Tight and Taut Submanifolds
MSRI Publications
Volume **32**, 1997

Geometry in Curvature Theory

NICOLAAS H. KUIPER

Abstract. This article is based on the Roever Lectures in Geometry given by Kuiper at Washington University, St. Louis, in January 1986. Although incomplete, it is an excellent exposition of the topics it does cover, starting with elementary versions of the notion of tightness and going through the analysis of topsets, the classification in low dimensions, the notions of total curvature for curves and surfaces in space, homological notions of tightness, the Morse inequalities, and Poincaré polynomials. It contains a detailed proof of Kuiper's remarkable result that a tight two-dimensional surface substantially immersed in $\mathbb{R}^5$ must be a Veronese surface.

Editors' Note. At the time of Kuiper's death in December, 1994, this paper existed in the form of an unfinished typescript. For inclusion in this volume, it was edited by Thomas Banchoff, Thomas Cecil, Wolfgang Kühnel, and Silvio Levy. A few Editors' Notes such as this one were included, mostly pointing to additional references. Several minor typos were corrected and the numbering was normalized for ease of reference; thus Sections 5 and 6 of the manuscript were renumbered 4 and 5, since there was no Section 4. The present illustrations were made by Christine Heinitz and by Levy, based on Kuiper's hand drawings.

1. Banchoff's two-piece property. Zero-tightness

Prerequisites and Notation. *Euclidean space* $E = E^N$ of dimension N is the real vector space $\mathbb{R}^N$ with norm $\|u\| = \sqrt{\Sigma_1^N (u^i)^2}$ for $u = (u^1, \dots, u^N) \in \mathbb{R}^N$ and distance $\|v-u\|$ for $u, v \in \mathbb{R}^N$. The identification $\kappa : \mathbb{R}^N \to E^N$ can be replaced by any other preferred Euclidean coordinate system $\kappa \circ g : \mathbb{R}^N \to E^N$, where g is an isometry:

$$g(u) = u_0 + u \cdot g_0, \quad \text{for } u_0 \in \mathbb{R}^N \text{ and } g_0 \in O(N) \text{ an orthogonal matrix.}$$

We use E^N to emphasize Euclidean space aspects and $\mathbb{R}^N$ for vector space aspects. A set $X \subset E$ is called *convex* if

$$u + \lambda(v-u) \in X \quad \text{for all } u, v \in X \text{ and } 0 \le \lambda \le 1.$$

The smallest convex set containing $X \subset E$ is its *convex hull*, denoted $\mathcal{H}X$. The smallest affine subspace (also a Euclidean space) that contains X is its *span*, denoted $\mathrm{span}(X)$. If $\mathrm{span}(X) = E$, then X is called *substantial* in E. The boundary of $\mathcal{H}X$ in $\mathrm{span}(X)$ is called the *convex envelope* $\partial\mathcal{H}X$ of $X \subset E$. If X is one point then $\mathcal{H}X = X$ and $\partial\mathcal{H}X = \varnothing$.

The subspaces $\{u : \|u\| < r\}$ and $\{u : \|u\| = r\}$, for $r > 0$, are called the *N-ball* B^N and the *$(N-1)$-sphere* S^{N-1}, respectively. As metric spaces they are called the *round ball* and the *round sphere*. Let $z : \mathbb{R}^N \to \mathbb{R}$ be a *linear function*, $z(u) = \sum_i \zeta_i u^i$ for $\zeta_i \in \mathbb{R}$, with $\|z\|^2 = \sum \zeta_i^2 > 0$. The subspaces $\{u : z(u) \geq c\}$, $\{u : z(u) > c\}$ and $\{u : z(u) = c\}$ of E are called the *half-space* h, its interior the *open half-space* $\mathring{h}$, and its boundary the *hyperplane* ∂h, respectively. The function z is often called a *height function*.

A metrizable topological space X is called *separated* if it is the disjoint union of two nonempty open and closed subsets, say X_1 and X_2. If $U \subset X$ contains points $x_1 \in X_1$ and $x_2 \in X_2$, then $U \cap X_1$ and $U \cap X_2$ are disjoint open and closed in U, and so U is also separated. The space X is called *connected* if it is not separated. A connected nonempty open closed subset of a metrizable space X is called a *topological component* of X.

Example. The plane set

$$\{(\xi, \eta) : \xi = 0 \text{ or } \eta = \sin(\xi^{-1})\}$$

is connected (but not pathwise connected).

Consequence. If Y is a metrizable space and for any two points $y_1, y_2 \in Y$ there is a connected space $W(y_1, y_2) \subset Y$ containing y_1 and y_2 (in other words, if "any two points $y_1, y_2 \in Y$ can be connected in Y"), then Y is connected. Indeed Y separated would show an immediate contradiction.

Definitions and General Theorems. For given compact spaces X, either embedded in $E = E^N$ or given independently, we are interested in embeddings or other continuous maps in E with nice properties that generalize convexity. We introduce the important notion and tool called a topset:

Definition. Suppose the half-space

$$h = h_z = \{u \in E : z(u) \geq c\}$$

for some linear function $z : E \to \mathbb{R}$ supports ("leans against") the compact set X, without containing it completely:

$$X \neq h \cap X = \partial h \cap X \neq \varnothing.$$

Then $X_z = h \cap X$ is called a (proper) *topset* of the set $X \subset E$. It is the set of points in X for which the function $z : X \to \mathbb{R}$ attains its maximal (top) value.

Note that $\mathring{h} \cap X = \varnothing$. More generally, if $f : X \to E$ is a continuous map of the compact space X into E, and $h \cap f(X)$ is a topset of $f(X) \subset E$, then

$$f^{-1}(h) = \{x \in X : f(x) \in h\} = f^{-1}(\partial h) \subset X$$

is called a *topset of the map* f. A topset of a topset is called a *top*^{2}*set*. A top^{j}set for some $j \geq 1$ is called a *top***set*. If the span of a top*set X' of X has dimension k, then X' is called an E^k-*top***set*.

REMARK. Let X be a substantial set of $N+1$ points $e_0, \dots, e_N$ in E^N. Then $\mathcal{H}X$, the convex hull, is an N-simplex σ_N. Any proper nonempty subset of X is a topset.

EXERCISE. Determine all topsets of a standard torus in E^3, obtained by rotating a circle around a disjoint line in its plane.

EXERCISE. Determine the topsets of the map $f : w \to w^3$ of the unit circle $\{w : |w| = 1\} \subset \mathbb{C}$ into $\mathbb{C} = \mathbb{R} \oplus \mathbb{R} = E^2$.

THEOREM 1.1. *The convex envelope of a compact set $X \subset E$ is the union of the convex hulls of its topsets*:

$$\partial \mathcal{H}X = \bigcup_z \mathcal{H}X_z.$$

The union may be taken only over all linear functions $z : E \to \mathbb{R}$ with norm $\|z\| = 1$. If X consists of one point, both sides are the empty set.

PROOF. We need to show the implications in both directions:

$$x \in \bigcup_z \mathcal{H}X_z \iff x \in \partial \mathcal{H}X.$$

Assume that the span of X is $\operatorname{span}(X) = E = E^N$. For a topset $X_z = h \cap X$, let $x \in \mathcal{H}X_z$. Since h supports X, it supports also $\mathcal{H}X$. Then $x \in h \cap \mathcal{H}X \subset \partial \mathcal{H}X$. Conversely if $x \in \partial \mathcal{H}X$, then there is a $\mathcal{H}X$-supporting half-space h_z containing x, and

$$x \in h_z \cap \partial \mathcal{H}X = h_z \cap \mathcal{H}X = h_z \cap \mathcal{H}X_z = \mathcal{H}X_z. \qquad \square$$

Now we propose a preliminary generalization of convexity:

DEFINITION (for compact sets). A connected compact set $X \subset E$ is said to have the *two-piece property*, or TPP [Banchoff 1971b], and is called 0-*tight*, in case any of the following equivalent conditions hold:

(a) $h \cap X$ is connected for every half-space h.
(b) $\mathring{h} \cap X$ is connected for every h.
(c) The set difference $X \setminus \partial h$ has at most two components for every h (this is the two-piece property).
(d) In terms of Čech homology and any coefficient ring, the homomorphism $H_0(h \cap X) \to H_0(X)$ is injective for every h.

We mention this last condition now for the sake of completeness, but defer the relevant discussion till later (page 35).

Proof of equivalence. (a) $\Rightarrow$ (b). If (a) holds then any two points in $\mathring{h} \cap X$ are contained for some half-space h_i in $h_i \cap X \subset \mathring{h} \cap X$, and they can be connected in $h_i \cap X$. Then $\mathring{h} \cap X$ is connected.

(b) $\Rightarrow$ (a). Suppose $h \cap X$ is not connected for some $h = \{u \in E : z(u) \geq c\}$, let and Y_1 and Y_2 be disjoint nonempty open closed subsets with union $Y_1 \cup Y_2 = h \cap X$. Let U_1 and $U_2 \subset X$ be disjoint nonempty open neighborhoods of the compact subsets Y_1 and Y_2. If $c - 2\varepsilon$ is the maximum of z on the compact set $X \setminus (U_1 \cup U_2)$ and $\mathring{h}_0 = \{u \in E : z(u) > c - \varepsilon\}$, then $\mathring{h}_0 \cap X$ is not connected. This contradicts (b).

The equivalence (b) $\Leftrightarrow$ (c) is tautological. $\square$

The same proof works for the equivalences in the following more general situation.

Definition (for maps). A continuous map $f : X \to E$ of a connected compact space X in E has the *two-piece-property* (TPP) and is called 0-*tight* if any of the following equivalent conditions hold:

(a) $f^{-1}(h)$ is connected for any half-space h.
(b) $f^{-1}(\mathring{h})$ is connected for any h.
(c) $f^{-1}(E \setminus \partial h)$ has at most two components for any h (two-piece-property).
(d) $H_0(f^{-1}(h)) \to H_0(X)$ is injective for any h.

Examples. The following are 0-tight sets:

(1) a convex body $X = \mathcal{H}X \subset E^N$, for $N \geq 0$;
(2) a convex hypersurface $X = \partial\mathcal{H}X$ substantial in E^N, for $N \geq 2$ (convex curve for $N = 2$);
(3) a hemisphere, $\{u \in \mathbb{R}^3 : \|u\| = 1,\ z(u) \leq 0\}$;
(4) the standard round torus in E^3 (see the first exercise on page 3);
(5) the solid round ring (solid torus) bounded by the standard torus;
(6) the 1-*skeleton* $\mathrm{Sk}_1(\sigma_N)$ of the N-simplex $\sigma_N \subset E^N$; this is by definition the union of all edges of σ_N, and as a topological space it is a *complete graph* on $N + 1$ vertices.

Remark. These are corollaries of the definition:

(1) 0-tightness is invariant under *linear embeddings* $i : \mathbb{R}^M \to \mathbb{R}^N$ and *projections* $p : \mathbb{R}^N \to \mathbb{R}^M$, where $M < N$. Indeed, if $f : X \to \mathbb{R}^M$ and $g : Y \to \mathbb{R}^N$ are 0-tight, then so are $i \circ f : X \to E^N$ and $p \circ g = Y \to E^M$.

An example of a 0-tight map (immersion) is the projection of the 1-skeleton $\mathrm{Sk}_1(\sigma_3)$ in E^3 onto the union of edges and diagonals of a convex 4-gon in a plane in E^3. Note that $f : X \to \text{point} \in E^N$ is 0-tight for any connected compact X.

(2) 0-tightness is an *affine* and even a *projective property* in the following sense. Let P^N be a real projective N-space and P^{N-1} a hyperplane. Then $P^N \setminus P^{N-1}$

can be identified with E^N, and this identification is natural up to affine transformations $u \mapsto u_0 + gu$, where $u_0 \in \mathbb{R}^N$, $g \in \mathrm{GL}(n, \mathbb{R})$. Given $X \subset E^N$, let $\eta : P^N \to P^N$ be a projective transformation such that $\eta(X) \subset E^N \subset P^N$. Suppose $f : X \to E^N$ is 0-tight. Then also $\eta \circ f : X \to E^N$ is 0-tight by condition (c), which is expressed in terms of hyperplane sections.

THEOREM 1.2. *Any topset X_z of a* 0*-tight set $X \subset E$ or of a* 0*-tight map $f : X \to E$ is itself* 0*-tight. So is any top*set.*

PROOF. We deal with the case of a set; the proof for a map is the same. Suppose $X \subset E$ has a topset that is not 0-tight, say $X_z = h_z \cap X$, where $h_z = \{u \in E : z(u) \geq c_1\}$. See Figure 1. Then there exists a half-space $h_0 = \{u \in E : w(u) \geq c_2\}$ such that

$$\varnothing \neq h_0 \cap X_z \neq X_z,$$

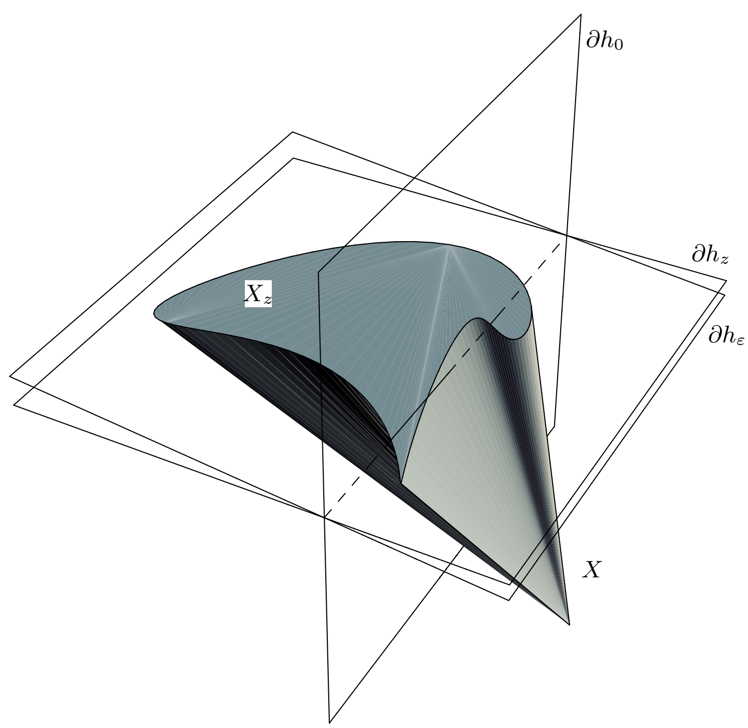


Figure 1. If a topset X_z of X is not 0-tight, some hyperplane ∂h_0 cuts X_z into more than two pieces. By tilting slightly the support hyperplane ∂h_z of X_z around the hinge $\partial h_z \cap \partial h_0$, one can obtain a hyperplane that cuts X into more than two pieces, so X is not 0-tight. (In the figure, the half-space h_0 is to the right of, and the half-spaces h_z and h_ε are above, their bounding planes.)

and $h_0 \cap X_z = h_0 \cap h_z \cap X = X_1 \cup X_2$ is separated with X_1 and X_2 open and closed in the relative topology, compact, and disjoint. There is a (small) open neighborhood U in X of $X_1 \cup X_2$ that is also separated:

$$U = U_1 \cup U_2, \quad U_1 \cap U_2 = \varnothing, \quad U_1 \supset X_1, \quad U_2 \supset X_2.$$

For small ε, the half-space

$$h_\varepsilon = \{u \in E : (z(u) - c_1) + \varepsilon(w(u) - c_2) \geq 0\}$$

meets X in $h_\varepsilon \cap X \supset h_0 \cap h_z \cap X$ and

$$h_\varepsilon \cap X \subset U = U_1 \cup U_2.$$

Therefore $h_\varepsilon \cap X$ is not connected, so X is not 0-tight. □

THEOREM 1.3 (CLASSIFICATION OF PLANE 0-TIGHT SETS). *Any plane compact 0-tight substantial set $X \subset E^2$ can be obtained from its convex hull $\mathcal{H}X$ by deleting a countable family of disjoint open convex subsets from the interior*:

$$X = \mathcal{H}X \setminus \bigcup_{i=0}^{r} U_i, \quad \textit{where } 0 \leq r \leq \infty.$$

PROOF. Let X be substantial and 0-tight in E^2. Every topset X_z of X is 0-tight and lies in a supporting line ∂h_z, so it is either a point or a line segment, and in any case convex. By Theorem 1.1 we have $\bigcup_z X_z = \bigcup_z \mathcal{H}X_z = \partial\mathcal{H}X$, and $\partial\mathcal{H}X$ is contained in X. The set X is then obtained from $\mathcal{H}X$ by deleting disjoint open connected sets (holes) U. Any embedded circle in a hole U does not separate X, and can be contracted inside U to a point. So each hole U is contractible and homeomorphic to an open disc.

Now suppose U is not convex, so there are distinct $u_1, u_2 \in U$ and $0 < \lambda < 1$ such that $u_1 + \lambda(u_2 - u_1) \notin U$. The smallest such value λ for given u_1 and u_2 yields a point $x = u_1 + \lambda(u_2 - u_1)$ in X. Let h and h^- be the two half-planes having as common boundary the line $u_1u_2 = \partial h = \partial h^-$. Connect u_1 and u_2 by an embedded polygonal arc $\beta \subset U$ that meets the line ∂h transversally in every intersection point (see Figure 2). Since X is 0-tight, $h \cap X$ and $h^- \cap X$ are connected. Then x and $h \cap \partial\mathcal{H}X$ can be connected in $h \cap X$ and (even better) in the component of $(h \cap \mathcal{H}X) \setminus \beta$, which contains $h \cap X$. In $h \cap X$, the points x and $h \cap \partial\mathcal{H}X$ can be connected by a polygonal arc α, which meets ∂h only in the point x.

There is another such polygonal arc α^- in $(h^- \cap \mathcal{H}X) \setminus \beta$ connecting x with $h^- \cap \partial\mathcal{H}X$. The union $\alpha \cup \alpha^-$ lies in $\mathcal{H}X \setminus \beta$ and divides the segment $\partial h \cap \mathcal{H}X$, as well as $\mathcal{H}X$, into two parts, one containing u_1 and the other containing u_2. This contradicts the existence of the arc β from u_1 to u_2. So all holes U are open convex discs.

Any collection of disjoint open sets in the plane is countable. □

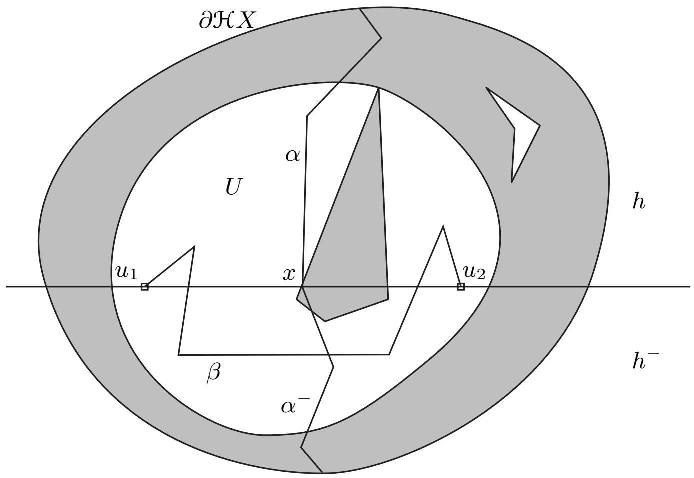


Figure 2. The holes in a 0-tight plane set must be convex (see the proof of Theorem 1.3).

Example. For later reference we mention a curious example of 0-tight plane set, the *limit Swiss cheese*. A Swiss cheese is a round 2-ball (disc) in E^2 from which a union of disjoint open round discs is deleted; see Figure 3. Touching of discs in their boundaries is permitted. If the union of the open discs is everywhere dense, the resulting 0-tight set is called a *limit Swiss cheese*.

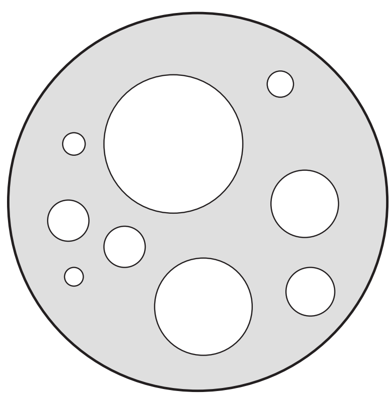


Figure 3. A Swiss cheese has the TPP.

Definition. The subspace $Y \subset X \subset E^N$ is a *local topset* of X if Y has an open neighborhood U in X such that Y is a topset of $\bar{U} \subset E$ (here as elsewhere the bar indicates closure). That means

$$X \neq Y = h \cap \bar{U} = \partial h \cap \bar{U} \neq \varnothing$$

for some half-space h. Note that Y is then open and closed in $\partial h \cap X$.

Theorem 1.4 (topsets). *The connected compact set $X \subset E$ is* 0*-tight if and only if every local topset of X is a connected topset of X; equivalently, if and only if every height function z has one (connected) maximum on X.*

The generalization for maps is as follows:

Definition. The subset $Y \subset X$ is a *local topset* of the map $f : X \to E$ if Y has an open neighborhood U in X such that Y is a topset of the restriction $f|_{\bar{U}} : \bar{U} \to E$, that is

$$X \neq Y = (f|_{\bar{U}})^{-1}h = (f|_{\bar{U}})^{-1}(\partial h) \neq \varnothing,$$

for some half-space h.

Theorem 1.4 (maps). *Let X be connected and compact. The map $f : X \to E$ is* 0*-tight if and only if every local topset Y of f is a connected topset of f; that is, if and only if every z has one connected maximum on X.*

Remarks. If the continuous function $f : X \to \mathbb{R}$ has exactly one local topset, and so does the function $-f$, then f is 0-tight. The map $z \mapsto z^3$ for $|z| = 1$ (see the second exercise on page 3) is not 0-tight.

Proof of Theorem 1.4 for sets. If X is not 0-tight, there is a half-space $h' = \{u \in E : z(u) \geq c'\}$ for which $h' \cap X$ is separated and is the disjoint union of two open closed subsets X_1 and X_2. Let $c \geq c'$ be the maximal value for which $h \cap X_1$ and $h \cap X_2$ are both nonempty, where $h = \{u \in E : z(u) \geq c\}$. One at least of $h \cap X_1$ and $h \cap X_2$ is then a local topset and not a connected topset.

Conversely, if $Y \subset X$ is a local topset in $\partial h \cap X$ and $\partial h \cap X$ is not a connected topset, then $\partial h \cap X = Y \cup Z$ is the disjoint union of Y and Z and $h \cap X$ is the disjoint union of Y and $h \cap X \setminus Y \supset Z$, both open and closed. So $h \cap X$ is separated and X is not 0-tight. $\square$

Exercise. Prove Theorem 1.4 for maps.

Example. Let $\sigma_4 = \mathcal{H}(\{e_1, \ldots, e_5\}) \subset E^4$ be a four-simplex, and let M be the union of five triangles $\mathcal{H}(\{e_i, e_{i+1}, e_{i+2}\})$, for $i = 1, \ldots, 5$ (indices being taken modulo 5). Then M is a 0-tight Möbius band, substantial in E^4. Observe that M contains the 1-skeleton $\mathrm{Sk}_1(\sigma_4)$. Figure 4 shows a projection in E^3 (a 0-tight embedding), as well as a projection in E^2 (a 0-tight map) with folds along the edges e_1e_2, e_2e_3, e_3e_4, e_4e_5, and e_5e_1. The boundary of M is the polygon $e_1e_3e_5e_4e_2e_1$. To prove that $M \subset E^4$ is 0-tight, we observe that any local topset Y of M contains at least one vertex and cannot cut any opposite edge in σ_N transversally. Then Y lies in a supporting half-space h and $Y = h \cap M = \partial h \cap M$.

Remark. For the same reason any union X of $\mathrm{Sk}_1(\sigma_N) \subset E^N$ with some of the simplices of σ_N of various dimensions is 0-tight. In particular, $\mathrm{Sk}_i(\sigma_N)$ is 0-tight, for $1 \leq i \leq N$.

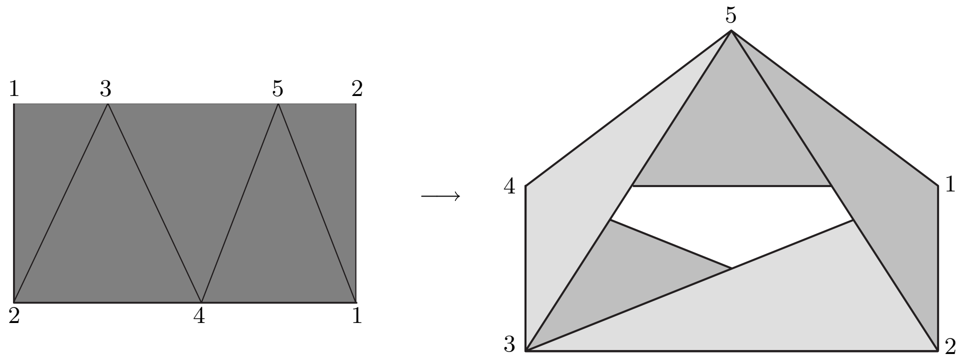


Figure 4. A Möbius band made from five of 2-faces of a tetrahedron. The embedded in E^3 is 0-tight, as is the projection onto the plane of the paper.

Zero-Tight Balls and Spheres of Dimension 1 and 2. We now prove some classification theorems with the help of our tool, the topsets.

A 0-tight embedded arc (1-ball) $X \subset E^N$ in E^N is necessarily a straight line segment. If not, there is a point $y \in X$ not on the line connecting the endpoints x_0 and x_1, and a half-space h containing x_0 and x_1 but not y. So $h \cap X$ is separated and X is not 0-tight.

THEOREM 1.5 (0-TIGHT CIRCLES, SPHERES, BALLS).

(i) *A* 0-*tight embedded closed curve in* E^N *is a plane convex curve.*

(ii) *A* 0-*tight substantial* 2-*sphere in* E^N *is a convex surface* $X = \partial\mathcal{H}X$ *in* 3-*space* E^3.

(iii) [Lastufka 1981] *A* 0-*tight substantial* 2-*ball* (*disc*) *in* E^N *is either*

(a) *a convex plane disc in* E^2, *or*

(b) $X = \partial\mathcal{H}X \setminus U$ *in* E^3, *where the deleted set* U *is an open disc of a plane convex topset* $(\partial\mathcal{H}X)_z = \mathcal{H}X_z$ (see Figure 5).

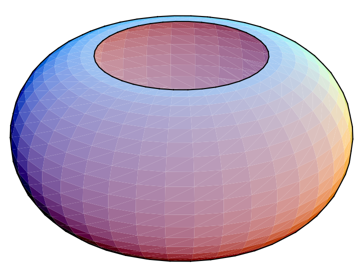


Figure 5. Zero-tight disc in E^3.

PROOF. If $X \subset E$ is 0-tight, then by Theorem 1.2 so are any of its top*sets. An E_0-top*set of X is a point; an E_1-top*set of X is a line segment; possible E_2-top*sets of X are described in Theorem 1.3.

(i) If X is a 0-tight closed curve, or "circle" for short, in E^2, then

$$\partial\mathcal{H}X = \bigcup_z \mathcal{H}X_z = \bigcup_z X_z$$

is a circle embedded in X, hence equal to X, and X is a plane convex curve.

Now let X be a 0-tight substantial circle X in E^N, where $N \geq 3$. Choose a nonplanar 4-gon with vertices u_1, u_2, u_3, u_4 on X, cyclically ordered on this circle. Let h be a half-space whose boundary ∂h passes through the midpoints of u_1u_2, u_2u_3, u_3u_4, and u_4u_1, and such that h does not contain u_1. Then h contains u_2 and u_4 and not u_1 and u_3; thus $h \cap X$ is separated and X cannot be 0-tight.

(ii) Let X be a 0-tight 2-sphere in E^N. First let $N = 3$. Any topset of X is a point, a line segment, or an E_2-topset $Y = X_z$. We show that in the latter case Y has to be convex. Indeed, X contains the convex envelope $\partial\mathcal{H}Y \subset Y \subset X$. If Y is not convex then $\mathcal{H}Y \setminus Y$ contains at least one hole U as component and the open half-space $\mathring{h} = E^3 \setminus h_z$ intersects X in $X \setminus Y$, which has nonempty open pieces in X, separated by Y, also separated by the circle $\partial U \subset X = S^2$. This contradicts 0-tightness. So all topsets X_z of X are convex and the "convex" surface $\partial\mathcal{H}X$ is contained in the 2-sphere X. Then X is equal to this convex surface $\partial\mathcal{H}X$.

Next suppose X is a 0-tight substantial 2-sphere in E^N for $N > 3$. Let Y be as before a nonconvex E_2-top*set. Then $\partial\mathcal{H}Y \subset Y \subset E^2 \subset E^N$. Choose x_1 and x_2 in $X \setminus Y \subset X \setminus \partial\mathcal{H}Y$ in different components of $X \setminus Y$. There exists a half-space h such that $x_1, x_2 \in h$, but $h \cap Y = \varnothing$. Then $h \cap X = h \cap (X \setminus Y)$ is separated, contradicting 0-tightness of X. So all E^2-top*sets are convex. If Y is an E^3-top*set, then all topsets of Y are convex and the 2-sphere $\partial\mathcal{H}Y \subset Y \subset X$ must be equal to X. It cannot be substantial in E^N, for $N > 3$. So there is no E_3-top*set. Let k be either N or the smallest number $k < N$ with $k > 3$ for which there is an E_k-top*set Y. Then all topsets of X and of Y, are convex. So $\partial\mathcal{H}X \subset X$ and $\partial\mathcal{H}Y \subset Y \subset X$. But the dimension of $\partial\mathcal{H}Y$ is $k-1 > 2$. This is absurd for dimension reasons. The desired result follows.

(iii) A 0-tight 2-disc in E^2 is a convex disc by Theorem 1.2. We can therefore assume the 0-tight disc X embedded in E^N, where $N \geq 3$. First let $N = 3$. A 0-tight nonconvex top*set X_z is necessarily an E_2-topset X_z, obtained from the convex disc $\mathcal{H}X_z$ by deleting one or more open sets U. If U is one of them and ∂U is not the boundary ∂X of the disc X, then $X \setminus X_z = X \setminus h_z$ contains at least two points x_1 and x_2, which are separated by ∂U. As $X \setminus h_z = (E \setminus h_z) \cap X$ is then separated and $E \setminus h_z$ is an open half-space, this contradicts 0-tightness. There is only *one* boundary ∂X for X, so that only one topset of X is nonconvex and it has only one convex hole U. This is the conclusion of the theorem for $N = 3$.

Finally let X be a 0-tight substantial 2-disc in E^N, where $N > 3$. Let Y be a nonconvex E^2-top*set. As before, $\mathcal{H}Y \setminus Y$ contains at least one hole U as component. If ∂U is not the boundary ∂X of X, then $X \setminus Y$ has at least two points x_1 and x_2 that are separated in $X \setminus Y \subset X \setminus \partial U$. As in (ii) we find a contradiction. Also there can be at most one hole U. Since ∂U must be the boundary of X, we can fill in U to obtain $U \cup X$, a 0-tight embedded S^2 in E^N, for $N > 3$. This led to a contradiction in (ii), and part (iii) of Theorem 1.5 is proved. □

Problem. Let Y be a plane 2-disc from which a number of disjoint open 2-discs are deleted. Determine all continuous 0-tight embeddings of Y in E^N. Lastufka [1981] found that all 0-tight immersions are embeddings.

Background: Manifolds and the Topological Classification of Surfaces. Although we are mainly interested in surfaces, we make some general remarks on higher-dimensional manifolds.

A compact *n-dimensional topological manifold* (or simply *n-manifold*) is a metrizable topological space M such that each point $x \in M$ can be assigned a homeomorphism or *chart*

$$\kappa_x : U_x \to \kappa_x(U_x) \subset h^n = \{u \in \mathbb{R}^n : u^n \geq 0\} \subset \mathbb{R}^n$$

of some open neighborhood U_x onto $\kappa_x(U_x)$, an open set in h^n. The maps

$$\kappa_{yx} = \kappa_y \kappa_x^{-1} : \kappa_x(U_x \cap U_y) \to \kappa_y(U_x \cap U_y),$$

for $x, y \in M$, are local homeomorphisms in h^n. The pieces $\kappa_x^{-1}(\partial h^n)$ constitute the *boundary* ∂M of M. If $\partial M = \varnothing$, we say that M is a *closed manifold*, that is, compact without boundary. If $\partial M \neq \varnothing$, it is easy to see that ∂M is a closed $(n-1)$-manifold.

If choices for κ_x are given such that every change-of-coordinate map κ_{yx} is smooth (by which we mean C^∞, that is, such that all derivatives exist), we can consider the set of *all* functions locally generated from functions $u^i \circ \kappa_x : U_x \to \mathbb{R}, i = 1, \ldots, n$ by smooth compositions ψ (smooth functions of n variables $u^1, \ldots, u^n$): $\psi(u^1 \circ \kappa_x, \ldots, u^n \circ \kappa_x)$. This set of functions is called a *smoothing* of M, and M with this smoothing is called a *smooth manifold*. Any smooth closed manifold determines a topological manifold by forgetting the smoothness. It is known that for $n \leq 3$ every closed topological n-manifold can so be obtained from some smooth n-manifold, which smooth n-manifold is moreover unique but for differentiable equivalence. For $n \geq 4$ there are topological closed manifolds for which existence fails, and others for which existence holds but not uniqueness. (The result for $n = 4$ is due to deep work of M. Freedman [1982] and S. Donaldson [1983; 1986].)

A *finite simplicial complex* W can be defined as a union of affine subsimplices $\mathcal{H}(e_{i_0}, \ldots, e_{i_r})$ of various dimensions $r \geq 0$ of an N-simplex $\sigma_N = \mathcal{H}(e_0, \ldots, e_N)$ in E^N for some N. A homeomorphism $\lambda : W \to X$ of W onto a given topological space X is called a *triangulation* of X. If $\nu : W_1 \to W$ is a triangulation of W that is affine on each simplex of W_1 and restricts to a triangulation onto each simplex of W, then the triangulation $\lambda \circ \nu : W_1 \to X$ is called a *subdivision* of $\lambda : W \to X$.

A triangulation of a closed n-manifold is called *Brouwer* if the union of all n-simplices with a common vertex, admits a homeomorphic embedding into E^n for which the image of each n-simplex is an affine n-simplex in E^n. One can prove that every smooth n-manifold M has a Brouwer triangulation with smooth embeddings for each simplex, which is unique modulo subdivision and modulo diffeomorphisms of M.

A topological n-manifold has, for $n \leq 3$, a unique Brouwer triangulation [Moise 1952]. This is not so for $n \geq 4$. Some subdivision of a Brouwer triangulation of a closed n-manifold has in any case a compatible smoothing for $n \leq 7$, which is unique for $n \leq 6$, but not so for larger dimensions. The n-sphere has a few-vertex triangulation as the boundary of the n-simplex $\sigma_n \subset E^n$. This is a Brouwer triangulation. A few-vertex triangulation of the Möbius band was seen in Figure 4. In these examples the number of vertices is minimal.

An *orientation* of a manifold M at a point $x \in M$ is a choice of one of the two generators of the homology group $H_0(M,\, M\backslash x;\, \mathbb{Z}) \approx \mathbb{Z}$. For an embedded arc $I \subset M$ with end points x_0 and x_1, the axioms of homology [Spanier 1966, p. 294 ff.] give natural isomorphisms by inclusions of spaces

$$\mathbb{Z} \approx H_0(M,\, M\backslash x_0;\, \mathbb{Z}) \leftarrow H_0(M,\, M\backslash I;\, \mathbb{Z}) \to H_0(M,\, M\backslash x_1;\, \mathbb{Z}),$$

which permit unique transport of an orientation from x_0 to x_1. Transport along homotopic paths gives the same result. If the result is the same for any paths connecting two points x_0 and x_1 then it defines a global orientation on M, now assumed connected, and M is called *orientable* and *oriented* by this choice. The local orientations determine a two-point bundle (double covering of M) and M is orientable if and only if this bundle is a product bundle. The Möbius band is not orientable.

Let B_x be an embedded n-ball around $x \in M$, with boundary the embedded $(n-1)$-sphere $\partial B_x = S_x$. Then there are natural isomorphisms (by excision and exactness)

$$H_0(M,\, M\backslash x;\, \mathbb{Z}) = H_0(B_x, S_x; \mathbb{Z}) \to H_0(S_x; \mathbb{Z}) \approx \mathbb{Z}.$$

Thus an orientation of M at x determines a generator of $H_0(S_x; \mathbb{Z}) = \mathbb{Z}$.

Given connected closed topological n-manifolds M_1 and M_2, one constructs a new n-manifold $M_1 \# M_2$ called the *connected sum* as follows (Figure 6). For $i = 1, 2$, delete from M_i, a small open ball U_i interior of an n-ball $B_i \subset M_i$ with boundary an $(n-1)$-sphere $S_i = \partial B_i$, and glue $M_1 \setminus U_1$ to $M_2 \setminus U_2$ by a homeomorphism $\lambda : \partial U_1 \to \partial U_2$. If M_1 and M_2 are orientable and oriented, one chooses λ moreover in such a way that for $x_1 \in (M_1 \setminus B_1)$ and $x_2 \in (M_2 \setminus B_2)$, there is

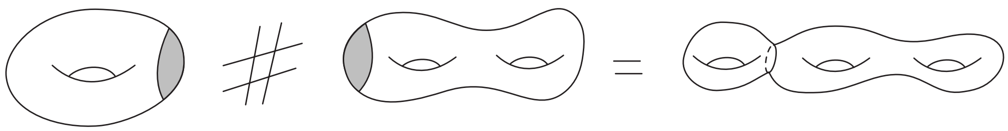


Figure 6. The connected sum of two closed surfaces.

an orientation on $M_1 \# M_2$ such that local generators that define orientations correspond in the sequence of isomorphisms:

$$\begin{aligned}\mathbb{Z} \longleftrightarrow H_0(M_1,\, M_1\backslash x_1;\, \mathbb{Z}) &\longleftrightarrow H_0(M_1 \# M_2,\, M_1 \# M_2\backslash x_1;\, \mathbb{Z})\\ &\longleftrightarrow H_0(M_1 \# M_2,\, M_1 \# M_2\backslash x_2;\, \mathbb{Z}) \longleftrightarrow H_0(M_2,\, M_2\backslash x_2;\, \mathbb{Z}_2).\end{aligned}$$

It is known (only recently for $n = 4$) that the connected sum for topological closed manifolds so defined is unique up to equivalence. If at least one of M_1 and M_2 is nonorientable, then we get always the same (unique) connected sum. For closed oriented n-manifolds M_1 and M_2 and $n \geq 3$, the orientations may give two nonhomeomorphic results. For any two closed surfaces ($n = 2$) the connected sum is however a unique closed surface.

Remark. A continuous map $f : X \to Y$ is called an *embedding* if $f : X \to f(X)$ is a homeomorphism. It is called an *immersion* if $f|_{U_x} : U_x \to Y$ is an embedding for some neighborhood U_x of any $x \in X$. If X and Y are manifolds of dimension n and $N \geq n$, the immersion f is called *tame* in case for any $x \in X$ there is a commutative diagram concerning neighborhoods $U_x \subset X$ and $U_y \subset Y$, $y = f(x)$, charts κ_x and κ_y, and a linear embedding i :

$$\begin{tikzcd} U_x \arrow[r, "f"] \arrow[d, "\kappa_x"] & U_y \arrow[d, "\kappa_y"] \\ R^n \arrow[r, "i"] & R^N. \end{tikzcd}$$

These definitions for topological manifolds have natural counterparts in the smooth context. Most of these facts were known for dimension $n \neq 4$ around 1971. For $n = 4$ the breakthrough came since 1982 with the work of Casson [1986], Freedman [1982] and Donaldson [Donaldson 1983; 1986].

Every orientable closed surface is the (repeated) connected sum of a two-sphere S^2 and some number $g \geq 0$ of tori T:

$$M_g = T \# T \cdots \# T, \qquad M_0 = S^2.$$

We say M_g has *genus* g and has *Euler characteristic* $\chi = 2 - 2g$. Every nonorientable closed surface is a connected sum of the form

$$P \# M_g, \qquad \text{for some } g \geq 0,$$

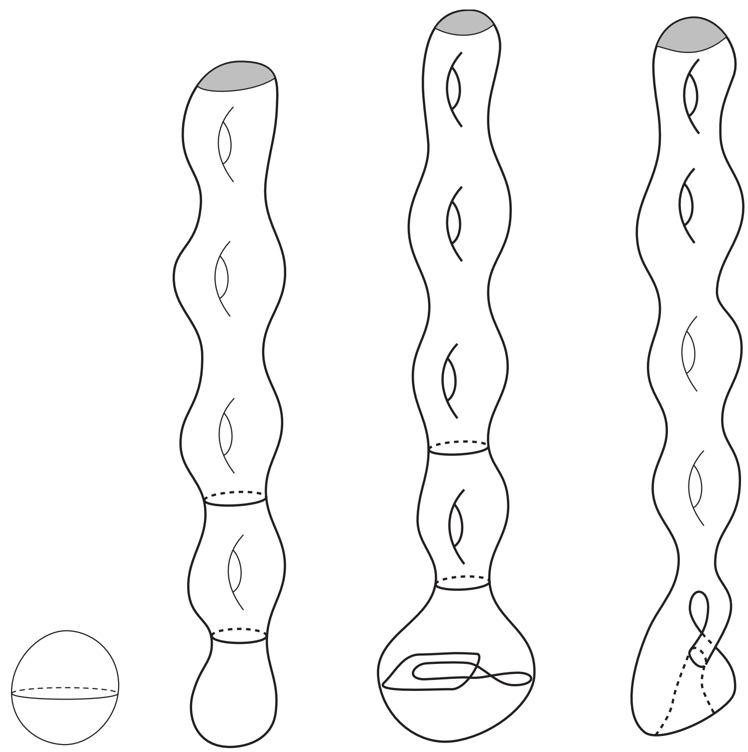


Figure 7. The classification of closed surfaces.

or of the form

$$P \# P \# M_g = K \# M_g, \qquad \text{for some } g \geq 0,$$

where P is the real projective plane and $K = P \# P$ is the Klein bottle. Surfaces of these types have Euler characteristic $\chi = 1 - 2g$ and $\chi = -2g$, respectively. See Figure 7.

Any compact surface with boundary is obtained from a unique closed surface by deleting the interiors of a finite number $r \geq 1$ of disjoint embedded 2-balls (discs). The same classifications for surfaces hold in the smooth as in the triangulated (modulo subdivision) context.

The Möbius band is a projective plane from which the interior of one open 2-ball is deleted, and the projective plane is obtained from the Möbius band by closing it with a disc.

Zero-Tight Immersions of Surfaces in the Plane. If a compact surface $(M, \partial M)$ admits an immersion $f : M \to E^2$ in the plane, it is orientable (because it gets, by virtue of the immersion, a unique orientation from an orientation of E^2) and has at least one boundary component (embedded circle), because any point on the boundary of $f(M)$ must be an image of boundary points of M by definition of immersion. For any such surface $M_g^{(r)}$ of genus g, with $r \geq 1$ interiors of disjoint discs deleted, there is an immersion in the plane, as seen in the examples in Figure 8.

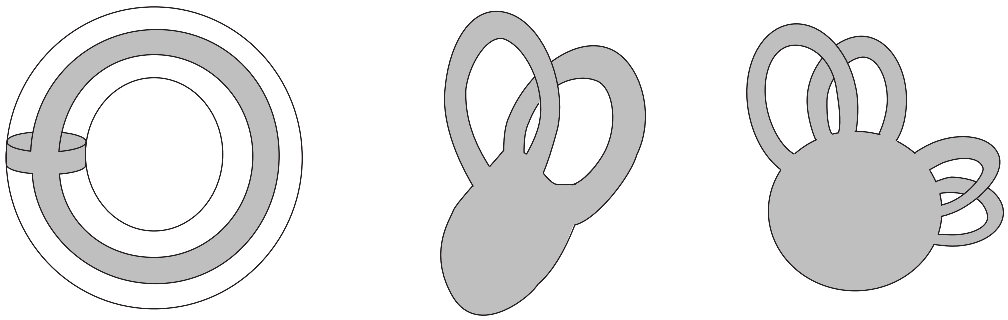


Figure 8. Left: The surface with boundary obtained by removing an open disk from the torus $T^2 = M_1$ is denoted $M_1^{(1)}$. Middle: Immersion of $M_1^{(1)}$ in E^2. Right: Immersion of $M_2^{(1)}$ (genus-two surface minus a disk) in E^2.

Next let $f : X = M_g^{(r)} \to E^2$ be a 0-tight immersion. By Theorem 1.2 any topset $X_z = f^{-1}(\partial h_z)$ is connected and immersed in the line ∂h_z. So it is embedded as a convex set (a point or a segment). The union of these topsets, $f^{-1}\partial \mathcal{H} f(X)$, is then a topological circle in X embedded into the convex curve $\partial \mathcal{H} f(X)$, and this circle is one boundary component $\partial_1 X$ of $X = M_g^{(r)}$. If there are no other boundary components ($r = 1$), then $f : X \to E$ is an immersion onto the convex disc $\mathcal{H} f(X)$, and every image-point $Y \in \mathcal{H} f(X)$ is covered by the same finite number of points by the immersion f. This number is 1 because that's what it is on the boundary. Then f is an embedding onto $\mathcal{H} f X$ and X is a disc: $X = M_g^{(r)} = M_0^{(1)}$. In particular, $g = 0$. For $g \geq 1$, the 0-tight immersed surface must have at least two boundary circles, $\partial X = \partial_1 X \cup \cdots \cup \partial_r X$, and the first $f \partial_1 X = \partial \mathcal{H} f X$ is embedded and convex.

Next consider a boundary point x on another component $\partial_i X$, where $i \geq 2$. Let

$$\kappa_x : U_x \to h^2 = \{(u^1, u^2) \in \mathbb{R}^2 : u^2 \geq 0\}$$

be a chart for a connected neighborhood U_x of x in X, so small that the restriction $f|_{U_x}$ is an embedding into E. (See Figure 9.) We can assume that $f\big(\kappa_x^{-1}(\partial h^2)\big)$ and $\kappa_x^{-1}(\partial h^2)$ are connected. The first of these sets is an open arc (1-manifold). By 0-tightness of f, no straight line segment $[y_1, y_2] \subset E$ with end-points y_1 and y_2 in $f\big(\kappa_x^{-1}(\partial h^2)\big)$ can divide $f(U_x)$. That is, $\big(f(U_x) \setminus [y_1, y_2]\big)$ must be connected. Components of $U_x \setminus [y_1, y_2]$ that touch $[y_1, y_2]$ in interior points must then be convex. Therefore $f\big(\kappa_x^{-1}(\partial h^2)\big)$ is nonconvex (call it *concave*) with respect to the interior of $f(U_x)$. We conclude that f immerses each boundary circle $\partial_i(X)$ for $i \geq 2$ in a locally concave manner.

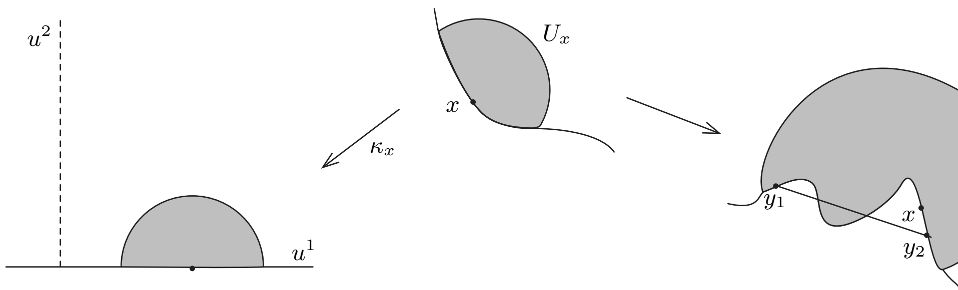


Figure 9. A 0-tight immersion of a surface with boundary in E^2 maps all but one boundary component in a concave manner.

Figure 10 shows 0-tight immersions of $M_1^{(2)}$ and $M_2^{(2)}$. Following the pattern of this figure we can conclude (and summarize):

THEOREM 1.6. *There are* 0*-tight immersions of the compact surfaces* $X = M_g^{(r)}$ (*genus* g, $r \geq 1$ *holes*) *in the plane if* $g = 0$, $r \geq 1$ *and if* $g \geq 1$, $r \geq 2$, *but not if* $g \geq 1$, $r = 1$. *The boundary components of* $\partial X = \partial_1 X \cup \cdots \cup \partial_r X$ *have one convex image* $f(\partial_1 X)$ *and further locally concave immersions* $f : \partial_i X \to E$ *for* $i \geq 2$.

EXERCISE. In Figure 10, left, the 0-tight immersion of $\mathrm{Sk}_1(\sigma_3)$ "embeds in" the 0-tight immersion of $M_1^{(2)}$.

Zero-Tight Immersions of Closed Surfaces in Space. In Figure 11 we show examples of 0-tight immersed surfaces in E^3. Every height function z is seen to have, and must have, exactly one local top set, a maximum, and one local minimum (maximum of $-z$). This characterizes 0-tightness by Theorem 1.4. We now prove for smooth immersions:

THEOREM 1.7 (smooth version). *If* $f : M \to E^3$ *is a* 0*-tight smooth immersion of a closed surface* $M \neq S^2$, *there is a decomposition of* M *as a union of surfaces with boundary* $M = M^+ \cup M^-$, *where* M^+ *is connected*, $M^- = \bigcup_{j=1}^r M_j^-$, *and*

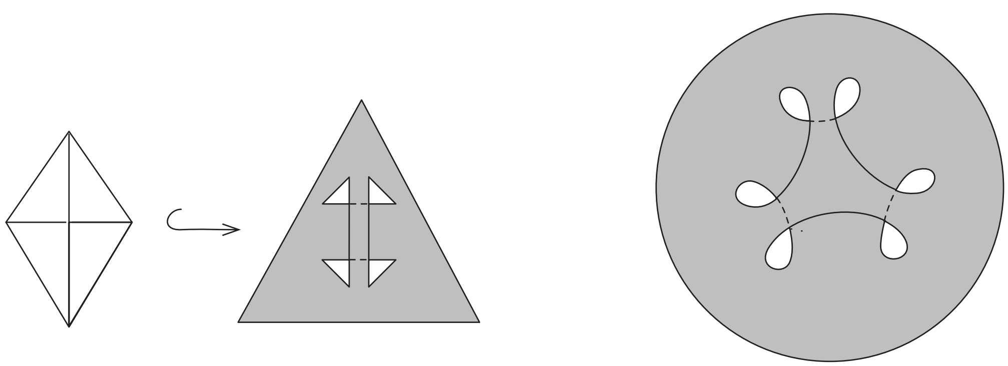


Figure 10. Zero-tight immersions of $M_1^{(2)}$ and $M_2^{(2)}$.

Figure 11. A 0-tightly smoothly immersed closed surface in E^3 splits into one component M^+ of positive curvature and one or more components M_j^- of negative curvature. These pieces join along plane curves X_i, called windows.

the boundary $\partial M^+ = M^+ \cap M^- = \bigcup_{i=1}^r X_i$ *is a union of* $r \geq 2$ *closed curves with the following properties*:

(i) $M^+ \to fM^+ \subset E^3$ *is embedded into the convex surface* $\partial \mathcal{H} fM$ *as the complement of the union of the interiors of* r *disjoint plane convex discs*, *called windows. The Gauss curvature* $K(p)$ *is nonnegative for* $p \in M^+$ *and nonpositive for* $p \in M^-$.

(ii) *No local topset* $A \subset M$ *is contained completely in* $\mathring{M}^-$.

(iii) *Each closed curve* X_i *carries an essential cycle of* $H_1(M, \mathbb{Z}_2)$.

(iv) $\chi(M) = \chi(M^+) + \sum_{j=1}^r \chi(M_j^-)$, $\chi(M^+) \leq 0$, *and* $\chi(M_j^-) \leq 0$.

(v) $M_{K>0} = \partial \mathcal{H}(M)_{K>0}$.

(For convenience we will write M^+ for fM^+ and X_i for fX_i.)

Proof. The 0-tight restriction of f to an E^k-top*set X is a 0-tight immersion of a connected set onto one point for $k = 0$, and onto a straight segment for $k = 1$. So for $k = 0$ and $k = 1$, it is an embedding onto a convex set. An E^2-topset $X = M_z \subset M$ immerses 0-tightly into a subset of the plane convex set $\mathcal{H} fX$ and restricts to an embedding on (each of the points of) the circle $X' = f^{-1} \partial \mathcal{H} fX \subset X \subset M$. Suppose the circle X' bounds in M, that is $M \setminus X'$ has two components. If both components have image points under the topset level of z, then $z \circ f$ has two minima on M, and f is not 0-tight. So one of the components immerses into $\mathcal{H} fX$ and its boundary into $\partial \mathcal{H} fX$. Such an immersion is a topological covering of $\mathcal{H} fX$. As it covers each boundary point once, it is an embedding onto $\mathcal{H} fX$, and $X = f^{-1} \mathcal{H} fX$ is an E^2-topset disc $X \subset M$. The union of all convex topsets $X = fX \subset M$ so far discussed is, by definition, the set $M^+ = fM^+ \subset \partial \mathcal{H} fM$. It is bounded by plane convex curves $X' = fX' = \partial \mathcal{H} fX$, for which $X' \subset M$ is not bounding in M. The circle X' then carries a generator of $H_1(M, \mathbb{Z}_2)$ and is called an (essential) *top cycle*. This proves (iii). As (ii) is obvious, there remains the proof of the last part of (iv). Suppose M^+ or M_j^- has only one boundary component. Then this boundary component bounds in M and is a top cycle, a contradiction. $\square$

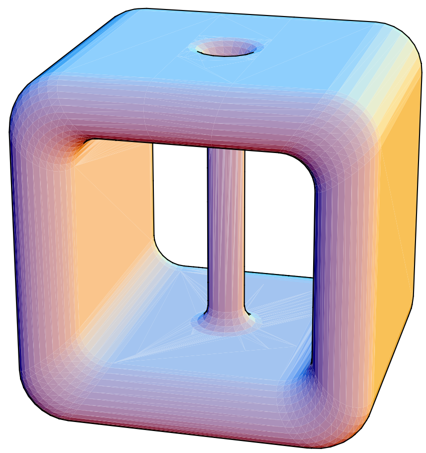


Figure 12. A 0-tight smooth immersion of the nonorientable surface of Euler characteristic -2 ($K \# T$). There are three windows (front, back, and top) and one curve of self-intersections, where the central column goes through the "ceiling".

Corollary. There is no smooth tight immersion of the projective plane P into E^3.

Proof. By (iv) $\chi(M) \leq 0$, but $\chi(P) = 1$. Another proof: Every closed embedded curve in P which is nonbounding has a Möbius band as neighborhood. Every top cycle for a smooth 0-tight surface in E^3 can have only a (trivial) band as neighborhood. "The M-normal bundle is trivial along the top cycle." $\square$

For topological immersions, we mention without proof a more subtle result:

Theorem 1.8 (topological version). *If $f : M \to E^3$ is a 0-tight C^0-immersion of a closed surface M, there is a decomposition of M as $M = M^+ \cup \mathring{M}^-$, $M^+ \cap \mathring{M}^- = \varnothing$, $\mathring{M}^- = \cup \mathring{M}_j^-$, M^- an open surface, satisfying furthermore the following properties*:

(i) *$M^+ = fM^+ \subset E^3$ is an embedded subset of M, embedded in the convex surface $\partial \mathcal{H} fM = \partial \mathcal{H} fM^+$, as the complement of the union of r disjoint interiors of plane convex sets $\mathcal{H} fX_j$. Here $X_j \subset M^+ \subset M$ is an embedded circle.*

(ii) *No local topset $A \subset M$ is completely contained in the open set $\mathring{M}^-$.*

(iii) *Each closed curve X_j compactifies one end of $\mathring{M}^-$ as a boundary.*

(iv) $\chi(M) = \chi(M^+) + \sum_{j=1}^{r} \chi(M_j^-)$, $\chi(M^+) \leq 0$, $\chi(M_j^-) \leq 0$.

See Figure 13 for an example, a 0-tight torus.

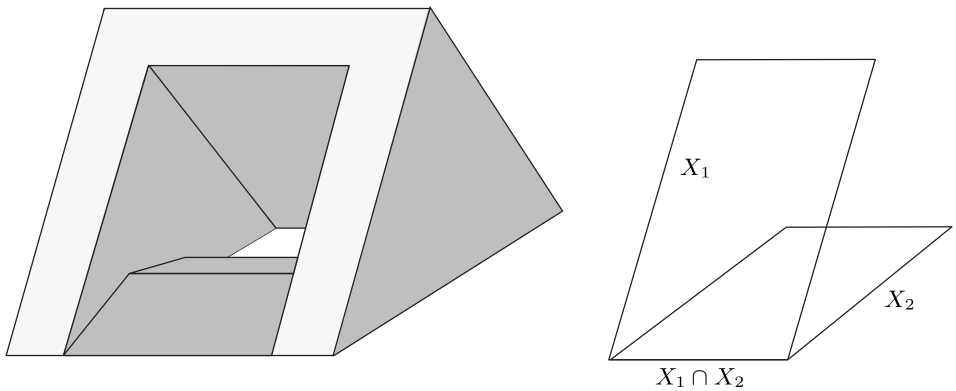


Figure 13. A 0-tightly embedded PL torus, showing that windows can intersect (Theorem 1.8).

2. Curvature and Knots

Definitions of Curvature. A *closed curve* $\gamma : S^1 \to \mathbb{R}^N$ in Euclidean space is a continuous immersion of the circle

$$S^1 = \{(\cos 2\pi t, \sin 2\pi t) : t \in \mathbb{R}\}.$$

It can be parametrized and oriented by t mod 1. For convenience, we only discuss the case $N = 3$.

A *polygon* P (n-gon γ_n) is a closed curve γ_n with vertices $u_i = \gamma_n(t_i) \in \mathbb{R}^3$, i mod n, for $0 \le t_1 < t_2 < \ldots < t_n < 1$, connected by straight line segments $\gamma_n(t)$, $t_{i-1} \le t \le t_i$.

The curvature on a curve is a measure. For a polygon the measure is concentrated in the vertices. Let $\alpha_i \in [0, \pi)$ be the angle between two successive edges meeting at u_i, that is the angle between the unit vectors v_{i-1} and v_i, where

$$v_j = \frac{u_j - u_{j-1}}{\|u_j - u_{j-1}\|} \in S^2, \quad j \text{ taken mod } n.$$

Choose an open interval U_i on γ_n between u_{i-1} and u_{i+1} that covers u_i. Then the *normalized curvature* on U_i is $\tau(\gamma_n, U_i) = \alpha_i/\pi$. The (normalized) *total curvature* of γ_n is $\tau(\gamma_n) = \sum_i (\alpha_i/\pi)$. We connect the points v_i and v_{i+1} in S^2 by a geodesic segment of length α_i in S^2 and obtain a "tangential image", a continuous image of an oriented circle. This leads to the following equivalent definition:

Definition 2.1. The (normalized) *total absolute curvature* $\tau(\gamma_n)$ is $1/\pi$ times the length $L(\tan \gamma_n)$ of the tangential image $\tan \gamma_n : S^1 \to S^2$ in S^2.

Next we consider the normal unit vectors along the open edge from u_{i-1} to u_i. At each point they form a unit circle, which we carry over by parallel displacement to $0 \in \mathbb{R}^3$ to obtain a great circle $S^1(v_i)$ in S^2 orthogonal to $v_i \in S^2$. We rotate

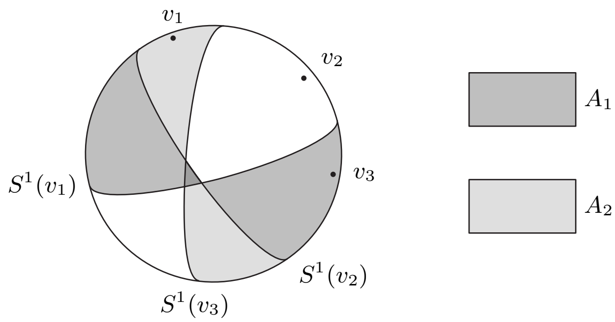


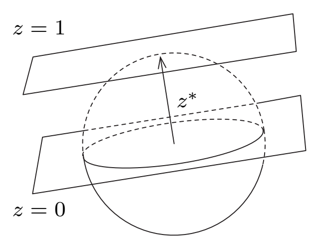
Figure 14. At a vertex of a polygonal line the direction changes from v_i to v_{i+1}. The smaller pair of segments bounded by the circles of normals $S^1(v_i)$ and $S^1(v_{i+1})$ is denoted A_i.

$S^1(v_i)$ into $S^1(v_{i+1})$ through circles $S^1(v)$, orthogonal to v, moving along the geodesic arc with ends v_i and v_{i+1} in S^2. The point set swept by the circles is denoted A_i. It consists of two congruent diametrical sectors (see Figure 14). Let $|A_i|$ be the area of A_i and $|S^2|$ the area of S^2. Clearly $|A_i|/|S^2| = \alpha_i/\pi$. So we have the following equivalent definition:

Definition 2.2. Let $\tau(\gamma_n, U_i) = |A_i|/|S^2|$ be the swept-out area of the unit normal bundle on the interval U_i. The total absolute curvature is

$$\tau(\gamma_n) = \sum_i \tau(\gamma_n, U_i).$$

Critical Points. Let $z : \mathbb{R}^3 \to \mathbb{R}$ be a linear function satisfying $\|z\| = 1$, and denote the gradient of z by $z^* \in S^2$ (see right). The restriction $z|_{U_i}$ may or may not have a maximum or minimum on U_i. We assign to z (and to z^*) the number $\mu_z(\gamma_n, U_i)$ of such maxima and minima. Clearly $\mu_z(U_i) = 1$ if $z^* \in \mathring{A}_i$, the interior of A_i, and $\mu_z(U_i) = 0$ if $z^* \notin A_i$. Functions z that have a constant value on an edge of γ_n are called *degenerate*; the set of $z^* \in S^2$ for which this occurs has measure zero and can be neglected. Let $\mathcal{E}_z$ denote the expectation or mean value with respect to the standard SO(3)-invariant measure for z^* on S^2. If $|A_i|$ and $|S^2|$ denote the area of A_i and S^2, respectively, then clearly



$$\mathcal{E}_z(\mu_z(\gamma_n, U_i)) = |A_i|/|S^2| = \alpha_i/\pi.$$

Therefore we have another equivalent definition:

Definition 2.3. The total absolute curvature is the mean number of critical points of z^* on γ, with respect to the standard measure on S^2.

$$\tau(\gamma_n, U_i) = \mathcal{E}_z \mu_z(\gamma_n, U_i), \qquad \tau(\gamma_n) = \mathcal{E}_z \mu_z(\gamma_n).$$

Total Curvature for General Closed Curves. Let $\gamma : S^1 \to \mathbb{R}^N$ be a closed continuous immersed curve. A polygon P with vertices $u_i = \gamma(t_i)$, i mod n, $0 \leq t_1 \leq \ldots \leq t_n < 1$ is called an *inscribed polygon*; we denote this by $P \prec \gamma$. We propose:

Definition 2.4. The *total curvature* of a continuous closed immersed curve $\gamma : S^1 \to \mathbb{R}^N$ is the least upper bound of $\tau(P)$ for inscribed polygons

$$\tau(\gamma) = \sup_{P \prec \gamma} \tau(P) \leq \infty.$$

In this section we will recover the important Definition 2.3 for general closed curves thanks to the following result:

Theorem 2.5. *If the total curvature $\tau(\gamma)$ of a continuous immersed curve $\gamma : S^1 \to \mathbb{R}^N$ is finite, it equals the mean number of maxima and minima*:

$$\tau(\gamma) = \mathcal{E}_z \mu_z(\gamma).$$

To begin the proof, we first prepare the definition of $\mu_z(\gamma)$ and again elaborate only the case $N = 3$. The function $z|_\gamma$ is called *degenerate* if it is constant on some maximal arc on γ. If this arc is on a straight line in $\mathbb{R}^3$, it "belongs to" at most a great circle of points $z^* \in S^2$. Such arcs can be counted by their nonincreasing lengths in the parameter t. The corresponding circles on S^2 are countable in number, and their union has measure zero in S^2. Any nonlinear planar arc "belongs to" two points z^* and $-z^*$ in S^2. Such arcs can be counted as well. It follows that the points z^* that belong to degenerate $z|_\gamma$ have a union of measure zero in S^2.

Definition. *One maximum* of $z|_\gamma$ is a point or a maximal arc σ on γ in which z is constant, and such that some open neighborhood of σ gives no greater values for $z|_\gamma$. One maximum of $-z|_\gamma$ is called *one minimum* of $z|_\gamma$. If $z|_\gamma$ is constant we count one maximum and one minimum.

The total number of maxima and minima of $z|_\gamma$ is denoted $\mu_z(\gamma) \leq \infty$. It is twice the number $b_z(\gamma)$ of maxima. Note that $b_z(\gamma)$ can be infinite for a nondegenerate function, when the isolated (maximal) points have one or more accumulation points.

For the proof of Theorem 2.5 we need some lemmas.

Lemma 2.6. *If $P \prec \gamma$ is an inscribed polygon of γ, then $\mu_z(\gamma) \geq \mu_z(P)$ for all z.*

Proof. Suppose $z|_P$ attains a maximum on P in one point, the vertex $u_i = \gamma(t_i)$. Then

$$z(\gamma(t_{i-1})) < z(\gamma(t_i)), \qquad z(\gamma(t_{i+1})) < z(\gamma(t_i)).$$

Therefore z has a maximum on the open interval $\{\gamma(t) : t_{i-1} < t < t_{i+1}\}$. If $z|_P$ attains a maximum at the maximal segment from u_i to u_{i+l} of P, then

$$z(\gamma(t_{i-1})) < z(\gamma(t_i)) = \ldots = z(\gamma(t_{i+l})) > z(\gamma(t_{i+l+1})),$$

and z has a maximum on the open interval $\{\gamma(t) : t_{i-1} < t < t_{i+l+1}\}$.

If z is constant on P the conclusion is obvious. The same applies to minima of z and the lemma follows. $\square$

Corollary 2.7. *If $P \prec P'$ is an inscribed polygon in the polygon P' (obtained, for example, by deleting one vertex of P'), then $\mathcal{E}_z\mu_z(P') \geq \mathcal{E}_z\mu_z(P)$, and so $\tau(P') \geq \tau(P)$.*

Corollary 2.8. $\mu_z(\gamma) \geq \sup_{P\prec\gamma} \mu_z(P)$.

Let $\{t_i : i = 1, 2, \ldots\}$, with $0 \leq t_i < 1$ and $t_j \neq t_h$ for $j \neq h$, be a countable dense subset of $[0, 1)$, and let $P_j \prec \gamma$ be the inscribed polygon with vertices $\gamma(t_1), \ldots, \gamma(t_j)$. Then P_j is said to *converge* to $\lim_{j\to\infty} P_j = \gamma$. From the definition of $\tau(\gamma)$ and Corollary 2.7, follows another equivalent definition of curvature:

Definition 2.9. $\tau(\gamma) = \lim_{P\to\gamma} \tau(P) \leq \infty$.

Corollary 2.10. $\mu_z(\gamma) \geq \lim_{P\to\gamma} \mu_z(P)$.

Next suppose $z|_\gamma$ is nondegenerate. Then $z|_{P_j}$ has a maximum (minimum) as near as we please to any finite number of isolated maxima (minima) of $z|_\gamma$, for j sufficiently large. Therefore:

Lemma 2.11. *If $z|_\gamma$ is nondegenerate, then $\mu_z(\gamma) \leq \lim_{P\to\gamma} \mu_z(P) \leq \infty$.*

Theorem 2.5 follows from Corollary 2.10 and Lemma 2.11.

Exercise. Define suitably the total curvature of an immersed closed curve in $\mathbb{R}^N$ and an immersed arc $\gamma : \{t : 0 \leq t \leq 1\} \to \mathbb{R}^N$, so that Theorem 2.5 applies also to arcs. *Hint:* Use μ_z(interior of arc γ) instead of $\mu_z(\gamma)$ to avoid curvature at the end points.

Exercise. Consider the arc $\gamma \subset \mathbb{R}^2$ defined by

$$u = (u^1, u^2) = \begin{cases} (t, 0) & \text{for } -1 \leq t \leq 0, \\ (t^\alpha, \sin 2\pi t^{-1}) & \text{for } 0 < t \leq 1. \end{cases}$$

Prove that:

(a) $\gamma(t)$ has a tangent for all t if and only if $\alpha > 1$.
(b) The tangent depends continuously on t if and only if $\alpha > 2$.
(c) $\tau(\gamma) < \infty$ if and only if $\alpha > 3$.
(d) γ has continuous second derivatives if and only if $\alpha > 4$.

Hint: Consider by way of comparison the inscribed polygons P_j, $j \to \infty$, with vertices for $t = 0$, $t = -1$, $t^{-1} = 1$, and for $t^{-1} = (1 + 2i)/4$ for $i = 4, 5, \ldots, j$.

As a corollary of Theorem 2.5 we have Fenchel's Theorem:

Theorem 2.12 [Fenchel 1929]. *The total curvature of a continuous closed immersed curve $\gamma : S^1 \to \mathbb{R}^N$ is $\tau(\gamma) \geq 2$. Equality is attained if and only if γ is plane convex.*

Fenchel proved this for smooth curves in $\mathbb{R}^3$, Borsuk [1947] for $N \geq 3$.

PROOF. Any continuous function on the circle has at least one maximum and one minimum. So $\mu_z \geq 2$ for all $z^* \in S^2$, therefore

$$\tau(\gamma) = \mathcal{E}_z \mu_z(\gamma) \geq \mathcal{E}_z(2) = 2.$$

If $\tau(\gamma) = 2$, then $\tau(P) = 2$ for any inscribed polygon P as well. So $\mu_z(P) = 2$ and P has Banchoff's two-piece property, is 0-tight, hence plane convex by Theorem 1.5. Adding vertices yields polygons, say P_j, converging to γ. All P_j are convex in the same plane. So γ is plane convex, $N = 2$. □

Next we prepare the generalization of the curvature as π^{-1} times the length of a tangential image, a result due to van Rooij [1965]. If $u(t) \in \mathbb{R}^3$ is an immersed arc or a closed curve in $\mathbb{R}^3$ and

$$\frac{\big(u(t) - u(t_0)\big)(t - t_0)}{\big\|u(t) - u(t_0)\big\|\, |t - t_0|}$$

converges to a unit vector $v(t_0)$ for $t \to t_0+$, $t \to t_0-$ or $t \to t_0$, then $v(t_0)$ is called a *right*, *left* or *general tangent unit vector*, respectively. It is denoted by $\dot{u}_+(t_0), \dot{u}_-(t_0)$, and $\dot{u}(t_0)$, respectively. Here the dot does not (yet) mean differentiation! The parallel lines through $u(t_0)$ are called *right*, *left* and *general tangents*, respectively. A plane convex curve has a right and left tangent at every point.

LEMMA 2.13. *Let* $\gamma(t) = u(t)$, $0 \leq t < 1$, t *mod* 1, *be a continuous closed curve in* $\mathbb{R}^3$ *with* $\tau(\gamma) < \infty$. *Then*:

(i) *γ has a right and left tangent at every point t.*
(ii) *The set of points t for which $\dot{u}_+(t) \neq \dot{u}_-(t)$ is countable.*
(iii) *The right* (*left*) *tangent vector is continuous on the right* (*left*).

PROOF. (i) If $u(t)$ has no right tangent for $t = 0$, at say $u(0) = 0 \in \mathbb{R}^3$, there exists a sequence $t_1 > t_2 > \ldots > 0$ converging to 0 and unit vectors v and $w \neq v \in S^2$ such that

$$u(t_j)/\|u(t_j)\| \to \begin{cases} v & \text{for } j = 2i - 1 \to \infty, \\ w & \text{for } j = 2i \to \infty. \end{cases}$$

We can assume moreover that

$$\|u_{j+1}\| < \tfrac{1}{100}\|u_j\| \quad \text{ for } j \geq 1.$$

Then the inscribed polygon P_k with vertices $\gamma(t_1), \ldots, \gamma(t_{k-1})$ and $\gamma(0) = 0$ has curvature $\tau(P_k)$ converging to ∞ for $j \to \infty$, a contradiction.

(ii) If $\dot{u}_-(t_0)$ and $\dot{u}_+(t_0)$ form an angle $\alpha(t_0) > 0$, then a polygon with vertices at $\gamma(t_0 - \delta)$, $\gamma(t_0)$ and $\gamma(t_0 + \delta)$ will contribute, in the limit for $\delta \to 0$, an amount $\alpha(t_0)/\pi$ to $\tau(\gamma)$. The sum of such amounts is bounded by $\tau(\gamma)$. We can count such vertices by their nonincreasing contributions to $\tau(\gamma)$. The set of values t

for which $\dot u_+(t_0) \neq \dot u_-(t_0)$ is therefore countable. All other points have a unique tangent $\dot u(t) = \dot u_+(t) = \dot u_-(t)$.

(iii) If $\dot u_+(t)$ is not continuous on the right at $u(0) = 0 \in \mathbb{R}^3$ for $t = 0$, then there exist v and $w \neq v \in S^2$ and $t_j > t_{j+1} > \ldots > 0$ converging to 0, with limits

$$\dot u_+(t_{2i-1}) \to v, \quad \dot u_+(t_{2i}) \to w, \quad \text{for } i \to \infty,$$

and we can assume

$$\left\| \frac{u(t_{2i}) - u(t_{2i-1})}{\|u(t_{2i}) - u(t_{2i-1})\|} - \dot u_+(t_{2i-1}) \right\| < \tfrac{1}{10}\|v - w\| \quad \text{for all } i \geq 1.$$

The inscribed polygons P_j with vertices $u(t_i)$, $i = 1, \ldots, j$, and $u(0) = 0 \in \mathbb{R}^3$ have unbounded curvatures $\tau(P_j)$ for $j \to \infty$, a contradiction. □

Let $\gamma(t) = u(t)$ be as before and assume $\tau(\gamma) < \infty$. Let $u(t_i)$, for $i = 1, \ldots, h$, $0 \leq h \leq \infty$, $0 < t_i < 1$, be the points where $\dot u_-(t_i)$ and $\dot u_+(t_i)$ form a positive angle. Insert an interval of length 2^{-i} at t_i in the parameter space of $t \in \mathbb{R}$ mod 1, so as to obtain a parameter space with a new parameter $t' \in \mathbb{R}$ modulo $(1+\sum_{i=1}^h 2^{-i})$, $0 \leq h \leq \infty$. We define as follows a map $\tan\gamma$ that generalizes the tangential map for polygons. It has as embedded image the geodesic arc in S^2 between $\dot u_-(t)$ and $\dot u_+(t)$ for an inserted interval at $t = t_i$, where $\dot u_-(t) \neq \dot u_+(t)$. For other points, it has as image $\dot u(t) = \dot u_-(t) = \dot u_+(t)$. It is continuous in t' and is again called the *tangential map* $\tan\gamma : S^1 \to S^2$. We recover Definition 2.1 in this result:

THEOREM 2.14 [van Rooij 1965]. *If $\tau(\gamma) < \infty$, there is a continuous tangential map $\tan\gamma : S^1 \to S^2$. The length of the image $L(\tan\gamma)$ is bounded, and the total curvature $\tau(\gamma)$ is $\tau(\gamma) = L(\tan\gamma)/\pi$.* (Here the length $L(\tan\gamma)$ is defined to be the least upper bound of the length of inscribed polygons.)

PROOF. Let γ have the parameter t, $0 \leq t \leq 1$. Consider a sequence of j-gons $P_j \prec \gamma$ with vertices at $u(t_{2i-1,j})$ for $i = 1, \ldots, j$, with $0 < t_{2i-1,j} < t_{2i+1,j} < 1$. Suppose P_j converges to γ for $j \to \infty$. For each P_j add vertices and obtain a $2j$-gon $\bar P_{2j}$ with vertices at $u(t_{i,j})$ for $i = 1, \ldots, 2j$, $0 < t_{i,j} < t_{i+1,j} < 1$ and such that

$$\left\| \frac{u(t_{2i,j}) - u(t_{2i-1,j})}{\|u(t_{2i,j}) - u(t_{2i-1,j})\|} - \dot u_+(t_{2i-1,j}) \right\| < \varepsilon_j.$$

The points $\dot u_+(t_{2i-1,j})$ are image points of $\tan\gamma$. They determine an "inscribed polygon" T_j for $\tan\gamma$ with length $L(T_j)$. Note that some edges of T_j may be degenerated to a point. Clearly $L(T_j)$ has as limit

$$\lim_{j\to\infty} L(T_j) = L(\tan\gamma).$$

If ε_j is chosen very small, half of the points of $\tan P_{2j}$ are as near as we please to the vertices of T_j in S^2. Therefore we can assume

$$(1 - 2^{-j}) L(T_j) \leq L(\tan \bar P_{2j}) = \tau(\bar P_{2j}),$$

and in the limit

$$L(\tan\gamma) \le \tau(\gamma) < \infty.$$

To prove the reverse inequality we start by observing that the properties (i)–(iii) in Lemma 2.11 imply that $\gamma(t)$ is rectifiable: it has length $L(\gamma) = \lim_{P\to\gamma} L(P)$ and can be parametrized by arclength $s(t)$ such that the right-hand derivative of $u(s)$ exists and equals $\big(du(s)/ds\big)_+ = \dot u_+(t)$, $s = s(t)$.

For convenience suppose $L(\gamma) = 1$, $s \equiv t$. Let $Q_j(t)$ be the unique polygonal arc in $\mathbb{R}^3$ for $0 \le t \le 1$, with vertices for $t = i2^{-j}$ for $i = 0, \dots, 2^j$ with edges of length 2^{-j} and such that the initial point is

$$Q(0) = u(0) = \gamma(0)$$

and

$$Q_j((i+1)2^{-j}) = Q_j(i2^{-j}) + 2^{-j}\dot u_+(i2^{-j}) \quad \text{for } i = 0, 1, \dots, 2^{j-1}.$$

If $z|_\gamma$ is nondegenerate, one can verify that

$$\lim_{j\to\infty} \mu_z(Q_j) \ge \mu_z(\gamma).$$

But $\mathcal{E}_z\mu_z(\gamma) = \tau(\gamma)$, and $\mathcal{E}_z\mu_z(Q_j) = \tau(Q_j) = L(Q_j)$ which for $j \to \infty$ converges to $L(\tan\gamma)$. Therefore

$$L(\tan\gamma) \ge \tau(\gamma).$$

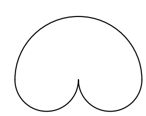
This completes the proof of van Rooij's Theorem 2.14. $\square$

Corollary. Let $\gamma(t) = u(t)$ be a closed curve of class C^2. The usual definition of total curvature and our definition are equivalent:

$$\frac{1}{\pi}\int \left\|\frac{d^2u}{ds^2}\right\| ds = \frac{1}{\pi}L(\tan\gamma) = \tau(\gamma) = \mathcal{E}_z\mu_z(\gamma).$$

Proof. If s is the arclength on γ, then $(\tan\gamma)(s) = du/ds \in S^2$, and the length of the curve $\tan\gamma$ is $\int \|d^2u/ds^2\|\, ds$. $\square$

Exercise. Use each of the definitions to calculate $\tau(\gamma)$ for the plane closed curve on the right, consisting of three semicircles (see Definition 2.4, Theorem 2.5 and Theorem 2.12).



Editors' Note. For results on the total curvature of knots, see [Fáry 1949; Milnor 1950; 1953]. For knotted surfaces, see [Kuiper and Meeks 1983; 1984; 1987].

3. Smooth Submanifolds

Definitions of Curvature. In Section 1 we defined topological n-manifolds M with charts $\kappa_x : U_x \to \kappa_x U_x \subset h^n \subset \mathbb{R}^n$. A two-manifold is called a *surface*. If the homeomorphisms

$$\kappa_{yx} = \kappa_y \kappa_x^{-1} : \kappa_x(U_x \cap U_y) \to \kappa_y(U_x \cap U_y)$$

all belong to the pseudogroup C^k (continuous k-th derivatives, for $k \leq \infty$), C^ω (analytic), or C^{Nash} (the graph of κ_{yx} in $\mathbb{R}^n \times \mathbb{R}^n$ is a nonsingular part of a real algebraic variety given by polynomial equations), then M has a *C^k-structure*, and is called a *C^k-manifold.* (Here k can be a nonnegative integer, ∞, ω or Nash.) Two C^k-manifolds M_1 and M_2 are *C^k-equivalent* if there is a homeomorphism $\varphi : M_1 \to M_2$ that is expressed in C^k-charts in a commutative diagram

$$\begin{array}{ccc}
x \in U & \xrightarrow{\kappa_x} & \kappa_x(U) \subset \mathbb{R}^n \\
\Big\downarrow \quad \Big\downarrow \varphi & & \Big\downarrow \psi \\
y \in V & \xrightarrow{\kappa_y} & \kappa_y(V) \subset \mathbb{R}^n
\end{array}$$

where ψ is C^k. If $l < k$ then $M \in C^k$ determines a unique (up to equivalence) C^l-structure. Here $0 < 1 < \ldots < \infty < \omega <$ Nash.

Another pseudogroup on $\mathbb{R}^n$, called PL, consists of homeomorphisms $\psi : U_1 \to U_2$, with U_1, U_2 open subsets in $\mathbb{R}^n$, that are affine on each affine simplex of a finite triangulation of some compact part of $\mathbb{R}^n$. Manifolds with such a structure are called *piecewise linear manifolds* or *PL-manifolds*. Obviously $0 <$ PL.

For closed surfaces all C^k-equivalence classes for $k < \omega$ correspond one-to-one to all C^0-equivalence classes. Recall from Section 1 the classification of smooth closed surfaces, summarized in Figure 7.

The $\mathbb{Z}_2$-Betti numbers $\beta_i = \operatorname{rank} H_i(M, \mathbb{Z}_2)$ of the surface are $\beta_0 = \beta_2 = 1$, $\beta_1 = 2 - \chi$. The alternating sum $\beta_0 - \beta_1 + \beta_2$ is χ, the Euler characteristic. The sum of the Betti numbers is $\beta = \beta_0 + \beta_1 + \beta_2 = 4 - \chi$. Note that the height function z on the smooth surfaces of Figure 7 has exactly β nondegenerate critical points, the minimum possible number. Note also that the surface obtained from M by deleting an open disc can be contracted over itself to a wedge of β_1 circles $S^1 \vee S^1 \vee \cdots \vee S^1$. These circles carry generators for $H_1(M, \mathbb{Z}_2)$. A C^k-map $f : M^n \to \mathbb{R}^N$ of an n-manifold is called a *C^k-immersion* for $k \geq 1$ if its derivative has maximal rank n. It is then clearly a tame immersion in the topological sense.

Remark. There is no smooth (or even topological) embedding of a closed nonorientable surface in $\mathbb{R}^3$.

PROOF. Let $M \subset \mathbb{R}^3$ be such a smooth embedding. Take a small orthogonal vector v at a point $p \in M$. Move p along a closed embedded curve γ in M and drag v along, so that on returning to p the vector has the opposite direction. Move γ away from M in the direction of the transported vectors v. The end points at p of the curve γ' so obtained are connected by a straight segment orthogonal to M. Now you have a closed curve in $\mathbb{R}^3$ that meets M in one (odd number) point. Make it a smooth curve and move it far away into space $\mathbb{R}^3$, but in such a way that the number of intersection points changes each time by two ("transversal move"). The final curve γ'' meets M in an odd number of points, but also in no point, a contradiction. $\square$

If $f : M \to \mathbb{R}^3$ is a smooth immersion, there is a choice of charts near $p \in M$ and orthogonal Euclidean coordinates z, x, y for $E^3 = \mathbb{R}^3$ so that the surface has near P an equation

$$z = \tfrac{1}{2}(k_1 x^2 + k_2 y^2) + O(x^2 + y^2)^{3/2}.$$

The numbers k_1 and k_2 are the *principal curvatures* and their product is the *Gaussian curvature* (a density) $K = k_1 k_2$. It can be positive, zero or negative.

The *normalized absolute curvature* of an open set $U \subset M$ is defined by

$$\tau(U) = \int_U \frac{|K\,d\sigma|}{2\pi}$$

where $d\sigma$ is the volume element (a two-form) induced by the immersion in $\mathbb{R}^3$. It differs from the classical curvature by a normalization factor.

Consider all unit vectors (p, z^*) at points $p \in U$ orthogonal at $f(p)$ to $f(U) \subset \mathbb{R}^3$, for some small U for which $f : U \to f(U)$ is an embedding. These vectors, displaced parallel to $(p, 0)$, form a surface $\mathcal{N}(U) \subset M \times S^2 \subset M \times \mathbb{R}^3$ consisting of two parts, one for each side of $f(U)$ in $\mathbb{R}^3$. There is a natural projection

$$\gamma : \mathcal{N}(U) \to S^2.$$

The surface $\mathcal{N}(U)$ is called the *unit normal bundle* of M on U. If $K > 0$ on U (or $K < 0$ on U) then γ sweeps out a set $A \subset S^2$ consisting of two diametrically opposite parts. See Figure 15. Each is called a *Gauss-map image* of U. It is well

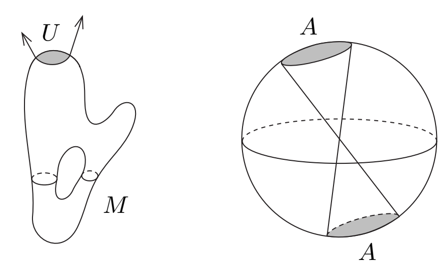


Figure 15. The Gauss map.

known that its total area $|A|$ determines the curvature:

$$\tau(U) = \int_U \frac{|K\,d\sigma|}{2\pi} = |A|/4\pi = |A|/|S^2|.$$

Another expression for the measure τ is

$$\tau(U) = c^{-1} \int_{\mathcal{N}(U)} |\nu^*(\omega)|,$$

where $c = |S^2| = 4\pi$, and $\nu^*(\omega)$ is the pull-back of the volume form ω on S^2. The unit normal bundle space $\mathcal{N}$ of f over M,

$$\mathcal{N} = \mathcal{N}(f) = \{(p, z^*) : z^* \text{ orthogonal to } f(M) \text{ at } f(p)\} \subset M \times \mathbb{R}^3$$

is a double covering of the surface M. It is connected if and only if M is nonorientable. We thus derive an equivalent definition of the total absolute curvature, in terms of the swept-out area of unit normal bundle:

DEFINITION 3.1. The *total absolute curvature* of a smooth immersion $f : M \to \mathbb{R}^3$ is

$$\tau(f) = c^{-1} \int_{\mathcal{N}} |\nu^*(\omega)| = \int_M |K\,d\sigma|/2\pi.$$

REMARK. A smooth C^∞-map ν such as our $\nu : \mathcal{N} \to S^2$ is called *critical* at a point $q \in \mathcal{N}$ if its derivative $d\nu$ does not have maximal rank. An image of such a point q is called a *critical value*. The theorem of Sard says that the set of critical values (here in S^2) has measure zero. A generalization by A. P. Morse [1939] says more, namely that the same conclusion holds if ν is a C^1-map between manifolds of the same dimension.

Morse's Lemma and Inequalities. A smooth function $\varphi : M \to \mathbb{R}$ is *nondegenerate* at a critical point $p \in M (d\varphi(p) = 0)$ if $d^2\varphi(p)$ has maximal rank. (Recall that $d^2\varphi(p)$ is expressed in coordinates by the matrix $\{\partial^2\varphi/\partial x^i \partial x^j\}$.) By linear algebra one has for some coordinate chart near p:

$$\varphi = \varphi(p) - \sum_{j=1}^{i} (x^j)^2 + \sum_{k=i+1}^{n} (x^k)^2 + O(\|x\|^2).$$

The number i is called the *index* of the critical point. We can do even better with the various results knows as *Morse's Lemma*, after M. Morse:

LEMMA 3.2 (C^∞ VERSION [Milnor 1963, p. 6]). *In suitable C^∞-coordinates, φ is expressed near a nondegenerate critical point $p \in M$ of index i by*

$$\varphi = \varphi(p) - \sum_{j=1}^{i} (x^j)^2 + \sum_{k=i+1}^{n} (x^k)^2.$$

LEMMA 3.2 (C^h VERSION [Kuiper 1966]). *A C^h-function $(2 \le h \le \omega)$ φ on M^n can with suitable C^{h-1} coordinates be expressed near a nondegenerate critical point $p \in M$ of index i by*

$$\varphi = \varphi(p) - \sum_{j=1}^{i} (x^j)^2 + \sum_{k=i+1}^{n} (x^k)^2.$$

So for C^h-functions $(2 \le h < \infty)$ one loses one class of differentiability!

The function $\varphi : M \to \mathbb{R}$ is called *nondegenerate* if it is so at every critical point.

The C^h-Morse lemma for $2 < h < \omega$ also holds on infinite-dimensional manifolds modelled on Hilbert space [Palais 1963].

Let $f : M^2 \to \mathbb{R}^3$ be our smooth immersion. Consider as before linear functions $z : \mathbb{R}^3 \to \mathbb{R}$, $\|z\| = 1$. The real function $z \circ f : M \to \mathbb{R}$ is the *restriction* of z to M. It is clearly nondegenerate at points $p \in M$ where $K(p) \neq 0$. Take a small disc U on M where $K(p) > 0$ and count, for every point $z^* \in S^2$, the number of critical points $\mu_{iz}(U)$ of index i of the function z on U. See Figure 16. The sum $\mu_z(U) = \sum_i \mu_{iz}(U)$ is one or zero. It is one in the case where $z^* \in \mathring{A}$ and zero in the case where $z^* \notin A$. The value of the function $\mu_z(U)$ for $z^* \in S^2$ has therefore a mean (expectation) of

$$\tau(U) = \mathcal{E}_z \mu_z(U) = c^{-1} \int_{S^2} \mu_z(f)|\omega| = |A|/4\pi.$$

In this way we get positive contributions near points $p \in M$ where $K > 0$, or where $K < 0$. These points form an open set on M. We can now state the definition of total curvature in terms of the mean number of critical points:

DEFINITION 3.3. The total curvature of the smooth (or C^2) immersion $f : M \to \mathbb{R}^3$ is

$$\tau(f) = \mathcal{E}_z \mu_z(f) = c^{-1} \int_{S^2} \mu_z(f)|\omega|. \tag{3.1}$$

To show that this is equivalent to Definition 3.1, observe that the set $z^* \in S^2$ for which $z \circ f$ is a degenerate smooth (or C^2) function is just the set of critical values of the C^∞ (or C^1) map $\nu : \mathcal{N} \to S^2$. This set has measure zero and can

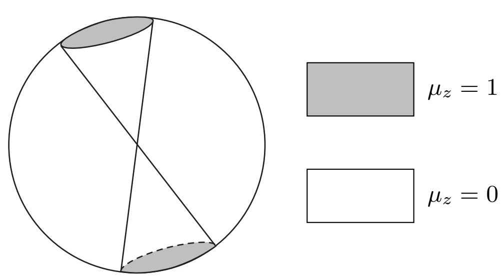


Figure 16. The number of critical points of z in a neighborhood $U \subset M$, as a function of z^*.

be neglected in the integral. The equality then follows from our discussion of the parts $K > 0$ and $K < 0$ of M.

We now formulate Definitions 3.1 and 3.3 for dimension n, but leave the proof of their equivalence to the reader.

Let $f : M \to \mathbb{R}^N$ be a smooth (or C^2) immersion of a closed manifold, $\mathcal{N} = (p, z^*) \subset M \times S^{N-1}$ the unit normal bundle, an S^{N-n-1}-bundle over M, and $\nu : \mathcal{N} \to S^{N-1}$ the projection into $S^{N-1} \subset \mathbb{R}^N$.

Definition 3.4. The total (absolute) curvature of the smooth immersion $f : M^n \to \mathbb{R}^N$ is

$$\tau(f) = c^{-1} \int_{\mathcal{N}} |\nu^*(\omega)| = \int_M |\tau(p)\, d\sigma|, \qquad c = |S^{N-1}|.$$

Here $\tau(p)$ is a continuous curvature density function, and $d\sigma$ is the volume element of M induced by the immersion f. $\tau(p)$ is obtained by integration over the fibres. For hypersurfaces it corresponds to $K = k_1 \cdots k_{N-1}$, with principal curvatures $k_1, \ldots, k_{N-1}$.

Definition 3.5. The total curvature of the smooth immersion $f : M^n \to \mathbb{R}^N$ is

$$\tau(f) = \mathcal{E}_z \mu_z(f) = c^{-1} \int_{z^* \in S^{N-1}} \mu_z(f) |\omega|.$$

Clearly the definition implies

$$\tau(f) \geq \min_z \mu_z(f),$$

where the minimum is taken over all z such that $z \circ f$ is nondegenerate. If $\gamma^{(N)} = \gamma(M^n, \mathbb{R}^N)$ is the minimum of $\min_z \mu_z(f)$ for all immersions f and γ is the minimal number of critical points a nondegenerate function $\varphi : M^n \to \mathbb{R}$ can have, then $\min_z \mu_z(f) \geq \gamma^{(N)} \geq \gamma^{(N+1)} = \gamma \geq \beta$. Here $\gamma^{(N+1)} = \gamma$ because we can introduce the optimal φ as the $(N{+}1)$-st coordinate for $\mathbb{R}^{N+1}$. By the Morse inequalities (see Theorem 3.8 below) one has $\mu_z(f) \geq \beta$ for $\beta = \beta(F) = \operatorname{rank} H_*(M, F)$ for any coefficient field F. For surfaces in $\mathbb{R}^3$ we saw in Figure 7 immersions with $\mu_z(f) = \beta$, for z pointing in the vertical direction. Thus we can conclude part (i) of next theorem:

Theorem 3.6. (i) *The total curvature of a smooth immersion $f : M^n \to \mathbb{R}^N$ is*

$$\tau \geq \gamma^{(N)} \geq \gamma^{(N+1)} = \gamma \geq \beta.$$

For surfaces, $\gamma \geq \beta$.

(ii) *The greatest lower bound of $\tau(f)$ for all immersions $f : M^n \to \mathbb{R}^N$ is*

$$\inf_f \tau = \gamma^{(N)} \geq \gamma \geq \beta.$$

For surfaces in $\mathbb{R}^N$, with $N \geq 3$, this is

$$\inf_f \tau = \beta = 4 - \chi.$$

REMARK. Sharpe [1989] proved that $\gamma^{(N)} = \gamma$ for $n > 5$.

Part (ii) of Theorem 3.6 follows from the next result:

LEMMA 3.7 [Kuiper 1959; Wilson 1965]. *Let $z \circ f$ be nondegenerate with k critical points for the smooth immersion $f : M^n \to \mathbb{R}^N$ and let $f_\varepsilon = g_\varepsilon \circ f$ where for $\varepsilon > 0$ we define g_ε by*

$$g_\varepsilon(x, z) = g(\varepsilon x, z), \quad x = (u^1, \ldots, u^{N-1}) \in \mathbb{R}^{N-1}, \quad z = u^N, \quad (x, z) \in \mathbb{R}^N.$$

Then $\lim_{\varepsilon \to 0} \tau(f_\varepsilon) = k$.

Motivation: for $\varepsilon \to 0$, it seems that all curvature gets concentrated around the critical points of $z \circ f_\varepsilon$ (which correspond, one to one, to those of $z \circ f$) and they contribute k in $\tau(f_\varepsilon)$.

PROOF OF LEMMA 3.7. Any unit vector in S^{N-1} can be written

$$\cos\theta\, z^* + \sin\theta\, w^*, \quad \text{for some } w^* \in S^{N-2} \subset \mathbb{R}^{N-1} \subset \mathbb{R}^N.$$

The volume element on S^{N-1} is

$$\omega_{N-1} = \omega_{N-2} \wedge \cos^{N-1}\theta\, d\theta.$$

There exists θ_0 near $\pi/2$ such that $\mu_z(f) = k$ for all z^* with $|\theta(z^*)| > \theta_0$.

The mapping g_ε induces a map of tangent hyperplanes. A normal vector $\cos\theta\, z^* + \sin\theta\, w^*$, is replaced by a new normal vector $\cos\theta'\, z^* + \sin\theta'\, w^*$, where $\tan\theta' = \varepsilon \tan\theta$; hence

$$\frac{d\theta'}{\cos^2\theta'} = \varepsilon \frac{d\theta}{\cos^2\theta}$$

and

$$\cos^{N-1}\theta'\, d\theta' = \varepsilon \frac{\cos^{N+1}\theta'}{\cos^{N+1}\theta} \cos^{N-1}\theta\, d\theta.$$

For $|\theta| < \theta_0$ one finds for the volume element ω on S^{N-1}:

$$|\omega'| = \varepsilon \left(\frac{\cos\theta'}{\cos\theta}\right)^{N+1} |\omega| < \frac{\varepsilon}{(\cos\theta_0)^{N+1}} |\omega|.$$

Consider the contributions of the two parts defined by $|\theta| > \theta_0$ and $|\theta| < \theta_0$ in the integral $\tau(f)$ in (3.1). The first part is replaced by an integral of the constant k over the part $|\theta'| > \theta_0'$. As θ_0' converges to zero this converges, for $\varepsilon \to 0$, to k. The second part is replaced by an integral which is smaller than $\varepsilon/\cos^{N+1}\theta_0$ times what it was. It converges to zero for $\varepsilon \to 0$. This proves the lemma

$$\lim_{\varepsilon \to 0} \tau(f_\varepsilon) = k. \qquad \square$$

Theorem 3.8 (Morse inequalities). *Let $\varphi : M^n \to \mathbb{R}$ be a smooth nondegenerate function with μ_k critical points of index k and with Betti numbers $\beta_k =$ rank $H_k(M,F)$ for a field F (we mainly use $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$). Suppose the critical points have different values. Then*

$$\frac{\sum_0^n \mu_k t^k - \sum_0^n \beta_k t^k}{1+t} = \sum_0^n K_k t^k$$

is a polynomial in t with nonnegative integer coefficients $K_k \geq 0$. In particular,

(i) $\mu_i \geq \beta_i$;
(ii) $\mu \geq \beta \quad (\mu = \sum \mu_i, \beta = \sum \beta_i)$;
(iii) $\sum (-1)^i \mu_i = \sum (-1)^i \beta_i = \chi \quad (t = -1)$;
(iv) $\sum_0^n (\mu_i - \beta_i) t^i (1 - t + t^2 - \cdots) = \sum_l K_l t^l$;
(v) $(\mu_l - \beta_l) - (\mu_{l-1} - \beta_{l-1}) + \cdots + (-1)^l (\mu_0 - \beta_0) = K_l \geq 0$.

Proof. Let $B \supset A$ be a pair of spaces with finite-dimensional homology groups. Then we have the exact triangle

$$\begin{tikzcd} H_*(A) \arrow[rr, "i"] & & H_*(B) \arrow[dl, "\alpha"] \\ & H_*(B,A) \arrow[ul, "\partial"] & \end{tikzcd}$$

or, in detail,

$$H_k(A) \xrightarrow{i_k} H_k(B) \xrightarrow{\alpha_k} H_k(B,A) \xrightarrow{\partial_k} H_{k-1}(A) \xrightarrow{i_{k-1}} H_{k-1}(B)$$

where $H_*(A) = \bigoplus_k H_k(A)$, etc. Exactness means that $\operatorname{Im} i = \operatorname{Ker} \alpha$, etc. The commutative groups then split into direct sums

$$\begin{aligned} H_k(A) &\simeq \operatorname{Ker} i_k \oplus \operatorname{Ker} \alpha_k, \\ H_k(B) &\simeq \operatorname{Ker} \alpha_k \oplus \operatorname{Ker} \partial_k, \\ H_k(B,A) &\simeq \operatorname{Ker} \partial_k \oplus \operatorname{Ker} i_{k-1} \end{aligned}$$

Define Poincaré polynomials in t:

$$\begin{aligned} P(A) &= \sum_k \dim H_k(A) t^k, \\ P(B) &= \sum_k \dim H_k(B) t^k, \\ P(B,A) &= \sum_k \dim H_k(B,A) t^k, \end{aligned}$$

$$P(\operatorname{Ker} i) = \sum_k \dim(H_k(A) \cap \operatorname{Ker} i)t^k,$$
$$P(\operatorname{Ker} \alpha) = \sum_k \dim(H_k(B) \cap \operatorname{Ker} \alpha)t^k,$$
$$P(\operatorname{Ker} \partial) = \sum_k \dim(H_k(B, A) \cap \operatorname{Ker} \partial)t^k.$$

Note that ∂ decreases the dimension from k to $k-1$. We calculate:

$$P(A) = P(\operatorname{Ker} i) + P(\operatorname{Ker} \alpha),$$
$$P(B) = P(\operatorname{Ker} \alpha) + P(\operatorname{Ker} \partial),$$
$$P(B, A) = P(\operatorname{Ker} \partial) + tP(\operatorname{Ker} i).$$

Therefore

$$P(B, A) - \big(P(B) - P(A)\big) = (1+t)P(\operatorname{Ker} i). \tag{3.2}$$

This shows that:

Lemma 3.9 (Preliminary Morse formula). *The expression*

$$\frac{P(B, A) - \big(P(B) - P(A)\big)}{1+t} = P(\operatorname{Ker} i)$$

is a polynomial $\sum K^k t^k$ *with integer coefficients* $K_k \geq 0$.

Next we discuss a smooth function $\varphi : M \to \mathbb{R}$ on the closed n-manifold M. Recall that the critical set is defined as $\operatorname{Cr}(\varphi) = \{p \in M : d\varphi = 0\} \subset M$. The set of critical values $\varphi(\operatorname{Cr}(\varphi))$ has measure zero on $\mathbb{R}$ by the Theorem of Sard (this is also true for a C^r-function for $r \geq n$). Assume, as in the case of Theorem 3.8, that the critical values form a finite set. This is also the case when φ is an algebraic function on an algebraic manifold.

We define the *croissant*

$$M_t = \{p \in M : \varphi(p) \leq t\}$$

and its interior

$$M_{t^-} = \{p \in M : \varphi(p) < t\}.$$

Let $c_1 < c_2 < \cdots < c_L$ be the critical values of φ. Consider an arbitrary Riemannian metric on M. Then φ has as differential the one-form $d\varphi$, whose dual with respect to the metric is the gradient grad $\varphi = *d\varphi$. It vanishes at p if and only if p is a critical point. Integral curves of the vector field grad φ are transversal (not tangent) to the levels of φ, except at critical points. Then we see that, *if $c_k < a < b < c_{k+1}$, the croissant M_b is diffeomorphic to M_a and*

$$H_k(M_b, M_a) = 0, \quad H_*(M_b) = H_*(M_a).$$

In other words, the diffeomorphy type of M_t can only change at critical values. There are at most $2L$ types, namely for $t < c_1$, $t = c_1$, $c_1 < t < c_2$, $\ldots$, $t = c_L$. The first type is the empty set, the last type is that of M. The open manifolds

M_{t^-} have at most $L{+}1$ types, namely for $t \le c_1$, $c_1 < t \le c_2, \dots, c_{N-1} < t \le c_N$, $t \ge c_N$. They carry at most $L+1$ homotopy types. Note that the real function $\varphi : u \to u^3$ on $M = \mathbb{R}$ has only one type for $M_{t^-} = \{p : \varphi(p) < t\}$. Let $t_0 < c_1 < t_1 < \cdots < c_L < t_L$, and consider the sequence of spaces

$$\varnothing = M_{t_0} \overset{i_1}{\subset} M_{t_1} \cdots \overset{i_L}{\subset} M_{t_L} = M.$$

By (3.2) we have

$$\frac{P(M_{t_j}, M_{t_{j-1}}) - \bigl(P(M_{t_j}) - P(M_{t_{j-1}})\bigr)}{1+t} = P(\operatorname{Ker} i_j).$$

Summation for $j = 1, \dots, L$ yields

$$\frac{\sum_{j=1}^{L} P(M_{t_j}, M_{t_{j-1}}) - P(M)}{1+t} = \sum_{j=1}^{L} P(\operatorname{Ker} i_j). \tag{3.3}$$

We now take the assumptions of Theorem 3.8; there is exactly one nondegenerate critical point in $\{p \in M : t_{j-1} < \varphi(p) < t_j\}$. Suppose it has index k. In local coordinates near p the Morse lemma gives on an n-ball $B_\delta = \sum x_j^2 < \delta^2$ the expression

$$\varphi = c_j - \sum_1^k x_j^2 + \sum_{k+1}^n x_l^2.$$

Then

$$\begin{aligned}
H_*(M_{t_j}, M_{t_{j-1}}) &= H_*(p \cup M_{c_j^-}, M_{c_j^-}) \\
&= H_*(p \cup \{q \in B_\delta : \varphi < c_j\}, \{q \in B_\delta : \varphi < c_j\}) \\
&= H_*\left(\left\{q \in B_\delta : \sum_{k+1}^n x_l^2 = 0\right\}, \left\{q \in \partial B_\delta : \sum_{k+1}^n x_l^2 = 0\right\}\right) \\
&= H_k\left(\left\{q \in B_\delta : \sum_{k+1}^n x_l^2 = 0\right\}, \left\{q \in \partial B_\delta : \sum_{k+1}^n x_l^2 = 0\right\}\right) \\
&= F.
\end{aligned}$$

Thus $P(M_{t_j}, M_{t_{j-1}}) = t^k$. Counting critical points of all indices yields

$$\sum_{j=1}^{L} P(M_{t_j}, M_{t_{j-1}}) = \sum_k \mu_k t^k.$$

Since $P(M) = \sum_k \beta_k t^k$, (3.3) yields Theorem 3.8. □

Editors' Note. See [Kuiper 1962; 1971] for more on Morse relations and tightness.

Minimal Total Curvature and Tightness. An embedding of topological spaces $f : A \subset B$ is called *injective in homology* over $\mathbb{Z}_2$, or H_**-injective*, in case the induced homomorphism in $\mathbb{Z}_2$-homology

$$f_* : H_*(A) \to H_*(B)$$

is injective. It is easy to show that:

LEMMA 3.10. *If $A \subset B \subset C$, and $A \subset C$ as well as $B \subset C$ are H_*-injective, then so is $A \subset B$.*

REMARK. *Čech cohomology* theory and *Čech homology* theory have been introduced for spaces whose singular homology does not give satisfactory answers. For manifolds and CW-complexes they give the same answer. If X is a compact subset of a manifold or a CW-complex Y, then $H_*^{\text{Čech}}(X)$ is the inverse limit of $H_*^{\text{sing}}(U_i(X))$ for $U_i \supset U_{i+1} \supset \cdots \supset X$ any nested sequence of open sets of Y converging to X. For such cases this can be used as definition. In general it is known that if the compact space X is embedded in any space Y, then $H_*^{\text{Čech}}(X)$ is the inverse limit of $H_*^{\text{Čech}}(Y_i)$ for any nested sequence of subspaces $Y \supset Y_i \supset Y_{i+1} \supset \cdots \supset X$ converging to X. In our applications we will not use more than these facts from the notion of Čech homology. We will use $\mathbb{Z}_2$-Čech homology, denoted H_*. We remark also that Čech cohomology has better general properties than Čech homology and for this reason is used more.

EXERCISE. If

$$X = \{(u, v) \in \mathbb{R}^2 : u = \sin(1/v) \text{ for } v > 0 \quad \text{or} \quad u = 0 \text{ for } v \leq 0 \quad \text{or} \quad v = 0\}$$

then $H_0^{\text{Čech}}(X) = \mathbb{Z}_2$ but $H_0^{\text{sing}}(X) = \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

In Theorem 3.6 we saw that the total curvature of a smooth immersion $f : M \to \mathbb{R}^N$ of a closed n-manifold M has total curvature $\tau(f) \geq \beta(M)$. We now study equality:

THEOREM 3.11. *The smooth immersion $f : M \to \mathbb{R}^N$ has total curvature $\tau(f) = \beta(M)$ if and only if the embeddings $f^{-1}(h) \subset M$ are homology injective for every half-space $h \subset \mathbb{R}^N$. The same statement holds if we replace half-spaces h by open half-spaces $\mathring{h}$.*

PROOF. Let $\tau(f) = \mathcal{E}_z \mu_z(f) = \beta$. Then $\mu_z(f) = \beta$ and $\mu_{kz}(f) = \beta_k$ for all indices k, of any nondegenerate height function z with $\|z\| = 1$. The croissant

$$M_t = \{p \in M : zf(p) \leq t\}$$

equals $f^{-1}(h)$ for the half-spaces $h = \{q \in \mathbb{R}^N : z(q) \leq t\}$. Assume that the initial values $c_1, \ldots, c_L$ are all different ($L = \beta(M)$) and let noncritical values t_j be chosen such that $t_0 < c_1 < t_1 < c_2 \cdots < c_L < t_L$, as in the proof of Equation (3.3). Any given noncritical value t can be assumed to be t_j for some j. As $\mu_k = \beta_k$ for all k, $K_k = 0$ for all k. So all the polynomials vanish in (3.3), $M_{t_{j-1}} \subset M_{t_j}$ is H_*-injective and so is the composition $M_t = M_{t_j} \subset M$. The space $f^{-1}(h) \subset M$ is for every half-space h the limit of a nested sequence of spaces

$$M \supset f^{-1}(h_i) \supset f^{-1}(h_{i+1}) \cdots \supset f^{-1}(h)$$

for which $f^{-1}(h_i) \subset M$ is already known to be H_*-injective. Then the same follows for $f^{-1}(h) \subset M$. In fact the limit $H_*(f^{-1}(h)) = H_*(f^{-1}(h_{i_0}))$ is already attained for some integer i_0, because $H_*(M) = \oplus_1^{\beta} \mathbb{Z}_2$ is finite. For any open half-space $\mathring{h}$ we find a sequence of half-spaces h_j such that the nested sequence

$$f^{-1}(h_i) \subset f^{-1}(h_{i+1}) \cdots \subset f^{-1}(\mathring{h}) \subset M$$

exhausts $f^{-1}(\mathring{h})$. The limit $H_*(f^{-1}(\mathring{h})) = H_*(f^{-1}(h_{i_0}))$ is again already attained for some integer i_0. □

Definition. We call a continuous map $f : X \to \mathbb{R}^N$ of a compact connected metrizable space X *tight* or *perfect* if $f^{-1}(h) \subset X$ is H_*-injective for every half-space $h \subset \mathbb{R}^N$. We call f *k-tight* if $f^{-1}(h)$ is H_i-injective for all $i \leq k$ and all h.

Exercise. A map $f : X \to \mathbb{R}^N$ into one point is tight.

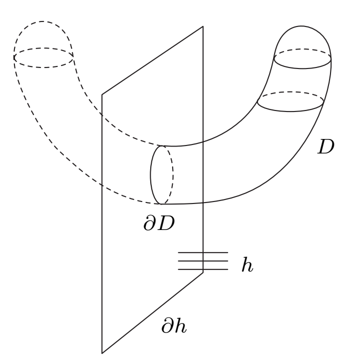
For $k = 0$ our definition agrees with Section 1 where 0-tightness, the two-piece property, was studied. We can now reformulate Theorem 3.11 with a slight extension as follows:

Theorem 3.12. *The smooth immersion $f : M \to \mathbb{R}^N$ of a closed n-manifold has minimal total curvature equal to $\tau(f) = \beta(M)$ if and only if f is k-tight, where $n - 2 \leq 2k \leq n - 1$.*

Proof. By Poincaré duality $\beta_i = \beta_{n-i}$. For any nondegenerate z one has $\mu_{n-i}(z) = \mu_i(-z)$. So if f is k-tight, then $\mu_i = \beta_i$ for $i \leq k$ and for $n - i \leq k$. There only remains to show for n even the equality $\mu_{k+1} = \beta_{k+1}$. This follows from $\sum_0^n (-1)^i \mu_i = \sum_0^n (-1)^i \beta_i = \chi$. □

Corollary. A smooth immersion $f : M^2 \to \mathbb{R}^N$ of a closed surface M^2 has minimal total curvature $\tau(f) = \beta(M^2)$ if and only if f is 0-tight.

Example. A 0-tight smooth embedding $f : S^3 \to \mathbb{R}^4$ need not have have minimal total curvature (need not be tight). For example, consider the surface $D \subset h \subset E^3 \subset E^4$ with boundary $\partial D \subset \partial h = E^2$ shown on the right. Assume D orthogonal to ∂h along ∂D. Rotate (D, h) in E^4 around the (fixed point set) plane $\partial h = E^2$ to form the three-sphere $M \subset E^4$. See [Kuiper 1970, pp. 221–224] for more detail.



Exercise. Prove that the limit Swiss cheese in E^2 (page 7) is tight.

Open Question. Is there a 0-tight smooth S^3 embedded substantially in $\mathbb{R}^5$?

For topological or PL embeddings (see also [Banchoff 1965]) the possibilities are richer:

Theorem 3.13 [Banchoff 1971a]. *There is for every $N > n \geq 3$ a substantial 0-tight polyhedral embedding $f : S^n \to E^N$.*

4. Tight Smooth Surfaces in Codimensions 3 and Higher

This section is devoted to the proof of the following result:

Theorem 4.1 [Kuiper 1962]. *Let $f : M \to E^N$ be a 0-tight smooth closed embedded surface substantial in E^N. Then $N \leq 5$. For $N = 5$, M is the real projective plane and f is an embedding onto a Veronese surface.*

Tightness and "high" codimension (at least 3) therefore dramatically impose uniqueness and rigidity up to projective transformation.

Proof. We prove the theorem in a sequence of steps.

(a) $N \leq 5$.

Proof. Let f fulfill the conditions of the theorem and let z be a height function with nondegenerate maximum at $p \in M$. Then p will be called a *nondegenerate top point*. Choose Euclidean coordinates $u_1 \dots u_{N-2}, v_1, v_2$ vanishing at $p = 0 \in \mathbb{R}^N$, and such that $u_1, \dots, u_{N-2}$ vanish on the tangent space $T_p(M)$, and u_1 is nondegenerate maximal at p. Then v_1 and v_2 (or more precisely $v_1 f$ and $v_2 f$) are coordinates for M near p. The height functions that vanish at p form a vector space

$$W = \{w_1 u_1 + \dots + w_{N-2} u_{N-2} : (w_1, \dots, w_{N-2}) \in \mathbb{R}^{N-2}\}.$$

Denote $\mathcal{J}^2 z = \mathcal{J}^2(zf)$ ($\mathcal{J}^2$ for 2-jet) the part up to degree two of the Taylor series of zf, in terms of v_1, v_2. It is a quadratic function. The induced map $\mathcal{J}^2 : W \to Z$ into the 3-space Z of quadratic functions in v_1 and v_2 is a homomorphism.

Suppose $\mathcal{J}^2 z = 0$ for some height function z with $\|z\| = 1$ and $z \neq \pm u_1$. Then $\mathcal{J}^2(u_1 + \lambda z) = \mathcal{J}^2 u_1$ is negative definite and $u_1 + \lambda z$ has a nondegenerate isolated maximum at p for all λ. If $z(p_1) \neq 0$ for some $p_1 \in M$ then the half-space

$$h = \{q \in \mathbb{R}^N : u_1(q) + \lambda z(q) \geq 0, \quad u_1(p_1) + \lambda z(p_1) = 0\}$$

meets M in an isolated point p and in another part containing p_1, contradicting tightness. So $z(p_1) = 0$ for all $p_1 \in M$, contradicting the substantiality of f. It follows that $\mathcal{J}^2$ must be injective and $\dim W \leq \dim Z = 3$, so $N \leq 5$. ☐

Remark. The codimension of a 0-tight smooth closed n-submanifold substantially in E^n is $N - n \leq \frac{1}{2} n(n+1)$, by the same argument [Kuiper 1970].

From now on assume $N = 5$.

(b) *Preferred affine coordinates at* $p \in M$, *and the topsets*

$$M_\theta = M_\theta(p) \subset E^3_\theta \subset E^4_\theta.$$

Since $N = 5$, $\mathfrak{J}^2 : W \to Z$ is bijective onto Z. Choose $u'_1, u'_2, u'_3 \in W$ such that $\mathfrak{J}^2(u'_1) = -v_1^2 - v_2^2$, $\mathfrak{J}^2(u'_2) = -v_1^2 + v_2^2$, $\mathfrak{J}^2(u'_3) = 2v_1v_2$. Denote them by u_1, u_2, u_3 and consider u_1, u_2, u_3, v_1, v_2 as an orthonormal basis of Euclidean space E. Let

$$z = w_1u_1 + w_2u_2 + w_3u_3 = \sin\varphi u_1 + \cos\varphi\cos\theta\, u_2 + \cos\varphi\sin\theta\, u_3.$$

Then

$$\mathfrak{J}^2 z = w_1(-v_1^2 - v_2^2) + w_2(-v_1^2 + v_2^2) + w_3 2v_1v_2,$$

with determinant

$$\begin{vmatrix} -w_1 - w_2 & w_3 \\ w_3 & -w_1 + w_2 \end{vmatrix} = w_1^2 - w_2^2 - w_3^2 = \sin^2\varphi - \cos^2\varphi = -\cos 2\varphi.$$

Therefore z is nondegenerate at p and its index is

$$\begin{array}{lll} 2 \text{ (maximum)} & \text{if and only if} & \pi/4 < \varphi \le \pi/2, \\ 1 & \text{if and only if} & -\pi/4 < \varphi < \pi/4, \\ 0 \text{ (minimum)} & \text{if and only if} & -\pi/2 < \varphi < -\pi/4. \end{array}$$

The height function z has, for $\varphi = \pi/4 - \varepsilon$ with $\varepsilon > 0$ small, a critical point of index 1 at p. Then we see that the half-space

$$h = h(\varphi, \theta) = \{q \in E^5 : z = z_{\varphi,\theta} \ge 0\}$$

contains in $h \cap M$ an essential one-cycle of M, and so does the limit for $\varepsilon = 0$, $\varphi = \pi/4$.

We now denote the half-space $h(\pi/2, \theta)$ by $h(\theta)$ and conclude that the topset $M_\theta = M \cap h(\theta)$ carries an essential 1-cycle of M. The boundary $\partial h(\theta)$ is the 4-plane with equation $u_1 + \cos\theta\, u_2 + \sin\theta\, u_3 = 0$; see Figure 17.

The half-spaces $h(\theta)$ envelop a solid cone with equations

$$u_1 \le 0, \quad u_1^2 \ge u_2^2 + u_3^2$$

with boundary the quadratic 4-dimensional cone with equation

$$u_1 \le 0, \quad u_1^2 = u_2^2 + u_3^2. \tag{4.1}$$

It contains as u_1-topset the whole tangent plane $T_p(M)$ with equation $u_1 = u_2 = u_3 = 0$. The half-space $h(\theta)$ and the 4-plane $\partial h(\theta)$ support the cone in a 3-plane E^3_θ with equation

$$\begin{aligned} u_1 + \cos\theta\, u_2 + \sin\theta\, u_3 &= 0, \\ -\sin\theta\, u_2 + \cos\theta\, u_3 &= 0. \end{aligned} \tag{4.2}$$

The second equation is obtained from the first by differentiation with respect to θ.

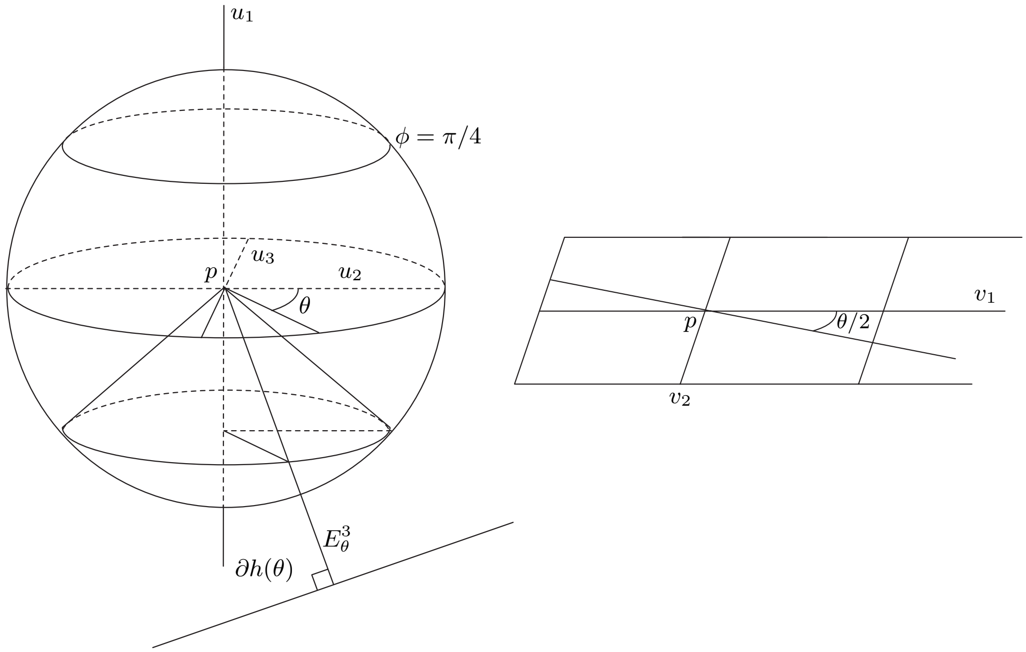


Figure 17. Left: normal space at p. The cone has equation (4.1). The 3-plane E^3_θ has equation (4.2); dimensions v_1 and v_2 are not shown. E^3_θ is the intersection of the cone with E^4_θ, the 4-plane generated by E^3_θ and the u_1-axis. Moreover $E^3_\theta = E^4_\theta \cap \partial h(\theta)$. Right: tangent space at p.

The 4-plane with equation $-\sin\theta\, u_2 + \cos\theta\, u_3 = 0$ is denoted E^4_θ. It contains E^3_θ as well as $E^3_{\theta+\pi}$. In summary,

$$p \in M_\theta \subset E^3_\theta \subset E^4_\theta.$$

(c) *M_θ has a tangent line at p with equation*

$$\cos(\theta/2)\, v_1 - \sin(\theta/2)\, v_2 = 0, \qquad u_1 = u_2 = u_3 = 0. \tag{4.3}$$

Proof. Consider the function on M

$$u_1 + \cos\theta\, u_2 + \sin\theta\, u_3 = (-v_1^2 - v_2^2) + \cos\theta(-v_1^2 + v_2^2) + \sin\theta 2 v_1 v_2 + O(\|v\|^3).$$

Here $O(\|v\|^3)$ means that the remainder $R(v_1, v_2)$ is such that $R(v_1, v_2)/\|v\|^3$ is bounded for small $\|v\| = \sqrt{v_1^2 + v_2^2}$. As the expression vanishes on M_θ we obtain after calculation

$$-\big(\cos(\theta/2)\, v_1 - \sin(\theta/2)\, v_2\big)^2 + O(\|v\|^3) = 0.$$

Hence

$$\cos(\theta/2)\, v_1 - \sin(\theta/2)\, v_2 = O\big(\|v\|^{\frac{3}{2}}\big)$$

and
$$\left|\cos(\theta/2)\big(v_1/\|v\|\big) - \sin(\theta/2)\big(v_2/\|v\|\big)\right| = O\big(\sqrt{\|v\|}\big).$$
The unit vector $\|v\|^{-1}(v_1, v_2)$ goes in the limit to $\pm(\sin(\theta/2), \cos(\theta/2))$, for $(v_1, v_2) \in M_\theta, \|v\| \to 0$. $\square$

(d) *M_θ is a smooth curve near p.*

Proof. It suffices to prove this for $\theta = 0$. The equations (4.2) for $\theta = 0$ are
$$u_1 + u_2 = 0, \quad u_3 = 0,$$
and we obtain
$$v_1 = O(\|v\|^{\frac{3}{2}}), \quad \frac{v_1}{\|v\|} = O\sqrt{\|v\|}, \quad \frac{v_2}{\|v\|} \to 1,$$
and
$$2v_1v_2 - \varphi(v_1, v_2) = 0. \tag{4.4}$$
The function φ is C^∞ with $\mathcal{J}^2(\varphi) = 0$ at $(v_1, v_2) = (0, 0)$:
$$\varphi(v_1, v_2) = O(\|v\|^3).$$
Consider for any fixed v_2 the function in v_1 given by
$$\eta : v_1 \mapsto v_1 - (2v_2)^{-1}\varphi(v_1, v_2).$$
Since $d\eta/dv_1 \neq 0$ for small $\|v\|$, the map η is a C^∞-diffeomorphism for any small value v_2. Call its inverse η^{-1}. Then the solution of (4.4) is
$$v_1 = \eta^{-1}(0).$$
Its dependence on v_1 is expressed as
$$v_1 = \psi(v_2).$$
The inverse function theorem says that ψ is also C^∞. $\square$

We obtain for every tangent line (4.3) a smooth curve. The 4-planes E^4_θ cut out these curves in pairs near p, M_θ and $M_{\theta+\pi}$, with tangent lines
$$\cos(\theta/2)\, v_1 - \sin(\theta/2)\, v_2 = 0, \quad \sin(\theta/2)\, v_1 + \cos(\theta/2)\, v_2 = 0.$$

(e) *Any component of $M_\theta(p) \cap \mathcal{U}$ is a plane convex curve.*

Proof. If z has a nondegenerate maximum, then so has any height function z_1 near to z. The nondegenerate top points therefore form an open set $\mathcal{U}$ in M. We study the levels of the function $\theta = \arctan(u_3/u_2)$ on M in a small neighborhood $U(p) \subset U$ of p (but p is excluded as θ is not defined there). By Sard's theorem, the critical values of θ on $M \setminus E_{2,3}, (E_{2,3} : u_2 = u_3 = 0)$ have measure zero. For a regular value θ_0 the level $\{q \in M \setminus E_{2,3} : \arctan(u_3(q)/u_2(q) = \theta_0\}$ is a smooth curve. Its intersection with $\partial h(\theta_0)$ can be completed with the point p to obtain the top set M_{θ_0}, a tight connected closed curve. By Fenchel's Theorem (Theorem 2.12) it is a convex plane curve. As θ_0 is a regular value, the function

θ is not critical on M_{θ_0}, hence on some neighborhood of M_{θ_0} in M. Then θ is also a regular value on that neighborhood, and M_θ is also a plane convex curve for $|\theta - \theta_0|$ small. So the set of θ for which the topset M_θ is plane convex is open and dense in $\mathbb{R}$ mod 2π. But inside the small neighborhood $U(p)$ of p in M the limit of a set of plane curves M_θ for $\theta \to \theta_1$ must be a plane curve as well. So every topset M_θ has a plane convex curve part in $U(p) \subset \mathcal{U}$. The same arguments apply for any $p \subset \mathcal{U}$, hence to $\mathcal{U}$. $\square$

(f) *Any component of $M_\theta(p) \cap \mathcal{U}$ is part of a conic.*

PROOF. Consider $q \in M_{\theta_1}(p)$ and $M_{\theta_2}(q) = \gamma \neq M_{\theta_1}(p)$. There is an open interval of regular values θ near to θ_1 giving rise to curves $M_\theta(p) \cap U$ that meet γ in some open interval near q. That interval lies in the intersection of the plane of γ and the quadratic cone for p. This proves statement (f) locally, hence globally. $\square$

(g) The remaining part of the proof belongs to classical projective geometry. We only indicate the main ideas. Take the above situation in some neighborhood $U(q_0)$ of

$$q_0 \in M_{\theta_1}(p), \qquad M_{\theta_2}(q_0) = \gamma \neq M_{\theta_1}(p).$$

Consider an interval of top conic sections $M_\theta(p)$, $\theta_1 - \delta < \theta < \theta_1 + \delta$, and a two-parameter family of top conic sections $M_\omega(p)$ parametrized by $q \in M_{\theta_1}(p)$, q near to q_0, and ω, $\theta_2 - \delta < \omega < \theta_2 + \delta$, such that every $M_\theta(p)$ meets every $M_\omega(q)$ in exactly one point inside $U(q_0) \subset U$. We consider $E^5 = P^5 \setminus P^4$ as the complement of a projective 4-plane P^4 in real projective space P^5. We project $U(p)$ from the center $P \in E^5$ into P^4. The point p is itself excluded from this projection, which is denoted $\hat{p}$. We see that:

(i) $\hat{p}(T_p(M) \setminus p)$ is a line $L_0 \subset P^4$.
(ii) $\hat{p}(M_\theta(p))$ is a line $L_\theta \subset P^4$, and $L_0 \cap L_\theta$ is a point.
(iii) $\hat{p}(M_\omega(q)$ is a conic $\gamma(\omega, q) \subset P^4$ in a plane $\Pi(\omega, q)$.
(iv) Every line L_θ meets every conic $\gamma(\omega, q)$ in one point in $\Pi(\omega, q)$.

These are very strong conditions on the lines L_θ and planes $\Pi(\omega, q)$.

In P^4 the family of all 2-planes that meet 4 lines in general position is a two-dimensional family of planes all meeting one more (fifth) line. However, if the four lines L_θ, for $\theta = \theta_1, \theta_2, \theta_3, \theta_4$, are in general position but for the fact that they all meet one and the same line L_0, then there is a complete one-parameter family of such lines [Kuiper 1941] and a degenerate Veronese surface $\mathcal{V}_1$, that can be described as follows.

Take three points $\theta_1, \theta_2, \theta_3$ on a projective line $L_0 \subset P^4$. Take a disjoint plane Π and in it a conic $\mathcal{E} \subset \Pi$. Let $\rho : L_0 \to \mathcal{E}$ be a birational correspondence. $\mathcal{V}_1$ is the union of all lines L_θ from $\theta \in L_0$ to $\rho(\theta) \in \mathcal{E}$.

From our description it follows that $\mathcal{V}_1$ has no projective invariants. A Veronese surface $\mathcal{V}$ can be found in E^5 which passes through p, has the same tangent plane $T_p(\mathcal{V}) = T_p(M)$, and yields the projection $\hat{p}\mathcal{V} = \mathcal{V}_1$. As other points $p' \in U(q_0)$ also give algebraic projections $\hat{p}'(U(q_0))$ in a projective surface $\mathcal{V}_1'$ we can succeed in finding a unique Veronese surface $\mathcal{V}$ which contains all of $U(p)$ and then by continuation all of U. By extension we also see that the set of nondegenerate top-points $U \subset M$ is not only open but also closed in $\mathcal{V}$. Then $f(M) = U = \mathcal{V}$, the Veronese surface. This ends the proof of Theorem 4.1. $\square$

Details of the last part can be found in [Kuiper 1980], which contains Theorem 4.3 below.

Generalization. The conclusion of Theorem 4.1 remains true if we assume $f : M \to E^5$ to be an immersion. We mention, without proof, a deep generalization of Theorem 4.1:

THEOREM 4.2 [Little and Pohl 1971]. *Let $f : M^n \to \mathbb{R}^N$ be a 0-tight smooth substantial immersion of a closed n-manifold $n \geq 2$. Then $N - n \leq \frac{1}{2}n(n+1)$. If we assume equality, $N - n = \frac{1}{2}n(n+1)$, then M is real projective n-space and $f(M^n)$ is the standard Veronese n-manifold (up to projective transformation). Note that only 0-tightness is used.*

Hard to prove is:

THEOREM 4.3 [Kuiper 1980]. *Let $f : M^{2d} \to E^{3d+2}$ be a smooth substantial tight embedding of a manifold M like a projective plane (that is, $\beta_0 = \beta_d = \beta_{2d} = 1$, $\beta = \sum \beta_i = 3$) into E^{3d+2}, $d = 1, 2, 4$ or 8. Then $f(M^{2d})$ is an algebraic variety. For $d = 1$, we have Theorem 4.1. For $d = 2$, $f(M^4)$ is the standard model $\mathcal{V}(\mathbb{C})$ of the complex projective plane, up to real projective transformation in E^8.*

In both theorems the assumptions lead to a high degree of rigidity of the embedding.

OPEN QUESTION (perhaps not difficult). It is *conjectured* that the conclusions of Theorem 4.3 for obtaining the standard models for projective planes over $\mathbb{C}$ also hold for quaternion planes (or even the same conclusion for 3-connected 8-manifolds) and for the Cayley-plane (or even for 7-connected 16-manifolds).

EDITORS' NOTE. See also [Kuiper and Pohl 1977], in which it is shown that a TPP topological embedding of the real projective plane into E^5 is either a Veronese surface or Banchoff's piecewise linear embedding with six vertices [Banchoff 1974].

5. Tightness of Topsets

THEOREM 5.1. *If $Y \subset \mathbb{R}^N$ is compact and tight, then so is every topset X of Y, and the inclusion $X \subset Y$ is homology injective. In particular, every essential cycle in X carries an essential cycle in Y.*

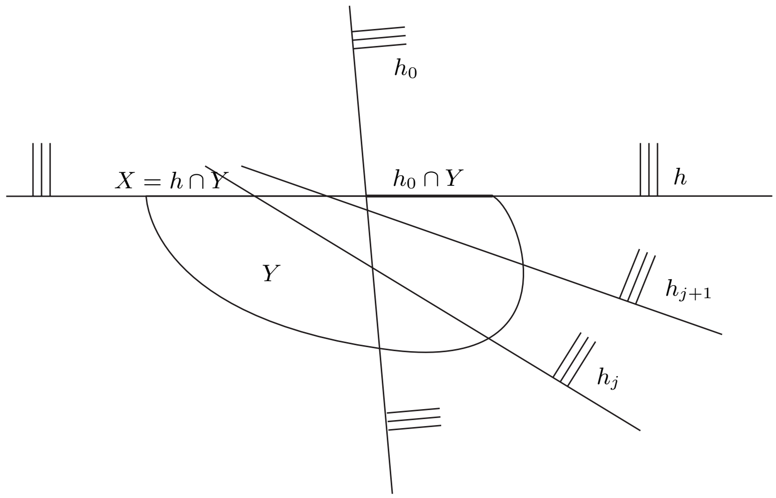


Figure 18. Proving that the topset X of a tight set Y is tight.

THEOREM 5.2. *If $f : M \subset \mathbb{R}^N$ is a smooth embedding of a closed manifold M with minimal total curvature $\tau(f) = \beta(M)$, then every topset $X \subset M$ is tight, and every essential cycle in $H_*(X)$ carries an essential cycle in $H_*(M)$.*

Note that there is no reason for X to be a manifold! This is one motivation for our generalization. Theorem 5.1 is a special case of the next result:

THEOREM 5.3. *If $f : Y \to \mathbb{R}^N$ is a tight map of a compact connected metrizable space Y and if $X \subset Y$ is any topset, then $f : X \to \mathbb{R}^N$ is tight and $X \subset Y$ is homology injective. In particular every essential cycle in X carries an essential cycle in Y.*

PROOF. (Compare the examples of tight spaces given so far). If the conclusion holds for any topset of a tight map, it holds for a topset of this topset by composition. Then, it holds for any top^{k}set for $k \geq 1$ by induction. It suffices therefore to prove the theorem for a topset $X \subset Y$, say

$$\varnothing \neq X = f^{-1}(h) = f^{-1}(\partial h) \neq Y,$$

h a supporting half-space. Let h_0 be another half-space. Say it "meets" X in $(f|_X)^{-1} = X \cap f^{-1}(h_0) = f^{-1}(h \cap h_0) \neq \varnothing$. See Figure 18. There is a sequence of half-spaces $h_1, h_2, \ldots$ in $\mathbb{R}^N$ such that

$$f^{-1}(h_1) \supset f^{-1}(h_2) \supset \cdots \supset f^{-1}(h_j) \supset \cdots \supset f^{-1}(h \cap h_0) = \bigcap_j f^{-1}(h_j) = X$$

is a nested sequence of subspaces of Y, *converging to* X.

The inclusions $X = f^{-1}(h) \subset Y$ and $f^{-1}(h_j) \subset Y$ are H_*-injective by tightness of f. The inclusion $f^{-1}(h \cap h_0) \subset Y$ is H_*-injective because $f^{-1}(h \cap h_0)$ is the inverse limit of $\{f^{-1}(h_j)\}$. Then the first inclusion in the sequence

$$(f|_X)^{-1}(h_0) = f^{-1}(h \cap h_0) \subset f^{-1}(h) = X \subset Y$$

is H_*-injective by the easy Lemma 3.10. This means that the restriction $f|_X$ is tight. $\square$

Our theorem is proved for topsets and holds then for top*sets in general.

Exercise. Formulate and prove the analogous theorems concerning k-tightness.

Tight Balls and Spheres. A compact space X with $H_*(X) = H_0(X) = \mathbb{Z}_2$ is called a *homology point space*.

Theorem 5.4. *If $f : X \subset E^N$ is a substantial and tight embedding of a homology point space X in E^N then $X = f(X)$ is a convex set.*

Proof. The conclusion is true for $N = 0$. Suppose it is true for all $N < k$. Let $f : X \to E^k$ be substantial. All topsets X' of X are H_*-injective in X. So they are homology points, and convex by induction. Therefore $\partial\mathcal{H}X \subset X$ is a $(k-1)$-sphere in X. It is bounding because $H_{k-1}(X) = 0$. Then, no point of $\mathcal{H}X$ can be not in X, so $X = \mathcal{H}X$. By induction the theorem follows. $\square$

We next prove the analogous theorem for maps.

Theorem 5.5. *If $f : X \to E^N$ is a tight substantial map of a homology point space X, then $f(X)$ is convex.*

Proof. Assume the theorem true for all $N < k$. It is true for $n = 0$. Let $f : X \to E^k$ be substantial tight. All topsets are homology injective in X, so they are homology point spaces, their images are convex sets by induction. Their union is contained in fX, i.e., $\partial\mathcal{H}fX \subset fX$. Denote $W = f^{-1}\partial\mathcal{H}fX \subset X$. Let $R \subset W \times S^{k-1}$ be the compact relation

$$R = \{(w, z^*) : \max_{p \in X} zf(p) = zf(w)\}.$$

Given z we see that the set of w for which $(w, z^*) \in R$ is just the topset determined by the maximal value of zf on X. It is a homology point space. Therefore the projection $p_S : W \times S^{k-1} \to S^{k-1}$ induces a projection $p_S : R \to S^{k-1}$ whose fibers are all homology-point spaces. Then by a theorem of Vietoris and Begle [Spanier 1966, p. 344], p_S induces an isomorphism in homology:

$$H_*(R) \xrightarrow{\sim} H_*(S^{k-1}).$$

Given $w \in W$, the set of $z^* \in S^{k-1}$ for which $(w, z^*) \in R$ is nonempty and convex by geometry (convexity of $\mathcal{H}fX$ containing $f(w)$). Therefore, the fibers of the projection $p_W : R \to W$ are also homology point spaces, and again the theorem of Vietoris and Begle reads that $H_*(R) \to H_*(W)$ is an isomorphism. By composition, we see that

$$H_{k-1}(W) = H_{k-1}(S^{k-1}) = \mathbb{Z}_2.$$

As $W \subset X$ and $H_*(X) = H_0(X) = \mathbb{Z}_2$, the space W is bounding in X and so is its image $\partial\mathcal{H}fX$ in fX. Then $fX = \mathcal{H}fX$ is convex. $\square$

THEOREM 5.6 [Kuiper 1980]. *Let $f : S^n \to E^N$ be a substantial tight map. Then either $N = n+1$ and $f(S^n)$ is a convex hypersurface or $N \leq n$ and $f(S^n)$ is a convex set.*

COROLLARY [Chern and Lashof 1957]. Let $f : M^n \to E^N$ be a substantial smooth immersion of a closed n-manifold M^n with total curvature $\tau(f) = 2$. Then M^n is the n-sphere with standard smooth structure (not exotic), $N = n+1$, and $f(M)$ is a convex hypersurface.

PROOF OF COROLLARY. Since $\tau(f) = \mathcal{E}_z\mu_z(f) = 2$, we have $\mu_z(f) = 2$ and M must be homeomorphic to a n-sphere. There is no immersion in R^N for $N \leq n$, so f is a smooth immersion onto a smooth convex hypersurface in $\mathbb{R}^{n+1}$. Then, it is an embedding and $f(S^n)$ has the standard smooth structure. $\square$

PROOF OF THEOREM 5.6. Given f one observes that every topset $X \subset S^n$ misses at least one point of S^n and it is homology injective in S^n. Therefore $H_*(X) = H_*(\text{point})$. So by Theorem 5.5 every topset is a homology point space. As in the proof of Theorem 5.5, we get an isomorphism

$$H_*(W) \simeq H_*(\partial\mathcal{H}fS^n) \simeq H_*(S^{N-1}),$$

where $\partial\mathcal{H}fS^n \subset fS^n, W = f^{-1}(\partial\mathcal{H}fS^n)$. As $W \subset S^n$, then $N-1 \leq n$. If $N-1 < n$, then $W \subset S^n$ bounds in S^n, that is $H_{N-1}(W) \to H_{N-1}(S^n)$ maps into zero. Then also

$$0 \neq H_{N-1}(fW) = H_{N-1}(\partial\mathcal{H}fX) \to H_{N-1}(fX)$$

maps into zero. Then $f(X) = \mathcal{H}f(X)$ is convex.

If $N-1 = n$ then $W \subset S^n$ induces an isomorphism $H_n(W) \to H_n(S^n)$ and so $W = S^n$ and $f(W) = \partial\mathcal{H}fS^n$, a convex hypersurface in E^{n+1}. $\square$

Application: Tight Projective Planes. A closed manifold P^{2d} that contains a tame embedded d-sphere S^d so that $P^{2d} \setminus S^d$ is homeomorphic to $\mathbb{R}^{2d}$ is called a *manifold like a projective plane*. Such manifolds were studied in [Eells and Kuiper 1962]. P^{2d} is a CW-complex with three cells; $\beta = \beta_0 + \beta_d + \beta_{2d}$. Necessarily $d = 1, 2, 4, 8$. Examples are the standard projective planes over $\mathbb{R}, \mathbb{C}, \mathbb{H}$ (quaternions) and Ca (octaves or Cayley numbers), for $d = 1, 2, 4, 8$ respectively. The self intersection of the essential d-cycle in P^{2d} is one. Here we prove:

THEOREM 5.7. *Let $f : P \subset E^N$ be a tight substantial embedding of a manifold like a projective plane $P = P^{2d}$. Then $N \leq 3d+2$.*

PROOF. Let k be the smallest number for which there exists an E^k-top*set X that contains an essential cycle in $H_*(X)$. If there is no such k, let $k = N$. Clearly, $k > 1$ because every E^0- or E^1-top*set is convex.

Every topset Y of X is homology injective in X and is therefore a homology point space. Then Y is convex. Consequently, the $(k-1)$-sphere $\partial\mathcal{H}X$, a union of such convex topsets, is in X. If $k-1 \neq d$ or $2d$, then $\partial\mathcal{H}X$ bounds in X because by tightness if not, it also does not bound in P. Then every point inside $\partial\mathcal{H}X$ must belong to X, and $X = \mathcal{H}X$, contradicting the assumption on k.

Now assume that $k-1 = d$. Then $\partial\mathcal{H}X$ carries the d-cycle of P^{2d}. Now we project E^N parallel to $E^k = \mathrm{span}(X)$ into a Euclidean space E^{N-k} orthogonal to E^k. Every height function z which is constant on E^k has $\mu_z(X) = 3$. There is a minimum $\mu_{0z}(X) = 1$ for some $z \leq z(X)$, a maximum $\mu_{2dz}(X) = 1$ for some $z \geq z(X)$ and no other critical point for $z \neq z(X)$. This is so because any d-cycle in $\{p \in X : z(p) < z(X)\}$ has to meet geometrically the d-cycle in X by nonzero self intersection in homology. The space $P \setminus X$ can be exhausted by a nested sequence of half-space sections $f^{-1}(h_j) \subset f^{-1}(h_{j+1}) \ldots \subset P \setminus X$. As they do not meet X they are (open) homology point spaces all homeomorphic to $\mathbb{R}^{2d}$. The one point compactification $P/(X = \text{point})$ is then a sphere S^{2d} with a tight map into E^{n-k} :

$$\begin{array}{ccc} P & \longrightarrow & E^N \\ \downarrow & & \downarrow \\ P/X & \longrightarrow & E^{N-k}. \end{array}$$

By Theorem 5.5 then $N - k = N - d - 1 \leq 2d + 1$, and $N \leq 3d + 2$.

If $k-1 \neq d$ we are in the case $k-1 = 2d$. Then $\partial\mathcal{H}X$ is a $2d$-sphere in $X \subset P$. This is a contradiction. □

Remarks. Equality $N = 3d + 2$ is attained for the standard smooth models of standard projective spaces. It has also been attained for polyhedral embeddings for the real projective space, for the standard complex projective space in [Kühnel and Banchoff 1983] and for some P^8 (perhaps $\mathbb{HP}(2)$) in [Brehm and Kühnel 1992]. By projection one finds tight embeddings for $N = 3d + 1$. There are no embeddings for $N \leq 3d$ by characteristic class obstructions.

By projections one finds tight maps for all $N \leq 3d + 2$.

Open Question. Let $f : P^{2d} \to \mathbb{R}^N$ be a tight substantial *map*. Is $N \leq 3d+2$? For $N = 2d + 1$, is $f(P)$ a convex hypersurface in E^{2d+1}? For $N \leq 2d$, is $f(P)$ a convex set?

Counting Critical Points. A continuous function $\varphi : M \to \mathbb{R}$ on a compact space M is said to have a *Poincaré polynomial* in case there exists for every value t, a number $\varepsilon > 0$, such that if $t - \varepsilon < r < t \leq s < t + \varepsilon$, the following conditions are satisfied:

(a) the homomorphisms

$$H_*(M_{t-\varepsilon}) \to H_*(M_r) \quad \text{and} \quad H_*(M_s) \to H_*(M_{t+\varepsilon})$$

induced by inclusions, are bijective.

(b) the group $H_*(M_{t+\varepsilon}, M_{t-\varepsilon}) = H_*(M_t, M_{t-})$ is finitely generated.

Here $M_t = \{p \in M : \varphi(p) \leq t\}$.

The value t is called *critical* in case this group is nonzero. Clearly any real algebraic function on a compact real algebraic manifold $M \subset E^N$ has a Poincaré polynomial.

Call the finitely many critical values $c_1 < c_2 \ldots < c_L$. Choose noncritical values $t_0, \ldots, t_L$ such that

$$t_0 < c_1 < t_1 < \ldots < c_L < t_L.$$

Definition. The *Poincaré polynomial* of φ is

$$P(\varphi) = \sum_{j=1}^{L} P(M_{t_j}, M_{t_{j-1}}) = \sum_k \mu_k(\varphi) t^k.$$

It is independent of the choice of $t_0, \ldots, t_L$. As in equation (3.3), it obeys the Morse inequalities:

$$\frac{P(\varphi) - P(M)}{1+t} = \frac{\sum_k (\mu_k(\varphi) - \beta_k) t^k}{1+t} = \sum_k K_k t^k$$

is a polynomial with integer coefficients $K_k \geq 0$.

In particular, putting $\mu_k = \mu_k(\varphi)$, we deduce all the inequalities of Theorem 3.8 for this more general case.

The function φ is *perfect* or *tight* in case $K_k = 0$ for all k, because this holds if and only if the inclusion $M_t \subset M$ is homology-injective for all t. This is the case if and only if $\mu_k(\varphi) = \beta_k(M)$ for all k. Equivalently $M_{c-\varepsilon} \subset M_c$ is homology injective for small $\varepsilon > 0$ and every critical value $c = c_j$.

Lemma 5.8 (the lacunary principle). *Suppose* $\mu_{2_l+1}(\varphi) - \beta_{2l+1}(M) = 0$ *for all* l. *Then* φ *is perfect. In particular* φ *is perfect in case* $\mu_{2l+1} = 0$ *for all* l.

Proof. Write μ_i for $\mu_i(\varphi)$. By the Morse inequalities (Theorem 3.8) we have

$$(\mu_{2l} - \beta_{2l}) - (\mu_{2l-1} - \beta_{2l-1}) + \ldots + (\mu_0 - \beta_0) = K_{2l} \geq 0$$

and

$$(\mu_{2l+1} - \beta_{2l+1}) - (\mu_{2l} - \beta_{2l}) \ldots = (\mu_{2l+1} - \beta_{2l+1}) - K_{2l} = 0 - K_{2l} = K_{2l+1} \geq 0.$$

Therefore $K_{2l} = K_{2l+1} = 0$ for all l. □

Remark. The limit Swiss cheese M (page 7) is tight and its height functions are tight as well but they have no Poincaré polynomials, since $H_1(M)$ has no finite basis.

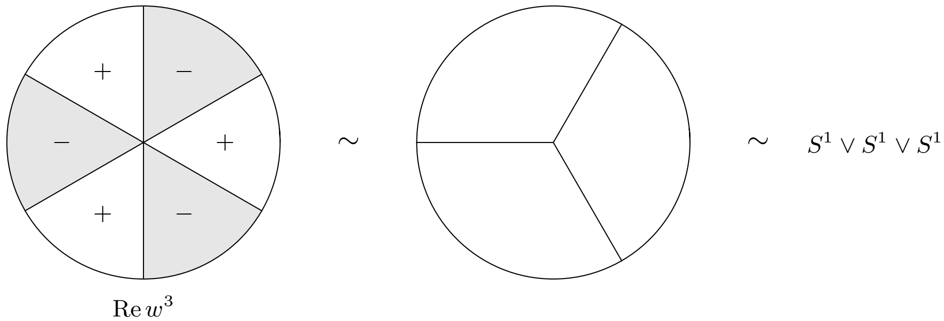


Figure 19. There is homotopy equivalence (indicated by $\sim$) between the set of (5.1) and a bouquet of r circles.

Definition. If $P(M_c, M_{c-}) = \sum_k \mu_{kc} t^k$, the critical value c counts for μ_{kc} critical points of index k, with total number

$$P(M_c, M_{c-})_{t=1} = \Sigma_k \mu_{kc} = \mu_c$$

critical points for the critical value c.

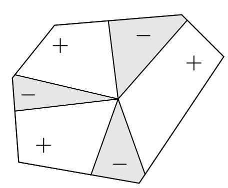
Exercise. Let the smooth function $\varphi : M^2 \to \mathbb{R}$ have at most one critical point p at level $\varphi^{-1}(p)$ (given by $d\varphi(p) = 0$) which in some topological chart is locally expressed by $(u, v) \to \mathrm{Re}(u + iv)^{r+1}$. Then our count is $\mu_1 = r$ for $r \geq 1$. The contribution in $P(\varphi)$ is t^r. The case $r = 1$ gives a nondegenerate critical point, $r = 2$ a monkey saddle, $r \geq 3$ an octopus saddle.

Proof. (See Figure 19, where $r + 1 = 3$.) Look at the set of points

$$\{w = u + iv : |w| \leq 1,\ \mathrm{Re}\, w^{r+1} \leq 0 \text{ or } |w| = 1\} \tag{5.1}$$

modulo the set $\{w : |w| = 1\}$. □

Exercise. Let φ be a PL-function on a PL-surface with one critical point p in the level $\varphi^{-1}(p)$. The level curve of φ near p is then in some chart either one point (minimum $\mu_0 = 1$ or maximum $\mu_2 = 1$; in both cases we have $\mu = 1$) or it is the union of $2(r + 1) \geq 4$ straight segments ending in p. In a suitable topological chart, we obtain the figure on the right. The contribution in $P(\varphi)$ at level $\varphi(p)$ is t^r [Kuiper 1971].



References

[Banchoff 1965] T. F. Banchoff, "Tightly embedded 2-dimensional polyhedral manifolds", *Amer. J. Math.* **87** (1965), 462–472.

[Banchoff 1971a] T. F. Banchoff, "High codimensional 0-tight maps on spheres", *Proc. Amer. Math. Soc.* **29** (1971), 133–137.

[Banchoff 1971b] T. F. Banchoff, "The two-piece property and tight n-manifolds-with-boundary in E^n", *Trans. Amer. Math. Soc.* **161** (1971), 259–267.

[Banchoff 1974] T. F. Banchoff, "Tight polyhedral Klein bottles, projective planes, and Möbius bands", *Math. Ann.* **207** (1974), 233–243.

[Borsuk 1947] K. Borsuk, "Sur la courbure totale des courbes fermées", *Ann. Polon. Math.* **20** (1947), 251–265.

[Brehm and Kühnel 1992] U. Brehm and W. Kühnel, "15-vertex triangulations of an 8-manifold", *Math. Ann.* **294**:1 (1992), 167–193.

[Casson 1986] A. J. Casson, "Three lectures on new-infinite constructions in 4-dimensional manifolds", pp. 201–244 in *À la recherche de la topologie perdue*, edited by L. Guillou and A. Marin, Progr. Math. **62**, Birkhäuser Boston, Boston, MA, 1986. With an appendix by L. Siebenmann.

[Chern and Lashof 1957] S.-s. Chern and R. K. Lashof, "On the total curvature of immersed manifolds", *Amer. J. Math.* **79** (1957), 306–318.

[Donaldson 1983] S. K. Donaldson, "An application of gauge theory to four-dimensional topology", *J. Differential Geom.* **18**:2 (1983), 279–315.

[Donaldson 1986] S. Donaldson, "Connections, cohomology, and the intersection forms of 4-manifolds", *J. Diff. Geom.* **24** (1986), 275–341.

[Eells and Kuiper 1962] J. Eells, Jr. and N. H. Kuiper, "Manifolds which are like projective planes", *Publ. Math. Inst. Hautes Études Sci.* **14** (1962), 5–46 (181–222 in year).

[Fáry 1949] I. Fáry, "Sur la courbure totale d'une courbe gauche faisant un nœud", *Bull. Soc. Math. France* **77** (1949), 128–138.

[Fenchel 1929] W. Fenchel, "Über die Krümmung und Windung geschlossener Raumkurven", *Math. Ann.* **101** (1929), 238–252.

[Freedman 1982] M. H. Freedman, "The topology of four-dimensional manifolds", *J. Diff. Geom.* **17**:3 (1982), 357–453.

[Kühnel and Banchoff 1983] W. Kühnel and T. F. Banchoff, "The 9-vertex complex projective plane", *Math. Intelligencer* **5**:3 (1983), 11–22.

[Kuiper 1941] N. H. Kuiper, "Lijnen in R_4", *Nieuw Archief voor Wiskunde* (2) **21** (1941), 124–143.

[Kuiper 1959] N. H. Kuiper, "Immersions with minimal total absolute curvature", pp. 75–88 in *Colloque Géom. Diff. Globale* (Bruxelles, 1958), Centre Belge Rech. Math., Louvain, 1959.

[Kuiper 1962] N. H. Kuiper, "On convex maps", *Nieuw Arch. Wisk.* (3) **10** (1962), 147–164.

[Kuiper 1966] N. H. Kuiper, "C^r-functions near non-degenerate critical points", mimeoraphed notes, Warwick Univ., Coventry, 1966.

[Kuiper 1970] N. H. Kuiper, "Minimal total absolute curvature for immersions", *Invent. Math.* **10** (1970), 209–238.

[Kuiper 1971] N. H. Kuiper, "Morse relations for curvature and tightness", pp. 77–89 in *Proceedings of Liverpool Singularities Symposium* (Liverpool, 1969/1970), vol. 2, edited by C. T. C. Wall, Lecture Notes in Math. **209**, Springer, Berlin, 1971.

[Kuiper 1980] N. H. Kuiper, "Tight embeddings and maps: Submanifolds of geometrical class three in E^n", pp. 97–145 in *The Chern Symposium* (Berkeley, 1979), edited by W.-Y. Hsiang et al., Springer, New York, 1980.

[Kuiper and Meeks 1983] N. H. Kuiper and W. H. Meeks, III, "Sur la courbure des surfaces nouées dans R^3", pp. 215–217 in *Third Schnepfenried geometry conference* (Schnepfenried, 1982), vol. 1, Astérisque **107-108**, Soc. Math. France, Paris, 1983.

[Kuiper and Meeks 1984] N. H. Kuiper and W. H. Meeks, III, "Total curvature for knotted surfaces", *Invent. Math.* **77**:1 (1984), 25–69.

[Kuiper and Meeks 1987] N. H. Kuiper and W. H. Meeks, III, "The total curvature of a knotted torus", *J. Differential Geom.* **26**:3 (1987), 371–384.

[Kuiper and Pohl 1977] N. H. Kuiper and W. F. Pohl, "Tight topological embeddings of the real projective plane in E^5", *Invent. Math.* **42** (1977), 177–199.

[Lastufka 1981] W. S. Lastufka, "Tight topological immersions of surfaces in Euclidean space", *J. Diff. Geom.* **16** (1981), 373–400.

[Little and Pohl 1971] J. A. Little and W. F. Pohl, "On tight immersions of maximal codimension", *Invent. Math.* **13** (1971), 179–204.

[Milnor 1950] J. W. Milnor, "On the total curvature of knots", *Ann. of Math.* (2) **52** (1950), 248–257.

[Milnor 1953] J. W. Milnor, "On the total curvature of closed space curves", *Math. Scand.* **1** (1953), 289–296.

[Milnor 1963] J. Milnor, *Morse theory*, Annals of Mathematical Studies **51**, Princeton U. Press, Princeton, 1963.

[Moise 1952] E. E. Moise, "Affine structures in 3-manifolds V: The triangulation theorem and Hauptvermutung", *Ann. of Math.* (2) **56** (1952), 96–114.

[Morse 1939] A. P. Morse, "The behavior of a function on its critical set", *Ann. Math.* **40** (1939), 62–70.

[Palais 1963] R. S. Palais, "Morse theory on Hilbert manifolds", *Topology* **2** (1963), 299–340.

[van Rooij 1965] A. C. M. van Rooij, "The total curvature of curves", *Duke Math. J.* **32** (1965), 313–324.

[Sharpe 1989] R. W. Sharpe, "A proof of the Chern-Lashof conjecture in dimensions greater than five", *Comment. Math. Helv.* **64**:2 (1989), 221–235.

[Spanier 1966] E. H. Spanier, *Algebraic Topology*, McGraw-Hill, New York, 1966.

[Wilson 1965] J. P. Wilson, "The total absolute curvature of an immersed manifold", *J. London Math. Soc.* **40** (1965), 362–366.

Nicolaas H. Kuiper
⋆ June 28, 1920
† December 12, 1994