

A Criterion for Flatness in Minimal Area Metrics that Define String Diagrams

K. Ranganathan*

Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Received April 1, 1991

Abstract. It has been proposed that the string diagrams of closed string field theory be defined by a minimal area problem that requires that all nontrivial homotopy curves have length greater than or equal to 2π . Consistency requires that the minimal area metric be flat in a neighbourhood of the punctures. The theorem proven in this paper, yields a criterion which if satisfied, will ensure this requirement. The theorem states roughly that the metric is flat in an open set, U if there is a unique closed curve of length 2π through every point in U and all of these closed curves are in the same free homotopy class.

1. Introduction and Motivation

Consistency Conditions on String Field Theory. The fundamental idea of string theory is to define an n string amplitude as an integral over the moduli spaces of n punctured Riemann surfaces of all genera. For on-shell states the integrand arises from a conformal field theory. In fact the integrand is just the correlation function of n vertex operators inserted at the location of the punctures. The operator formulation of conformal field theory, which defines correlation functions as vacuum expectation values of a string of quantum fields, does not tell us how to extend the definition of amplitudes to off-shell states. This is done naturally in the path integral approach where we have a prescription for calculating general amplitudes (off and on shell). Fundamental to this prescription is the notion of a string diagram. A string diagram is a Riemann surface together with a choice of local coordinate (up to phase ambiguity) around each puncture. In fact for the purposes of this paper it is sufficient to say that the problem of defining off-shell string amplitudes is one of defining string diagrams [SoZ]. It is necessary that the local coordinates be ambiguous up to phase if the string diagrams are to be globally well defined on moduli space. We want amplitudes that are independent of the phase that we choose around each

* Supported in part by funds provided by the US Department of Energy (DOE) under contract # DE-AC02-76ER03069

puncture. This requires that instead of general off-shell states we restrict to those that satisfy the $L_0 - \bar{L}_0$ condition. Any globally defined choice of string diagrams will not give off-shell amplitudes satisfying with the requirements of symmetry, unitarity and factorization. Ensuring these properties requires a consistent set of string diagrams (described in [SoZ]). This however is only the beginning of the story. We further want to see that these amplitudes arise from a string field theory. In string field theory the amplitude is given as a sum over various Feynman diagrams. Each Feynman diagram corresponds to a set of string diagrams whose contributions we sum over. This allows us to reinterpret the integral done in computing a single Feynman diagram as an integral over a region of moduli space. With this picture in mind there is an additional consistency condition on string field theory. This requires that the string diagrams defined by string field theory smoothly cover the moduli spaces of all genera exactly once (henceforth called the smooth covering condition). The smooth covering of the moduli spaces requires in particular that there are no discontinuities as we go between Feynman diagrams that border each other in moduli space.

String Diagrams from Metrics. Certain metrics on a Riemann surface can define string diagrams. In a local parameter defined on a region U , a metric is equivalent to a positive real valued function on U , say $\rho(z)$, via the relation $dl = \rho(z)|dz|$. If the metric around the punctures looks like a semiinfinite cylinder going off to infinity we can define a string diagram. More precisely this condition means that we can choose a local coordinate around each puncture in which the metric is $1/r$. This prescription defines for us a class of local coordinates related to each other by an arbitrary complex scaling factor. We had required earlier that the phase of the local coordinate be undefined. However the scale of the coordinate system remains to be fixed and is done as follows. In one of the allowed local parameters (defined above) find the circle of largest radius centered at the origin with the property that the metric is flat at all points within it. Scale the coordinate system so that this circle becomes the circle $|z| = 1$ in the new local coordinate. This new local coordinate (and those related to it by a phase factor) serve to define a string diagram.

Metrics from Minimal Area Problems. Having seen that certain metrics can define string diagrams we move on to the proposal [Z2] that the metric solving the following minimal area problem (problem M) be used to construct a string field theory. This minimal area problem cannot have more than one metric as its solution. In fact such uniqueness is characteristic of minimal area problems on a Riemann surface [St].

Definition of Problem M

- a) R is a compact Riemann surface with a finite set of marked points (punctures).
- b) Y is the set of non-trivial homotopy curves on $R - \{\text{punctures}\}$.
- c) The length condition L has $L(\gamma) = 2\pi \quad \forall \gamma \in Y$.
- d) The allowed metrics are those in the conformal class defined by R .

If $l_\rho(\gamma)$ is the length of the curve γ in the metric ρ and $A(\rho)$ is the area of ρ (we use the regularized area; see [Z1]) then the solutions to problem M are the metrics in the allowed class which minimize A and satisfy $l_\rho(\gamma) \geq L(\gamma) \quad \forall \gamma \in Y$.

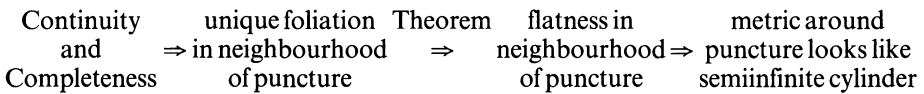
Properties Required of the Minimal Area Metric. The minimal area metric, ρ_m will define a string diagram if the metric looks like a semiinfinite cylinder going off to infinity (see discussion earlier on metrics and string diagrams). If ρ_m does define string diagrams then it was shown [Z2] that this set of string diagrams would be a consistent set (i.e. would ensure symmetry, unitarity and factorization of the amplitude). In addition a choice of propagator and procedure for constructing vertices was suggested that would give rise to string diagrams arising from the minimal area metric. This method would be correct (give a single covering of moduli space with string diagrams arising from the minimal area metric) if two crucial properties of ρ_m hold. One is the property that was used earlier, stating that the metric around a puncture looks like a semiinfinite cylinder going off to infinity. To understand the other, referred to as the amputation property, consider the bordered surface got by amputating the Riemann surface, along circles of constant radius in the local coordinates around the punctures. The amputation property is the requirement that the minimal area metric, under the condition that all nontrivial closed curves on the bordered surface have length greater or equal 2π , be the restriction of ρ_m to the amputated surface. One way in which we could verify that these two properties hold is to check that the metric ρ_m arises from a Jenkins–Strebel quadratic differential. Jenkins–Strebel quadratic differentials are those for which all of the horizontal trajectories are closed. In fact in the class of problems in which length conditions are applied only on a set of homotopy classes for which we can simultaneously choose a set of non-intersecting representatives (an admissible set of classes), the minimal area metric arises from a Jenkins–Strebel quadratic differential [St][Z1]. This does not tell us if ρ_m arises from a Jenkins–Strebel quadratic differential since the set of all nontrivial homotopy classes is not an admissible set. A partial result was obtained, stating that Jenkins–Strebel quadratic differentials did yield ρ_m for all the genus zero surfaces [Z1]. Further in many of the higher genus surfaces Jenkins–Strebel quadratic differentials yield ρ_m , but there are indications that this will not be true for all the higher genus surfaces. We must therefore try a more direct method to establish the truth of the two properties.

What This Paper Does and Does Not Show. Establishing that the minimal area metric, ρ_m looks like a semiinfinite cylinder going off to infinity around the punctures will be our concern in this paper. Let us remember that this property came up at two points in the discussion above. This property follows from the flatness of the metric in a neighborhood of the puncture if for any point P in that neighborhood the length of the shortest closed curve, among those that encircle the puncture and pass through p , is 2π . The latter fact will be established in the paragraph below. We will establish a criterion, stated in the form of the theorem below, that ensures flatness in an open set on the Riemann surface. Loosely stated the criterion requires that there be a unique closed curve of length 2π through every point $p \in U$ and that the free homotopy class of these closed curves be independent of p . More precisely let Δ_p be the set of all homotopy classes of closed curves passing through a point p on the Riemann surface. Let $\alpha_p \in \Delta_p$ be a nontrivial homotopy class of closed curves passing through p . Let $S^{\alpha_p} = \inf_{\gamma \in \alpha_p} l(\gamma)$.

Theorem. *Let U be an open set on the Riemann surface R . Let ρ_m be the minimal area metric solving problem M . Assume that the minimal area metric is C^1 -continuous.*

Assume there exists a free homotopy class of closed curves, v such that for every $p \in U$ there is a closed curve of length 2π in the class v_p (this implies that $S^{v_p} = 2\pi$). Assume further that $\forall p \in U$ and $\forall \alpha_p \neq v_p, \alpha_p \in \Delta_p, S^{\alpha_p} > l' > 2\pi$. We can then conclude that ρ_m is flat in U .

The conditions of this theorem are in fact satisfied in a sufficiently small open set enclosing a puncture if ρ_m has the properties of continuity and completeness. After defining continuity and completeness we will indicate roughly why these properties ensure the conditions stipulated in the theorem. Continuity of the metric is continuity of the real valued function ρ_m in every local parameter. Completeness is the requirement that any curve having one of its end points at a puncture be of infinite length. It was shown in [Z2] that in the minimal area metric ρ_m the infimum of lengths of closec curves through a point p is 2π (i.e. $\inf_{\alpha_p \in \Delta_p} S^{\alpha_p} = 2\pi$). Completeness tells us that for any point p , in a sufficiently small neighbourhood of a puncture, U , all closed curves through p not in the homotopy class going around the puncture, have length greater than 2π . In fact when we combine these two facts with the continuity of the metric we can show the existence of a unique closed geodesic encircling the puncture for every point p , in U . The arrow diagram below indicates the logical context for the above theorem.



While we cannot claim to have proved flatness around punctures until continuity and completeness of ρ_m are established, we believe that the reduction of the result to the verification of these two properties is a useful step. The authors of [WZ] have an alternate proof of the theorem above and a proof of the amputation property. They are investigating issues of existence, completeness and continuity.

2. Structure of Proof

In the first part of this section the crucial idea of localization is discussed and leads to a short sketch of the arguments to follow. This should be useful even though some of the terms used will be defined later in this section. Next, to bring out the logic involved in implementing this idea, three questions are asked and answered. An understanding of these without their justification will give the reader an idea of the gross features of the proof. The answer to the second question is justified in full in this section. The answers to the first and third question are justified in Sects. 3 and 4 respectively.

Localization. The idea of the proof is to localize the global minimal area problem to a local one. The length condition in problem M is applied on “large” curves (non-trivial homotopy curves). Our intuition tells us that in a neighbourhood D , of a point p , $\rho_m|_D$ is the solution to a local minimal area problem in which a length condition is applied on the curves in D that run between two points on its boundary. We now give the precise definition of a local minimal area problem followed by the definition of two length conditions l_i and l_e .

Definition of Local Minimal Area Problem

D: *D* is a subset of R^2 which is a conformal image of the unit disc $|z| \leq 1$.

B: *B* is the boundary of *D*.

P: *P* is the set of pairs of points on the boundary (i.e. $P = \{(b_1, b_2) : b_1, b_2 \in B\}$).

l: *l* is a length condition (i.e. $l: P \rightarrow R$).

ρ : In a local parameter the metric ρ is a positive real valued function on *D* (i.e. $\rho: D \rightarrow R^+$).

We seek the minimal area metric, ρ_e (denoted ρ_e, ρ_f , etc. to indicate that the length condition in the problem is l_e, l_f , etc.) such that any curve between b_1 and b_2 is longer than $l_e(b_1, b_2)$.

Definition of l_i and l_e

$l_i(b_1, b_2)$ = length of shortest path between b_1 and b_2 .

$l_e(b_1, b_2) = [2\pi - (\text{length of shortest path in any noninternal homotopy class, between } b_1 \text{ and } b_2)]$. Non-internal homotopy class here means that the curve is a nontrivial closed curve on the space obtained by identifying points in the region *D*. Notice that l_e could be negative for certain values of b_1 and b_2 .

One might suspect that the length condition l_e (and/or l_i) would have $\rho_m|_D$ as the associated minimal area metric. This could be established if l_e (and/or l_i) was known to satisfy an inequality (triangle inequality). However it is not apparent why l_e (and/or l_i) should satisfy this inequality. Even if they did yield $\rho_m|_D$ it would not help since it is not clear why they should give rise to minimal area metrics that are flat. We will solve the problem of localization by finding a class of length conditions giving rise to flat metrics. All length conditions in this class will be found to satisfy the triangle inequality. In this class of length conditions we will find one that has $\rho_m|_D$ as its solution.

Q1. Is there a class of length conditions on the disc for which the minimal area metric is flat?

ANS. Yes. It is shown in Sect. 3 that the minimal area metric, ρ_u is flat if $l_u(b_1, b_2) = |u(b_1) - u(b_2)|$, where u is a real valued continuous function on the boundary *B* with only two extrema. It will be useful later to recognize that the triangle inequality $l_u(b_1, b_3) \leq l_u(b_1, b_2) + l_u(b_2, b_3)$ is satisfied by l_u for any choice of the function u .

Q2. What condition on u ensures $\rho_u = \rho_m|_D$?

ANS. We prove now that if $l_e(b_1, b_2) \leq l_u(b_1, b_2) \leq l_i(b_1, b_2)$ then $\rho_u = \rho_m|_D$. This inequality will be central to our arguments and will be referred to as the central inequality. Our method will be to establish first that, $\rho_m|_D$ is admissible to the local minimal area problem defined by l_u and secondly that $A(\rho_u) = A(\rho_m|_D)$. These two facts coupled with the known uniqueness of solutions to the minimal area problem defined by l_u implies that $\rho_u = \rho_m|_D$. The first step is easy since $l_u(b_1, b_2) \leq l_i(b_1, b_2)$ implies that $\rho_m|_D$ is admissible for the problem defined by l_u . In addition we have already advanced one step in proving $A(\rho_u) = A(\rho_m|_D)$ since the last statement implies $A(\rho_u) \leq A(\rho_m|_D)$. All we need show to complete the argument is $A(\rho_u) \geq A(\rho_m|_D)$. To do this define a metric ρ_{new} on the Riemann surface by replacing ρ_m by

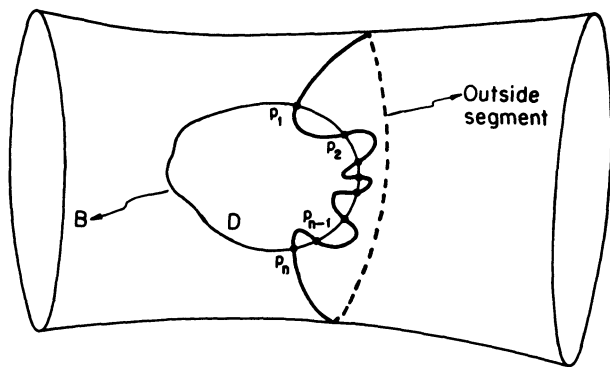


Fig. 1. Proving that ρ_{new} is admissible to Problem M

ρ_u in the region D and leaving it unchanged elsewhere. If ρ_{new} is admissible to problem M then we can conclude that $A(\rho_u) \geq A(\rho_m|_D)$.

We now show that ρ_{new} is admissible to problem M . Consider an arbitrary non-trivial homotopy curve on R , say γ , which intersects D . As we move from $t = 0$ to $t = 1$ the curve intersects the boundary at the points $b_1 \cdots b_n$ (assumed finite). Of the n segments $b_1b_2, b_2b_3, \dots, b_nb_1$ at least one is an open curve in an external homotopy class (referred to as the outside segment with length l_{out}). For the remaining segments we can say that excepting for endpoints they lie either in the interior of D or its exterior and will be referred to as interior and exterior segments respectively. See Fig. 1.

The length of γ in the new metric, denoted $l_{new}(\gamma)$, is the sum of the length of its segments. Denoting the length of segments $b_1b_2, b_2b_3, \dots, b_nb_1$ as $l_{12}, l_{23}, \dots, l_{n1}$ we can write

$$l_{new}(\gamma) = l_{12} + l_{23} \cdots + l_{n1}. \tag{1}$$

Separate out the contributions of the outside, interior and exterior segments. Use now the trivial fact that the exterior segments cannot be shorter than the shortest curves between their endpoints (i.e. those defining l_i length condition) to obtain

$$l_{new}(\gamma) \geq \sum_{int} l_{new} + \sum_{ext} l_i + l_{out}, \tag{2}$$

Now use $l_u(b_1, b_2) \leq l_i(b_1, b_2)$ to reestimate the length of the exterior segments. For the interior segments notice that their lengths are defined in terms of the metric ρ_u and therefore must satisfy the l_u length condition. As a result

$$l_{new}(\gamma) \geq \sum_{int} l_u + \sum_{ext} l_u + l_{out}. \tag{3}$$

The l_u length condition satisfies a triangle inequality of the form $l_u(b_1, b_3) \leq l_u(b_1, b_2) + l_u(b_2, b_3)$. This, used repeatedly, implies that $l_u(b_1, b_n) \leq \sum_{int} l_u + \sum_{ext} l_u$. As a result we get

$$l_{new}(\gamma) \geq l_u(b_1, b_n) + l_{out}. \tag{4}$$

If we use the fact that $l_e(b_1, b_n) \leq l_u(b_1, b_n)$ together with the definition of $l_e(b_1, b_n)$ as in Eq. 5 we see that the length of γ is not less than 2π ,

$$l_{\text{new}}(\gamma) \geq l_e(b_1, b_n) + l_{\text{out}} \geq 2\pi. \tag{5}$$

Q3. Can we construct a u satisfying the above conditions?

ANS. Yes. We have assumed that through every point $p \in D$ there is a unique geodesic of length 2π , called a saturating geodesic. These geodesics form a foliation. Using the tangent vector to a geodesic we can define a vector at every point in D . This vector field has zero curl. This allows us to integrate the vector field to get a real valued function u . The restriction of this function to the boundary B , is the function that we seek. The detailed argument proving this is given in Sect. 4 but now we give a sketch of the argument. Observe first that $l_e(b_1, b_2) \leq l_i(b_1, b_2)$ follows directly from the fact that ρ_m is the solution to problem M . To understand why l_u defined in this way satisfies the central inequality notice that the length conditions $l_e(b_1, b_2), l_i(b_1, b_2)$ and $l_u(b_1, b_2)$ are the lengths of certain geodesic segments between b_1 and b_2 . Since geodesics are solutions to an extremal length problem, the change in their length as their endpoints are changed can be computed in terms of line integrals of vector fields along any pair of paths from the initial to the final endpoints (see Lemma 3). To establish the central inequality for a pair of points (b_1, b_2) we first choose a point \tilde{b}_2 such that the central inequality holds for the pair (b_1, \tilde{b}_2) (e.g. $\tilde{b}_2 = b_1$). Then we choose a path from \tilde{b}_2 to b_2 . As mentioned earlier the difference in value of either of the three length conditions between pairs (b_1, \tilde{b}_2) and (b_1, b_2) is the line integral of certain vector fields along this path. We will divide the path into two parts. On the first part the vector fields to be integrated will be so oriented that they will satisfy an infinitesimal form of the central inequality (an inequality on scalar products). This implies that the contribution from the first part does not alter the central inequality. On the second part of the path the scalar product inequality will not be valid but the contribution of the second part to the length conditions can be shown to be small enough to ensure that the central inequality is still valid.

3. Length Conditions Yielding Flat Metric in a Disc

We justify here the answer to the first question. In other words we establish that ρ_u , the minimal area metric such that any curve between b_1 and b_2 is longer than $l_u(b_1, b_2) = |u(b_1) - u(b_2)|$, is flat. This problem is a generalization of the problem where u is defined and constant on disconnected segments of the boundary and is left undefined elsewhere $[A]$. A criterion by Beurling gives us a sufficient condition for a metric to be of minimal area $[A]$. We will now use this criterion to prove that $\rho_u = |\nabla u|$, where u is the solution to the Dirichlet problem defined by the boundary value $u(B)$. As can be verified easily such a metric is flat. To make the argument simple we will make the technical assumption that u has only two extrema. We believe however that the result could be established for more general $u(B)$.

Firstly the metric $\rho = |\nabla u|$ is admissible since if $\eta(t), t \in [0, 1]$ is a curve with $\eta(0) = b_1$ and $\eta(1) = b_2$ then

$$l(\eta) = \int_0^1 |\nabla u(t)| |dr(t)| \geq \left| \int_0^1 \nabla u(t) \cdot \vec{dr}(t) \right| = l_u(b_1, b_2). \tag{6}$$

Next we note that the curves that saturate the length condition are the integral curves of the vector field ∇u (Fig. 2). To see this let b_1 and b_2 be two points on the boundary lying on the same integral curve γ . Then we see that

$$l(\gamma) = \int_0^1 \nabla u(t) |dr(t)|; = \int_0^1 |\nabla u(t) \cdot \vec{dr}(t)| = \left| \int_0^1 \nabla u(t) \cdot \vec{dr}(t) \right| = l_u(b_1, b_2). \tag{7}$$

Denote the set of saturating curves by Γ . See Fig. 2. The Beurling criterion requires us to check that $\int_D h \rho dx dy \geq 0$ for every real valued function h on D satisfying

$$\int_{\gamma} h |dr| \geq 0 \quad \forall \gamma \in \Gamma \tag{8}$$

To do this consider v , the harmonic function conjugate to u . It is well defined since D is a simply connected domain. Let $z = x + iy$ then we can show that the analytic function $f(z) = u + iv$ is univalent in D . In fact consider the closed curve which is the image of B under $f(z)$. The argument principle in the theory of complex variables tells us that if $f(z)$ is not univalent then $f(B)$ must wind around some point more than once. However this would require that u has more than two extrema. Univalence of $f(z)$ tells us that u and v define a coordinate system on D . We now compute the metric in this coordinate system.

$$dudv = |\nabla u|^2 dx dy = \rho^2 dx dy.$$

This change of variables means that in the coordinate system defined by u and v the metric is identically one. If we rewrite the Beurling criterion in terms of the new variables it is clear that it is satisfied.

$$\int_D h \rho dx dy = \int_D \frac{h}{\rho} dudv = \int dv \int du \frac{h}{\rho} \geq 0. \tag{9}$$

In the last step above we have used the fact that along the curves in Γ (which are the constant v curves) $du = \rho |dz|$ which implies that $\int_{\gamma \in \Gamma} \frac{h}{\rho} du = \int_{\gamma \in \Gamma} h |dr| \geq 0$.

4. Central Inequality and Extrema of $u(B)$

At this stage flatness can be established at a point $p \in U$ if there is a region D containing p , on the boundary of which a function u , with two extrema, can be

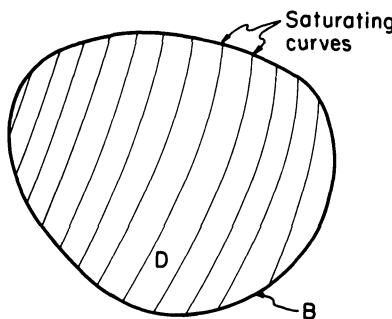


Fig. 2. Beurling criterion

defined satisfying the central inequality $l_e(b_1, b_2) \leq l_u(b_1, b_2) \leq l_i(b_1, b_2)$. In the series of lemmas in the first subsection we set up the picture of foliating geodesics, understand the geodesics defining the various length conditions and define u . This will set up the groundwork and the language needed later. The second subsection is devoted to obtaining a region D with properties that will be necessary in the proof of the central inequality. At the end of the second subsection we will verify that u restricted to the boundary has exactly two extrema. The last section will establish the central inequality.

4.1 The Groundwork

Definition 1 (Saturating geodesics and the associated vector field). A closed geodesic of length 2π , is called a saturating geodesic and it defines a vector field $\vec{u}(p) = \rho_m(p)\vec{\tau}(p)$, where $\vec{\tau}(p)$ is the unit tangent (in the local parameter) to the foliating geodesic through p .

Lemma 1 (Geodesic foliation). Given an open set $U \subset \mathbb{R}$ satisfying the conditions of the theorem we define an annular region Λ_U as the set of points through which the saturating geodesics through U pass. The C^1 continuity of the metric implies [WZ] that $\forall q \in \Lambda_U$ there is a unique saturating geodesic in the homotopy class v_q and these geodesics are non-trivial Jordan curves in Λ_U . The saturating geodesics in Λ_U (denoted λ_1, λ_5 etc) are totally ordered and the ordering will be denoted, e.g. $\lambda_1 < \lambda_5$ (See Fig. 3). This ordering is obtained by mapping the annular region to an annulus centered at the origin of the complex plane. Since saturating geodesics never intersect each other λ_1 is always inside λ_5 ($\lambda_1 < \lambda_5$) or λ_1 is always outside λ_5 ($\lambda_1 > \lambda_5$).

Lemma 2 (Geodesics defining l_i & l_e). There exists an open set $V \subset U$ such that $p \in V$ and $\forall p_1, p_2 \in V$ there are only three types of paths between them that are relevant in the definition of l_i and l_e . These are $\mathfrak{P}^{int}(p_1, p_2)$, $\mathfrak{P}^{ext}_{lh}(p_1, p_2)$ and $\mathfrak{P}^{ext}_{rh}(p_1, p_2)$ referring to the shortest paths (hence geodesics) in the internal, left external and right external homotopy class of paths from between p_1 and p_2 . The path defining $l_e(b_1, b_2)$ is either $\mathfrak{P}^{ext}_{lh}(p_1, p_2)$ or $\mathfrak{P}^{ext}_{rh}(p_1, p_2)$. Since these paths lie completely in Λ_V (Λ_V is defined analogous to Λ_U above) we will be able to restrict our attention to Λ_V for the arguments that follow this lemma. See Fig. 3.

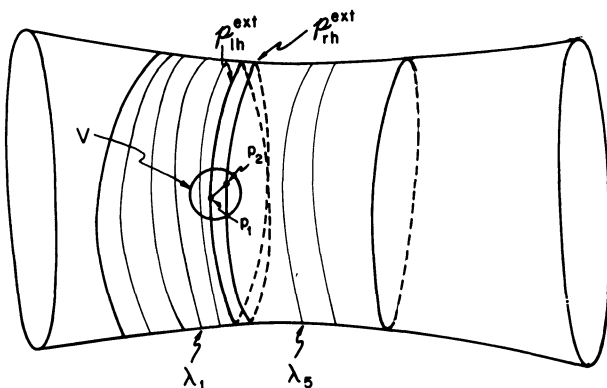


Fig. 3. Geodesics defining l_i and l_e

Proof. While the existence of the three paths can be established easily, proving that $\mathfrak{P}_{ih}^{ext}(p_1, p_2)$ and $\mathfrak{P}_{rh}^{ext}(p_1, p_2)$ are the only external paths relevant in defining l_e requires more work. We know that for any $p \in U$, $S^{\alpha_p} \geq l' > 2\pi \forall \alpha_p \neq \nu_p$. Choose a sufficiently small open set $V \subset U$ to ensure that any two points of V can be joined by a curve of length less than $\frac{l' - 2\pi}{2}$. This ensures that the shortest external path between p_1 and p_2 other than $\mathfrak{P}_{ih}^{ext}(p_1, p_2)$ and $\mathfrak{P}_{rh}^{ext}(p_1, p_2)$ will always be longer than $l' - \frac{l' - 2\pi}{2} = \frac{l' + 2\pi}{2}$. However since $\mathfrak{P}_{ih}^{ext}(p_1, p_2)$ is the shortest path in its homotopy class $l(\mathfrak{P}_{ih}^{ext}(p_1, p_2)) < l(\mathfrak{P}_{ih}^{ext}(p_1, p_1)) + l(\mathfrak{P}^{int}(p_1, p_2)) < 2\pi + \frac{l' - 2\pi}{2} < \frac{l' + 2\pi}{2}$. Similarly $l(\mathfrak{P}_{rh}^{ext}(p_1, p_2)) < \frac{l' + 2\pi}{2}$. This implies that $\mathfrak{P}_{ih}^{ext}(p_1, p_2)$ and $\mathfrak{P}_{rh}^{ext}(p_1, p_2)$ are shorter than any other external paths between p_1 and p_2 .

Lemma 3 (Geodesic length and endpoint variation). *Let $\gamma(s)$ ($s \in [i, f]$) be a family of geodesics with endpoints $\gamma_0(s)$ and $\gamma_1(s)$. Let the unit tangent vectors (in the local parameter) at the endpoints be $\vec{t}_0(s)$ and $\vec{t}_1(s)$. Define $\vec{v}(\gamma_0(s), \hat{\gamma}_1(s)) = \rho_m(\gamma_1(s))\vec{t}_1(s)$. The ordering of the endpoints in the parentheses indicates that the geodesic $\gamma(s)$ is to be parameterized from γ_0 to γ_1 . The caret above one of the endpoints indicates that we choose the tangent vector at that endpoint. See Fig. 4. We assert that*

$$l(\gamma(f)) - l(\gamma(i)) = \int_i^f ds [\vec{v}(\gamma_0(s), \hat{\gamma}_1(s)) \cdot \vec{d}\gamma_1 - \vec{v}(\hat{\gamma}_0(s), \gamma_1(s)) \cdot \vec{d}\gamma_0]. \tag{10}$$

Proof. This follows from the fact that geodesics are the solution of a minimal length problem. The lagrangian for this problem would be $\rho_m(\gamma(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$. The first order variation in length of a geodesic receives contributions only from the end point variation. In symbols $\delta l = \vec{p} \cdot \delta \vec{r}$, where the canonical momentum $\vec{p} = \rho_m \vec{t}$. The lemma is just an integrated form of this result.

Lemma 4 (u is well defined). $u(q) = \int_E^q \vec{u}(q) \cdot \vec{d}r$ is a well defined function of the point q . The integration is done along any path from E to q . The vector field $\vec{u}(q)$ is defined in Definition 1.

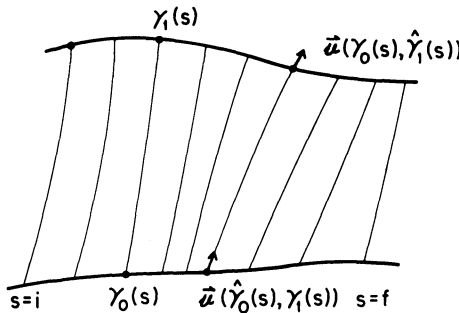


Fig. 4. Computing changes in the length of geodesics

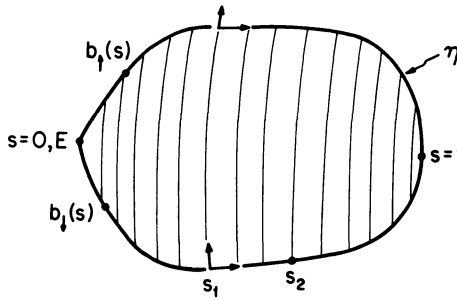


Fig. 5. Proving that u is a well defined function

Proof. Consider an arbitrary closed curve in D , say η . Without loss of generality we choose η so that the saturating geodesics don't intersect the boundary more than twice. See Fig. 5. We need to show that

$$\oint_{\eta} \vec{u}(q) \cdot \vec{dr} = 0. \tag{11}$$

Associate with a parameter $s \in [0, 1]$ a saturating geodesic interesting η at $b_1(s)$ and $b_2(s)$. We require that the saturating geodesics satisfy $\lambda(s_1) < \lambda(s_2)$ if $s_1 < s_2$. Let

$$l(s) = \int_E^{b_1(s)} \vec{u}(q) \cdot \vec{dr} - \int_E^{b_2(s)} \vec{u}(q) \cdot \vec{dr}. \tag{12}$$

Lemma 3 tells us that $l(s)$ is the length of the shortest curve running between $b_1(s)$ and $b_2(s)$. Since $\oint_{\eta} \vec{u}(q) \cdot \vec{dr} = l(1) = 0$ we have the required result.

4.2. Construction of D Satisfying Required Properties

We now construct the region D with a view to ensuring that it satisfies a set of properties. These properties which are defined shall be used in proving the assumptions on the function u that we had made earlier. In order to proceed to the next subsection, where the actual proof of the inequality is presented, one could just understand these properties and skip everything else in this subsection. The fact that u has only two extrema is a simple corollary of these properties and this is shown at the end of this subsection. First we define the idea of Conjugate points and Uncrossed pairings which will be required in stating the properties of the region D .

Definition 2 (Conjugate points). *Kfl*($\mathfrak{P}_{rh}^{ext}(p_1, p_2)$) = $l(\mathfrak{P}_{ih}^{ext}(p_1, p_2))$ ($p_1, p_2 \in D$) then p_1 and p_2 are said to be conjugate to each other.

Definition 3 (Uncrossed pairings). *A pairing is a collection of pairs of distinct points on the boundary such that every point is contained in exactly one pair. Consider two pairs of points on the boundary, B . Choose one of the pairs. It cuts the boundary into two segments. If both points in the other pair lie in the same segment then the two*

pairs are uncrossed. A pairing is an uncrossed pairing if any two pairs that it defines is uncrossed.

Let us call the direction defined by the $\bar{u}(p)$ to be the vertical and the direction perpendicular to this to be the horizontal. Choose a local parameter whose origin is at the point p and aligned such that the x axis is along the horizontal and $\bar{u}(p)$ points along the positive y axis. The local parameter allows us to use the concepts of planar geometry such as straight line segments and circles. These should be distinguished from the entities intrinsic to the metric such as geodesic segments. In addition we can define the angle between two vectors based at different points, $\text{Ang}(\bar{w}(p_1), \bar{v}(p_2))$, as the lesser of the two angles that the axes defined by them make. This definition implies that Ang will always be less than $\frac{\pi}{2}$. Choose a circular disc W (in this local parameter), of sufficiently small radius to ensure that the following constraints hold. (The value of ε will be fixed later on. For the present it will be sufficient if $\varepsilon < \frac{\pi}{8}$.)

Constraint 1: $\text{Ang}(\bar{u}(p), \bar{u}(q)) \leq \varepsilon, \forall q \in W$.

Constraint 2: At p_1 and p_2 the geodesics $\mathfrak{P}_{ih}^{\text{ext}}(p_1, p_2)$ and $\mathfrak{P}_{rh}^{\text{ext}}(p_1, p_2)$ define vectors ${}_{rh}\bar{v}^e(\hat{p}_1, p_2), {}_{ih}\bar{v}^e(\hat{p}_1, p_2)$ etc. by the prescription given earlier defining vectors from geodesics (see definition of \bar{u} in Subsect. 4.1). We use the generic symbol $\bar{v}^e(\hat{p}_1, p_2)$ for any of these vectors. We require $\text{Ang}(\bar{u}(p), \bar{v}^e(\hat{p}_1, p_2)) \leq \varepsilon, \forall p_1, p_2 \in W$.

Constraint 3: The geodesic segment (shortest path) between two points in W should be “sufficiently straight.” More precisely we require that for any two points on the path, the angle between the tangent vectors to the path be less than ε .

Let $\tilde{l}_i(p_1, p_2)$ be the length of the geodesic segment and $l(p_1, p_2)$ the length of the straight line segment between p_1 and p_2 ; the lengths of these curves being computed using a metric that is identically one on the disc. A consequence of Constraint 3 which will be useful in understanding Constraint 4 below is that $l(p_1, p_2) \leq \tilde{l}_i(p_1, p_2) \leq \sec(\varepsilon) \cdot l(p_1, p_2)$.

Constraint 4. The length $l_i(p_1, p_2)$, of the geodesic segment connecting p_1 and p_2 satisfies $(1 - \varepsilon) \cdot \rho(p) \cdot l(p_1, p_2) \leq l_i(p_1, p_2) \leq (1 + \varepsilon) \cdot \rho(p) \cdot l(p_1, p_2)$.

Let us see why Constraint 4 can be satisfied. Clearly the definition of lengths in the metric ρ implies $\tilde{l}_i(p_1, p_2) \cdot \rho_{\min} \leq l_i(p_1, p_2) \leq \tilde{l}_i(p_1, p_2) \cdot \rho_{\max}$, where ρ_{\min} and ρ_{\max} are the minimum and maximum values of the metric on the disc W . Our comment (below Constraint 3) can now be used to conclude that $l(p_1, p_2) \cdot \rho_{\min} \leq l_i(p_1, p_2) \leq l(p_1, p_2) \cdot \sec(\varepsilon) \cdot \rho_{\max}$. Continuity of the metric at the point p implies that for sufficiently small disc W the last inequality yields Constraint 4.

Estimate 1. Consider two points p_1 and p_2 in W which are the conjugates of each other. We ask for the maximum angle, θ_{\max} that the straight line segment $p_1 p_2$ can make with the horizontal. Constraint 2 tells us that for all q on the segment $p_1 p_2$, $\theta > \varepsilon$ implies that ${}_{rh}\bar{v}^e(\hat{q}, p_1)$ and ${}_{ih}\bar{v}^e(\hat{q}, p_1)$ have opposite sign for their scalar product with the tangent to the straight line segment $p_1 p_2$. If we now integrate along the straight line segment $p_1 p_2$ to compute $l(\mathfrak{P}_{ih}^{\text{ext}}(p_1, p_2))$ and $l(\mathfrak{P}_{rh}^{\text{ext}}(p_1, p_2))$ we see that

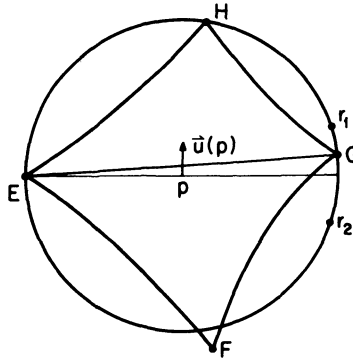


Fig. 6. Construction of the region D

one of them increases monotonically whereas the other decreases monotonically. We conclude that $l(\mathfrak{P}_{ih}^{\text{ext}}(p_1, p_2)) \neq l(\mathfrak{P}_{rh}^{\text{ext}}(p_1, p_2))$ if $\theta > \varepsilon$ and hence that $\theta_{\max} = \varepsilon$.

Construction of EFGH. The horizontal through p intersects the boundary of W at two points. Choose one of these to be the point E . Any point conjugate to E lies in a cone of angle 2ε with apex at E . Let the cone intersect the circle W at two points r_1 and r_2 . Notice that $l(\mathfrak{P}_{ih}^{\text{ext}}(E, r_1)) \geq l(\mathfrak{P}_{rh}^{\text{ext}}(E, r_1))$ but $l(\mathfrak{P}_{ih}^{\text{ext}}(E, r_2)) \leq l(\mathfrak{P}_{rh}^{\text{ext}}(E, r_2))$. Constraint 2 implies that $l(\mathfrak{P}_{ih}^{\text{ext}}(E, b))$ and $l(\mathfrak{P}_{rh}^{\text{ext}}(E, b))$ vary monotonically as b moves along the boundary from r_1 to r_2 . This shows that there is a unique point G on the boundary of W conjugate to E . Now send out a pair of geodesics from both E and G at 45° to the straight line segment EG as shown in Fig. 6. Let them meet at F and H as shown in Fig. 6. D is then the geodesic quadrilateral $EFGH$. It can be verified that the construction ensures that p lies in the interior of D .

Property 1. D is geodesically convex. This means that the geodesic connecting two points b_1 and b_2 on the boundary B , in the metric ρ_m , lies completely within the region D .

Proof. The edges of the quadrilateral are geodesic segments. Choose any edge. The entire quadrilateral lies on the side of the geodesic defining this edge. This implies geodesic convexity.

Property 2. The saturating geodesics intersect B at precisely two points except for the ones passing through E and G alone. This defines the f -pairing. If b_1 and b_2 are paired in this manner then we write $b_2 = b_1 \downarrow$. If $b_1 \in EFG$ then $b_1 \downarrow \in GHE$ and vice versa. Further the f -pairing is uncrossed.

Proof. Let b_1 and b_2 be on the same saturating geodesic. Using the Constraint 1 on $\vec{u}(p)$ we find that the maximum angle that the line segment $b_1 b_2$ makes with the vertical is ε . This shows that the pairing is well defined. The pairing is uncrossed since otherwise the saturating geodesics would intersect and this is impossible.

Property 3. Conjugate points on the boundary can be paired and this will be called the c -pairing. This pairing excludes the points F and H . If two points b_1 and b_2 on the boundary are paired then we write $b_2 = b_1 \uparrow$. If $b_1 \in HEF$ then $b_1 \uparrow \in HGF$ and vice versa. In addition the c -pairing is uncrossed.

Proof. We will first establish that the c -pairing is well defined. The maximum angle that a pair of conjugate points can make with the horizontal is ε (Estimate 1). If $b_1 \in HEF$, the geometry of the region D , requires that any point conjugate to it lie in HGF . In fact an argument similar to the one given in the paragraph entitled “Construction of EFGH” tells us that there is a unique point $b_1^* \in HGF$ conjugate to b_1 . It is useful to observe that $E^* = G$ by the construction of D . Having established the existence of the c -pairing we will now check that it is an uncrossed pairing. Consider the conjugate pair b_1, b_1^* . If we move downwards (upwards) from b_1 to b_2 then to get b_2^* we must move downwards (upwards) from b_1^* to ensure that equality is regained between $l(\mathfrak{B}_{ih}^{ext}(p_1, p_2))$ and $l(\mathfrak{B}_{rh}^{ext}(p_1, p_2))$. This implies that the c -pairing is uncrossed.

We will establish now that $u(B)$ has exactly two extrema. Constraint 1 tells us that $\vec{u}(b) \cdot \vec{dr}(b) \neq 0$ at any point $b \in B$. Further $\vec{u}(b) \cdot \vec{dr}(b)$ changes sign only at F and H . This implies that the function u (see Lemma 4 for its definition) restricted to B has extrema only at F and H .

4.3. *The Central Inequality, $l_e(b_1, b_2) \leq l_u(b_1, b_2) \leq l_i(b_1, b_2)$*

Sketch. It is useful to notice that $l_e(b_1, b_2) \leq l_i(b_1, b_2)$ follows directly from the fact that ρ_m is the solution to problem M . It is only sandwiching l_u between l_e and l_i that requires detailed argument below. Choose a point b_1 . Associated to this point are the vector fields (depending on the variable b) $\vec{v}^e(\hat{b}, b_1)$, $\vec{u}_{b_1}(b)$ and $\vec{v}^i(b_1, \hat{b})$. The vector field $\vec{u}_{b_1}(b)$ is related to the vector field $\vec{u}(b)$ (see the next page for details). The length conditions $l_e(b_1, b_2)$, $l_u(b_1, b_2)$ and $l_i(b_1, b_2)$ are the line integrals of these vector fields.

$$\begin{aligned}
 l_e(b_1, b_2) &= \int_{b_1}^{b_2} \vec{v}^e(\hat{b}, b_1) \cdot \vec{dr}(b), & l_u(b_1, b_2) &= \int_{b_1}^{b_2} \vec{u}_{b_1}(b) \cdot \vec{dr}(b), \\
 l_i(b_1, b_2) &= \int_{b_1}^{b_2} \vec{v}^i(b_1, \hat{b}) \cdot \vec{dr}(b). & &
 \end{aligned}
 \tag{13}$$

We can say roughly that the central inequality is a consequence of the inequality on scalar products, $\vec{v}^e(\hat{b}, b_1) \cdot \vec{dr} \leq \vec{u}(b) \cdot \vec{dr} \leq \vec{v}^i(b_1, \hat{b}) \cdot \vec{dr}$ (see Fig. 7). There are four vectors involved in the scalar product inequality. First we will give conditions, on the configuration of the four vectors, that are sufficient to ensure the scalar product inequality. Next we discuss the nature of the three vector fields with regard to these sufficiency conditions. The concluding section of the proof considers two differing choices for the path of integration, depending on the relative configuration of the points b_1 and b_2 . For either configuration we will in general need to break the path of integration into two segments. Only on one of these segments will the scalar product inequality be valid. We will therefore need to establish that the contribution from the other segment does not make the central inequality invalid.

Inequality on Scalar Products. The scalar product inequality, $\vec{v}^e(\hat{b}, b_1) \cdot \vec{dr} \leq \vec{u}(b) \cdot \vec{dr} \leq \vec{v}^i(b_1, \hat{b}) \cdot \vec{dr}$, is the crux of the proof of the central inequality. We will now establish conditions that are sufficient to ensure the scalar product inequality. Let $p_1 < p_2$ ($p_1 > p_2$). See Fig. 7. The scalar product inequality will be valid at a point q on the path of integration from p_1 to p_2 if

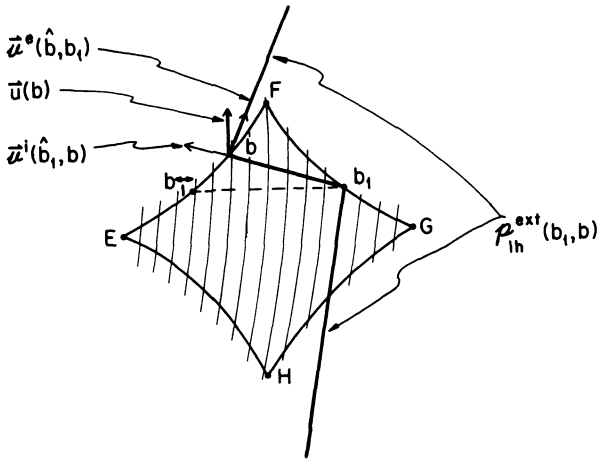


Fig. 7. Vector fields and the scalar product inequality

- a) the tangent vector at q (to the path) points in the direction of increasing (decreasing) grading. The grading referred to is the one given to the saturating geodesics in Subsect. 4.1.
- b) $\vec{v}^e(\hat{b}, b_1)$, $\vec{u}_{b_1}(b)$ and $\vec{v}^i(b_1, \hat{b})$ all lie on the same side of the tangent to the path.

We leave the proof of this simple result to the reader. Let us look ahead and see whether the two conditions above will hold for the paths chosen in the proof. The path chosen will never pass through the points E and G . This will ensure the first requirement that the tangent vectors to the path always point the right way. In addition for all points b on the path $\vec{v}^e(\hat{b}, b_1)$ and $\vec{v}^i(b_1, \hat{b})$ will be outward pointing. However $\vec{u}_{b_1}(b)$ will not in general be outward pointing for all points b on the path.

Vector Field for l_e . The geodesic defining the vector field for l_e switches between $\mathfrak{P}_{1h}^{ext}(b_1, b)$ and $\mathfrak{P}_{rh}^{ext}(b_1, b)$. In fact as b varies along the boundary (with b_1 fixed) this switching occurs only at $b = b_1^{\leftrightarrow}$. We can verify using Property 3 that if $b_1 \in GHE$ ($b_1 \in EFG$) then the vector field defining l_e is outward pointing along the boundary segments $b_1 H b_1^{\leftrightarrow}$ ($b_1 F b_1^{\leftrightarrow}$) and EFG (GHE). Later we will see that the path chosen for performing the line integrals above will lie within the segments on which this vector field is outward pointing.

Vector Field for l_u . We need to be more careful with $l_u(b_1, b_2) = |u(b_1) - u(b_2)|$ because of the absolute value symbol in its definition. As an example of the kind of behavior we expect consider keeping b_1 fixed and varying b . If $|u(b) - u(b_1)|$ becomes zero then we will have to reverse the vector field we have been using because of the absolute value sign. This kind of reasoning leads us to assert that $\vec{u}_{b_1}(b) = \pm \vec{u}(b)$, where we choose the plus or minus sign at a point b depending on whether $\int_{b_1}^b \vec{u}(b) \cdot \vec{dr}$ is greater than or less than zero. This rule for the choice of sign will later determine that on the chosen path of integration $\vec{u}_{b_1}(b)$ is outward pointing on the first segment of the path and inward pointing on the second segment.

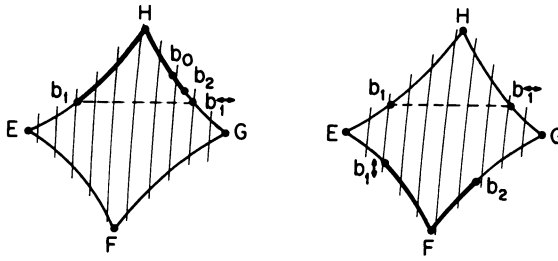


Fig. 8. The path of integration and its division into segments

Vector Field for l_i . $\vec{v}^i(b_1, \hat{b})$ is outward pointing for any two points b_1 and b since the region D is a geodesically convex.

Case 1. $b_1, b_2 \in GHE$ (or $b_1, b_2 \in EFG$) Property 3 implies that either $b_2 \in b_1 F b_1^{\leftarrow}$ or $b_1 \in b_2 F b_2^{\leftarrow}$ but not both. Choose b_1 as the base point if the first statement holds, b_2 otherwise. Say b_1 is the base point. Choose the path $\gamma_{12} \subset B$ from b_1 to b_2 to exclude b_1^{\leftarrow}, E and G . See Fig. 8a. With b_1 as the base point integrate from b_1 to b_2 along γ_{12} in Eqs. 13. Assume that $l_u(b_1, b)$ never becomes zero for any point b on γ_{12} . This will mean that $\vec{u}_{b_1}(b)$ is outward pointing on all of γ_{12} and hence that $\vec{v}^e(\hat{b}, b_1) \cdot \vec{dr} \leq \vec{u}(b) \cdot \vec{dr} \leq \vec{v}^i(b_1, \hat{b}) \cdot \vec{dr}$ for all points b , on γ_{12} . Since all of the length conditions are zero for $b = b_1$ the inequality on scalar products implies $l_e(b_1, b_2) \leq l_u(b_1, b_2) \leq l_i(b_1, b_2)$.

If however $l_u(b_1, b)$ does become zero at $b = b_0$ we cannot assert that \vec{u} is outward pointing on all of γ_{12} . We break γ_{12} into two segments. See Fig. 8a. On the first segment (b_1 to b_0), \vec{u} will be outward pointing and hence the inequality on scalar products will be valid. As a consequence we can assert that $l_e(b_1, b_0) \leq l_u(b_1, b_0) \leq l_i(b_1, b_0)$. Now we must compute the contribution from the segment $b_0 b_2$. On this segment $\vec{v}^e(\hat{b}, b_1) \cdot \vec{dr} \leq \vec{u}(b) \cdot \vec{dr}$ is valid as can be seen by checking the signs of these scalar products. Hence we see that $l_e(b_1, b_2) \leq l_u(b_1, b_2)$ will be true. Let us now prove that $l_u(b_1, b_2) \leq l_i(b_1, b_2)$. Let the length of $b_0 b_2$ in the metric ρ_m be μ . We see that $l_u(b_1, b_2) \leq \mu$ and $l_i(b_1, b_2) \geq l_i(b_1, b_0) - \mu$ since these arise from integrals (Eqs. 13) in which the maximum magnitude of the integrand is $\rho(b)$. This tells us that $l_u(b_1, b_2) \leq l_i(b_1, b_2)$ if $\mu \leq l_i(b_1, b_0) - \mu$; that is if $\mu \leq \frac{1}{2} l_i(b_1, b_0)$. Constraints 1 and 2 imply that the maximum angle that the straight line segments $b_1 b_0$ and $b_1 b_2$ can make with each other at the point b_1 is 2ϵ . This would lead us to expect that $\mu \approx 2\epsilon \cdot l_i(b_1, b_0)$. In fact we can use Constraints 3 and 4 to establish the rigorous bound $\mu \leq g(\epsilon) \cdot l_i(b_1, b_0)$, where $g(\epsilon) = \frac{1 + \epsilon \sin(2\epsilon)}{1 - \epsilon \sin(\pi/4 - 4\epsilon)}$. If we choose ϵ so that $g(\epsilon) < \frac{1}{2}$ then $\mu \leq \frac{1}{2} l_i(b_1, b_0)$ as was to be shown.

Case 2. $b_1 \in EHG$ and $b_2 \in EFG$ (or viceversa). Now we choose the path from $b_1 \uparrow$ to b_2 that does not pass through E or G . See Fig. 8b. Since $l_e(b_1, b_1 \uparrow) = l_u(b_1, b_1 \uparrow) = l_i(b_1, b_1 \uparrow)$ we see that the relative ordering of $l_e(b_1, b_2)$, $l_u(b_1, b_2)$ and $l_i(b_1, b_2)$ is the same as the relative ordering of the contributions from the path. An argument similar to that of Case 1 gives the desired result.

Acknowledgements. I am indebted to B. Zwiebach for suggesting this problem, for numerous discussions and for help in making this a more readable paper.

References

- [SoZ] Sonoda, H., Zwiebach, B.: Closed string field theory loops with symmetric factorizable quadratic differentials. Nucl. Phys. **B331**, 592 (1990)
- [Z1] Zwiebach, B.: How covariant closed string field theory solves a minimal area problem. Commun. Math. Phys. **136**, 83–118 (1991); Consistency of closed string polyhedra from minimal area. Phys. Lett. **B241**, 343 (1990)
- [St] Strebel, K.: Quadratic differentials. Berlin, Heidelberg, New York: Springer 1984
- [Z2] Zwiebach, B.: Quantum closed strings from minimal area. Mod. Phys. Lett. A, **5**, No. 32, 2753 (1990)
- [WZ] Wolf, M., Zwiebach, B.: Works in preparation
- [A] Ahlfors, L. V.: Conformal invariants. McGraw-Hill Series in Higher Mathematics. New York: McGraw-Hill 1973

Communicated by S.-T. Yau

