphere. In addition, we define a 2-dimensional \mathbb{Z}_2 -homology class w_2 with the following \mathbb{Z}_2 -cycle for S^2 : Take all parametrized great circles which start from the half equator $\{(\cos 2\pi s, \sin 2\pi s, 0); 0 \leq s \leq 1/2\}$ with initial direction tangent to the closed upper half sphere.

In Proposition 4 we have shown that there exists a closed geodesic, call it now c_0 , with $E(c_0) = \lambda_0 = \inf E|\tilde{\Lambda}_0$. Similarly, we now consider

$$(3.2) \quad \lambda_1 = \inf\{(\max E|\tilde{w}_1); \text{ all cycles } \tilde{w}_1 \text{ representing } w_1\}.$$

Using the standard methods, which in our case are presented in Klingenberg [2002], it is not difficult to show that there exists a closed geodesic, call it c_1 , with $E(c_1) = \lambda_1$. This is the counterpart of Proposition 4.

It remains to show that there also exists a simple closed geodesic. Here, we argue similarly as in 2.7 and 2.8. More precisely, put

$$(3.3) \quad \begin{aligned} \tilde{\lambda}_1 = \inf\{(\max E|\tilde{w}_1); \text{ all cycles } \tilde{w}_1 \text{ representing } w_1 \\ \text{and belonging to } \tilde{\Lambda}_0/(\mathbb{Z}_2 \times \mathbb{Z}_2)\}. \end{aligned}$$

Let us assume that this infimum is assumed by the E -value of a family C_1 of simple closed geodesics. This is the counterpart of the set C_0 and with the same arguments as used in the proof of Proposition 5 one shows that C_1 is bounded away from $\partial\tilde{\Lambda}_0$. One now continues as in 2.8 and thus gets a simple closed geodesic c_1 which is different from an appropriately chosen Poincaré geodesic c_0 , always neglecting the parametrization, of course. Repeating everything with the cycles in the \mathbb{Z}_2 -homology class w_2 , one gets a third simple closed geodesic c_2 , different from c_0 and c_1 .

So far, we have restricted ourselves to a convex surface M . The property $K > 0$ here is essential to make sure that the function Ω on $\tilde{\Lambda}$ has regular values only. If we now consider a general oriented closed surface M of genus 0, Poincaré's method for the construction of a simple closed geodesic does not work. In this case, one has to use a method introduced by Birkhoff [1917] who considers the cycles in the generating 1-dimensional homology class of the relative space (Λ, Λ^0) , where Λ^0 is formed by the constant curves having E -value 0. A detailed exposition of this method using the manifold $\Lambda = {}_k\Lambda M$ of p.g. closed curves with a fixed number k of corners is given in Klingenberg [2002]. In this way, we get a closed geodesic c_0^* on M . To show that a refinement of this method yields a simple closed geodesic c_0 one proves the counterpart to Proposition 5. Note that in the proof we never used the property $K > 0$. One then continues with arguments similar to the ones in 2.8. Here we need a family $\{M_s, 0 \leq s \leq 1\}$ of surfaces where M_0 is a sphere and M_1 is the given surface M . The existence of such a family is a standard result.

After we got in this way a simple closed geodesic c_0 on M , additional simple closed geodesics c_1 and c_2 are constructed using a 2-dimensional and a 3-dimensional \mathbb{Z}_2 -homology class of the relative space (Λ, Λ^0) modulo the $(\mathbb{Z}_2 \times \mathbb{Z}_2)$ -action described above. The use of homology classes of this type goes back to Lusternik-Schnirelmann [1929]. The most delicate point again is to make sure that c_1 and c_2 can be assumed to be simple. It is here that an analogue of Proposition 5 and the arguments used in 2.8 come into play.

REFERENCES

- [1917] G. D. Birkhoff, *Dynamical systems with two degrees of freedom*. Trans. Amer. Math. Soc. **18**, 199-300 (1917).
- [1930] W. Blaschke, *Vorlesungen über Differentialgeometrie I*, 3. Aufl. Julius Springer, Berlin (1930). MR **17**:888g
- [1983] R. Bott, *Lectures on Morse Theory, Old and New*. Proc. Sympos. Pure Math. vol. **39**, part 2, 3-30 (1983). MR **84m**:58026b
- [1982] C. Croke, *Poincaré's Problem and the Length of the Shortest Closed Geodesic on a Convex Hypersurface*. J. Differential Geom. **17**, 595-634 (1982). MR **84f**:58034
- [2002] W. P. A. Klingenberg, *Klassische Differentialgeometrie*. <ftp://www.math.uni-bonn.de/people/klingenb/>.
- [1929] L. Lusternik and L. Schnirelmann, *Sur le problème des trois géodésiques fermées sur les surfaces de genre 0*. C. R. Acad. Sci. Paris Sér. I Math. **189**, 269-271 (1929).
- [1905] H. Poincaré, *Sur les lignes géodésiques sur les surfaces convexes*. Trans. Amer. Math. Soc. **17**, 237-274 (1909).
- [1958] S. Smale, *Regular curves on Riemannian manifolds*. Trans. Amer. Math. Soc. **87**, 2492-502 (1958). MR **20**:1319
- [1992] I. A. Taĭmanov, *On the existence of three nonselfintersecting closed geodesics on manifolds homeomorphic to the 2-sphere*. Russian Math. (Iz. VUZ) **40**, 565-590 (1993). (Translation of a paper which appeared 1992). MR **93k**:58051

MATHEMATISCHES INSTITUT DER UNIVERSITÄT BONN, BERINGSTRASSE 1, 53115 BONN, GERMANY

E-mail address: `klingenb@math.uni-bonn.de`