



Pietro Donatini · Patrizio Frosini

## Natural pseudodistances between closed surfaces

Received October 13, 2005 and in revised form October 28, 2005

**Abstract.** Let us consider two closed surfaces  $\mathcal{M}, \mathcal{N}$  of class  $C^1$  and two functions  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$ ,  $\psi : \mathcal{N} \rightarrow \mathbb{R}$  of class  $C^1$ , called measuring functions. The natural pseudodistance  $d$  between the pairs  $(\mathcal{M}, \varphi)$ ,  $(\mathcal{N}, \psi)$  is defined as the infimum of  $\Theta(f) := \max_{P \in \mathcal{M}} |\varphi(P) - \psi(f(P))|$  as  $f$  varies in the set of all homeomorphisms from  $\mathcal{M}$  onto  $\mathcal{N}$ . In this paper we prove that the natural pseudodistance equals either  $|c_1 - c_2|$ ,  $\frac{1}{2}|c_1 - c_2|$ , or  $\frac{1}{3}|c_1 - c_2|$ , where  $c_1$  and  $c_2$  are two suitable critical values of the measuring functions. This shows that a previous relation between the natural pseudodistance and critical values obtained in general dimension can be improved in the case of closed surfaces. Our result is based on a theorem by Jost and Schoen concerning harmonic maps between surfaces.

**Keywords.** Natural pseudodistance, measuring function, harmonic map

---

### Introduction

The *natural pseudodistance* is a new variational approach to the comparison of manifolds endowed with real-valued functions defined on them. In [2] we proved a result about the values that such a pseudodistance  $\delta$  can take in general dimension. In this work we focus on the 2-dimensional case, showing that the previous result can be improved in the case of closed surfaces. Assuming that two homeomorphic closed manifolds  $\mathcal{M}$  and  $\mathcal{N}$  of class  $C^1$  are given together with two functions  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$ ,  $\psi : \mathcal{N} \rightarrow \mathbb{R}$  of class  $C^1$  (called *measuring functions*), we consider the value

$$\delta((\mathcal{M}, \varphi), (\mathcal{N}, \psi)) := \inf_{f \in H(\mathcal{M}, \mathcal{N})} \max_{P \in \mathcal{M}} |\varphi(P) - \psi(f(P))|,$$

where  $H(\mathcal{M}, \mathcal{N})$  denotes the set of all homeomorphisms from  $\mathcal{M}$  onto  $\mathcal{N}$ . The number  $d = \delta((\mathcal{M}, \varphi), (\mathcal{N}, \psi))$  is called the *natural pseudodistance* between the pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$  (called *size pairs*).

The closeness of  $d$  to zero means that there are homeomorphisms for which the difference between the values taken by the measuring functions at corresponding points is

---

P. Donatini: Dipartimento di Matematica, Università di Bologna, Piazza Porta San Donato, 5, I-40126 Bologna, Italy; e-mail: donatini@dm.unibo.it

P. Frosini: Advanced Research Centre for Electronic Systems “Ercole De Castro”, Università di Bologna, via Toffano, 2/2, I-40135 Bologna, Italy; e-mail: frosini@dm.unibo.it

*Mathematics Subject Classification (2000):* Primary 58E20, 53A05; Secondary 58C05, 49Q10

arbitrarily small. On the other hand, if the infimum is large, we find that every homeomorphism between the manifolds considered must change the values taken by our measuring function considerably.

In [2] we proved (Theorem 6.2) that a suitable multiple of  $d$  by a positive integer  $k$  coincides with the distance between two critical values of the functions  $\varphi, \psi$ . It is interesting to observe that in every known example, the minimum possible value for  $k$  is 1 or 2. In this paper we shall show that, in the 2-dimensional case, the minimum value for the integer  $k$  is either 1, 2, or 3. We remark that an analogous statement has also been proved in [4] for curves, using different techniques, but in that case we are able to prove that only the values 1 and 2 are possible.

Besides its intrinsic interest from a purely mathematical point of view, the natural pseudodistance between closed surfaces associated with measuring functions can also be used for shape comparison purposes, together with the “twin” and strictly related concept of *size function*. For more theoretical details and examples of practical applications we refer to [1, 13, 17–20].

In Section 1 we sketch the main ideas of this paper. In Section 2 we give the main definitions and some examples, while in Section 3 further examples are presented, highlighting some characteristic phenomena. In Section 4 the concepts of train and minimal  $d$ -approximating sequence are illustrated, together with some related results. In Section 5 we prove our main result (Theorem 5.7) about the natural pseudodistance between closed surfaces endowed with measuring functions. In Section 6 open problems and further research are briefly described.

## 1. The point of this paper

As reported in the previous section, it was proved in [2] that the natural pseudodistance between size pairs always equals  $|c_1 - c_2|/k$ , where  $c_1, c_2$  are two suitable critical values of the measuring functions and  $k$  is an appropriate integer number. The minimum possible value for  $k$  is called the *analytic folding number*.

It is interesting to observe that in every known example, the analytic folding number is 1 or 2.

Two questions naturally arise: Are there examples with an analytic folding number strictly greater than 2? Is this question related to the dimension of our manifolds?

In this paper we take a first step towards answering these questions.

It is important to observe right now that the attempt to minimize the change  $\Theta(f) := \max_{P \in \mathcal{M}} |\varphi(P) - \psi(f(P))|$  in the measuring functions under the action of  $f$  does not, in general, lead to a homeomorphism, as we are going to show in the next section. Degeneracies can arise, and hence we cannot confine ourselves to studying a single optimal homeomorphism. Instead of a single homeomorphism, approximating sequences of homeomorphisms must be considered. In some sense, “optimal” approximating sequences  $(f_i)$  of homeomorphisms exist, converging to relations that represent the best way to map one manifold to another with respect to the change in the measuring functions. The study of these relations leads us to the concept of a “train of limit  $d$ -jumps”, describing some

degeneracies corresponding to the sequence  $(f_i)$ . As we are going to see in the following sections, the properties of these structures imply the properties of the analytic folding number.

How can we study these properties?

In [2] local deformations were used, based on the flow diffeomorphism of the gradient of the measuring functions but, unfortunately, this approach does not seem to be sufficient to answer the questions we posed. The main idea of this paper is to use the theory of harmonic maps to confront the bidimensional case. A result by Jost and Schoen allows us to study the case of surfaces.

We shall proceed this way. We shall consider each “optimal” sequence  $(f_i)$  of homeomorphisms between the manifolds  $\mathcal{M}$  and  $\mathcal{N}$  we are examining, where optimal means that  $\inf_i \max_{P \in \mathcal{M}} |\varphi(P) - \psi(f_i(P))|$  equals  $\delta((\mathcal{M}, \varphi), (\mathcal{N}, \psi))$ . Then we shall describe the degeneracies related to  $(f_i)$  using a train of limit  $d$ -jumps, and assume that the degeneracies of  $(f_i)$  are minimal with respect to a suitable order  $\preceq$  we are going to define.

Finally, we shall apply a local harmonization procedure to each  $f_i$  far away from the critical points, using Jost and Schoen’s theorem. The key remark will be that the change we are going to apply produces a new sequence that is “smaller” than  $(f_i)$  with respect to  $\preceq$ . Since  $(f_i)$  will already be minimal, some further information about the length of the trains of  $d$ -jumps for  $(f_i)$  will be derived, implying our main result.

Some technicalities will be necessary in order to use our ideas in practice, but the key point is simply the possibility (in some sense unexpected) of reducing the change of the measuring functions by locally decreasing the energy of the transformations we use between our manifolds. The following sections will formalize the ideas we have just described.

## 2. The natural pseudodistance

### 2.1. The main definition

The definition of natural pseudodistance can be introduced for  $n$ -dimensional manifolds. Let us consider the set  $\text{Size}_n$  of all pairs  $(\mathcal{M}, \varphi)$ , where  $\mathcal{M}$  is a closed  $n$ -manifold of class  $C^k$  and  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$  is a function of class  $C^k$ . We shall call  $(\mathcal{M}, \varphi)$  an  $(n$ -dimensional) *size pair* of class  $C^k$  and  $\varphi$  a *measuring function*.

Assume  $(\mathcal{M}, \varphi), (\mathcal{N}, \psi)$  are two size pairs.  $H(\mathcal{M}, \mathcal{N})$  will denote the set of all homeomorphisms from  $\mathcal{M}$  to  $\mathcal{N}$ .

**Definition 2.1.** If  $H(\mathcal{M}, \mathcal{N}) \neq \emptyset$ , the function  $\Theta : H(\mathcal{M}, \mathcal{N}) \rightarrow \mathbb{R}$  given by

$$\Theta(f) = \max_{P \in \mathcal{M}} |\varphi(P) - \psi(f(P))|$$

is called the natural size measure with respect to the measuring functions  $\varphi$  and  $\psi$ .

In other words,  $\Theta$  measures how much  $f$  changes the values taken by the measuring functions, at corresponding points.

**Definition 2.2.** *The natural size pseudodistance is the pseudodistance  $\delta : \text{Size}_n \times \text{Size}_n \rightarrow \mathbb{R} \cup \{+\infty\}$  defined by*

$$\delta((\mathcal{M}, \varphi), (\mathcal{N}, \psi)) = \begin{cases} \inf_{f \in H(\mathcal{M}, \mathcal{N})} \Theta(f) & \text{if } H(\mathcal{M}, \mathcal{N}) \neq \emptyset, \\ +\infty & \text{otherwise.} \end{cases}$$

In the following, the symbol  $d$  will denote the value of the natural pseudodistance  $\delta$  computed between the pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$  that we are considering. As previously explained, this pseudodistance gives a method for comparing two manifolds with respect to the measuring functions chosen.

We point out that  $\delta$  is not a distance, since two size pairs can have a vanishing pseudodistance without being equal. On the other hand, the symmetry property and the triangle inequality can be trivially proved.

**Remark 2.3.** The word “size” in our definitions is due to the link between the pseudodistance  $\delta$ , size functions and size homotopy groups (cf. [7, 15]). However, for the sake of simplicity, we shall often drop the word “size” in the expressions “natural size measure” and “natural size pseudodistance”. The term “natural” is used in order to distinguish the pseudodistance studied here from some pseudodistances we can define between submanifolds of the Euclidean space (cf. [6]) and from other pseudodistances between manifolds paired with measuring functions.

In spite of the considerable difficulty in computing natural size pseudodistances, the following result holds for the general dimension  $n$  (cf. [2]):

**Theorem 2.4.** *Assume that  $\mathcal{M}$  and  $\mathcal{N}$  are two homeomorphic closed manifolds of class  $C^1$  and that  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$  and  $\psi : \mathcal{N} \rightarrow \mathbb{R}$  are two functions of class  $C^1$ . Then, if  $d$  denotes the natural pseudodistance between the size pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$ , there exists a positive integer  $k$  for which one of the following properties holds:*

- (i)  $k$  is odd and  $kd$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ ;
- (ii)  $k$  is even and  $kd$  equals either the distance between two critical values of  $\varphi$  or the distance between two critical values of  $\psi$ .

The smallest positive integer  $k$  for which either (i) or (ii) of Theorem 2.4 holds is called the *analytic folding number* for the pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$ . It is interesting to observe that in every known example, the analytic folding number is 1 or 2.

In this paper we shall prove that in the case of two homeomorphic closed surfaces of class  $C^1$ , endowed with  $C^1$  measuring functions, the analytic folding number always equals either 1, 2 or 3. This fact, besides showing a particular property of the 2-dimensional case, simplifies a direct computation of natural pseudodistances for closed surfaces.

However, the hypothesis  $n = 2$  will not be used until Section 5.

In Section 3, we shall show that the infimum of  $\Theta(f)$  for  $f$  varying in  $H(\mathcal{M}, \mathcal{N})$  is not always attained. When it is, we shall say that each homeomorphism  $f \in H(\mathcal{M}, \mathcal{N})$  with  $d = \Theta(f)$  is an *optimal homeomorphism*.

In the case where an optimal homeomorphism exists, the following result holds (Theorem 6.3 in [2]).

**Theorem 2.5.** *Assume that  $\mathcal{M}$  and  $\mathcal{N}$  are two  $C^1$  closed homeomorphic manifolds and that  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$  and  $\psi : \mathcal{N} \rightarrow \mathbb{R}$  are of class  $C^1$ . If there exists an optimal homeomorphism  $f \in H(\mathcal{M}, \mathcal{N})$  between the size pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$ , then the natural pseudodistance  $d = \delta((\mathcal{M}, \varphi), (\mathcal{N}, \psi))$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ .*

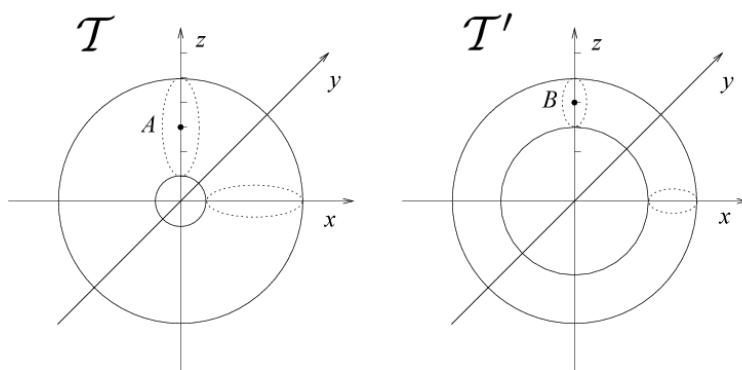
N.B.: For simplicity, throughout this paper we shall use the expression “closed surface” to mean a closed 2-manifold (we shall *not* require this manifold to be connected).

In order to simplify our notations, we shall assume that the manifolds  $\mathcal{M}$  and  $\mathcal{N}$  are disjoint, and that the corresponding measuring functions are obtained by restricting a function  $\omega : \mathcal{M} \cup \mathcal{N} \rightarrow \mathbb{R}$ , so that  $\varphi = \omega|_{\mathcal{M}}$  and  $\psi = \omega|_{\mathcal{N}}$ . In this way we can use just one symbol to denote both measuring functions. These hypotheses are not restrictive, since we can always replace the size pair  $(\mathcal{N}, \psi)$  with a new size pair  $(\hat{\mathcal{N}}, \hat{\psi})$ , having vanishing pseudodistance from the previous one and such that  $\mathcal{M} \cap \hat{\mathcal{N}} = \emptyset$ . Sometimes, when no confusion can arise, we shall use the symbol  $\omega$  to denote both  $\omega|_{\mathcal{M}}$  and  $\omega|_{\mathcal{N}}$ .

Moreover, it is easy to prove that, for every 2-dimensional size pair  $(\mathcal{M}, \omega)$  of class  $C^k$ , there exists an integer  $m$  and an embedding  $g : \mathcal{M} \rightarrow \mathbb{R}^m$  of class  $C^k$  such that  $x_m(P) = \omega(g^{-1}(P))$  for each point  $P \in g(\mathcal{M})$ . If  $\omega$  is Morse (i.e., smooth and having invertible Hessian at each critical point), we can assume that  $x_m$  is also Morse on  $g(\mathcal{M})$ . In other words, there is no loss of generality in assuming that the measuring functions associated with the closed surfaces  $\mathcal{M}, \mathcal{N}$  in question are obtained by restricting the  $x_m$ -coordinate in  $\mathbb{R}^m$ . Sometimes, when no confusion can arise, we shall use the symbol  $x_m$  to denote both  $x_m|_{\mathcal{M}}$  and  $x_m|_{\mathcal{N}}$  and use the expression “height of a point”. For the sake of clarity, in our examples and figures we shall often assume that our measuring function is the  $z$ -coordinate in  $\mathbb{R}^3$ .

**Example 2.6.** In  $\mathbb{R}^3$  consider the unit sphere  $\mathcal{S}$  of equation  $x^2 + y^2 + z^2 = 1$  and the ellipsoid  $\mathcal{E}$  of equation  $x^2 + 4y^2 + 9z^2 = 1$ . On  $\mathcal{S}$  and  $\mathcal{E}$  consider respectively the measuring functions  $\varphi$  and  $\psi$  that assign to every point of  $\mathcal{S}$  and  $\mathcal{E}$  the Gaussian curvature of the relevant manifold at that point. We have  $\delta((\mathcal{S}, \varphi), (\mathcal{E}, \psi)) = 35$ . In fact  $\varphi(\mathcal{S}) = \{1\}$ , while  $\psi(\mathcal{E}) = [4/9, 36]$ , and therefore  $\Theta(f) = 35$  for every  $f \in H(\mathcal{S}, \mathcal{E})$ .

**Example 2.7.** Consider the two tori  $\mathcal{T}, \mathcal{T}' \subset \mathbb{R}^3$  generated by the rotation around the  $y$ -axis of the circles lying in the  $yz$ -plane, with centres  $A = (0, 0, 3)$  and  $B = (0, 0, 4)$ , and radii 2 and 1, respectively (see Figure 1). As a measuring function  $\varphi$  (resp.  $\varphi'$ ) on  $\mathcal{T}$  (resp. on  $\mathcal{T}'$ ) we take the restriction to  $\mathcal{T}$  (resp. to  $\mathcal{T}'$ ) of the function  $\zeta : \mathbb{R}^3 \rightarrow \mathbb{R}$ ,  $\zeta(x, y, z) = z$ . We point out that, for both  $\mathcal{T}$  and  $\mathcal{T}'$ , the image of the measuring function is the closed interval  $[-5, 5]$ . We can easily prove that the natural size pseudodistance between  $(\mathcal{T}, \varphi)$  and  $(\mathcal{T}', \varphi')$  is 2 (for the proof involving size homotopy groups see [15]).



**Fig. 1.** In this case an optimal homeomorphism (i.e. one minimizing  $\Theta$ ) exists and  $d = 2$ ;  $d$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\varphi'$ .

Moreover, the homeomorphism  $f$ , taking each point of  $\mathcal{T}$  to the point having the same toroidal coordinates in  $\mathcal{T}'$ , has natural size measure  $\Theta(f) = 2$ .

In general,  $d$  is far from being easily computable as in the previous Examples 2.6 and 2.7. In Example 2.6, for every homeomorphism  $f \in H(\mathcal{S}, \mathcal{E})$ ,  $\Theta(f)$  equals the Hausdorff distance  $\delta_H(\varphi(\mathcal{S}), \psi(\mathcal{E}))$  between the sets  $\varphi(\mathcal{S})$  and  $\psi(\mathcal{E})$  in  $\mathbb{R}$ . Now it is clear that the natural size pseudodistance  $\delta((\mathcal{M}, \varphi), (\mathcal{N}, \psi))$  is always greater than or equal to  $\delta_H(\varphi(\mathcal{M}), \psi(\mathcal{N}))$  and therefore  $\Theta(f)$  must be the natural size pseudodistance we want to compute. We also point out that, in Example 2.6, the images of  $\varphi$  and  $\psi$  are different sets and so the natural size pseudodistance is trivially positive.

In Example 2.7 the natural size pseudodistance is strictly greater than the (vanishing) Hausdorff distance between the images of the two measuring functions.

Computing natural size pseudodistances is usually difficult. For this reason the concepts of *size function* and *size homotopy group* have been developed, making it easier to compute the value  $d$ , using some lower-bound theorems. Anyway, here we cannot illustrate these closely related concepts, and we refer to [6, 7, 13, 15] for more details.

### 3. Some interesting examples about curves and surfaces

For the sake of clarity, even if this paper focuses on the bidimensional case, we shall begin our formal treatment from 1-dimensional examples.

**Example 3.1.** The first example we give is shown in Figure 2. Here  $\mathcal{M}$  and  $\mathcal{N}$  are smooth closed curves in  $\mathbb{R}^3$ , embedded in the  $xz$ -plane. It is clear that the natural pseudodistance  $d$  between the size pairs  $(\mathcal{M}, z)$  and  $(\mathcal{N}, z)$  equals  $z(B) - z(A)$ , that is, the distance between a critical value of  $z|_{\mathcal{M}}$  and a critical value of  $z|_{\mathcal{N}}$ .

In this example no optimal homeomorphism exists, since it would have to map both the maximum points for  $z|_{\mathcal{M}}$  to  $A$ , contrary to injectivity.

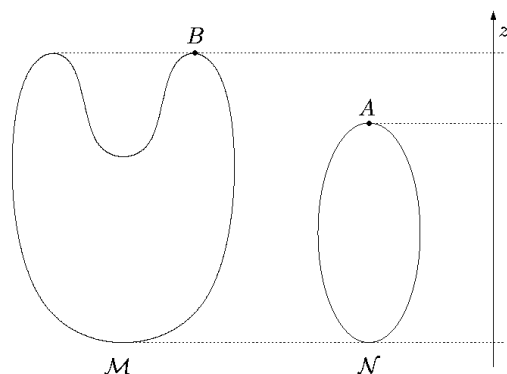


Fig. 2. The natural pseudodistance between the size pairs  $(\mathcal{M}, z)$  and  $(\mathcal{N}, z)$  is  $z(B) - z(A)$ .

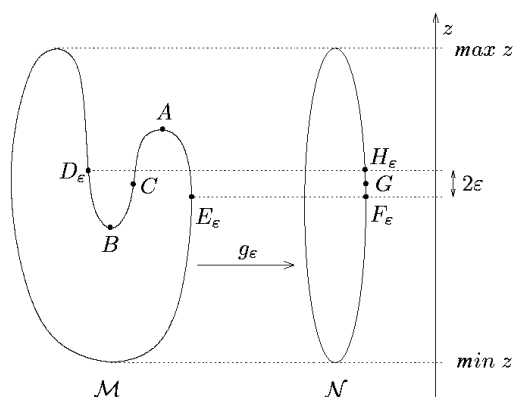


Fig. 3. Construction of the homeomorphism  $g_\epsilon$  for which  $\Theta(g_\epsilon) \leq d + \epsilon$ .

**Example 3.2.** Let us consider the smooth closed curves  $\mathcal{M}$  and  $\mathcal{N}$  in Figure 3. The points  $A$  and  $B$  are critical points of the function  $z$  and  $z(C) = \frac{1}{2}(z(A) + z(B)) = z(G)$ . We want to prove that the natural pseudodistance between the size pairs  $(\mathcal{M}, z)$  and  $(\mathcal{N}, z)$  takes the value

$$d = \frac{1}{2}(z(A) - z(B))$$

and that no optimal homeomorphism exists. In order to do that we shall construct a sequence  $(f_i)$  of homeomorphisms for which  $\lim_i \Theta(f_i) = \frac{1}{2}(z(A) - z(B))$ , and show that  $\Theta(f) > \frac{1}{2}(z(A) - z(B))$  for every homeomorphism  $f \in H(\mathcal{M}, \mathcal{N})$ .

Let us start by proving that, for every  $\epsilon > 0$ , there exists a homeomorphism  $g_\epsilon : \mathcal{M} \rightarrow \mathcal{N}$  such that  $\Theta(g_\epsilon) \leq \frac{1}{2}(z(A) - z(B)) + 2\epsilon$ . Consider the points  $D_\epsilon, E_\epsilon, H_\epsilon$  and  $F_\epsilon$  in Figure 3, satisfying  $z(D_\epsilon) = z(H_\epsilon) = z(C) + \epsilon$  and  $z(E_\epsilon) = z(F_\epsilon) = z(C) - \epsilon$ . We choose a homeomorphism  $g_\epsilon$ , taking the arc  $D_\epsilon C E_\epsilon$  to the arc  $H_\epsilon G F_\epsilon$  in such a way that

$g_\varepsilon(D_\varepsilon) = H_\varepsilon$  and  $g_\varepsilon(E_\varepsilon) = F_\varepsilon$ . Outside the arc  $D_\varepsilon C E_\varepsilon$  in  $\mathcal{M}$  we define  $g_\varepsilon$  by mapping every point  $P$  to a point  $g_\varepsilon(P)$  satisfying  $z(P) = z(g_\varepsilon(P))$ .

For every  $i \in \mathbb{N} - \{0\}$  we set  $f_i = g_{1/i}$ . It is easy to prove that

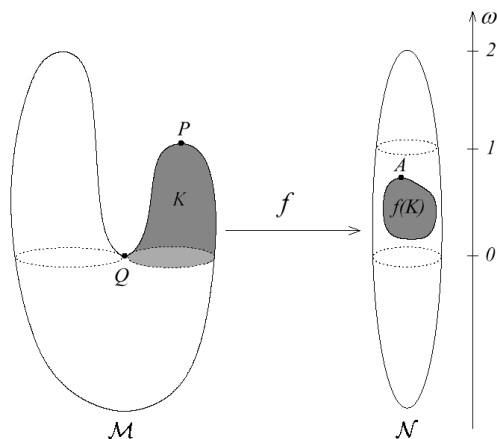
$$\lim_i \Theta(f_i) = \frac{1}{2}(z(A) - z(B)).$$

Now we only have to verify that  $\Theta(f) \leq \frac{1}{2}(z(A) - z(B))$  for no homeomorphism between  $\mathcal{M}$  and  $\mathcal{N}$ . If such a homeomorphism existed, for every  $P \in \mathcal{M}$  we would have

$$|z(P) - z(f(P))| \leq \frac{z(A) - z(B)}{2}$$

and hence  $z(f(A)) \geq z(G) \geq z(f(B))$ . Therefore we could easily find points  $P \in \mathcal{M}$  for which  $|z(P) - z(f(P))| > \frac{1}{2}(z(A) - z(B))$ , contradicting our assumption.

**Example 3.3.** Consider the size pairs  $(\mathcal{M}, \omega)$  and  $(\mathcal{N}, \omega)$  in Figure 4, where  $\mathcal{M}$  and  $\mathcal{N}$  are smooth surfaces embedded into  $\mathbb{R}^3$ . We want to prove that the natural pseudodistance between these size pairs is  $1/2$ .



**Fig. 4.** The natural pseudodistance between these size pairs is  $d = 1/2$ .

The critical points  $P, Q \in \mathcal{M}$  for which  $\omega(P) = 1$  and  $\omega(Q) = 0$  belong to the displayed closed set  $K \subset \omega^{-1}([0, 1])$ . First of all, we shall prove that  $d \geq 1/2$ , by showing that

$$\Theta(f) > \frac{1}{2}(\omega(P) - \omega(Q)) = \frac{1}{2}$$

for every homeomorphism  $f : \mathcal{M} \rightarrow \mathcal{N}$ . Suppose  $f(K)$  contains no point of  $\mathcal{N}$  that is critical for  $\omega$  (otherwise  $\Theta(f)$  would be at least 1 and our inequality would be satisfied). Let  $A$  be the point of  $f(K)$  at which the measuring function  $\omega|_{f(K)}$  attains its maximum. Since  $A$  belongs to the boundary of  $f(K)$ , we must have  $\omega(f^{-1}(A)) = 0$ , and as  $P$  is



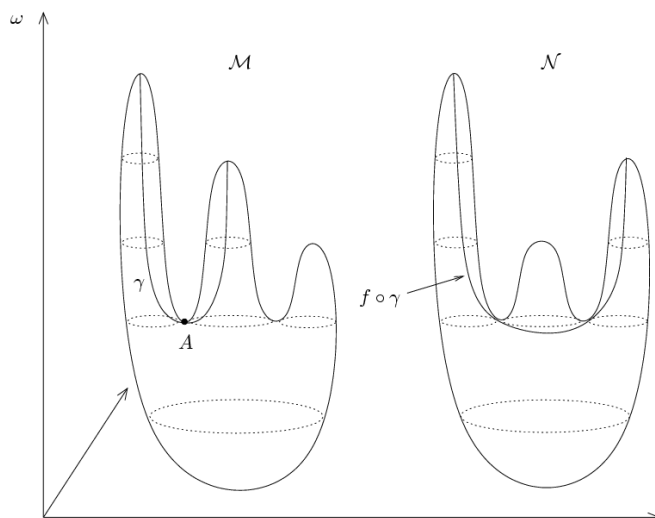
internal to  $K$ ,  $\omega(f(P)) < \omega(A)$ . Consequently,  $\Theta(f) \geq \omega(A) > \omega(f(P))$  and hence  $\Theta(f) \geq \omega(P) - \omega(f(P)) > \omega(P) - \Theta(f)$ . It follows that  $\Theta(f) > \omega(P)/2 = 1/2$ .

In order to complete our proof that the natural pseudodistance is really  $1/2$ , we still have to give a suitable sequence  $(f_i)$  of homeomorphisms such that

$$\lim_i \Theta(f_i) = 1/2.$$

Since the construction of such a sequence is conceptually similar to the one we gave for the previous example about curves, we skip its analytic expression.

**Example 3.4.** Consider the smooth surfaces  $\mathcal{M}$  and  $\mathcal{N}$  displayed in Figure 5 and the corresponding measuring function  $\omega$ . The dotted lines are level curves for the measuring function  $\omega$ .



**Fig. 5.** An example of vanishing natural pseudodistance.

**Property 1.** *The natural pseudodistance between the two size pairs is zero.*

It is easy to see that we can isotopically deform the left surface to the right one by “torsion”, exchanging the positions of the two smallest humps. This deformation can be performed by an arbitrarily small change in the values of the height  $\omega$ . Therefore, we can construct a sequence of homeomorphisms  $(f_i)$  from  $\mathcal{M}$  to  $\mathcal{N}$  such that  $\Theta(f_i) \rightarrow 0$ .

**Property 2.** *No optimal homeomorphism exists between the two size pairs.*

Suppose there exists a homeomorphism  $f$  such that  $\Theta(f) = 0$ . Consider a path  $\gamma$  as in Figure 5, chosen in such a way that, in the image of the path,  $\omega(P) = \omega(A)$  for no point  $P$  different from  $A$ . We can easily verify that the image of the path  $f \circ \gamma$  must contain more than one point at which  $\omega$  takes the value  $\omega(A)$ . This contradicts our assumptions, since  $\Theta(f) = 0$  implies  $\omega(f(P)) = \omega(P)$  for every  $P$  in the image of  $\gamma$ .

#### 4. Some technical tools and definitions

##### 4.1. The concept of “train of limit $d$ -jumps”

In order to prove our main theorem, we need some new definitions and technical results. Assume two size pairs  $(\mathcal{M}, \omega)$ ,  $(\mathcal{N}, \omega)$  are given.

The symbol  $S_H(\mathcal{M}, \mathcal{N})$  will denote the set of all sequences  $(f_i)$  of homeomorphisms in  $H(\mathcal{M}, \mathcal{N})$  such that  $\Theta(f_i) \rightarrow d$ . Every sequence in  $S_H(\mathcal{M}, \mathcal{N})$  will be called a  $d$ -approximating sequence from  $(\mathcal{M}, \omega)$  to  $(\mathcal{N}, \omega)$ .

Let  $(f_i) \in S_H(\mathcal{M}, \mathcal{N})$ . We shall say that a pair of points  $(P, Q) \in \mathcal{M} \times \mathcal{N}$  is in relation with respect to  $(f_i)$  if there exists a sequence  $(P_r)$  in  $\mathcal{M}$  and a strictly increasing sequence  $(i_r)$  in  $\mathbb{N}$  such that

$$(P, Q) = \lim_r (P_r, f_{i_r}(P_r)).$$

In this case we shall write either  $P \rho Q$  or  $Q \rho P$ .

In the remaining part of this section we assume that  $0 < d < +\infty$ . The following compact sets are defined for each  $d$ -approximating sequence  $(f_i)$ :

$$\begin{aligned} \mathbf{N}_{\mathcal{M}}^+ &= \mathbf{N}_{\mathcal{M}}^+((f_i)) = \{P \in \mathcal{M} \mid \exists Q \in \mathcal{N} : P \rho Q, \omega(Q) - \omega(P) = d\}, \\ \mathbf{N}_{\mathcal{M}}^- &= \mathbf{N}_{\mathcal{M}}^-((f_i)) = \{P \in \mathcal{M} \mid \exists Q \in \mathcal{N} : P \rho Q, \omega(P) - \omega(Q) = d\}, \\ \mathbf{N}_{\mathcal{N}}^+ &= \mathbf{N}_{\mathcal{N}}^+((f_i)) = \{Q \in \mathcal{N} \mid \exists P \in \mathcal{M} : P \rho Q, \omega(P) - \omega(Q) = d\}, \\ \mathbf{N}_{\mathcal{N}}^- &= \mathbf{N}_{\mathcal{N}}^-((f_i)) = \{Q \in \mathcal{N} \mid \exists P \in \mathcal{M} : P \rho Q, \omega(Q) - \omega(P) = d\}. \end{aligned}$$

In other words, the points  $P$  in  $\mathbf{N}_{\mathcal{M}}^+$  are those for which there exists a point  $Q \in \mathcal{N}$  such that the pair  $(P, Q)$  can be approximated arbitrarily well by a pair  $(P_r, f_{i_r}(P_r))$  whose “jump”  $\omega(f_{i_r}(P_r)) - \omega(P_r)$  is arbitrarily close to  $d$ . Hence, if we think of  $\omega$  as a “height” function (cf. the examples in the previous section), the points  $P_r$  have images with height approximated by  $\omega(P_r) + d$ . In  $\mathbf{N}_{\mathcal{M}}^+$ , the symbol  $\mathcal{M}$  recalls the manifold to which  $P$  belongs, while the  $+$  indicates that, by mapping  $P$  to  $Q$ , we increase the value of the measuring function, i.e. the “jump” starting from the node in  $\mathcal{M}$  is “upwards”. The notations for the other three sets are quite analogous. The symbol  $-$  indicates nodes from which “downward jumps” start (the starting node belonging to the manifold shown as subscript).

It is clear that, for every  $P \in \mathbf{N}_{\mathcal{M}}^+$ , there exists  $Q \in \mathbf{N}_{\mathcal{N}}^-$  such that  $P \rho Q$  (and vice versa), and that an analogous relation holds for the sets  $\mathbf{N}_{\mathcal{M}}^-$  and  $\mathbf{N}_{\mathcal{N}}^+$ . For every sequence of homeomorphisms in  $S_H(\mathcal{M}, \mathcal{N})$  the sets  $\mathbf{N}_{\mathcal{M}} = \mathbf{N}_{\mathcal{M}}^+ \cup \mathbf{N}_{\mathcal{M}}^-$  and  $\mathbf{N}_{\mathcal{N}} = \mathbf{N}_{\mathcal{N}}^+ \cup \mathbf{N}_{\mathcal{N}}^-$  are non-empty because of the compactness of the manifolds.

Now we define the concept of “train” for a  $d$ -approximating sequence:

**Definition 4.1.** Let  $(N_0, N_1, \dots, N_k)$  be an ordered  $(k + 1)$ -tuple of points in  $\mathcal{M} \cup \mathcal{N}$  with  $k \geq 1$  such that, for  $j = 0, \dots, k - 1$  the following properties hold:

- (a)  $\omega(N_{j+1}) = \omega(N_j) + d$ ;
- (b)  $N_j \rho N_{j+1}$ .

In this case the ordered set  $(N_0, N_1, \dots, N_k)$  will be called a train of limit  $d$ -jumps for the sequence  $(f_i)$  (or, for short, a train) and its points will be called nodes. The pairs  $(N_j, N_{j+1})$  will be known as the wagons of the train. The number  $k$  will be called the length of the train and each train that is not included (in the obvious sense) in any other train will be said to be maximal. If  $(N_0, \dots, N_k)$  is a maximal train, its wagons  $(N_0, N_1)$  and  $(N_{k-1}, N_k)$  will be called initial and final train wagons (respectively), while  $N_0$  and  $N_k$  will be the initial and final train nodes. The remaining nodes will be called internal nodes. The symbol  $W((f_i))$  will denote the set of all the train wagons (for all the existing trains).

Since each point belonging either to  $\mathbf{N}_{\mathcal{M}}$  or to  $\mathbf{N}_{\mathcal{N}}$  is a node for at least one train, the set of all trains is not empty. Notice that the point  $P$  is an initial node for at least one maximal train if and only if either  $P \in \mathbf{N}_{\mathcal{M}}^+ - \mathbf{N}_{\mathcal{M}}^-$  or  $P \in \mathbf{N}_{\mathcal{N}}^+ - \mathbf{N}_{\mathcal{N}}^-$ , whereas it is a final node if and only if either  $P \in \mathbf{N}_{\mathcal{M}}^- - \mathbf{N}_{\mathcal{M}}^+$  or  $P \in \mathbf{N}_{\mathcal{N}}^- - \mathbf{N}_{\mathcal{N}}^+$ .

In Figure 6 we provide a graphic representation of a maximal train  $(A, B, C, D)$ . In this particular case, we have  $A \in \mathbf{N}_{\mathcal{N}}^+$ ,  $B \in \mathbf{N}_{\mathcal{M}}^+ \cap \mathbf{N}_{\mathcal{M}}^-$ ,  $C \in \mathbf{N}_{\mathcal{N}}^+ \cap \mathbf{N}_{\mathcal{N}}^-$  and  $D \in \mathbf{N}_{\mathcal{M}}^-$ . Hence,  $A$  is the initial node and  $D$  is the final train node, while  $B$  and  $C$  are internal nodes. The three ordered pairs  $(A, B)$ ,  $(B, C)$ ,  $(C, D)$  are the three wagons in the train;  $(A, B)$  and  $(C, D)$  are its initial and final wagons, respectively.

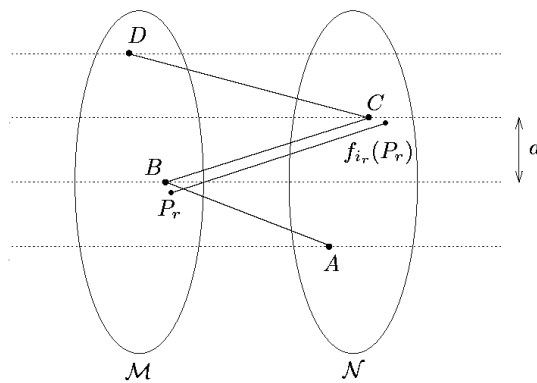


Fig. 6. A train of limit  $d$ -jumps given by the quadruple  $(A, B, C, D)$ .

In Figure 7 we can find the maximal train  $(B, G, A)$  associated with the  $d$ -approximating sequence we described in Example 3.2. In fact, we can easily prove that  $B\rho G, G\rho A$ ,  $z(G) - z(B) = d$  and  $z(A) - z(G) = d$ . Hence  $B \in \mathbf{N}_{\mathcal{M}}^+$ ,  $G \in \mathbf{N}_{\mathcal{N}}^+ \cap \mathbf{N}_{\mathcal{N}}^-$  and  $A \in \mathbf{N}_{\mathcal{M}}^-$ .

**Remark 4.2.** The example described in Figure 7 shows that the existence of a train of length 2 such that its initial node (in this case  $B$ ) and its final node (in this case  $A$ ) are critical points of the measuring function  $z$  guarantees that the natural pseudodistance  $d$  equals half the distance between two critical values of the measuring function.

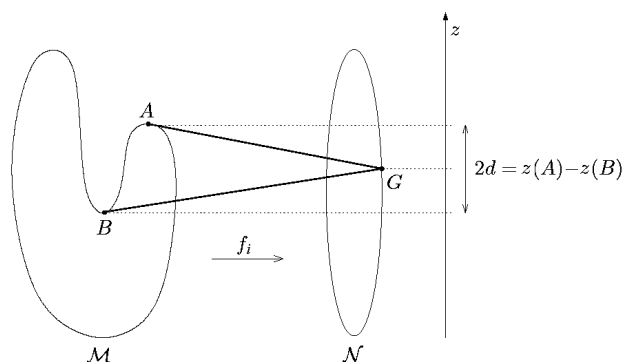


Fig. 7. An example of a train of limit  $d$ -jumps given by the triple  $(B, G, A)$ .

Our main goal will be to show that in the case of closed surfaces it is always possible to construct a sequence of  $d$ -approximating homeomorphisms for which we can demonstrate the existence of a train of length 1, 2 or 3, beginning and ending at critical heights for the measuring functions. We shall do that in the next subsection, 4.2, and in Section 5. The example we have just seen justifies our task, since it points out a simple relation between  $d$  and the critical values of  $z$ .

Now, in order to attain our goal, we need to introduce the concept of *minimal  $d$ -approximating sequence*.

#### 4.2. Minimal $d$ -approximating sequences

The concept of train that we have just introduced allows us to prove Theorem 2.4 cited in Section 2, and will be central in the following sections, devoted to the proof of the main result in this paper (Theorem 5.7). In this subsection we shall assume that  $\mathcal{M}$  and  $\mathcal{N}$  are smooth homeomorphic closed manifolds and  $\varphi$  and  $\psi$  are Morse measuring functions on  $\mathcal{M}$  and  $\mathcal{N}$ , respectively. We shall weaken these hypotheses at the end of this paper.

As explained in the introduction, the main goal of this paper is to show that the analytic folding number is either 1, 2 or 3 in the case of closed surfaces.

The idea is to extend the reasoning applied in Remark 4.2, about the example described in Figure 7. In order to do that, from a constructive point of view we need to take a  $d$ -approximating sequence and improve it by shortening its trains as much as possible, until we get a train of length 1, 2 or 3, beginning and ending at critical heights for the measuring functions.

This procedure will be carried out in two steps. The first will consist in a reduction of trains applicable in any dimension, which has been developed and applied in [2] (Lemma 4.6 in this paper) in order that only trains beginning and ending at critical points for the measuring functions remain.

The second step will be a reduction process, specifically developed for the case of surfaces, allowing us to get a further shortening of trains.

Our goal requires a formal definition of “improving” a  $d$ -approximating sequence.

Hence we need to define the following preordering  $\preceq$  on the set  $S_H(\mathcal{M}, \mathcal{N})$  of  $d$ -approximating sequences.

**Definition 4.3.** *If  $(f_i)$  and  $(g_i)$  are two  $d$ -approximating sequences, we write*

$$(g_i) \preceq (f_i) \quad (\text{or, equivalently, } (f_i) \succeq (g_i))$$

*if  $\varphi(\mathbf{N}_{\mathcal{M}}^+((g_i))) \subseteq \varphi(\mathbf{N}_{\mathcal{M}}^+((f_i)))$  and  $\varphi(\mathbf{N}_{\mathcal{M}}^-((g_i))) \subseteq \varphi(\mathbf{N}_{\mathcal{M}}^-((f_i)))$ .*

**Definition 4.4.** *Let  $(f_i)$  and  $(g_i)$  be two  $d$ -approximating sequences. We write  $(g_i) \prec (f_i)$  (or, equivalently,  $(f_i) \succ (g_i)$ ) if  $(g_i) \preceq (f_i)$  and either  $\varphi(\mathbf{N}_{\mathcal{M}}^+((g_i))) \neq \varphi(\mathbf{N}_{\mathcal{M}}^+((f_i)))$  or  $\varphi(\mathbf{N}_{\mathcal{M}}^-((g_i))) \neq \varphi(\mathbf{N}_{\mathcal{M}}^-((f_i)))$  (i.e., at least one of the two inclusions in Definition 4.3 is proper).*

*We shall say that  $(f_i) \in S_H(\mathcal{M}, \mathcal{N})$  is a minimal sequence if there exists no sequence  $(g_i) \in S_H(\mathcal{M}, \mathcal{N})$  such that  $(g_i) \prec (f_i)$ .*

**Remark 4.5.** The relations  $\preceq$  and  $\prec$  could be defined by referring to the nodes in  $\mathcal{N}$  in place of the nodes in  $\mathcal{M}$ . In fact, our definitions immediately imply that the inclusion  $\varphi(\mathbf{N}_{\mathcal{M}}^+((g_i))) \subseteq \varphi(\mathbf{N}_{\mathcal{M}}^+((f_i)))$  is equivalent to  $\psi(\mathbf{N}_{\mathcal{N}}^-((g_i))) \subseteq \psi(\mathbf{N}_{\mathcal{N}}^-((f_i)))$  and the inclusion  $\varphi(\mathbf{N}_{\mathcal{M}}^-((g_i))) \subseteq \varphi(\mathbf{N}_{\mathcal{M}}^-((f_i)))$  is equivalent to  $\psi(\mathbf{N}_{\mathcal{N}}^+((g_i))) \subseteq \psi(\mathbf{N}_{\mathcal{N}}^+((f_i)))$ . An analogous statement holds for proper inclusions.

We observe that, in our definition,  $(g_i) \preceq (f_i)$  does not mean that either  $(g_i) \prec (f_i)$  or  $(g_i) = (f_i)$ .

The minimal sequences for  $\prec$  are, in some way, the best sequences of homeomorphisms whose measure approximates the natural size pseudodistance, since they minimize the sets  $\varphi(\mathbf{N}_{\mathcal{M}}^+)$  and  $\varphi(\mathbf{N}_{\mathcal{M}}^-)$  (and hence also  $\psi(\mathbf{N}_{\mathcal{N}}^+)$  and  $\psi(\mathbf{N}_{\mathcal{N}}^-)$ , i.e. the sets of node heights for the four types of nodes we have considered). We shall see that it is always possible to construct a  $d$ -approximating sequence of homeomorphisms such that the sets  $\varphi(\mathbf{N}_{\mathcal{M}})$  and  $\psi(\mathbf{N}_{\mathcal{N}})$  are finite, and that this can also be done by using minimal sequences.

The existence of minimal sequences with respect to the preordering  $\prec$  will be important in Section 5.

The following lemma is the main tool used in [2] to prove Theorem 2.4 cited in this paper (for a proof of this lemma see [2, p. 710]).

**Lemma 4.6.** *Assume that  $0 < d < +\infty$  and the measuring functions  $\varphi, \psi$  are Morse. For every  $(f_i) \in S_H(\mathcal{M}, \mathcal{N})$  there exists  $(g_i) \in S_H(\mathcal{M}, \mathcal{N})$  such that all maximal trains begin and end at critical points of the measuring functions and  $W((g_i)) \subseteq W((f_i))$ .*

**Remark 4.7.** We observe that in Lemma 4.6 the relation  $(g_i) \preceq (f_i)$  follows easily from the inclusion  $W((g_i)) \subseteq W((f_i))$ .

The following proposition shows some properties of the minimal sequences we are going to use, under the hypotheses that our measuring functions are Morse. In the next pages the symbols  $K_\varphi$  and  $K_\psi$  will represent the sets of critical points of the measuring functions  $\varphi$  and  $\psi$ , respectively. The sets of critical values of  $\varphi$  and  $\psi$  will be denoted by  $\varphi(K_\varphi)$  and  $\psi(K_\psi)$ .

**Proposition 4.8.** *Assume that  $0 < d < +\infty$  and the measuring functions  $\varphi, \psi$  are Morse, and set  $\mathcal{A} = \{z \in \mathbb{R} \mid \exists c_1, c_2 \in \varphi(K_\varphi) \cup \psi(K_\psi), r, s \in \mathbb{N} : z - c_1 = rd, c_2 - z = sd\}$ . Then the following statements hold:*

- (a) *If a train for a  $d$ -approximating sequence begins and ends at critical points of the measuring functions, the heights of its nodes belong to the finite set  $\mathcal{A}$ .*
- (b) *For every  $d$ -approximating sequence  $(f_i)$ , there exists a minimal sequence  $(h_i) \preceq (f_i)$  whose maximal trains begin and end at critical points of the measuring functions.*
- (c) *If a  $d$ -approximating sequence  $(g_i)$  is minimal, the height of every node of its trains belongs to  $\mathcal{A}$ .*

*Proof.* (a) follows trivially from the definition of train. The finiteness of  $\mathcal{A}$  follows from the finiteness of  $K_\varphi$  and  $K_\psi$ , and hence of  $\varphi(K_\varphi)$  and  $\psi(K_\psi)$  (here we are using the hypothesis that the measuring functions are Morse).

(b) Lemma 4.6 ensures that we can take a sequence  $(g_i) \preceq (f_i)$  whose maximal trains begin and end at critical points of the measuring functions. The previous statement (a) and the definition of the relation  $<$  imply that no infinite descending chain  $(g_i) > (g_i^1) > (g_i^2) > \dots$  beginning at  $(g_i)$  can exist. Let us consider the last term  $(g_i')$  in a maximal descending chain beginning at  $(g_i)$ . Obviously,  $(g_i')$  is a minimal  $d$ -approximating sequence. Unfortunately, statement (b) is still not proved, since some maximal train of  $(g_i')$  could either begin or end at regular points of the measuring functions, as opposed to what happens for  $(g_i)$ . However, by applying Lemma 4.6 to  $(g_i')$  we get a new  $d$ -approximating sequence  $(h_i)$  that is still minimal and has the required property regarding maximal trains.

(c) By (b), there exists a minimal sequence  $(h_i) \preceq (g_i)$  whose maximal trains begin and end at critical points of the measuring functions. Since  $(g_i)$  is already minimal, it follows that  $\varphi(\mathbf{N}_{\mathcal{M}}^+(h_i)) = \varphi(\mathbf{N}_{\mathcal{M}}^+(g_i))$  and  $\varphi(\mathbf{N}_{\mathcal{M}}^-(h_i)) = \varphi(\mathbf{N}_{\mathcal{M}}^-(g_i))$  (and hence  $\psi(\mathbf{N}_{\mathcal{N}}^-(h_i)) = \psi(\mathbf{N}_{\mathcal{N}}^-(g_i))$  and  $\psi(\mathbf{N}_{\mathcal{N}}^+(h_i)) = \psi(\mathbf{N}_{\mathcal{N}}^+(g_i))$ ). Statement (a) ensures that  $\varphi(\mathbf{N}_{\mathcal{M}}(h_i)) \cup \psi(\mathbf{N}_{\mathcal{N}}(h_i))$  is included in the finite set  $\mathcal{A}$ , and therefore so is  $\varphi(\mathbf{N}_{\mathcal{M}}(g_i)) \cup \psi(\mathbf{N}_{\mathcal{N}}(g_i))$ .  $\square$

## 5. Our main result

In Section 2 we have recalled (Theorem 2.4) that the natural pseudodistance between two size pairs is related to the critical values of their measuring functions.

However, the examples we have displayed suggest that our results can be improved. In fact, in our examples the analytic folding number  $k$  is never greater than 2. In the first part of this section we shall prove (Theorem 5.4) that the analytic folding number is never greater than 3, under the assumption that  $\mathcal{M}$  and  $\mathcal{N}$  are two homeomorphic smooth closed surfaces and the measuring functions  $\varphi, \psi$  are Morse. These hypotheses will make our proofs easier from the technical point of view. In Subsection 5.1 we shall weaken our assumptions and return to the case of class  $C^1$  (Theorem 5.7).

Now we introduce two lemmas. The first one is trivial and clarifies the local nature of the concept of node.

**Lemma 5.1.** *Assume  $0 < d < +\infty$ . Let  $U$  be an open subset of  $\mathcal{M}$  and  $(f_i)$  and  $(g_i)$  be two  $d$ -approximating sequences such that, for every  $i \in \mathbb{N}$ ,  $f_i$  coincides with  $g_i$  in  $U$ . Then  $\mathbf{N}_{\mathcal{M}}^+(f_i) \cap U = \mathbf{N}_{\mathcal{M}}^+(g_i) \cap U$  and  $\mathbf{N}_{\mathcal{M}}^-(f_i) \cap U = \mathbf{N}_{\mathcal{M}}^-(g_i) \cap U$ .*

*Proof.* This follows immediately from the definitions of the sets  $\mathbf{N}_{\mathcal{M}}^+$  and  $\mathbf{N}_{\mathcal{M}}^-$ . □

A similar result obviously holds for an open subset  $V$  of  $\mathcal{N}$ , and can easily be obtained by interchanging the roles of the sequences  $(f_i)$ ,  $(g_i)$  and  $(f_i^{-1})$ ,  $(g_i^{-1})$  in Lemma 5.1.

The useful property described by the following key lemma justifies the introduction of the concept of minimal sequence in the case of closed surfaces.

**Lemma 5.2.** *Assume that  $\mathcal{M}, \mathcal{N}$  are smooth homeomorphic closed surfaces and  $\varphi, \psi$  are Morse measuring functions on  $\mathcal{M}$  and  $\mathcal{N}$ , respectively. Suppose that  $0 < d < +\infty$ , and  $(f_i)$  is a minimal  $d$ -approximating sequence from  $(\mathcal{M}, \varphi)$  to  $(\mathcal{N}, \psi)$ . If  $N \in \mathbf{N}_{\mathcal{M}}((f_i))$  and  $\varphi(N)$  is not a critical value for  $\varphi$ , then at least one of  $\varphi(N) - d, \varphi(N) + d$  is a critical value for  $\psi$ .*

In other words, under the hypotheses of the lemma (possibly by exchanging the roles of the two surfaces), if we consider the heights of three consecutive nodes in a train of a minimal sequence, at least one of them is a critical value. The proof of this property involves Jost and Schoen’s theorem about harmonic maps between surfaces and is the key to proving the main result of this paper (Theorem 5.7).

*Proof.* We shall prove that if  $\varphi(N)$  is a regular value for  $\varphi$  and both  $\varphi(N) - d$  and  $\varphi(N) + d$  are regular values for  $\psi$  then we can get a new  $d$ -approximating sequence  $(\tilde{f}_i)$  such that  $(\tilde{f}_i) < (f_i)$ , contradicting the assumption that  $(f_i)$  is minimal. So, in the following we assume that  $\varphi(N) \notin \varphi(K_\varphi)$  and  $\varphi(N) + d, \varphi(N) - d \notin \psi(K_\psi)$ .

Let us define the open sets  $\mathcal{D}_\varepsilon = \{P \in \mathcal{M} : |\varphi(P) - \varphi(N)| < \varepsilon\}$  and  $V_\varepsilon = \{Q \in \mathcal{N} : \min_{\bar{Q} \in K_\psi} |\psi(Q) - \psi(\bar{Q})| < \varepsilon\}$  (in other words  $\mathcal{D}_\varepsilon$  is the set of all points of  $\mathcal{M}$  whose height differs less than  $\varepsilon$  from the height of  $N$ , while  $V_\varepsilon$  is the set of all points of  $\mathcal{N}$  whose height differs less than  $\varepsilon$  from the height of a critical point of  $\psi$ ). Moreover, let us choose  $\varepsilon > 0$  so small that

- (1)  $\bar{\mathcal{D}}_\varepsilon$  does not contain critical points for  $\varphi$ ;
- (2)  $\partial \mathcal{D}_\varepsilon$  does not contain nodes belonging to  $\mathbf{N}_{\mathcal{M}}((f_i))$ ;
- (3) for  $i$  large enough, if  $Q \in f_i(\bar{\mathcal{D}}_\varepsilon)$  and  $|\varphi(N) - \psi(Q)| \geq d - 2\varepsilon$  then  $Q \notin \bar{V}_\varepsilon$ .

The existence of an  $\varepsilon > 0$  satisfying (1) and (2) is ensured by the assumption that  $\varphi(N) \notin \varphi(K_\varphi)$  and the fact that the set of heights of the nodes is finite (see Proposition 4.8(c)); recall that the measuring functions are Morse. As regards (3), if for every positive  $\varepsilon$  we could find an arbitrarily large  $i$  and a  $\bar{Q} \in \bar{V}_\varepsilon \cap f_i(\bar{\mathcal{D}}_\varepsilon)$  satisfying  $|\varphi(N) - \psi(\bar{Q})| \geq d - 2\varepsilon$ , then there would exist a wagon  $(\hat{N}, \hat{Q})$  for  $(f_i)$ , with  $\psi(\hat{Q})$  equal to a critical value of  $\psi$ ,  $\varphi(\hat{N}) = \varphi(N)$  and  $|\varphi(\hat{N}) - \psi(\hat{Q})| = d$ , since  $\lim_i \Theta(f_i) = d$ . Therefore either  $\varphi(N) + d$  or  $\varphi(N) - d$  would be a critical value for  $\psi$ , contrary to our hypothesis.

*Note.* As a matter of fact, the expression “for  $i$  large enough” in (3) can be replaced with “for  $i \geq \bar{i}$ ”, where  $\bar{i}$  is a natural number independent of  $\varepsilon$ . We just require  $\varepsilon$  to be strictly

less than  $\eta/3$ , with  $\eta$  being the minimum distance between the finite set of all critical values of the (Morse) function  $\psi$  and the set  $\{\varphi(N) + d, \varphi(N) - d\}$ . In order to proceed this way we only have to take  $\bar{i}$  so large that for every  $i \geq \bar{i}$  the inequality  $\Theta(f_i) \leq d + \eta/3$  holds, implying  $|\varphi(N) - \psi(Q)| \leq d + \eta/3 + \varepsilon < d + 2\eta/3$  for every  $Q \in f_i(\overline{\mathcal{D}_\varepsilon})$ . If we also assume that  $|\varphi(N) - \psi(Q)| \geq d - 2\varepsilon$ , then  $d - 2\eta/3 < |\varphi(N) - \psi(Q)| < d + 2\eta/3$  and hence the distance between  $\psi(Q)$  and the set  $\{\varphi(N) + d, \varphi(N) - d\}$  must be strictly less than  $2\eta/3$ . Therefore  $\psi(Q)$  is at a distance more than  $\eta/3 > \varepsilon$  from the set of all critical values of  $\psi$  and hence  $Q \notin \overline{V_\varepsilon}$ . Anyway, this change of statement is not necessary for our proof, and we maintain the simpler version of (3).

Now, we are going to prove that there exists a sequence  $(\tilde{f}_i) \in S_H(\mathcal{M}, \mathcal{N})$  such that  $\tilde{f}_i = f_i$  in the closed set  $\mathcal{M} - \mathcal{D}_\varepsilon$  and  $\mathbf{N}_{\mathcal{M}}(\tilde{f}_i) \cap \overline{\mathcal{D}_\varepsilon} = \emptyset$  (in other words, we can eliminate all wagons from  $\mathcal{M}$  to  $\mathcal{N}$ , beginning in  $\overline{\mathcal{D}_\varepsilon}$ ).

So, we start by setting  $\tilde{f}_i(P) = f_i(P)$  for  $P \in \mathcal{M} - \mathcal{D}_\varepsilon$ .

In order to define  $\tilde{f}_i$  in  $\mathcal{D}_\varepsilon$  we have to consider each connected component  $\mathcal{C}$  of  $\mathcal{D}_\varepsilon$ . Because of hypothesis (1),  $\mathcal{C}$  is (homeomorphic to) a cylinder. On  $\mathcal{C} \cong S^1 \times (\varphi(N) - \varepsilon, \varphi(N) + \varepsilon)$  let us define the product metric  $d\theta^2 + d\varphi^2$ , so that the function  $\varphi$  is linear in  $\mathcal{C}$  (i.e.  $\nabla^2\varphi \equiv 0$ ). Here,  $d\theta^2$  and  $d\varphi^2$  are an arbitrarily chosen Riemannian metric on  $S^1$  and the Riemannian metric on the interval  $(\varphi(N) - \varepsilon, \varphi(N) + \varepsilon)$  induced by the Euclidean distance, respectively (cf. [8]).

Then consider a Riemannian metric  $\mu_{\mathcal{N}}$  on  $\mathcal{N}$  such that the measuring function  $\psi$  is harmonic at each point of  $\mathcal{N} - \overline{V_\varepsilon}$ . In other words, we require that  $\psi$  is harmonic in  $\mathcal{N}$ , with the possible exception of the closure of the set of those points whose height has a distance smaller than  $\varepsilon$  from some critical height of  $\psi$ . We can get this by using the construction in the previous paragraph. The set  $\mathcal{N} - \overline{V_\varepsilon}$  is a union of cylinders, and the level sets of  $\psi$  slice each cylinder into circles. The construction in the last paragraph yields a metric so that  $\psi$  is harmonic on  $\mathcal{N} - \overline{V_\varepsilon}$ . We refer to [8] and [9] for alternative proofs of the existence of such a Riemannian metric.

In order to apply Jost and Schoen’s theorem we need to work with diffeomorphisms. This implies that we have to approximate our homeomorphisms  $f_i$  by diffeomorphisms, without changing the trains of our  $d$ -approximating sequence.

**Claim A.** *There exists a sequence  $(g_i)$  of diffeomorphisms such that  $W((g_i)) = W((f_i))$ .*

*Proof of Claim A.* Since  $\mathcal{M}$  and  $\mathcal{N}$  are smooth surfaces, for each index  $i$  we can find a diffeomorphism  $g_i : \mathcal{M} \rightarrow \mathcal{N}$  such that  $d_{\mathcal{N}}(f_i(P), g_i(P)) \leq 1/i$  for every  $P \in \mathcal{M}$ , where  $d_{\mathcal{N}}$  is the distance on  $\mathcal{N}$  induced by the Riemannian metric  $\mu_{\mathcal{N}}$  (cf., e.g., Corollary 1.18 in [21], and [16]). Hence  $P\rho Q$  with respect to  $(f_i)$  if and only if  $P\rho Q$  with respect to  $(g_i)$  (recall Subsection 4.1). This implies that  $W((g_i)) = W((f_i))$ .

Because of Claim A, we can assume without loss of generality that each  $f_i$  is a diffeomorphism.

The following theorem holds (cf. [12]):

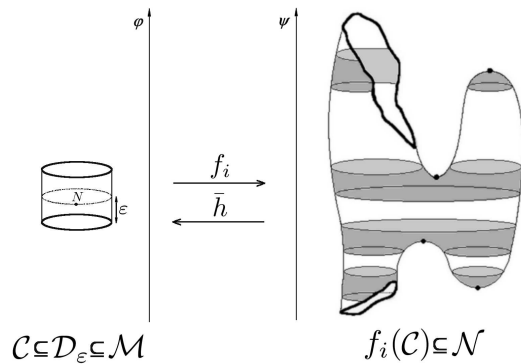
**Theorem 5.3 (Jost and Schoen).** *Let  $\Omega \subset M_1$  be a domain with non-empty boundary  $\partial\Omega$  consisting of  $C^1$  Jordan curves. Let  $h : \Omega \rightarrow M_2$  extend to a diffeomorphism of  $\overline{\Omega}$*



onto  $\overline{h(\Omega)}$ . Suppose the curves  $h(\partial\Omega)$  are of class  $C^{2+\alpha}$  and are locally convex with respect to  $h(\Omega)$ , i.e.  $h(\partial\Omega)$  has non-negative geodesic curvature with respect to the normal pointing into  $h(\Omega)$ . There exists a harmonic diffeomorphism  $\bar{h} : \Omega \rightarrow h(\Omega)$  which is homotopic to  $h$  and satisfies  $\bar{h} = h$  on  $\partial\Omega$ . Moreover,  $\bar{h}$  is of least energy among all diffeomorphisms homotopic to  $h$  and assuming the same boundary values.

Jost and Schoen’s theorem guarantees the existence of a diffeomorphism  $\bar{h} : \overline{f_i(\mathcal{C})} \rightarrow \overline{\mathcal{C}}$  that is harmonic in  $f_i(\mathcal{C})$  and coincides with  $f_i^{-1}$  at the boundary of  $f_i(\mathcal{C})$ .

Now, we are ready to define  $\tilde{f}_i(P)$  in the case  $P \in \mathcal{D}_\varepsilon$ , by setting  $\tilde{f}_i(P) = \bar{h}^{-1}(P)$  for every  $P \in \mathcal{C}$  and  $\mathcal{C}$  varying in the set of all the connected components of  $\mathcal{D}_\varepsilon$ . Practically, we are going to change  $f_i$  into  $\bar{h}^{-1}$  inside each cylinder  $\mathcal{C}$ . Notice that every  $\tilde{f}_i$  is a homeomorphism from  $\mathcal{M}$  to  $\mathcal{N}$  satisfying the equalities  $\tilde{f}_i(\overline{\mathcal{D}_\varepsilon}) = f_i(\overline{\mathcal{D}_\varepsilon})$  and  $\tilde{f}_i(\partial\mathcal{D}_\varepsilon) = f_i(\partial\mathcal{D}_\varepsilon)$ .



**Fig. 8.** The cylinder  $\mathcal{C}$  inside the set  $\mathcal{D}_\varepsilon$  and its image  $f_i(\mathcal{C})$ . The subset  $V_\varepsilon \cap f_i(\mathcal{C})$  is highlighted (in grey).

The key property of the new sequence  $(\tilde{f}_i)$  is that its “jumps” starting from  $\mathcal{D}_\varepsilon$  are controlled because of our hypotheses and the use of harmonic maps, and the “largest” jumps of  $(\tilde{f}_i)$  are not larger than the corresponding jumps of  $(f_i)$ . Formally, the following Claims B and C hold.

**Claim B.**  $\max_{f_i(\overline{\mathcal{D}_\varepsilon})} |\varphi \circ \tilde{f}_i^{-1} - \psi| = \max_{f_i(\partial\mathcal{D}_\varepsilon) \cup (\overline{V_\varepsilon} \cap f_i(\mathcal{D}_\varepsilon))} |\varphi \circ \tilde{f}_i^{-1} - \psi|$ .

*Proof of Claim B.* The key remark is that the (continuous) function  $\varphi \circ \tilde{f}_i^{-1} - \psi : \mathcal{N} \rightarrow \mathbb{R}$  is harmonic on  $f_i(\mathcal{D}_\varepsilon) - \overline{V_\varepsilon}$ . In fact, on the one hand, since  $\varphi$  is linear on  $\mathcal{D}_\varepsilon$  and  $\tilde{f}_i^{-1}$  is harmonic on  $f_i(\mathcal{D}_\varepsilon)$  it follows immediately that  $\varphi \circ \tilde{f}_i^{-1}$  is harmonic on  $f_i(\mathcal{D}_\varepsilon)$  (cf., e.g., Corollary 8.7.4 in [11]). On the other hand,  $\psi$  is harmonic on  $f_i(\mathcal{D}_\varepsilon) - \overline{V_\varepsilon}$ , by the choice of the Riemannian metric on  $\mathcal{N}$ . Hence, by the maximum principle, the restriction of  $\varphi \circ \tilde{f}_i^{-1} - \psi$  to  $\overline{f_i(\mathcal{D}_\varepsilon) - \overline{V_\varepsilon}}$  must take its maximum also at a point of  $\partial(f_i(\mathcal{D}_\varepsilon) - \overline{V_\varepsilon})$ . The same holds for the minimum value. This implies that the restriction of  $\varphi \circ \tilde{f}_i^{-1} - \psi$  to the compact set  $f_i(\overline{\mathcal{D}_\varepsilon})$  attains each extremum either in  $f_i(\partial\mathcal{D}_\varepsilon)$  or in  $\overline{V_\varepsilon} \cap f_i(\mathcal{D}_\varepsilon)$ . The conclusion of our claim follows immediately.

**Claim C.** For every  $i$  large enough, if  $\max_{\overline{\mathcal{D}}_\varepsilon} |\varphi - \psi \circ \tilde{f}_i| \geq d - \varepsilon$  then  $\max_{\overline{\mathcal{D}}_\varepsilon} |\varphi - \psi \circ \tilde{f}_i| \leq \max_{\partial \mathcal{D}_\varepsilon} |\varphi - \psi \circ f_i|$ . Therefore,  $\Theta(\tilde{f}_i) \leq \Theta(f_i)$ , and hence the new sequence  $(\tilde{f}_i)$  is a  $d$ -approximating sequence from  $(\mathcal{M}, \varphi)$  to  $(\mathcal{N}, \psi)$ .

*Proof of Claim C.* By Claim B, there exists  $\tilde{Q} \in f_i(\partial \mathcal{D}_\varepsilon) \cup (\overline{\mathcal{V}}_\varepsilon \cap f_i(\mathcal{D}_\varepsilon))$  such that  $|\varphi \circ \tilde{f}_i^{-1}(\tilde{Q}) - \psi(\tilde{Q})| = \max_{f_i(\overline{\mathcal{D}}_\varepsilon)} |\varphi \circ \tilde{f}_i^{-1} - \psi| = \max_{\overline{\mathcal{D}}_\varepsilon} |\varphi - \psi \circ \tilde{f}_i| \geq d - \varepsilon$ , under our hypothesis. Since  $\tilde{f}_i^{-1}(\tilde{Q}) \in \overline{\mathcal{D}}_\varepsilon$ , we have  $|\varphi \circ \tilde{f}_i^{-1}(\tilde{Q}) - \varphi(N)| \leq \varepsilon$ . It follows that  $|\varphi(N) - \psi(\tilde{Q})| \geq d - 2\varepsilon$  and hence  $\tilde{Q} \notin \overline{\mathcal{V}}_\varepsilon$  (for  $i$  large enough), because of the assumption (3) about  $\mathcal{D}_\varepsilon$ . Therefore  $\tilde{Q} \in f_i(\partial \mathcal{D}_\varepsilon)$ . Since the diffeomorphism  $\tilde{f}_i^{-1}$  coincides with  $f_i^{-1}$  in  $f_i(\partial \mathcal{D}_\varepsilon)$ ,  $\max_{\overline{\mathcal{D}}_\varepsilon} |\varphi - \psi \circ \tilde{f}_i| \leq \max_{\partial \mathcal{D}_\varepsilon} |\varphi - \psi \circ f_i|$ . So our claim is proved.

Lemma 5.1 (local nature of the concept of node) and the coincidence of  $\tilde{f}_i$  and  $f_i$  outside  $\overline{\mathcal{D}}_\varepsilon$  immediately imply the next claim.

**Claim D.** The set of all wagons from  $\mathcal{M} - \overline{\mathcal{D}}_\varepsilon$  to  $\mathcal{N}$  is the same for  $(f_i)$  and  $(\tilde{f}_i)$ . In particular,

$$\begin{aligned} \mathbf{N}_{\mathcal{M}}^+(\tilde{f}_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon) &= \mathbf{N}_{\mathcal{M}}^+(f_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon), \\ \mathbf{N}_{\mathcal{M}}^-(\tilde{f}_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon) &= \mathbf{N}_{\mathcal{M}}^-(f_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon). \end{aligned} \quad (5.0.1)$$

Finally, we can prove that under our hypotheses the new sequence  $(\tilde{f}_i)$  is “better” than  $(f_i)$  in the sense expressed by the following statement, saying that there exists no wagon from  $\mathcal{M}$  to  $\mathcal{N}$  beginning in  $\overline{\mathcal{D}}_\varepsilon$ .

**Claim E.** The set  $\mathbf{N}_{\mathcal{M}}((\tilde{f}_i)) \cap \overline{\mathcal{D}}_\varepsilon$  is empty.

*Proof of Claim E.* If an  $N' \in \mathbf{N}_{\mathcal{M}}((\tilde{f}_i)) \cap \overline{\mathcal{D}}_\varepsilon$  existed, then there would exist a sequence  $(P_r)$  of points of  $\mathcal{M}$  converging to  $N'$  and a strictly increasing sequence  $(i_r)$  in  $\mathbb{N}$  such that the sequence  $(\tilde{f}_{i_r}(P_r))$  converges and  $|\varphi(P_r) - \psi(\tilde{f}_{i_r}(P_r))| \rightarrow d$ . Since  $f_i$  and  $\tilde{f}_i$  coincide outside  $\mathcal{D}_\varepsilon$  and  $\partial \mathcal{D}_\varepsilon$  does not contain nodes for  $(f_i)$  (hypothesis (2) about  $\mathcal{D}_\varepsilon$ ), we can assume that all points  $P_r$  belong to  $\overline{\mathcal{D}}_\varepsilon$ . Then, by Claim C, there would also exist a converging sequence  $(B_r)$  such that  $|\varphi(B_r) - \psi(f_{i_r}(B_r))| \rightarrow d$ , where each  $B_r$  is in  $\partial \mathcal{D}_\varepsilon$ . This would imply the existence of a node for  $(f_i)$  belonging to  $\partial \mathcal{D}_\varepsilon$ , once more contradicting hypothesis (2).

In summary, we have seen that  $\overline{\mathcal{D}}_\varepsilon$  does not meet  $\mathbf{N}_{\mathcal{M}}((\tilde{f}_i))$ , but contains at least one node of  $\mathbf{N}_{\mathcal{M}}((f_i))$ , while  $\mathbf{N}_{\mathcal{M}}^+(\tilde{f}_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon) = \mathbf{N}_{\mathcal{M}}^+(f_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon)$  and  $\mathbf{N}_{\mathcal{M}}^-(\tilde{f}_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon) = \mathbf{N}_{\mathcal{M}}^-(f_i) \cap (\mathcal{M} - \overline{\mathcal{D}}_\varepsilon)$  (Claim D). It follows that  $(\tilde{f}_i) < (f_i)$ . This contradicts the hypothesis that  $(f_i)$  is a minimal sequence.  $\square$

Let us apply Lemma 5.2 to prove that the analytic folding number is never greater than 3 for closed surfaces.

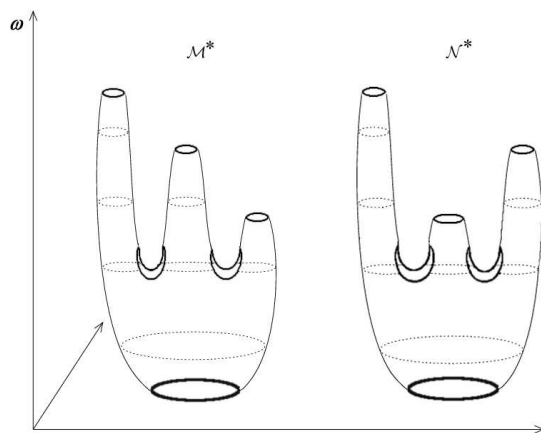
**Theorem 5.4.** Assume that  $\mathcal{M}$  and  $\mathcal{N}$  are two homeomorphic smooth closed surfaces and that  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$  and  $\psi : \mathcal{N} \rightarrow \mathbb{R}$  are two Morse functions. Then, if  $d$  denotes the natural pseudodistance between the size pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$ , at least one of the following properties holds:

- (i)  $d$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ ;
- (ii)  $d$  equals half the distance between two critical values of  $\varphi$ ;
- (iii)  $d$  equals half the distance between two critical values of  $\psi$ ;
- (iv)  $d$  equals one third of the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ .

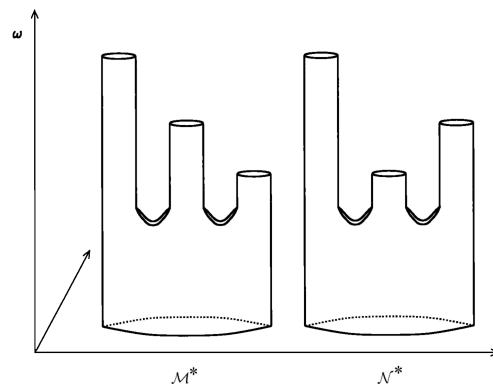
*Proof.* If  $d = 0$ , then  $\varphi$  and  $\psi$  have the same global minimum  $\mu$ . Hence,  $d = |\mu - \mu|$  and our assertion is trivial.

So let us assume  $d > 0$ . Let  $(h_i)$  be a minimal sequence whose maximal trains begin and end at critical points of the measuring functions (Proposition 4.8(b)) and suppose  $N_1 \in \mathcal{M}$  is the initial node of a maximal train (if no maximal train begins in  $\mathcal{M}$ , it is sufficient to exchange the roles of our surfaces; in this case (ii) and (iii) interchange in the following). Therefore  $\varphi(N_1)$  is a critical value for  $\varphi$ . Let  $N_2 \in \mathcal{N}$  be the next node in the train. If  $\psi(N_2)$  is a critical value for  $\psi$ , then condition (i) holds. Otherwise, let  $N_3 \in \mathcal{M}$  be the next node in the train ( $N_3$  exists because  $\psi(N_2)$  is not a critical value for  $\psi$ , and hence  $N_2$  is not the final node of the train). If  $\varphi(N_3)$  is a critical value for  $\varphi$ , then (ii) holds. Otherwise, let  $N_4 \in \mathcal{N}$  be the next node in the train ( $N_4$  exists because  $\varphi(N_3)$  is not a critical value for  $\varphi$ , and hence  $N_3$  is not the final node of the train). Lemma 5.2 applied for  $N = N_3$  ensures that  $\psi(N_4) = \varphi(N_1) + 3d$  is a critical value for  $\psi$ . Therefore  $d = \frac{1}{3}(\psi(N_4) - \varphi(N_1))$  and (iv) holds.  $\square$

**Remark 5.5.** It may be interesting to note that Example 3.4 can be used to show that the hypothesis of  $h(\partial\Omega)$  being locally convex with respect to  $h(\Omega)$  is necessary in Theorem 5.3. In fact, consider the open surfaces  $\mathcal{M}^*, \mathcal{N}^*$  displayed in Figure 9, obtained from the two surfaces  $\mathcal{M}, \mathcal{N}$  of Figure 5 by deleting suitable closed neighbourhoods of the critical points. On  $\mathcal{M}^*$  and  $\mathcal{N}^*$  consider two Riemannian metrics  $\mu_{\mathcal{M}^*}$  and  $\mu_{\mathcal{N}^*}$  such that  $\varphi$  and  $\psi$  are linear functions with respect to  $\mu_{\mathcal{M}^*}$  and  $\mu_{\mathcal{N}^*}$ , respectively (cf. [8]). The metrics  $\mu_{\mathcal{M}^*}$  and  $\mu_{\mathcal{N}^*}$  are the ones induced by the embeddings of  $\mathcal{M}^*$  and  $\mathcal{N}^*$  in



**Fig. 9.** There exists a diffeomorphism  $h : \mathcal{N}^* \rightarrow \mathcal{M}^*$  preserving the height of the boundary points (thick).



**Fig. 10.** The metrics  $\mu_{\mathcal{M}^*}$  and  $\mu_{\mathcal{N}^*}$  are the ones induced by the displayed embeddings of  $\mathcal{M}^*$  and  $\mathcal{N}^*$  into  $\mathbb{R}^3$ .

$\mathbb{R}^3$  displayed in Figure 10. Notice that  $\partial\mathcal{M}^*$  is *not* locally convex with respect to  $\mathcal{M}^*$ . We observe that there exists a diffeomorphism  $h : \overline{\mathcal{N}^*} \rightarrow \overline{\mathcal{M}^*}$  preserving the height of the boundary points. If there existed a diffeomorphism  $\bar{h} : \overline{\mathcal{N}^*} \rightarrow \overline{\mathcal{M}^*}$  harmonic in  $\mathcal{N}^*$  and coinciding with  $h$  at  $\partial\mathcal{N}^*$ , it should preserve the height of every point in  $\mathcal{N}^*$ , since the function  $\varphi - \psi \circ \bar{h} : \overline{\mathcal{N}^*} \rightarrow \mathbb{R}$  would be harmonic in  $\mathcal{N}^*$  and take its maximum and minimum at points of  $\partial\mathcal{N}^*$  where  $\varphi - \psi \circ \bar{h}$  vanishes. Moreover, we could easily extend  $\bar{h}$  to a diffeomorphism  $h' : \mathcal{N} \rightarrow \mathcal{M}$  that preserves the height of every point outside  $\mathcal{N}^*$ . Therefore, there would exist an optimal diffeomorphism between the size pairs  $(\mathcal{N}, \omega)$ ,  $(\mathcal{M}, \omega)$ , contradicting what we said in Example 3.4.

**Remark 5.6.** Lemma 5.2 may be considered analogous to Lemma 3.2 proved in [4] for curves, but the techniques used in the proof are substantially different, since here we have to handle harmonic maps in place of linear maps in dimension 1. As a consequence, because of the hypotheses required in Jost and Schoen's theorem, problems about the position of images of critical points arise after the harmonization process, since, in contrast to what happens in the case of curves, we do not know this position. As a result, in both the 1-dimensional and 2-dimensional cases we can prove that if we consider the heights of  $m$  consecutive nodes in a train of a minimal sequence, at least one of them is a critical value, but we have to set  $m = 2$  for curves and  $m = 3$  for surfaces, depending on the different techniques and dimensional constraints involved in our proofs. This difference explains why the conclusion of Theorem 3.4 in [4] (for curves) is stronger than the conclusion of Theorem 5.7 in this paper, concerning surfaces.

### 5.1. Weakening the hypotheses about the regularity of surfaces and measuring functions

Until now we have considered smooth closed surfaces and Morse measuring functions. By repeating the proofs used in [2] to weaken our hypotheses about regularity (see Section 6 in that paper), we can get our main result via an approximation procedure:

**Theorem 5.7.** *Assume that  $\mathcal{M}$  and  $\mathcal{N}$  are two homeomorphic closed surfaces of class  $C^1$  and that  $\varphi : \mathcal{M} \rightarrow \mathbb{R}$  and  $\psi : \mathcal{N} \rightarrow \mathbb{R}$  are two functions of class  $C^1$ . Then, if  $d$  denotes the natural pseudodistance between the size pairs  $(\mathcal{M}, \varphi)$  and  $(\mathcal{N}, \psi)$ , at least one of the following properties holds:*

- (i)  *$d$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ ;*
- (ii)  *$d$  equals half the distance between two critical values of  $\varphi$ ;*
- (iii)  *$d$  equals half the distance between two critical values of  $\psi$ ;*
- (iv)  *$d$  equals one third of the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ .*

## 6. Conclusions and further research

In this paper we have proved that for closed surfaces the relation between the natural pseudodistance and the critical values of the measuring functions is stronger than the one we proved in [2] for general dimension. In fact, Theorem 5.7 shows that the natural pseudodistance between two homeomorphic  $C^1$  closed surfaces associated with  $C^1$  measuring functions is always either the distance or half the distance or one third of the distance between two suitable critical values of the measuring functions.

Unfortunately, our techniques cannot be used for larger dimensions, since the statement of Jost and Schoen's theorem fails in dimension strictly greater than 2 (cf. [10, Section 5.8], and [5, Section 12]). Moreover, the application of this theorem requires the approximability of homeomorphisms by means of diffeomorphisms. This procedure is not available for dimensions strictly larger than 3 (cf., e.g., [14]). As a consequence we do not know if results analogous to Theorem 5.7 hold for dimensions strictly larger than 2. In other words, we wonder if there exist two  $n$ -manifolds equipped with regular measuring functions  $\varphi, \psi$  such that their pseudodistance equals neither  $D$  nor  $D/2$  nor  $D/3$ , for  $D$  varying in the set of all distances between the critical values of  $\varphi$  and  $\psi$ .

We intend to study this problem and the availability of new techniques for studying the general  $n$ -dimensional case.

However, it is interesting to note that we do not know of any examples where the analytic folding number equals 3, also in the bidimensional case. On the other hand, we are not able to improve our result by proving that the analytic folding number never equals 3, also in the case of surfaces (see Remark 5.6).

The difficulty in finding examples where the analytic folding number equals 3 deserves some further remarks. One technique that can be used for computing natural size pseudodistances is based on *size functions* (cf. [3]). The computation of size functions is usually easy and gives us a lower bound  $s$  for natural size pseudodistances. Obviously, when we are able to exhibit a sequence  $(f_i)$  of homeomorphisms for which  $\lim_i \Theta(f_i) = s$  we can claim that the natural size pseudodistance equals  $s$ . The key point is that the best lower bound  $s$  we can obtain is either the distance or half the distance between two suitable critical values of the measuring functions (cf. Theorem 2 in [3]). Therefore, if an example where the analytic folding number equals 3 really exists, we are not able to find

and recognize it using the previously described technique. Apparently, new techniques should be developed.

As regards the use of harmonic maps in our study, this corresponds to the property that the deformation due to tension fields decreases both the energy and the maximum change of the measuring functions, provided that we are far from their critical points. The use of different kinds of deformations (e.g. curvature evolution of level lines of the measuring functions) might be investigated. The main problem seems to be the possible birth of degeneracies.

Furthermore, it might be interesting to examine the possibility of moving from the study of trains of limit  $d$ -jumps to the study of relations obtained as limits of  $d$ -approximating sequences of homeomorphisms, with respect to the Hausdorff (or another more suitable) topology.

In conclusion, various interesting questions remain open and deserve further study.

*Acknowledgments.* We wish to thank M. Ferri for helpful discussions, and J. Jost and A. Ratto for providing some useful references. Anyway, the authors are solely responsible for any mistakes. Finally, Frosini wishes to thank I. Fossati, F. Battiato and A. Branduardi for their indispensable support.

This work was partially supported by MIUR (Italy), ARCES (Italy) and INdAM-GNSAGA (Italy). Investigation supported by the University of Bologna (funds for selected research topics).

This paper is dedicated to Natale Morea and Gabriele Simeoni.

## References

- [1] Dibos, F., Frosini, P., Pasquignon, D.: The use of size functions for comparison of shapes through differential invariants. *J. Math. Imaging Vision* **21**, 107–118 (2004) MR 2090127
- [2] Donatini, P., Frosini, P.: Natural pseudodistances between closed manifolds. *Forum Math.* **16**, 695–715 (2004) Zbl pre02144167 MR 2096683
- [3] Donatini, P., Frosini, P.: Lower bounds for natural pseudodistances via size functions. *Arch. Inequal. Appl.* **2**, 1–12 (2004) Zbl 1067.58008 MR 2043046
- [4] Donatini, P., Frosini, P.: Natural pseudodistances between closed curves. Submitted
- [5] Eells, J., Lemaire, L.: Another report on harmonic maps. *Bull. London Math. Soc.* **20**, 385–524 (1988) Zbl 0669.58009 MR 0956352
- [6] Frosini, P.: A distance for similarity classes of submanifolds of a euclidean space. *Bull. Austral. Math. Soc.* **42**, 407–416 (1990) Zbl 0707.53004 MR 1083277
- [7] Frosini, P.: Connections between size functions and critical points. *Math. Methods Appl. Sci.* **19**, 555–569 (1996) Zbl 0857.55005 MR 1385154
- [8] Frosini, P.: A note on the linearity of real-valued functions with respect to suitable metrics. *Geom. Dedicata* **108**, 105–110 (2004) Zbl 1067.58011 MR 2112667
- [9] Frosini, P., Landi, C.: Intrinsic harmonicity of Morse functions. *Mathematika* **50**, 167–170 (2003) Zbl 1075.57014 MR 2136359
- [10] Jost, J.: Harmonic Mappings between Riemannian Manifolds. *Proc. Centre Math. Anal. Austral. Nat. Univ.* 4, Canberra, 1983 Zbl 0542.58001 MR 0756629
- [11] Jost, J.: *Riemannian Geometry and Geometric Analysis*. Springer, Berlin (1995) Zbl 0828.53002 MR 1351009
- [12] Jost, J., Schoen, R.: On the existence of harmonic diffeomorphisms between surfaces. *Invent. Math.* **66**, 353–359 (1982) Zbl 0488.58009 MR 0656629

- [13] Landi, C., Frosini, P.: Size functions and morphological transformations. *Acta Appl. Math.* **49**, 85–104 (1997) Zbl 0883.68133 MR 1482881
- [14] Milnor, J.: On manifolds homeomorphic to the 7-sphere. *Ann. of Math. (2)* **64**, 399–405 (1956) Zbl 0072.18402 MR 0082103
- [15] Mulazzani, M., Frosini, P.: Size homotopy groups for computation of natural size distances. *Bull. Belg. Math. Soc.* **6**, 455–464 (1999) Zbl 0937.55010 MR 1721761
- [16] Munkres, J.: Obstructions to the smoothing of piecewise-differentiable homeomorphisms. *Ann. of Math. (2)* **72**, 521–554 (1960) Zbl 0108.18101 MR 0121804
- [17] Uras, C., Verri, A.: Invariant size functions. In: *Applications of Invariance in Computer Vision*, Lecture Notes in Comput. Sci. 825, Springer, 215–234 (1994)
- [18] Uras, C., Verri, A.: Metric-topological approach to shape representation and recognition. *Image Vision Comput.* **14**, 189–207 (1996)
- [19] Uras, C., Verri, A.: Computing size functions from edge maps. *Internat. J. Comput. Vision* **23**, 169–183 (1997)
- [20] Verri, A., Uras, C., Frosini, P. and M. Ferri: On the use of size functions for shape analysis. *Biol. Cybern.* **70**, 99–107 (1993) Zbl 0789.92030
- [21] Whitehead, J. H. C.: Manifolds with transverse fields in Euclidean space. *Ann. of Math.* **73**, 154–212 (1961) Zbl 0096.37802 MR 0124917